Journal of Medical Internet Research

Contents

Original Papers

Tutorial

Corrigenda and Addenda

XSL·FO

RenderX

Original Paper

Text Messaging for Exercise Promotion in Older Adults From an Upper-Middle-Income Country: Randomized Controlled Trial

Andre Matthias Müller[1], MA; Selina Khoo[1], PhD; Tony Morris[2], PhD

[1]Sports Centre, University of Malaya, Kuala Lumpur, Malaysia

[2]College of Sport and Exercise Science, Victoria University, Melbourne, Australia

Corresponding Author:
Andre Matthias Müller, MA
Sports Centre
University of Malaya
Jalan Universiti
Kuala Lumpur, 50603
Malaysia
Phone: 60 11 11556673
Fax: 60 3 795 9590
Email: andrematthiasmueller@gmail.com

Abstract

Background: Mobile technology to promote exercise is effective; however, most evidence is from studies of younger groups in high-income countries. Investigating if short message service (SMS) texting can affect exercise participation in older adults from an upper-middle-income country is important considering the proliferation of mobile phones in developing regions and the increased interest of older adults in using mobile phones.

Objective: The main objective was to examine the short- and long-term effects of SMS text messaging on exercise frequency in older adults. Secondary objectives were to investigate how SMS text messages impact study participants' exercise frequency and the effects of the intervention on secondary outcomes.

Methods: The Malaysian Physical Activity for Health Study ($_{my}$PAtHS) was a 24-week, 2-arm, parallel randomized controlled trial conducted in urban Malaysia. Participants were recruited via health talks in resident associations and religious facilities. Older Malaysians (aged 55-70 years) who used mobile phones and did not exercise regularly were eligible to participate in the study. Participants randomly allocated to the SMS texting arm received an exercise booklet and 5 weekly SMS text messages over 12 weeks. The content of the SMS text messages was derived from effective behavior change techniques. The non-SMS texting arm participants received only the exercise booklet. Home visits were conducted to collect outcome data: (1) exercise frequency at 12 and 24 weeks, (2) secondary outcome data (exercise self-efficacy, physical activity–related energy expenditure, sitting time, body mass index, grip and leg strength) at baseline and at 12 and 24 weeks. Intention-to-treat procedures were applied for data analysis. Semistructured interviews focusing primarily on the SMS text messages and their impact on exercise frequency were conducted at weeks 12 and 24.

Results: In total, 43 participants were randomized into the SMS texting arm (n=22) and the non-SMS texting arm (n=21). Study-unrelated injuries forced 4 participants to discontinue after a few weeks (they were not included in any analyses). Overall retention was 86% (37/43). After 12 weeks, SMS texting arm participants exercised significantly more than non-SMS texting arm participants (mean difference 1.21 times, bias-corrected and accelerated bootstrap [BCa] 95% CI 0.18-2.24). Interview analysis revealed that the SMS text messages positively influenced SMS texting arm participants who experienced exercise barriers. They described the SMS text messages as being encouraging, a push, and a reminder. After 24 weeks, there was no significant difference between the research arms (mean difference 0.74, BCa 95% CI –0.30 to 1.76). There were no significant effects for secondary outcomes.

Conclusions: This study provides evidence that SMS text messaging is effective in promoting exercise in older adults from an upper-middle-income country. Although the effects were not maintained when SMS text messaging ceased, the results are promising and warrant more research on behavioral mobile health interventions in other regions.

Trial Registration: Clinicaltrials.gov NCT02123342; http://clinicaltrials.gov/ct2/show/NCT02123342 (Archived by WebCite at http://www.webcitation.org/6eGSsu2EI).

Introduction

Population aging is a global phenomenon that is projected to continue [1,2]. In 2013, there were approximately 841 million people globally who were age 60 years or older (11.7%). This figure is expected to grow to more than 2 billion (21.1%) by 2050 [3]. Accelerated population aging takes place in places other than high-income countries (HICs) where more than 80% of the global older adult population will reside by 2050 [3]. Malaysia is an upper-middle-income country that shows rapid population aging [4,5]. The proportion of older adults in the total population is projected to almost double from 2015 (5.8%) to 2040 (11.4%) [6].

The global rise in the older adult population is linked to an increase in the prevalence of noncommunicable diseases (NCDs), disability, and other health problems which, in turn, increase the burden on public and private health care systems [1,2,7,8]. This is especially so in non-HICs where the increase of NCDs is far greater than in HICs [2]. Lifestyle factors, especially physical activity (PA) and particularly structured exercise, are commonly put forward as essential determinants of good physiological and psychological health in older age [2,9-11]. For example, exercising older adults enjoy increased brain plasticity and cognitive function [12], reduced risk of cardiovascular diseases [13], and reduced metabolic risks [14]. Additionally, older adults who start following a regular strength-training regimen are likely to reduce body fat and greatly increase muscular strength [13,15]. Despite these benefits, many older adults are not exercising enough [11,16]. A recent review of PA levels in older adults reported that between 0% and 17.2% of older adults are active for 150 minutes per week when measured objectively [16]. Because exercise is a subcategory of PA, exercise levels are even lower. Studies reporting on PA and exercise levels of older adults in non-HICs are rare [17]. However, it seems that older adults in non-HICs, such as Malaysia, are increasingly inactive [18-20]. This was confirmed in a recent study reporting that 88% of older Malaysians are not sufficiently active [20]. The unprecedented increase in the older adult population, especially in non-HICs, imposes a great burden on health care systems. Because PA and specifically exercise are essential for good health in older age, there is a need for innovative and cost-effective interventions to increase exercise levels.

Behavioral health interventions focusing on PA and/or exercise are increasingly delivered via mHealth approaches [21], particularly mobile phones [22,23]. However, most interventions are conducted with younger age groups in HICs [21,24,25]. This is surprising because mobile phone penetration is increasing rapidly in non-HICs primarily because of device affordability [26]. This is evident in the Asia-Pacific region where approximately 3.6 billion people own a mobile phone. This figure is likely to increase in coming years [27]. Mobile phone subscriptions are also increasing in older adults. A study of older adults in the United States showed that 75.9% of adults aged 65 and older own a mobile phone [28]. Research also indicated the willingness of older adults to use mobile phone features that provide beneficial information to them, especially for health [29,30]. Short message service (SMS) text messaging is particularly popular among older adults and is the most frequently used mobile phone feature because little technological expertise is required for sending and receiving SMS text messages [28,31]. Therefore, examining SMS text messaging to promote exercise in older adults residing in an upper-middle-income country (Malaysia) would be informative.

Prior Research

Text messaging has been shown to be successful in promoting PA and/or exercise in young adults [32], postnatal women [33,34], working women [35], and sedentary women [36] from HICs. Text messaging was either a stand-alone intervention or it constituted a major part within a multicomponent intervention. In contrast, in non-HICs, SMS text-messaging interventions were primarily implemented to address maternal, child, and sexual health as well as disease management [26,37-40]. To our knowledge, only one study addressed exercise using SMS text messages in a non-HIC (India). A lifestyle intervention primarily delivered via SMS text messages to prevent the onset of type 2 diabetes in men was not successful in increasing overall activity levels [41].

Only one study reported on the effects of SMS text messaging in older adults [42]. The authors recruited a small sample of older African-Americans in a 6-week SMS text-messaging trial and found that step counts and leisure-time PA increased significantly. The aim of our study is to examine if SMS text messaging can successfully impact exercise behavior in older adults residing in a non-HIC (Malaysia). Further, we aim to determine if the effects of SMS text messaging on exercise behavior are maintained when the SMS text messages are removed.

Current Intervention

The Malaysian Physical Activity for Health Study ($_{my}$PAtHS) is a randomized controlled trial (RCT) for older Malaysians who do not follow a regular exercise routine. We chose this age group because population aging is a great challenge for Malaysia [5] and the majority of older Malaysians do not exercise enough [20]. The tropical, hot, and humid climate and safety concerns prevent many older Malaysians from exercising outside (eg, brisk walking). Additionally, a lack of appropriate exercise facilities for older adults and insufficient knowledge on how to exercise contribute to low exercise levels in this age group. Consequently, we introduced specific exercises to our research participants and provided a printed home-based exercise booklet. Text messages served as an encouragement and reminder to follow the exercises. The majority of older Malaysians use a

mobile phone (analog or smartphone) and are familiar with the SMS text-messaging feature [43].

Our study aimed to (1) determine if older Malaysians receiving an exercise booklet and weekly SMS text messages exercise more than participants who only receive an exercise booklet, (2) examine if the effects of the SMS text messages are maintained when the SMS text messages are removed, (3) investigate how the SMS text messages support participants to exercise, and (4) investigate the effects of the SMS text messages on secondary outcomes (eg, exercise self-efficacy, weekly PA-related energy expenditure, daily sitting time, body mass index [BMI], grip strength, and lower body strength).

Methods

Study Design: Overview

The $_{my}$PAtHS is a RCT that uses a parallel study design. All participants were introduced to a set of exercises and received an exercise booklet. Participants randomized into the SMS text-messaging arm (SMS texting arm) received an additional 60 text messages over 12 weeks. Participants randomized into the other arm did not receive SMS text messages (non-SMS texting arm). After enrollment, the primary study outcome was assessed at weeks 12 and 24. The study design and protocol were approved by the Faculty of Medicine Ethics Committee, University of Malaya, and was registered (Clinicaltrials.gov NCT02123342). This trial is reported according to the CONSORT statement [44] and the CONSORT-EHEALTH extension [45] (Multimedia Appendix 1).

Settings

The study took place in urban Malaysia (Kuala Lumpur and Petaling Jaya) from June 2014 to January 2015. In Malaysia, 73% of the population lives in urban areas and Kuala Lumpur and Petaling Jaya are the most densely populated cities [46].

Eligibility Criteria for Participants

Eligible participants were English-speaking community-dwelling Malaysians aged between 55 and 70 years, who were not exercising regularly (no structured exercise more than once weekly), had no health conditions that would restrict moderate exercise, used a mobile phone with SMS text-messaging function, and were interested in health-promoting exercise.

Recruitment and Enrollment

Participants were recruited from local resident associations and religious facilities in April and May 2014. With the support of representatives from the respective organizations, one study team member conducted health talks for older adults within the recruitment area. The study was briefly introduced as an exercise for health program and eligibility criteria were described (SMS text messaging was not mentioned). Those who were interested in taking part were given an information sheet and asked to provide contact details so that a study team member could call them later. Approximately one week after the health talks, potential participants were called. During this call, eligibility criteria were checked, initial oral consent was obtained, and a baseline home visit was scheduled. Home visits were conducted because some participants did not have personal transportation

and the public transportation system is not easily accessible. During the home visits, final eligibility checks were conducted, study procedures were explained (eg, time lines, potential risks), informed consent was obtained, and enrollment finalized.

Randomization and Allocation Concealment

The overall sample was stratified into participants enrolling with their spouse and participants enrolling without a spouse. There is evidence that older adults enrolling in an exercise intervention with a spouse exercise significantly more than those who do not [47]. Participants in the with-spouse and without-spouse strata were randomized separately on the day of enrollment after baseline measurements were taken and informed consent was obtained. Within strata, restricted randomization into the SMS texting and the non-SMS texting arm was applied to achieve balanced sample sizes. Sealed opaque envelopes with chits indicating the study arm were prepared by a study team member. A different set of envelopes was prepared for individual participants and participants enrolling with their spouse (stratification procedure). The envelopes were shuffled and participants were asked to randomly select one of the envelopes (as a means of allocation concealment). Participants and investigators were not blinded to arms assignment; however, participants were not aware that there were 2 research arms.

Intervention

During the baseline home visit, all participants were introduced to a set of exercises and received an exercise booklet ($_{my}$PAtHS booklet) developed by one of the study team members, an exercise physiologist with experience in training older adults. This booklet contained information on the benefits of exercise, some safety instructions, and descriptions of 12 age-appropriate strengthening exercises that could be executed without any specific equipment. Brief warm-up and cool-down sections were included as well. The exercises targeted major muscle groups of the arms/shoulders, upper trunk/neck, and legs. They were described using pictures, explanations of key movements, and hints where the exercises should be felt. One practical exercise session was conducted during the initial home visit to ensure correct execution. Participants were advised to exercise as often as possible each week to increase health benefits, but no other formal recommendations were provided. To ensure that participants' mobile phones were operational and participants were competent using the SMS text message function, they were asked to confirm receipt of a text message sent before the baseline home visit.

During the 12 weeks following the baseline home visit, 60 SMS text messages were sent to SMS texting arm participants (during weekdays). Text messages were sent automatically via an online tool specifically developed for this study. This tool allowed the research team to schedule the SMS text messages for every participant and it was also used to confirm delivery of the SMS text messages (Multimedia Appendix 2). Text messages were scheduled for morning hours between 8 am and 11 am according to participant preference. Text messages were developed from previous research that identified behavior change techniques (BCTs) most successful for increasing exercise self-efficacy and exercise behavior [48,49]. These BCTs were providing

instructions to exercise and providing rewards/praise for efforts toward exercise behavior [49]. Text messages contained an instruction to exercise using the _{my}PAtHS exercise booklet and a statement that praised the participants' engagement (Textbox 1). Several unique messages were developed and participants received a different message every day. Participants were not required to reply to SMS text messages because this has been shown to not significantly increase effectiveness [50]. Text messages ceased after 12 weeks. For the non-SMS texting arm participants, all procedures were the same, except they did not receive SMS text messages during the 12-week period.

Textbox 1. Example of a SMS text message sent to SMS texting arm participants.

Tailoring: Hello Mr. Wong, I hope you are well.

Instruction: Please do the _{my}PAtHS exercises regularly.

Praise/Reward: All your efforts will impact your health.

Closing: Have fun!

Outcome Measures

In this study, a mixed methods approach was applied to collect outcome data. Quantitative data was supplemented by qualitative data from semistructured interviews.

Quantitative Data

The primary study outcome was weekly exercise frequency (exercise sessions using the exercise booklet). It was assessed immediately after the 12-week SMS texting intervention period and after 24 weeks. This outcome was measured with an exercise log appended to the exercise booklet. Participants were asked to record dates, times, and duration of exercise sessions. During the baseline home visit, participants were shown how to record their exercise routine and one trial was conducted to ensure correct data entry. Additionally, one example of a correct entry was provided (in the booklet). Completed logs were exchanged for new ones on subsequent follow-up home visits.

A number of secondary outcomes were assessed at baseline and at weeks 12 and 24. Exercise self-efficacy is strongly associated with exercise participation in older adults [51] and was a covariate controlled for in this study. Participants were assessed using the Exercise Self-Efficacy Scale (EXSE) [52]. Validity and reliability of the EXSE were established in studies with older adults [53,54]. Participants were asked to rate the level of confidence they had for exercising with the exercise program over the coming weeks (from week 1 to week 8). Response options ranged from 0% (not at all confident) to 100% (highly confident) with 10% increments. The mean was used for analysis.

Physical activity-related energy expenditure, in weekly Metabolic Equivalent of Task (MET) minutes (MET-minutes), and daily time spent sitting (in hours) were measured using the short form of the International Physical Activity Questionnaire (IPAQ). The validity and reliability of this instrument are well established and it is widely used [55]. The IPAQ short form consists of 7 items and was interviewer-administered in this study. Respondents were required to indicate frequency (days per week) and duration (hours and minutes per day) of vigorous PA, moderate PA, and walking during the previous week. Another item asked about the time spent sitting on a normal weekday.

In addition, BMI (in kg/m^2) was calculated from body height and weight using the Seca Clara 803 Digital Personal Scale (Seca GmbH & Co KG, Hamburg, Germany). We assessed maximum grip strength (in kg force) of the dominant hand using the North Coast Hydraulic Hand Dynamometer (North Coast Medical Inc, Morgan Hill, CA, USA). This device has been used in previous studies with different groups of older adults and provided valid and reliable data [56-58]. On a verbal cue ("ready, go"), participants were instructed to squeeze the device as strongly as possible. The test was performed in the standing position and the mean of 3 trials was used for analysis. Lower body strength was assessed with the 30-second chair-stand test. This valid and reliable test was specifically developed for studies with older adults and has been widely used [59-61]. The number of stands from a chair that can be completed in 30 seconds is the test score.

Qualitative Data

During the follow-up home visits, we also conducted semistructured interviews with all research participants to complement the quantitative data. The interviews lasted approximately 20 minutes. We were primarily interested in how the participants in the SMS texting arm perceived the SMS text messages. Questions about the impact of the SMS text messages, their content, and what was done with them were discussed (Multimedia Appendix 3).

Sample Size Calculation

A total of 36 participants (18 per arm) was estimated to provide 80% power at $\alpha=.05$ to detect a difference of one weekly exercise session between the arms at week 12, assuming a standard deviation of 1.1 session. We anticipated a dropout rate of 15%; hence, we aimed to include 42 participants (21 per arm) [34,42].

Statistical Analysis

Statistical analysis using SPSS version 21.0 included descriptive statistics of age, sex, education, employment status, health status, and marital status. The intention-to-treat principle framed the analyses. However, intervention-unrelated injuries resulted in dropouts at week 12 and no primary outcome data was collected from these participants (n=4), consequently invalidating the use of imputation procedures. None of these participants were included in any of the analyses. For all other analyses, we used the last observation carried forward procedure for missing data.

We also conducted a per protocol analysis for those participants with complete outcome data using the same procedures as in the intention-to-treat analysis (Multimedia Appendix 4).

Weekly exercise frequency at week 12 (SMS texting period) was compared between study arms using an independent *t* test. Additionally, an analysis of covariance (ANCOVA) was conducted to adjust for the effect of exercise self-efficacy at baseline (covariate) as a key predictor of exercise participation in older adults [54]. For the analysis of weekly exercise frequency at week 24 (analysis of outcome after removal of SMS text messages), mixed between-within subjects 2×2 (time×group) ANOVAs were conducted. Interaction effects were followed up with simple effects analysis. In addition, an independent *t* test was conducted to compare weekly exercise frequency at week 24.

For the secondary outcomes, data were converted into 2 change variables: one between baseline and week 12 and one between baseline and week 24. For each variable, an ANCOVA comparing the change scores between the arms at each time point was conducted with the baseline scores entered as a covariate. For each arm, we estimated model-adjusted means, 95% confidence intervals, and *P* values.

Interview Analysis

The first author transcribed (not verbatim) the interviews and categorized responses into broad predefined themes (eg, exercise program, SMS text message content, effects of the SMS text messages). Themes were further divided into subthemes that were partly derived from the responses of the participants (eg, exercise benefits, perception of how the SMS text messages affected exercise). Direct quotations from participants were extracted to exemplify the results derived from the interviews. Finally, a coauthor checked the interview analysis results for accuracy and discrepancies were resolved via discussion.

Results

Overview

Participants were recruited in April and May 2014, and follow-up data were collected until January 2015. Figure 1 depicts the flow of the participants through the study. Of the 89 individuals screened, 43 eligible participants were randomized into the SMS texting arm (n=22) and the non-SMS texting arm (n=21). Table 1 displays the baseline demographic data. Participants had a mean age of 63.3 years (SD 4.5, range 55-70 years). The majority of the participants were female (74%, 32/43), obtained a college or university degree (67%, 29/43), were married (81%, 35/43), not working (77%, 33/43), and reported good health (67%, 29/43). There were no significant differences between the research arms on categorical and continuous variables (*P*>.05). Follow-up assessments at weeks 12 and 24 were completed for 18 (82%, 18/22) SMS texting arm and 21 (100%, 21/21) non-SMS texting arm participants. The overall retention rate was 86% (37/43) from baseline.

Table 1. Demographic characteristics of the research participants.

Characteristic	SMS texting (n=22)	Non-SMS texting (n=21)	Total (N=43)
Age (years), mean (SD)	63.64 (4.58)	62.90 (4.48)	63.28 (4.50)
Sex, n (%)			
Male	6 (27)	5 (24)	11 (26)
Female	16 (73)	16 (76)	32 (74)
Highest education, n (%)			
Secondary	6 (27)	4 (19)	10 (23)
Postsecondary	1 (5)	3 (14)	4 (9)
College/university	15 (68)	14 (67)	29 (68)
Employment status, n (%)			
Working	6 (27)	4 (19)	10 (23)
Not working	16 (73)	17 (81)	33 (77)
Health status, n (%)			
Fair	4 (18)	3 (14)	7 (16)
Good	16 (73)	13 (62)	29 (67)
Very good or Excellent	2 (9)	5 (24)	7 (16)
Marital status, n (%)			
Single/Separated/Widowed	4 (18)	4 (19)	8 (19)
Married	18 (82)	17 (81)	35 (81)

Figure 1. Study participant flow.
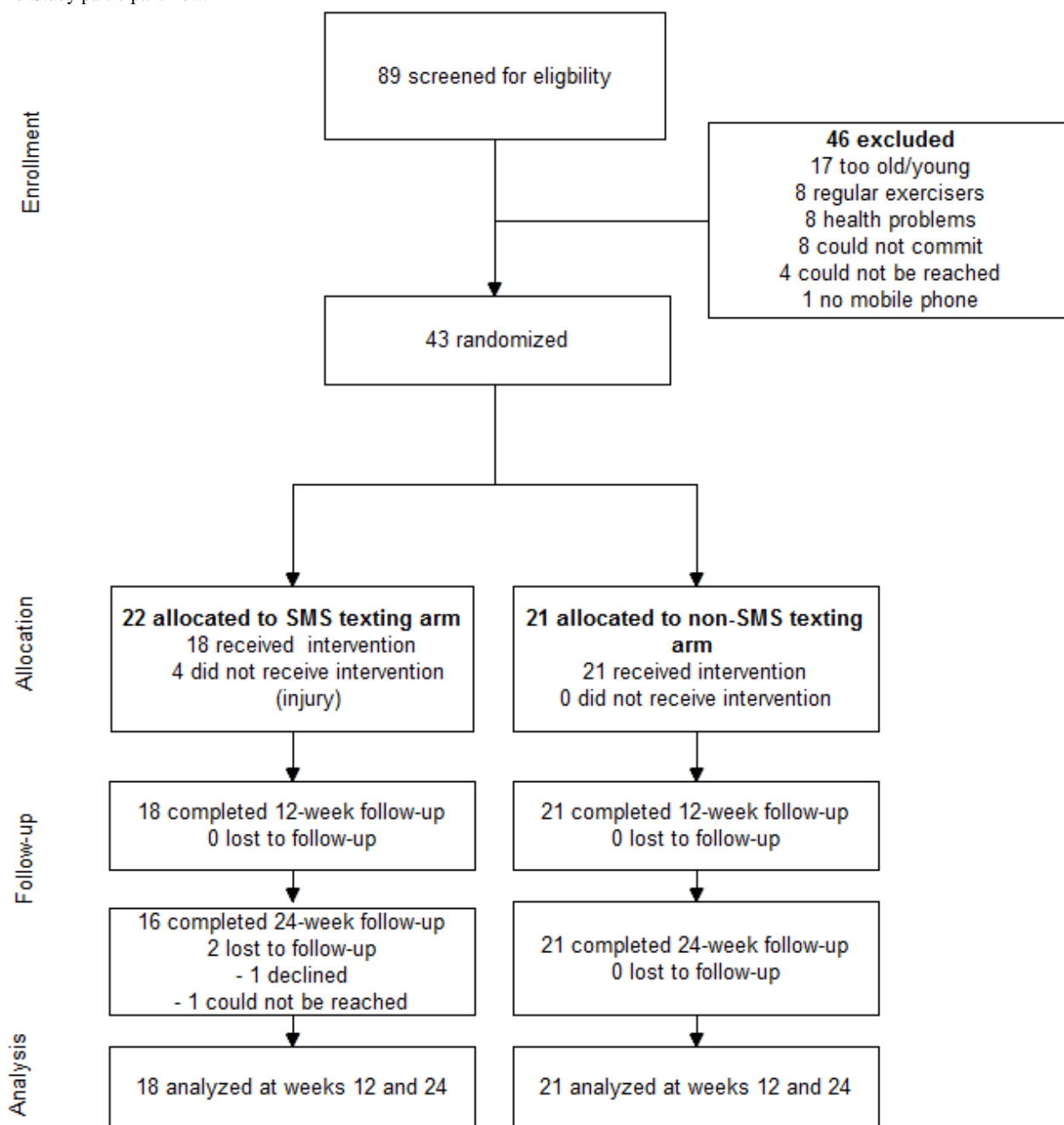
Outcome Assessments

Weekly Exercise Frequency

Over the 12-week intervention period, participants in the SMS texting arm exercised more frequently per week (mean 3.74, SD 1.34) compared to participants in the non-SMS texting arm (mean 2.52, SD 1.85). This difference (mean difference 1.21, bias-corrected and accelerated [BCa] 95% CI 0.18-2.24) was significant (t_{37}=2.30, P=.03, d=0.76). The effect of the SMS text messages on weekly exercise frequency was stronger after adjusting for baseline exercise self-efficacy with ANCOVA ($F_{1,36}$=6.81, P=.01).

Weekly exercise frequency decreased by 0.43 sessions (95% CI 0.12-0.74) from week 12 to week 24 in the overall sample ($F_{1,37}$=7.94, P=.008). There was no significant research arm by time interaction on weekly exercise frequency ($F_{1,37}$=2.46, P=.13). However, simple effects analysis revealed a significant decrease of 0.68 sessions within the SMS texting arm ($F_{1,37}$=8.93, P=.005), whereas no significant decrease was observed in the non-SMS texting arm ($F_{1,37}$ =0.85, P=.36). Further, an independent t test revealed that the SMS texting arm participants did not exercise more frequently (mean 3.07, SD 1.32) than the non-SMS texting arm participants (mean 2.33, SD 1.92) at week 24. The difference between the 2 arms (mean difference 0.74, BCa 95% CI –0.30 to 1.76) was not significant at week 24 (t_{37}=1.37, P=.18, d=0.45).

Secondary Outcomes

[Table 2](#) lists the effects of the SMS text messages on exercise self-efficacy, PA-related energy expenditure, daily sitting hours, BMI, grip strength, and lower body strength adjusted for the baseline values. There were no significant main or interaction effects ($P>.05$).

Table 2. Treatment effects on secondary outcomes.

Outcome	SMS texting, mean (SD) (n=18)			Non-SMS texting, mean (SD) (n=21)			Week 12		Week 24	
	Baseline	Change to week 12[a]	Change to week 24[a]	Baseline	Change to week 12[a]	Change to week 24[a]	Adj difference[a] (95% CI)	P	Adj difference[a] (95% CI)	P
Exercise self-efficacy score	81.94 (18.74)	−6.47 (19.67)	−3.02 (25.52)	81.55 (17.53)	−9.87 (19.65)	−14.32 (25.51)	−3.39 (−16.21, 9.43)	.60	−11.31 (−27.95, 5.34)	.18
PA-related energy expenditure (weekly MET-minutes)	662.29 (497.29)	383.43 (843.42)	434.69 (728.69)	968.71 (1479.10)	377.43 (842.56)	171.97 (728.22)	−5.99 (−558.14, 546.15)	.98	−262.72 (−739.77, 214.33)	.27
Daily sitting time (hours)	7.28 (3.39)	−0.36 (2.09)	−0.10 (1.90)	8.52 (2.20)	−1.17 (2.09)	−0.77 (1.89)	−0.81 (−2.19, 0.58)	.25	−0.67 (−1.92, 0.58)	.29
BMI (kg/m^2)	23.50 (3.47)	0.23 (0.51)	0.15 (0.68)	22.39 (2.81)	0.32 (0.51)	0.28 (0.68)	0.09 (−0.25, 0.42)	.60	0.13 (−0.32, 0.58)	.56
Grip strength (kg)	25.93 (8.70)	0.32 (2.01)	1.44 (2.55)	25.51 (6.34)	0.31 (2.01)	1.09 (2.55)	0.02 (−1.32, 1.29)	.98	−0.34 (−2.01, 1.32)	.68
Lower body strength (repetitions in 30-sec chair-stand test)	13.44 (3.42)	2.21 (3.40)	3.55 (2.80)	14.90 (3.81)	2.77 (3.39)	3.43 (2.79)	0.56 (−1.67, 2.80)	.61	−0.12 (−1.96, 1.72)	.90

[a] Adjusted for baseline.

Over the 24-week study period, a total of 4 adverse events occurred, all in the SMS texting arm (slipped disk: n=2; shoulder injury: n=1; hospitalization: n=1), none of which resulted directly from the study.

Qualitative Data

Week 12

The semistructured interviews at week 12 revealed that the participants from both arms were satisfied with the exercise program and faced few or no problems performing the exercises. They also reported that they improved their fitness, their health, and experienced elevated mood.

In the SMS texting arm, 9 participants reported few or no barriers to exercising (50%, 9/18). These participants indicated that the SMS text messages had limited impact and that they would have performed similarly without them. In contrast, 9 SMS texting arm participants (50%, 9/18), experienced a number of personal barriers (eg, laziness/tiredness, lack of motivation) to exercising. Despite these barriers, none of these participants discontinued exercise. They affirmed the value of the SMS text messages, which they described as very important and encouraging. They used words such as "cheering," "hopeful," and "inspiring" to express how they perceived the SMS text messages. For example, one participant said, "The text messages gave hope that I can do it." Four participants perceived the messages as an important push for them to exercise when they felt lazy. Interestingly, 4 participants reported feeling guilty when they received the SMS text messages on those days when they had no intention to exercise. They explained that the SMS text messages made them aware of their commitment and then they scheduled their exercise. Finally, 2 participants also reported that the SMS text messages served as a reminder on busy days.

In the non-SMS texting arm, there were also participants who experienced barriers to exercising (52%, 11/21). Six of these participants did not discontinue exercise, but suggested that an encouraging prompt would have been helpful. The remaining 5 participants exercised very infrequently before they discontinued (after 3 to 8 weeks). Three of them thought that reminders or prompts would have been important to help them continue exercising.

Week 24

The SMS texting arm participants who experienced few or no barriers to exercising during weeks 1 to 12 reported that they continued to exercise regularly, although some reported less exercise due to various reasons including traveling and busy schedules. Because the SMS text messages were not important to them from the beginning, they did not miss them.

Participants who experienced barriers to exercising during weeks 1 to 12 noted a decrease in barriers and 3 participants reported that the SMS text messages were no longer necessary. However, 2 participants said that it was very difficult for them to continue exercising without the SMS text messages. They reduced their exercise frequency: "I slowed down a little bit without it [text messages] because I did not get reminders." One participant reported that, without the SMS text messages, she exercised much less because she "did not feel the push and pressure from the text messages." In the non-SMS texting arm, participants

reported few or no changes during weeks 13 to 24 versus weeks 1 to 12.

Text Massages Exposure, Content, and Frequency

All SMS text messages were delivered to the SMS texting arm participants as scheduled, without technical problems. Participants who experienced barriers to exercising in the SMS texting arm (50%, 9/18) read all 60 SMS text messages and one participant saved them as well. In comparison, most participants who experienced few or no barriers to exercising ignored the SMS text messages after some time (78%, 7/9). The content of the SMS text messages was perceived as positive. Participants liked the encouragement the SMS text messages provided. One participant said that he "felt that his efforts were appreciated." Thirteen participants thought that the SMS text messaging frequency was too high (5 text messages per week), whereas 5 participants were in favor of the frequency. Some participants suggested sending more SMS text messages during the initial weeks and reducing the SMS text message frequency over time.

Discussion

Principal Results

This is the first RCT investigating a mHealth approach to promote exercise in older adults from a non-HIC. From the results, participants who received 60 encouraging SMS text messages over 12 weeks exercised significantly more than participants who did not receive such SMS text messages (mean difference 1.2 times per week). The SMS text messages were perceived as positive encouragement, especially for participants who experienced a number of barriers to exercising. Exercise frequency decreased significantly in the SMS texting arm when SMS text messages ceased. These findings suggest that SMS text messages have a strong impact on exercise participation in older adults, but the effect does not seem to be sustainable once they are removed.

Short-Term Effect of the Text Messages

In accordance with previous studies, we found that our SMS text messages had a marked short-term effect on exercise [32-36,42]. A number of reasons why SMS text messaging is effective in behavioral health interventions, particularly among older adults can be suggested. First, SMS text messaging is an easy-to-use mobile phone feature and older adults face fewer barriers to using this technology. This is a great advantage compared with more intricate features and technologies that tend to overwhelm them [28,31]. Our interview results reinforced this explanation because none of the participants experienced difficulties retrieving the SMS text messages. Second, SMS text messaging is generally perceived as a personal way to communicate [33]. This social component might be especially important to older adults, who might perceive a lack of personal interaction when using modern communication technology [28,29,62]. As a result, the attention paid to the intervention content is likely to be high and, in turn, affects intervention effectiveness. Third, some participants mentioned that they benefited from the SMS text messages because it reminded them to exercise.

Additionally, the call for designing behavioral mHealth interventions around effective BCTs has recently increased [63,64]. These BCTs are intended to actively and directly affect behavior [48]. The content of the SMS text messages in the current study was informed by the findings of a meta-analysis that identified the most effective BCTs for promoting exercise self-efficacy and exercise behavior [49]. The BCTs providing instructions to exercise and reinforcing effort of participants to exercise were perceived by a number of participants as influencing their behavior; thus, confirming the relevance of these BCTs.

Finally, our interview analysis revealed that the SMS text messages had a particularly strong impact on participants who experienced a number of barriers to exercise. This is an interesting finding that might explain why the SMS texting arm participants, on average, exercised more compared to the non-SMS texting arm participants. In each research arm, an equal proportion of participants experienced exercise barriers. However, participants who received SMS text messages continued to exercise, whereas a number of participants who did not receive SMS text messages discontinued exercise after some time. Researchers should examine the impact of SMS text messages in older adults who face exercise barriers.

Long-Term Effect of the Text Messages

Examining the long-term effects of a behavioral health intervention beyond its conclusion is important for research translation [65]. The results of our study indicate that the effect of the SMS text messages was not maintained when the SMS text messages ceased: after 24 weeks there was no difference between the SMS texting and non-SMS texting arms in exercise frequency. Similar findings were reported by Fjeldsoe et al [34], who implemented a 12-week SMS text-messaging intervention.

It is possible that the SMS text messages were not sent long enough to stabilize the acquired exercise routine [65]. Longer interventions might be especially important for older adults who have behavioral patterns that are well established and difficult to change [66]. However, instead of intervening longer with the same intensity, some researchers have also suggested sending prompts in the form of booster text messages to provide occasional support [34,67]. Our interview results do not provide a conclusive picture of the most appropriate behavioral maintenance method. Most participants suggested that the SMS text messages were unnecessary after the 12 weeks and only a few participants indicated reducing their exercise frequency in the absence of the SMS text messages. More research is needed to discern if booster SMS text messages can be an effective means to behavioral maintenance.

Finally, Fjeldsoe et al [65] in their systematic review found interventions with increased face-to-face contact led to more sustainable behavior change. This might be particularly relevant when working with older adults who value personal contact more than other age groups [29]. However, increased personal contact will also lead to increased costs and reduced outreach, decreasing some of the inherent advantages of SMS text messaging, especially when scaled up [67].

Secondary Outcomes

We did not observe any significant changes on secondary outcomes throughout the course of the study. That exercise self-efficacy did not change in the SMS texting arm compared to the non-SMS texting arm was particularly surprising considering that the BCTs incorporated in the SMS text message content were supposed to promote exercise self-efficacy [49]. One explanation for this is that exercise self-efficacy levels were already high at baseline (82/100) and increments were less likely to occur. McAuley et al [53] suggested measuring exercise self-efficacy approximately 3 weeks after the start of an exercise intervention because participants tend to be too optimistic at baseline and cannot accurately estimate how much effort it might take them to exercise. Measuring exercise self-efficacy after intervention exposure likely leads to a more accurate representation of baseline exercise self-efficacy levels that can serve as a reference for future assessments.

Limitations and Strengths

This study was limited by a lack of statistical power and the small sample size. Although our sample size calculation was based on the available literature [34,42], we did not expect that the standard deviations of the primary outcome would be as great as we observed. With this, the statistical power was less than the desired 80%. Dropout occurred only in the SMS texting arm. Four participants experienced a study-unrelated injury after a few weeks in the trial and could not continue exercising. We had conducted a rigorous randomization procedure leading to balanced research arms and this pattern of dropouts was unexpected. In this study, we obtained our primary outcome data with an exercise log. The bias of self-reporting is well documented in the literature [68], but using a log for data collection was most appropriate in our study. To ensure data accuracy and validity, participants practiced the data entry at baseline; during the interviews, they did not report any problems filling in the log.

A major strength of the current study was the investigation of behavioral change maintenance in older adults after the SMS text messages were removed; thereby, we filled an important gap in the evidence [42]. We also contribute to the growing body of literature on mHealth interventions in older adults who are less likely to be recruited into such interventions [69]. Further, the current study provides urgently needed evidence that shows that SMS text messaging is potentially effective in promoting health behavior in less developed regions where mobile phone proliferation is highest [37,38,70].

Conclusions

One of the great potentials of mHealth is that it can reach those most in need of health interventions, including people in non-HICs [38]. However, most knowledge about such interventions is generated in HICs and evidence from other regions is scarce [25,70]. In this study, older Malaysians exposed to SMS text messaging exercised more than those who did not receive SMS text messages, thus demonstrating the effectiveness of such an approach in a non-HIC. The effect of the SMS text messages was not maintained when they were removed, thus indicating a need for research on measures to increase sustainability.

Authors' Contributions

AMM and SK conceived the study, participated in its design and coordination, and drafted the manuscript. TM helped in designing the study, analyzing the data, and drafting the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Consort-EHealth checklist V1.6.1.

[PDF File (Adobe PDF File), 954KB - jmir_v18i1e5_app1.pdf]

Multimedia Appendix 2

Screenshots of the online tool used to send SMS text messages to SMS texting arm participants.

[PPTX File, 94KB - jmir_v18i1e5_app2.pptx]

XSL·FO

RenderX

Multimedia Appendix 3

Questions from the semistructured interviews (focus on the SMS text messages).

[PDF File (Adobe PDF File), 203KB - jmir_v18i1e5_app3.pdf]

Multimedia Appendix 4

Per protocol analysis.

[PDF File (Adobe PDF File), 402KB - jmir_v18i1e5_app4.pdf]

References

1. Rowe JW. Successful aging of societies. Daedalus 2015 Apr;144(2):5-12. [doi: 10.1162/DAED_a_00325]
2. World Health Organization. World Report on Ageing and Health. Geneva: World Health Organization; 2015. URL: http://apps.who.int/iris/bitstream/10665/186463/1/9789240694811_eng.pdf?ua=1 [WebCite Cache ID 6c7hbqN47]
3. United Nations, Department of Economic and Social Affairs Population Division. World Population Ageing 2013. New York: United Nations Publications; 2013. URL: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf [WebCite Cache ID 6cDQeitLc]
4. World Health Organization. World Health Statistics 2014. Geneva: World Health Organization; 2014. URL: http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf [WebCite Cache ID 6cDQoUfDn]
5. Tey NP, Siraj SB, Kamaruzzaman SB, Chin AV, Tan MP, Sinnappan GS, et al. Aging in multi-ethnic Malaysia. Gerontologist 2015 Nov 9:e1. [doi: 10.1093/geront/gnv153] [Medline: 26553738]
6. Department of Statistics Malaysia. Population Projections, Malaysia, 2010-2040. Putrajaya, Malaysia: Department of Statistics Malaysia; 2012.
7. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: The challenges ahead. Lancet 2009 Oct 3;374(9696):1196-1208 [FREE Full text] [doi: 10.1016/S0140-6736(09)61460-4] [Medline: 19801098]
8. Rechel B, Grundy E, Robine J, Cylus J, Mackenbach JP, Knai C, et al. Ageing in the European Union. Lancet 2013 Apr 13;381(9874):1312-1322. [doi: 10.1016/S0140-6736(12)62087-X] [Medline: 23541057]
9. Ziegelmann JP, Knoll N. Future directions in the study of health behavior among older adults. Gerontology 2015 Jan;61(5):469-476. [doi: 10.1159/000369857] [Medline: 25660128]
10. Sargent-Cox KA, Butterworth P, Anstey KJ. Role of physical activity in the relationship between mastery and functional health. Gerontologist 2015 Feb;55(1):120-131. [doi: 10.1093/geront/gnu042] [Medline: 24847845]
11. Macera CA, Cavanaugh A, Bellettiere J. State of the art review: Physical activity and older adults. Am J Lifestyle Med 2015 Feb 18:e1. [doi: 10.1177/1559827615571897]
12. Erickson KI, Gildengers AG, Butters MA. Physical activity and brain plasticity in late adulthood. Dialogues Clin Neurosci 2013 Mar;15(1):99-108 [FREE Full text] [Medline: 23576893]
13. Seco J, Abecia LC, Echevarría E, Barbero I, Torres-Unda J, Rodriguez V, et al. A long-term physical activity training program increases strength and flexibility, and improves balance in older adults. Rehabil Nurs 2013;38(1):37-47. [doi: 10.1002/rnj.64] [Medline: 23365004]
14. Petrella RJ, Lattanzio CN, Demeray A, Varallo V, Blore R. Can adoption of regular exercise later in life prevent metabolic risk for cardiovascular disease? Diabetes Care 2005 Mar;28(3):694-701. [Medline: 15735210]
15. American College of Sports Medicine, Chodzko-Zajko WJ, Proctor DN, Fiatarone SM, Minson CT, Nigg CR, et al. American College of Sports Medicine position stand. Exercise and physical activity for older adults. Med Sci Sports Exerc 2009 Jul;41(7):1510-1530. [doi: 10.1249/MSS.0b013e3181a0c95c] [Medline: 19516148]
16. Sun F, Norman IJ, While AE. Physical activity in older people: A systematic review. BMC Public Health 2013;13:449 [FREE Full text] [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
17. Macniven R, Bauman A, Abouzeid M. A review of population-based prevalence studies of physical activity in adults in the Asia-Pacific region. BMC Public Health 2012;12:41 [FREE Full text] [doi: 10.1186/1471-2458-12-41] [Medline: 22251660]
18. Teh CH, Lim KK, Chan YY, Lim KH, Azahadi O, Hamizatul AA, et al. The prevalence of physical activity and its associated factors among Malaysian adults: Findings from the National Health and Morbidity Survey 2011. Public Health 2014 May;128(5):416-423. [doi: 10.1016/j.puhe.2013.10.008] [Medline: 24726414]
19. Ibrahim S, Karim NA, Oon NL, Ngah WZ. Perceived physical activity barriers related to body weight status and sociodemographic factors among Malaysian men in Klang Valley. BMC Public Health 2013;13:275 [FREE Full text] [doi: 10.1186/1471-2458-13-275] [Medline: 23530696]
20. Kaur J, Kaur G, Ho BK, Yao WK, Salleh M, Lim KH. Predictors of physical inactivity among elderly Malaysians: Recommendations for policy planning. Asia Pac J Public Health 2015 Apr;27(3):314-322. [doi: 10.1177/1010539513517257] [Medline: 24425796]

XSL•FO

RenderX

21. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: A scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]

22. Silva BM, Rodrigues JP, de la Torre DI, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272. [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]

23. Okorodudu DE, Bosworth HB, Corsino L. Innovative interventions to promote behavioral change in overweight or obese individuals: A review of the literature. Ann Med 2015 May;47(3):179-185 [FREE Full text] [doi: 10.3109/07853890.2014.931102] [Medline: 25011006]

24. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]

25. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: A systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]

26. Déglise C, Suggs LS, Odermatt P. Short message service (SMS) applications for disease prevention in developing countries. J Med Internet Res 2012;14(1):e3 [FREE Full text] [doi: 10.2196/jmir.1823] [Medline: 22262730]

27. The World in 2014: ICT Facts and Figures. Geneva: International Telecommunication Union; 2014. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf [accessed 2015-10-12] [WebCite Cache ID 6cDRIG9eG]

28. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421. [doi: 10.1093/geront/gnt166] [Medline: 24379019]

29. Vroman KG, Arthanat S, Lysack C. "Who over 65 is online?" Older adults' dispositions toward information communication technology. Comput Hum Behav 2015 Feb;43:156-166. [doi: 10.1016/j.chb.2014.10.018]

30. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]

31. Zhou J, Rau PP, Salvendy G. Age-related difference in the use of mobile phones. Univ Access Inf Soc 2013 Sep 26;13(4):401-413. [doi: 10.1007/s10209-013-0324-1]

32. 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]

33. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: A randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]

34. Fjeldsoe BS, Miller YD, Graves N, Barnett AG, Marshall AL. Randomized controlled trial of an improved version of MobileMums, an intervention for increasing physical activity in women with young children. Ann Behav Med 2015 Aug;49(4):487-499. [doi: 10.1007/s12160-014-9675-y] [Medline: 25582987]

35. Gell NM, Wadsworth DD. The use of text messaging to promote physical activity in working women: A randomized controlled trial. J Phys Act Health 2015 Jun;12(6):756-763. [doi: 10.1123/jpah.2013-0144] [Medline: 25110303]

36. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]

37. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: A systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]

38. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: A systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]

39. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: A systematic review of the literature and strategic framework for research. Global Health 2014;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]

40. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: A review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]

41. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: A prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]

42. Kim BH, Glanz K. Text messaging to motivate walking in older African Americans: A randomized controlled trial. Am J Prev Med 2013 Jan;44(1):71-75. [doi: 10.1016/j.amepre.2012.09.050] [Medline: 23253653]

43. Hand Phone Users Survey 2014. Cyberjaya, Malaysia: Malaysian Communications and Multimedia Commission; 2014. URL: http://www.skmm.gov.my/skmmgovmy/media/General/pdf/Hand-Phone-User2014.pdf [WebCite Cache ID 6cpcGZfWq]

44. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]

45. Eysenbach G. 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]

46. Noncommunicable Diseases Country Profiles 2014. Geneva: World Health Organization; 2014. URL: http://apps.who.int/iris/bitstream/10665/128038/1/9789241507509_eng.pdf [WebCite Cache ID 6cDz5HitX]

47. Gellert P, Ziegelmann JP, Warner LM, Schwarzer R. Physical activity intervention in older adults: Does a participating partner make a difference? Eur J Ageing 2011 Jul 7;8(3):211-219. [doi: 10.1007/s10433-011-0193-5]

48. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]

49. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322 [FREE Full text] [doi: 10.1093/her/cyr005] [Medline: 21321008]

50. Head KJ, Noar SM, Iannarino NT, Grant HN. 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]

51. Koeneman MA, Verheijden MW, Chinapaw MJ, Hopman-Rock M. Determinants of physical activity and exercise in healthy older adults: A systematic review. Int J Behav Nutr Phys Act 2011;8:142 [FREE Full text] [doi: 10.1186/1479-5868-8-142] [Medline: 22204444]

52. McAuley E. Self-efficacy and the maintenance of exercise participation in older adults. J Behav Med 1993 Feb;16(1):103-113. [Medline: 8433355]

53. McAuley E, Mailey EL, Mullen SP, Szabo AN, Wójcicki TR, White SM, et al. Growth trajectories of exercise self-efficacy in older adults: Influence of measures and initial status. Health Psychol 2011 Jan;30(1):75-83 [FREE Full text] [doi: 10.1037/a0021567] [Medline: 21038962]

54. McAuley E, Jerome GJ, Elavsky S, Marquez DX, Ramsey SN. Predicting long-term maintenance of physical activity in older adults. Prev Med 2003 Aug;37(2):110-118. [Medline: 12855210]

55. 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]

56. Roberts HC, Denison HJ, Martin HJ, Patel HP, Syddall H, Cooper C, et al. A review of the measurement of grip strength in clinical and epidemiological studies: Towards a standardised approach. Age Ageing 2011 Jul;40(4):423-429 [FREE Full text] [doi: 10.1093/ageing/afr051] [Medline: 21624928]

57. Landi F, Cruz-Jentoft AJ, Liperoti R, Russo A, Giovannini S, Tosato M, et al. Sarcopenia and mortality risk in frail older persons aged 80 years and older: Results from ilSIRENTE study. Age Ageing 2013 Mar;42(2):203-209 [FREE Full text] [doi: 10.1093/ageing/afs194] [Medline: 23321202]

58. Taylor ME, Delbaere K, Lord SR, Mikolaizak AS, Brodaty H, Close JC. Neuropsychological, physical, and functional mobility measures associated with falls in cognitively impaired older adults. J Gerontol A Biol Sci Med Sci 2014 Aug;69(8):987-995. [doi: 10.1093/gerona/glt166] [Medline: 24149433]

59. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport 1999 Jun;70(2):113-119. [doi: 10.1080/02701367.1999.10608028] [Medline: 10380242]

60. Shumway-Cook A, Silver IF, LeMier M, York S, Cummings P, Koepsell TD. Effectiveness of a community-based multifactorial intervention on falls and fall risk factors in community-living older adults: A randomized, controlled trial. J Gerontol A Biol Sci Med Sci 2007 Dec;62(12):1420-1427. [Medline: 18166695]

61. Macfarlane DJ, Chou KL, Cheng YH, Chi I. Validity and normative data for thirty-second chair stand test in elderly community-dwelling Hong Kong Chinese. Am J Hum Biol 2006;18(3):418-421. [doi: 10.1002/ajhb.20503] [Medline: 16634026]

62. Silveira P, van de Langenberg R, van Het RE, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: A phase II preclinical exploratory trial. J Med Internet Res 2013;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]

63. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: The AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]

64. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]

65. 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]

66.   Floegel TA, Giacobbi PR, Dzierzewski JM, Aiken-Morgan AT, Roberts B, McCrae CS, et al. Intervention markers of physical activity maintenance in older adults. Am J Health Behav 2015 Jul;39(4):487-499. [doi: 10.5993/AJHB.39.4.5] [Medline: 26018097]

67.   Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: A systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]

68.   Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012;9:103 [FREE Full text] [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]

69.   Müller AM, Khoo S. Non-face-to-face physical activity interventions in older adults: A systematic review. Int J Behav Nutr Phys Act 2014;11(1):35 [FREE Full text] [doi: 10.1186/1479-5868-11-35] [Medline: 24612748]

70.   Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [Medline: 25361730]

Abbreviations

BCa: bias-corrected and accelerated bootstrap
BCT: behavior change technique
EXSE: Exercise Self-Efficacy Scale
HIC: high-income country
IPAQ: International Physical Activity Questionnaire
MET: Metabolic Equivalent of Task
myPAtHS: Malaysian Physical Activity for Health Study
NCD: noncommunicable disease
PA: physical activity
RCT: randomized controlled trial

Original Paper

The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review

Ghadah Alkhaldi[1], MPH; Fiona L Hamilton[1], PhD, FFPH, MRCGP; Rosa Lau[1], MSc; Rosie Webster[1], PhD; Susan Michie[2], DPhil, CPsychol, AcSS, FEHPsS, FBPsS; Elizabeth Murray[1], PhD, FRCGP, FRCPE

[1]eHealth Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom
[2]Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Ghadah Alkhaldi, MPH
eHealth Unit
Research Department of Primary Care and Population Health
University College London
Upper 3rd Floor Royal Free Hospital
Rowland Hill Street
London, NW3 2PF
United Kingdom
Phone: 44 020 7794050 ext 31399
Fax: 44 020 7794050
Email: Ghadah.alkhaldi.12@ucl.ac.uk

Abstract

Background: Digital interventions have been effective in improving numerous health outcomes and health behaviors; furthermore, they are increasingly being used in different health care areas, including self-management of long-term conditions, mental health, and health promotion. The full potential of digital interventions is hindered by a lack of user engagement. There is an urgent need to develop effective strategies that can promote users' engagement with digital interventions. One potential method is the use of technology-based reminders or prompts.

Objective: To evaluate the effectiveness of technology-based strategies for promoting engagement with digital interventions.

Methods: Cochrane Collaboration guidelines on systematic review methodology were followed. The search strategy was executed across 7 electronic databases: the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase, Web of Science, the Education Resources Information Center (ERIC), PsycINFO, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). Databases were searched from inception to September 13, 2013, with no language or publication type restrictions, using three concepts: randomized controlled trials, digital interventions, and engagement. Gray literature and reference lists of included studies were also searched. Titles and abstracts were independently screened by 2 authors, then the full texts of potentially eligible papers were obtained and double-screened. Data from eligible papers were extracted by one author and checked for accuracy by another author. Bias was assessed using the Cochrane risk of bias assessment tool. Narrative synthesis was performed on all included studies and, where appropriate, data were pooled using meta-analysis. All findings were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Results: A total of 14 studies were included in the review with 8774 participants. Of the 14 studies, 9 had sufficient data to be included in the meta-analyses. The meta-analyses suggested that technology-based strategies can potentially promote engagement compared to no strategy for dichotomous outcomes (relative risk [RR] 1.27, 95% CI 1.01-1.60, I^2=71%), but due to considerable heterogeneity and the small sample sizes in most studies, this result should be treated with caution. No studies reported adverse or economic outcomes. Only one study with a small sample size compared different characteristics; the study found that strategies promoting new digital intervention content and those sent to users shortly after they started using the digital intervention were more likely to engage users.

Conclusions: Overall, studies reported borderline positive effects of technology-based strategies on engagement compared to no strategy. However, the results have to be interpreted with caution. More research is needed to replicate findings and understand which characteristics of the strategies are effective in promoting engagement and how cost-effective they are.

KEYWORDS

systematic review; adherence; engagement; prompts; digital interventions

Introduction

Digital interventions (DIs) are programs that provide information and support—emotional, decisional, and/or behavioral—for physical and/or mental health problems via a digital platform (eg, website or computer) [1]. There has been substantial investment in DIs in developed countries, and they have been used in different health domains, including self-management of long-term conditions [2-4], promotion of healthy behaviors [1,5-7], and mental health [8]. The literature suggests that they can improve health behaviors and health outcomes [1-10]; however, systematic reviews of the effectiveness of DIs tend to report small effect sizes with a substantial level of heterogeneity [2,4,6,7,9]. One potential cause for the relatively small effect sizes is nonuse, or insufficient use, of the digital interventions [7]. Research has shown that there is a lack of engagement with DIs, and some studies have suggested a dose-response relationship between DIs' effectiveness and a user's level of engagement [11-16]. A review of DIs targeting physical activity showed that better engagement was associated with larger effects of the intervention [14]. Similar findings were seen in studies of DIs targeting fruit and vegetable consumption [11], weight loss [12], and smoking cessation [13,15]. Although it could be argued that the association between greater engagement and bigger positive effect is due to reverse causality (ie, the user experiences better outcomes so becomes more engaged), it is also plausible that better engagement leads to greater effectiveness [11-16]. Indeed, one systematic review of reviews looking at DIs aimed at health prevention reported, "One of the most substantial problems in online prevention is the low use of the interventions, a phenomenon seen across all behavior domains" [7]. Hence, one potential way of improving their effectiveness may be by promoting users' engagement.

In a three-round systematic Delphi experiment done by Brouwer et al [17], engagement was conceptualized into three phases. In the first phase, the user decides to first visit a DI to determine what it offers and whether he/she can benefit from it. In the second phase—prolonging the first visit—a user extends this visit and is exposed to part of the DI. In the third phase—revisiting the DI—the user returns to the DI after the first visit. The Brouwer et al study suggested that different factors impact on each phase. During the first phase, factors influencing the decision whether or not to visit the DI for the first time include user characteristics (eg, motivation and interest) and perceived relevance of the DI. In the second phase, the duration of the first visit is mostly determined by the characteristics of the DI (ie, whether it is tailored and easy to use). In the third phase, the decision whether to revisit is influenced by both user characteristics, such as motivation, and the presence or absence of reminders or prompts to revisit [17]. This systematic review targets the third phase by exploring the use of prompts as a method to promote revisiting DIs after the first visit [7,18-20]. Some systematic reviews have been published about technology-based prompts; however, these reviews have focused on the effect of prompts on the behavior addressed by the DI, rather than on the proximal effect on engagement [21-23]. There is some emerging evidence on design features, including use of prompts, that influence engagement [19,24]; one systematic review that performed qualitative analysis of the results of the included studies found that DIs that used email and phone contact with users were more likely to have better engagement [25].

To our knowledge, none of those reviews has focused specifically on the relationship between engagement, prompts, and the characteristics of prompts. Characteristics likely to influence effectiveness include timing (ie, when should a prompt be used), duration (ie, for how long should it be used) [18,25-27], frequency [22], mode of delivery (eg, email, text message, or telephone call [23]), sender [28,29], content [30], and theoretical underpinning [23]. It has been shown that an intervention based on theory is more effective than one that is not [23,31].

A review of digital interventions found that those that used more behavior change techniques (BCTs) were more effective than those that used fewer BCTs [23]. Therefore, this review attempted to code the content of the prompts using a BCT taxonomy [32], the same one used by the previously mentioned review [23]. The BCT taxonomy, comprised of 93 BCTs, has been rated, grouped, and agreed on by international behavior experts in a Delphi-type study; these BCTs are defined as "observable, replicable, and irreducible components of an intervention designed to alter or redirect causal processes that regulate behavior" [32]. This taxonomy can help identify the active ingredients that the intervention contains and, thus, the mechanism of action, which allows for a theory-based explanation of how to develop prompts that are effective in promoting engagement. The BCT taxonomy includes the prompt/cue techniques that "introduce or define environmental or social stimulus with the purpose of prompting or cueing the behavior." Thus, the term *strategy* was used in this review as it is more comprehensive and adaptable, and a strategy's content can include the BCT prompt/cue or more components.

The aim of this systematic review was to evaluate the effectiveness of technology-based strategies, defined in this review as digital and analog technology methods used to promote the user's regular interaction with all or part of the DI. These include, but are not limited to, emails, text messages, multimedia messages, telephone calls, automated voice calls, or faxes. Specific objectives of the review were to (1) describe technology-based strategies to promote engagement with DIs, (2) assess the effectiveness of technology-based strategies in promoting engagement with DIs, (3) explore whether different characteristics such as timing, duration, frequency, mode of delivery, sender, content, or use of theory are associated with differential effectiveness, and (4) to describe the cost of technology-based strategies to promote engagement with DIs.

Methods

This review followed Cochrane methodological guidance for systematic reviews [33] and the protocol with the full details about the methodology has been published [34].

Data Sources and Search Methods

The search was performed in 7 electronic databases: the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase, Web of Science, the Education Resources Information Center (ERIC), PsycINFO (including studies and dissertation abstracts), and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). Databases were searched from inception to September 13, 2013, with no language or publication type restrictions, using three concepts: randomized controlled trials (RCTs) *and* digital interventions *and* engagement (see Multimedia Appendix 1 for the MEDLINE search strategy). The search also included screening grey literature (Conference Proceedings Citation Index, formerly ISI Proceedings), references of the included studies, issues of key journals such as the Journal of Medical Internet Research (JMIR), and using Google Scholar to screen any papers citing included or other key papers [18,20,22,23].

Article Screening and Selection

All citations identified by the search strategy were deduplicated and downloaded into Endnote X5 (Thomson Reuters). Titles and abstracts were screened by one author (GA) and were double-screened by one of 3 other coauthors (EM, FH, or RW). Full texts of potentially eligible articles were screened by 2 authors (EM and GA). Any disagreement was resolved through discussion, referencing the eligibility criteria. If consensus could not be achieved, a third author (FH) was consulted. Justifications for exclusion were recorded and tabulated. All reviewers had training in systematic review methodology.

Inclusion Criteria

Participants

Participants were adults aged 18 years old or over. There were no limitations on gender, socioeconomic status, ethnicity, or health status. All settings were included for digital intervention; for technology-based strategies, the setting was online.

Interventions

The interventions of interest were technology-based strategies to promote engagement with digital interventions. To be included, the interventions had to meet the following definitions:

1. Digital interventions were defined as programs that provide information and support—emotional, decisional, and/or behavioral—for physical and/or mental health problems via a digital platform (eg, a website or a computer) [1].

2. Technology-based engagement-promoting strategies were defined as digital and analog technology methods used to promote the user's regular interaction with all or part of the DI, including, but not limited to, telephones calls, text messages, multimedia messages, emails, automated voice calls, or faxes. Examples of interventions that were included were a computerized treatment program with mobile phone text

messages that reminded the user to visit the program, and a blood pressure self-monitoring website that sent email prompts to users to enter their pressure readings on the website.

Comparisons

Three groups of comparators were defined: (1) minimal or inactive comparators, such as no strategy, (2) nontechnological strategies, such as printed materials or face-to-face contact, and (3) alternative technology-based strategies, for example, where the effects of email prompts are compared to the effects of text message prompts. Some studies tested the cumulative effect of multiple strategies; for example, both arms received prompts by email with one arm also receiving additional prompts by telephone call.

Outcomes

Primary Outcomes

The primary outcome was engagement with the DI, which was recorded as the number of log-ins/visits, number of pages visited, number of sessions completed, time spent on the DI, and number of DI components/features used. These measures were determined in advance before screening included studies [34].

Secondary Outcomes

Two types of secondary outcomes were selected:

1. Adverse outcomes, such as users feeling frustrated or irritated by email prompts, or experiencing a loss of self-esteem due to not being able to engage with the DI.

2. Economic outcomes, which were costs associated with strategies promoting engagement to inform future cost-effectiveness analysis.

Study Designs

RCTs were included; these were either trials of DIs that used strategies promoting engagement or trials evaluating strategies specifically. Economic evaluations were to be included if they were conducted alongside the main trial.

Exclusion Criteria

The following were the exclusion criteria:

1. Interventions targeted exclusively at health professionals (eg, computer-based decision aids to assist health professionals in making decisions with regard to treatments).

2. Trials where attrition from the trial and disengagement from the DI are nondistinguishable.

3. Trials where the effect of the DI components cannot be separated from the effect of the engagement-promoting strategy (eg, trials where the DI is not compared to another DI, such as a website to lose weight with email prompts compared with dietician face-to-face sessions with emails from the dietician; or when the difference between the 2 arms included different DIs as well as differential engagement strategies).

In the protocol, it was stated that quasi-RCTs would be included; however, upon further reflection, and due to the reasonable

number of eligible RCTs and the high risk of bias associated with quasi-RCTs, they were excluded.

Data Extraction

Data were extracted from included papers using an adapted version of the Cochrane Consumers and Communication Review Group data extraction template. One author (GA) extracted all the included papers and another coauthor (FH) verified the accuracy of the extraction; any disagreement was resolved through discussion. If no agreement was reached, a third author (EM) was consulted. Authors were contacted for more information about the characteristics of the strategy and any missing outcome data. The taxonomy for the BCTs [32] was used; strategy contents were coded by one author (GA) during data extraction and verified by another author (RW), who is an experienced user of the taxonomy.

Critical Appraisal Techniques

An assessment of risk of bias was done based on the Cochrane risk of bias assessment tool [33]. The following criteria were used:

1. Was the allocation sequence adequately generated?

2. Was allocation adequately concealed?

3. Was knowledge of the allocated interventions adequately prevented during the study (ie, blinding)?

4. Were incomplete outcome data adequately addressed?

5. Were study reports free of suggestion of selective outcome reporting?

6. Was the study free of other problems that could put it at risk of bias? These problems included, but were not limited to, baseline characteristic differences between groups, validity and reliability of outcome measures, sample size, and power.

The papers [11,35-47] were categorized as having low, high, or unclear risk of bias (ie, when the study did not provide enough information to judge the different aspects of trial quality). A risk of bias summary (see Multimedia Appendix 2A) and a risk of bias graph (see Multimedia Appendix 2B) were generated. The bias assessment was done by one author (GA) and was checked by another author (FH). Any discrepancies were resolved by a third author (EM).

Data Synthesis

Selection of Outcomes

Outcome measures were categorized as dichotomous or continuous engagement outcomes:

1. Dichotomous engagement outcome: any dichotomous measure of how participants engaged with the DI, such as proportion of participants who visited the DI, or proportion of participants who completed a prespecified number of modules.

2. Continuous engagement outcome: any continuous measure of how participants engaged with the DI, such as number of visits or page views.

Even within the categories of dichotomous and continuous outcomes, authors often reported more than one outcome. After

discussion with coauthors and for the purpose of analysis, one outcome was selected based on the following prespecified criteria:

1. The number of participants who visited the DI (ie, logged in to the website) or the number of visits/log-ins was selected, as these are the most appropriate indicators for engagement strategies [25,48].

2. The primary outcome defined or stated by the author.

3. The outcome reported separately for the control and intervention group, rather than lumped together.

4. The highest standard for engagement (ie, the authors report the number of participants who completed all the sessions rather than the number of participants who completed no sessions or a specific number of sessions).

5. Data from the longest measured follow-up period were chosen, as it is important to demonstrate sustained change.

Data Analysis

Results were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [49] and analyzed according to Cochrane guidelines [33]. Data from included studies were tabulated to allow for a narrative description of the results. Data on characteristics of engagement strategies were tabulated and all authors of included studies were contacted for clarification about their strategies, of whom 4 replied [35-38].

A meta-analysis was performed and continuous and dichotomous data from RCTs were pooled separately using a random effects model. The appropriate effect measures were determined depending on the type of data. For dichotomous outcomes, relative risks (RRs) and their 95% confidence intervals were used. For continuous outcomes, standardized mean differences (SMDs) with 95% confidence intervals were used. Due to the variable nature of the interventions, heterogeneity was expected and it was assessed using the I^2 statistic.

A sensitivity analysis was intended to be undertaken, as recommended by the Cochrane handbook, by excluding trials of poor quality to determine their effects on the study results, as well as a funnel plot to assess publication bias. However, there were insufficient studies to allow for a meaningful assessment. To investigate heterogeneity, a post hoc sensitivity analysis was conducted by removing one study [46] on the basis of visual inspection of the forest plots (see Multimedia Appendix 3).

Results

Summary of Search Results

Searching the electronic databases yielded a total of 18,881 records. After removing all duplicates (manually and using Endnote X5), 10,133 records remained for title and abstract screening. Of these, 93 went forward for full-text assessment, supplemented by 3 studies identified from reference tracking. A total of 77 papers were excluded at full-text screening for various reasons, the most common being that the engagement

XSL•FO

RenderX

strategy or DI did not meet the definition in this review, or that engagement was not measured in the study. There were 4 ongoing studies with only protocols available, and one study was a conference abstract. Figure 1 shows the results of the initial searches, screening, and selection processes.

Figure 1. PRISMA flow diagram.
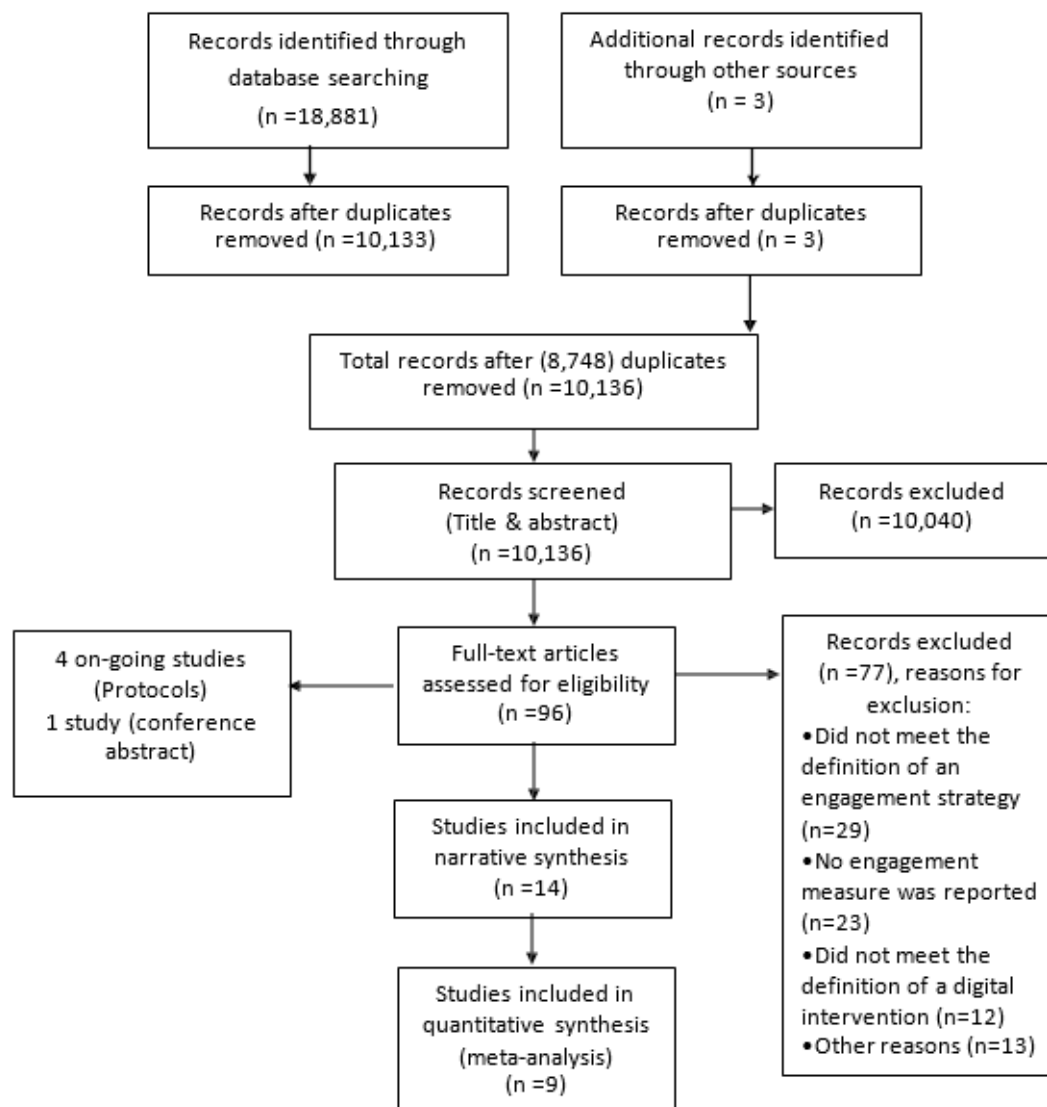
Included Studies

A total of 14 studies with 8774 participants were included in the systematic review; their characteristics are described in Table 1, with full details shown in Multimedia Appendix 4. The sample sizes ranged from 43 to 3448. One study was published in 2005 and the rest were published between 2009 and 2013; all studies were published in English. More than half of the studies [11,35,36,39-43] had more than 2 arms, but, with the exception of one study [39], only 2 arms met the inclusion criteria (see Table 1). One study was a factorial RCT where half of the participants received an engagement strategy while the other half did not [37]. One study had 7 arms assessing the effect of different timing and content of strategies [44]. The remaining 4 studies were RCTs with 2 arms [38,45-47].

All of the studies were conducted online and some studies specifically mentioned the location of the participants: the Netherlands [44,46], Australia [36,43,47], the United States [11,35,37,38,41], and Switzerland and Germany [40]. Six of the studies aimed to evaluate the effect of adding the strategy on the effectiveness of the DIs [35,39-42,47], 3 of the studies aimed to evaluate the effect of the technology-based strategies on promoting engagement with the DI [37,45,46], and 2 studies aimed to evaluate the effect of the strategy on digital intervention outcomes and engagement [36,43]. One study evaluated the effect of different timing and content of strategies on engagement [44], one study evaluated the effect of adding online peer coaching on increasing participation with a DI [38], and the final study explored the qualities of engagement with a DI [11].

Table 1. Characteristics of included studies.

Study	Digital intervention	Study design, engagement strategy, and comparator
Berger et al [39]	Internet-based self-help guide targeting social phobia	Three-arm RCT[a] (75 participants included in review) Arm 1: Emails (n=24); Arm 2: Emails + telephone (n=25); Comparator: No strategy (n=26)
Berger et al [40]	Internet-based self-help program targeting depression	Three-arm RCT (50 participants included in review), one arm excluded Arm 1: Emails (n=25); Comparator: No strategy (n=25)
Clarke et al [35]	Pure self-help program targeting depression	Three-arm RCT (155 participants included in review), one arm excluded Arm 1: Telephone (n=80); Comparator: Mail (n=75)
Couper et al [11]	Tailored Web program targeting health promotion (ie, intake of fruits and vegetables)	Three-arm RCT (1677 participants included in review), one arm excluded Arm 1: Emails (n=838); Comparator: No strategy (n=839)
Farrer et al [36]	Web intervention targeting depression	Four-arm RCT (83 participants included in review), 2 arms excluded Arm 1: Telephone (n=45); Comparator: No strategy (n=38)
Greaney et al [41]	Website targeting self-monitoring of physical activity, red meat intake reduction, fruit and vegetable consumption, daily multivitamin use, and smoking cessation	Two-arm RCT and one nonrandomized arm excluded (86 participants included in review) Arm 1: Emails + telephone (n=50); Comparator: Emails (n=36)
McClure et al [37]	Internet intervention targeting smoking cessation	Randomized factorial trial (1865 participants included in review) Arm 1: Emails (n=933); Comparator: No strategy (n=932)
Muñoz et al [42]	Web-based intervention targeting smoking cessation	Four-arm RCT (498 participants included in review), 2 arms excluded Arm 1: Emails (n=251); Comparator: No strategy (n=247)
Proudfoot et al [43]	Online psychoeducation program targeting bipolar disorder	Three-arm RCT (273 participants included in review), one arm excluded Arm 1: Emails (n=134); Comparator: No strategy (n=139)
Santucci et al [45]	An entirely automated and tailored Web-based intervention targeting anxiety and depression	Two-arm RCT (43 participants included in review) Arm 1: Emails (n=21); Comparator: No strategy (n=22)
Schneider et al [46]	Computer-tailored program targeting multiple health behaviors: physical activity, fruit and vegetable intake, smoking cessation, and decreasing alcohol consumption	Two-arm RCT (3448 participants included in review) Arm 1: Emails (n=1790); Comparator: No strategy (n=1658)
Schneider et al [44]	Internet-delivered computer-tailored program targeting multiple health behaviors: physical activity, fruit and vegetable intake, smoking cessation, and decreasing alcohol consumption	Seven-arm RCT (240 participants included in review) Arms 1-3: Email at 2, 4, or 6 weeks (n=34, 34, and 35, respectively); Arms 4-6: Email with updated content at 2, 4, or 6 weeks (n=36, 35, and 32, respectively); Comparator: No strategy (n=34)
Simon et al [38]	An interactive online program targeting bipolar disorder	Two-arm RCT (118 participants included in review) Arm 1: Email (n=64); Comparator: No strategy (n=54)
Titov et al [47]	A computer-delivered treatment targeting social phobia	Two-arm RCT (163 participants included in review) Arm 1: Telephone + emails + text messages (n=81); Comparator: Emails + text messages (n=82)

[a]RCT: randomized controlled trial.

Digital Interventions

The digital interventions targeted different health behaviors and conditions. Eight DIs were designed to target different mental health conditions, including social phobia [39,47], depression [35,36,40,45], anxiety [45], and bipolar disorder [38,43]. The rest of the DIs promoted a variety of health behaviors, including smoking cessation (n=4), decreasing alcohol consumption (n=2), self-monitoring of healthy behaviors (n=1), physical activity (n=2), and healthy diet (n=3) [11,37,41,42,44,46]. Most of the studies included detailed descriptions of the DIs. Two DIs were described as self-help guides with modules presented in a sequential order and participants could complete the whole program at once or over time [39,40]. Six DIs were composed

of sessions that were presented in a sequential and phased order [11,36,43-46]. There were 2 studies that updated their DIs with new information [44,46], and 2 described their DIs as interactive [35,38].

Technology-Based Engagement-Promoting Strategies and Their Characteristics

Timing

Four studies used their strategies at different time points. One engagement strategy was used at weeks 2 and 3 from baseline [41], one was used for the first 2 months postenrollment [37], one was used once on the third month from baseline and measured engagement at month 4 from baseline [46], and the

last study tested the use of the strategy at multiple time points (ie, second, fourth, or sixth week from baseline) [44].

Duration

Strategies were used either for the duration of the DI [11,36,38-40,42,43,45,47] or at specific times [35,37,41,44,46].

Frequency

Most of the studies reported using engagement strategies on a regular basis. Six studies used the strategy at least once per week [36,37,39,40,45,47], one used it for 2 weeks [41], one used it three times [35], and one used it to encourage users to complete sessions with up to 4 email prompts for each session [11]. Three studies reported variable frequencies [38,42,43] and 2 studies used a strategy once only [44,46].

Mode of Delivery

Email was the most commonly used mode of delivery among the different studies [11,37,38,40,42-46]. Telephone calls were used in 2 studies [35,36] and 3 studies used different modes of delivery: either telephone calls in addition to emails [39,41] or telephone calls, emails, and text messages [47].

Sender

Other characteristics that were identified were the type of sender or provider and whether the strategies were automated [38,42] or human supported. For the latter, therapists or counsellors [11,36,39,40], nonclinical staff [35], research staff [45,47], trained coaches [41], and trained peers [43] were usually the senders or providers.

Content

The content of the strategies was classified into 5 types: offering assistance with the DI [35,36,39-41], advertising or describing DI content [35,44,46], linking users to specific DI pages or sections [38,42,43], reminding or inviting users to complete their DI sessions [37,44-47], and providing support and feedback on the health behavior/health problem or engagement with the DI [11,39,40,43]. Some studies described the content of their strategies in a way that enabled coding them as BCTs. The BCTs used were *social support (unspecified)* [37,39,40,43,47], *prompts/cues* where strategies explicitly prompted the users to revisit the DI [37,42,45,46], *providing feedback on behavior* (ie, engagement) [39-41], using *social reward* in the form of written encouragement and praise on participants' progress in the DI [39,40,47], *providing feedback on the outcome of behavior* (ie, engagement) in terms of the improvement in their health [39,40], and *providing instructions on how to perform the behavior* (ie, engage with a DI, such as how to log in) [35].

Use of Theory

No paper provided information about any underlying theoretical framework for the use, delivery, or content of strategies.

Tailoring

Tailoring was reported in 3 studies. In one study, participants received reports about the frequency of their usage of the DI via emails [41], and in 2 studies, participants were sent emails with personalized greetings [44,46]. Four studies described strategies that can potentially be labeled as tailored: 2 studies

sent personalized feedback about progress in DI sessions to their participants [39,40], one reported using peer coaches to provide personalized advice via email to participants on how to use the materials provided through the DI [43], and one sent emails to users keyed to their smoking quit dates [42].

Quality of Studies

The studies differed in the way they were conducted and some did not provide sufficient information to judge their quality. All studies reported randomization but only 9 reported adequate sequence generation process [35,37,39,40,42-44,46,47]. Ten studies had adequate allocation concealment [35-37,39,40,42-44,46,47]. One study reported that participants and researchers were blinded [43]. Engagement measures were prespecified in 11 studies [11,36-38,40-46], however, 3 studies out of these did not report some engagement outcomes for the intervention and control group separately [11,37,42]. Engagement measures were measured objectively, so no bias was identified for any of the studies in terms of incomplete outcome data except for one study where engagement measures were not reported for 6 participants who dropped out [39]. Protocols were only reported in 3 studies [36,37,46].

Evaluating the Effectiveness of Technology-Based Engagement-Promoting Strategies

Technology-Based Engagement Strategies Compared to Minimal or Inactive Comparators

Data suitable for meta-analysis were only available for the comparison of a technology-based engagement strategy with no strategy. Two meta-analyses were performed, using dichotomous and continuous outcomes. The outcome measures of the studies included in the meta-analyses were number of DI modules/sessions/lessons completed, number of participants who completed DI modules/sessions/lessons, and number of participants who logged in/visited the DI; the outcome measures for the rest of the studies can be found in Multimedia Appendix 5.

Eight studies with 6120 participants reported sufficient data to be included in the meta-analyses, comparing a technology-based engagement strategy to no strategy using dichotomous outcomes (Analysis 1.1) (see Figure 2). This analysis showed that participants using DIs who received technology-based strategies were found to be significantly more likely to engage with the DI compared to those who did not receive any strategy (RR 1.27, 95% CI 1.01-1.60). However, the analysis demonstrated substantial heterogeneity between the findings of the included trials (I^2=71%), implying that the results from the included studies differed more than would be expected by chance. Visual inspection of the forest plot suggested that the Schneider et al study [46] was an outlier. This trial had a single email prompt at 3 months, which was much later than strategies used in other studies [46]. Sensitivity analysis, excluding the Schneider et al study [46] from the forest plot, reduced the heterogeneity (I^2=39%) and the effect of the technology-based strategy (RR 1.16, 95% CI 1.01-1.33) as shown in Multimedia Appendix 3.

Figure 3 shows the results of the meta-analysis for a technology-based engagement strategy compared to no strategy

using continuous outcomes (Analysis 1.2). Four studies with 226 participants were included, 3 of which were included in the previous meta-analysis, and no statistically significant difference was found in engagement with a DI between participants who received technology-based strategies compared to those who did not receive any strategy (SMD 0.19, 95% CI -0.11 to 0.48). Heterogeneity was low ($I^2=20\%$). There is an overlap in these meta-analyses, as 3 out of the 4 studies in Analysis 1.2 were also included in Analysis 1.1; however, the direction of effect in both meta-analyses was similar.

Figure 2. Analysis 1.1. Technology-based engagement strategy compared to no strategy: dichotomous outcomes.
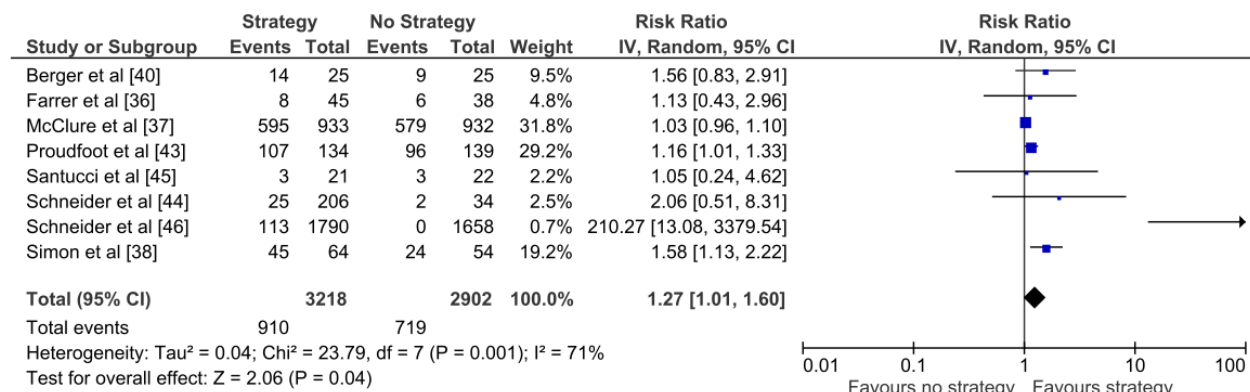
Figure 3. Analysis 1.2. Technology-based engagement strategy compared to no strategy: continuous outcomes.
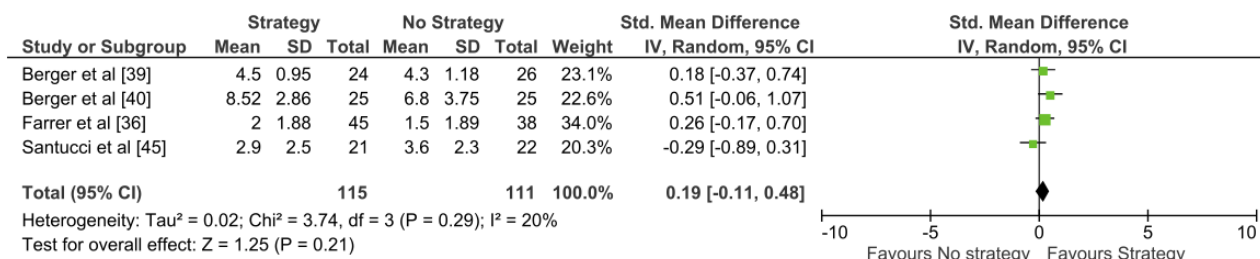
Technology-Based Engagement Strategies Compared to Nontechnological Strategies and Multiple Strategies

For the other comparator types, for which a meta-analysis was not performed, one study compared technology-based engagement strategies to nontechnological means of engagement (ie, comparing telephone calls to postal mail). The postal mail group had an average of 5.9 visits and the telephone call group had an average of 5.6 visits (mean difference = 0.3 visits, *P*=.65), suggesting no statistically significant difference in outcome between the groups [35].

As for the multiple strategies group, 3 studies had 2 arms with the same technology-based engagement strategy and one of the arms received an extra strategy delivered through telephone calls. None of the studies reported a significant difference in the effect of using multiple strategies on engagement [39,41,47]. However, no conclusions can be drawn for either comparator types, as meta-analysis was not possible due to the low number of studies.

Characteristics of Technology-Based Engagement Strategies

No conclusions can be drawn about the effect of the different characteristics, as only one study compared the effects of timing and content of strategies on engagement with a DI. The study found that strategies sent early and those that showed DIs' updated content were more likely to engage users [44].

Adverse and Economic Outcomes

Data on adverse and economic outcomes were intended to be extracted; however, none of the included studies reported these outcomes.

Unpublished Data

All authors were contacted to provide and confirm information about missing or unclear engagement outcome information or characteristics of strategies, and 4 authors replied. Farrer et al provided the mean and standard deviation of BluePage visits and time spent, and more information about the strategy, including the fact that it was not tailored [36]. McClure et al provided the exact number of people allocated to the strategy and the fact that the strategy was used for 12 months [37]. Clarke et al [35] and Simon et al [38] both confirmed the accuracy in categorizing their strategies' characteristics.

Discussion

Principal Findings

Technology-based strategies to promote engagement are an emerging field of research as shown by the number of included studies and their dates of publication. Generally, studies report borderline small-to-moderate positive effects of technology-based strategies on engagement compared to using no strategy, which support the use of technological strategies to improve engagement. However, this result should be treated

with caution due to the high heterogeneity, small sample sizes, and the lack of statistical significance in the analysis of continuous outcomes. There were insufficient studies to effectively explore reasons for heterogeneity. No firm conclusions were drawn about which characteristics of strategies were associated with effectiveness, and due to the absence of data, no conclusions could be drawn about costs or cost-effectiveness. Although the review aimed to investigate the cost-effectiveness of engagement strategies, none of the included papers reported cost data.

To our knowledge, this is the first systematic review that evaluated technology-based engagement-promoting strategies, using website metrics as outcome measures. Other systematic reviews [21-23] investigated the effect of technological engagement strategies of DIs on behavior change and some looked at engagement-promoting features of DIs, including the use of emails and telephone calls on the change in website metrics [25]. All of these systematic reviews reported a potentially positive effect of engagement strategies on changing health behavior and engagement. However, Brouwer et al, who used similar outcome measures, did not do a meta-analysis due to the heterogeneity of the outcome measures [25].

The findings in this review agree with previous reviews that technology-based strategies may potentially promote engagement, but that there is substantial heterogeneity, potentially due to the different outcome measures used [16,25,50,51], characteristics of the DI, and engagement strategies. In this systematic review, the measures were categorized into continuous and dichotomous outcomes, and outcomes were selected for meta-analysis using prespecified criteria. This allowed for performing two meta-analyses that shared similar studies but different measures. The two meta-analyses showed a similar direction of effect.

Authors often report multiple measures of engagement, and these often vary between studies. As measures of engagement are likely to vary depending on the research question, characteristics of the engagement strategy, and the DI, clear guidance for the optimal reporting of engagement is urgently needed. Researchers need to describe and detail clearly how a DI is intended to achieve its outcomes, the level of engagement intended or desired, and the rationale for that. For example, consider a structured and session-based DI targeting a mental disorder with an email prompting users to complete all the sessions to benefit from the DI, and the research question measuring how many participants completed all the sessions—an appropriate engagement measure would be the number of participants completing all the sessions rather than number of visits or time spent on the DI.

Authors should also clearly define their concept of optimal engagement in future studies, specifying a primary outcome for engagement and the rationale for choosing it. This is supported by the fact that the other systematic reviews of engagement reported that one of the most common reasons for excluding studies is a lack of reported engagement outcomes [19,25]. Another issue related to engagement measures is the extent/duration or level of engagement that defines whether a user is successfully engaging with a DI or not. One attempt to

quantify engagement was done by Kelders et al in a systematic review, which stated that a typical DI will have 50% of users engaged in it, using it at least once a week and up to 10 weeks. More research is needed to identify whether an outcome such as duration/level of engagement is enough to produce a positive effect size that justifies the cost of developing and implementing DIs [19].

This review identified themes in terms of characteristics of strategies to enable future research to selectively evaluate the different characteristics. Future primary studies that aim to determine the effectiveness of technological strategies on engagement with DIs should include a detailed description of the characteristics of engagement strategies, specifically the content of these strategies, and whether using different BCTs influence effectiveness. For this description, researchers could use the categories in this review, or expand on them. Researchers should also report the context (eg, characteristics of the DI) and outcome measures that contribute to heterogeneous results. This can help when conducting meta-analyses of future systematic reviews [52]. In addition, researchers should report multiple measures of outcome over the duration of the DI and not only report the engagement measure postintervention.

Researchers should also differentiate between attrition from the trial (ie, dropout attrition or loss to follow-up) and disengagement from the DI (ie, nonusage attrition), because studies have shown that the relation between these different types of attrition are complex and they do not share the same associated factors [18,20]. Disengagement is likely to impact on the effectiveness of the DI. It may be related to characteristics of the intervention (eg, design, usability, and perceived effectiveness) or to characteristics of the user (eg, motivation, self-efficacy, and resources). Loss to follow-up affects the ability of the study to answer the research question posed, with poor follow-up rates negatively impacting both the precision and the robustness of any estimate of effect.

Methodological Issues

The main strengths of this review are the rigorous and systematic methodology, which followed Cochrane methodological guidance, and the comprehensive and extensive search strategy. Furthermore, screening, extraction, and risk of bias assessment were independently conducted or reviewed by at least two authors. The review also includes meta-analyses to measure the effect of using the strategies compared to no strategies. In addition, the published, peer-reviewed protocol provides transparency.

The systematic review included RCTs as the most rigorous method for evaluating strategies, however, it is increasingly being recognized that the inclusion of other types of research is important. Policy makers and researchers are facing complex questions that the rigid and quantitative types of studies might not answer most appropriately. Rather, qualitative studies might be more equipped to fill in the gaps that RCTs cannot provide an answer for, such as the experiences of participants, the possible contradiction in some outcomes, and theory development [53]. In the case of engagement, certain issues can only be answered through conducting qualitative studies rather than quantitative ones [54]. These issues may include

understanding what outcomes mean for the user (eg, DI visits, page views, and time spent on the DI), what the experience of the engaged user is compared to the disengaged user, and the preference of users.

The limited search of the grey literature might be considered a limitation; however, in the case of this emerging field of research, the risk of significant publication bias is probably low because both negative and positive findings are of interest. A funnel plot could have been used to estimate the degree of publication bias; however, this was not possible because of the low number of studies, and the possibility of funnel plot asymmetry due to the different methodological qualities of the studies regardless of the existence of publication bias [33]. Another possible limitation might be that the use of the current Cochrane bias assessment guidelines might be more suitable for generic drug trials as opposed to DIs. For example, sequence generation is not an issue as judged in this review, as it is made easier with the use of online randomization programs. Blinding of staff and participants might not be possible as the control and intervention groups may be aware of receiving strategies sent by the staff. Criteria for traditional outcome assessment might not be suitable for reviewing studies of engagement, as it has to be tailored to how engagement is measured (eg, by automatic website metrics). For most of the studies, the description provided was not sufficient to judge the different aspects of trial quality. Authors and developers of DIs can benefit from using the enhanced CONSORT-EHEALTH reporting guide, published by JMIR. It can help clarify what authors need to report and describe in their studies to enable readers and reviewers to judge a study's quality [55].

Conclusions

Technology-based strategies may promote engagement compared to using no strategy; however, this finding should be interpreted with caution as only a small number of eligible studies were identified for the meta-analysis and the results were heterogeneous. The field of engagement strategies is an emerging field, as indicated by the number and dates of the studies; more research is needed to understand what strategy characteristics are effective and how cost-effective they are.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE search strategy.

[PDF File (Adobe PDF File), 91KB - jmir_v18i1e6_app1.pdf]

Multimedia Appendix 2

(A) Risk of bias summary. (B) Risk of bias graph.

[PDF File (Adobe PDF File), 170KB - jmir_v18i1e6_app2.pdf]

Multimedia Appendix 3

Comparison engagement strategy versus no engagement strategy: dichotomous outcomes sensitivity analysis after removing the study by Schneider et al [46].

[PNG File, 139KB - jmir_v18i1e6_app3.png]

Multimedia Appendix 4

Detailed characteristics of included studies.

[PDF File (Adobe PDF File), 381KB - jmir_v18i1e6_app4.pdf]

Multimedia Appendix 5

Main engagement outcomes and findings reported in included studies.

[PDF File (Adobe PDF File), 144KB - jmir_v18i1e6_app5.pdf]

XSL•FO

RenderX

References

1. Bailey J, Murray E, Rait G, Mercer C, Morris R, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]

2. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: A systematic review. J Med Internet Res 2014;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]

3. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]

4. Pal K, Eastwood S, Michie S, Farmer A, Barnard M, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:CD008776. [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]

5. Foster C, Richards J, Thorogood M, Hillsdon M. Remote and Web 2.0 interventions for promoting physical activity. Cochrane Database Syst Rev 2013 Sep 30;9:CD010395. [doi: 10.1002/14651858.CD010395.pub2] [Medline: 24085594]

6. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]

7. Kohl LF, Crutzen R, de Vries NK. 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]

8. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: Systematic review. Br J Psychiatry 2008 Sep;193(3):181-184 [FREE Full text] [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]

9. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: Web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]

10. Wantland D, Portillo C, Holzemer W, Slaughter R, McGhee E. 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]

11. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: Measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]

12. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of Internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]

13. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored Web-based smoking cessation program: Randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]

14. 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]

15. Zbikowski SM, Jack LM, McClure JB, Deprey M, Javitz HS, McAfee TA, et al. Utilization of services in a randomized trial testing phone- and Web-based interventions for smoking cessation. Nicotine Tob Res 2011 May;13(5):319-327 [FREE Full text] [doi: 10.1093/ntr/ntq257] [Medline: 21330267]

16. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

17. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to Internet-delivered behavior change interventions aimed at adults: A Delphi study approach. J Med Internet Res 2008;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]

18. 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]

19. 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]

20. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: Adherence and retention in a Web-based alcohol trial. J Med Internet Res 2013;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]

21. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: Systematic review. J Med Internet Res 2014;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]

22. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: Systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

XSL•FO

RenderX

23.   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]

24.   Lehto T, Oinas-Kukkonen H. Persuasive features in Web-based alcohol and smoking interventions: A systematic review of the literature. J Med Internet Res 2011;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]

25.   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]

26.   Glasgow RE, Boles SM, McKay HG, Feil EG, Barrera M. The D-Net diabetes self-management program: Long-term implementation, outcomes, and generalization results. Prev Med 2003 Apr;36(4):410-419. [Medline: 12649049]

27.   Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]

28.   Johnston L, Titov N, Andrews G, Spence J, Dear B. A RCT of a transdiagnostic Internet-delivered treatment for three anxiety disorders: Examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]

29.   Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: A randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]

30.   Michie S. What works and how? Designing more effective interventions needs answers to both questions. Addiction 2008;1036:886-887.

31.   Noar SM. Behavioral interventions to reduce HIV-related sexual risk behavior: Review and synthesis of meta-analytic evidence. AIDS Behav 2008 May;12(3):335-353. [doi: 10.1007/s10461-007-9313-9] [Medline: 17896176]

32.   Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

33.   Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. London, UK: The Cochrane Collaboration; 2011 Mar. URL: http://handbook.cochrane.org/ [accessed 2013-04-23] [WebCite Cache ID 6dsJoR5lQ]

34.   Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: A systematic review protocol. JMIR Res Protoc 2015;4(2):e47 [FREE Full text] [doi: 10.2196/resprot.3990] [Medline: 25921274]

35.   Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): A randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]

36.   Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: Randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]

37.   McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The effect of program design on engagement with an Internet-based smoking intervention: Randomized factorial trial. J Med Internet Res 2013;15(3):e69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]

38.   Simon GE, Ludman EJ, Goodale LC, Dykstra DM, Stone E, Cutsogeorge D, et al. An online recovery plan program: Can peer coaching increase participation? Psychiatr Serv 2011 Jun;62(6):666-669 [FREE Full text] [doi: 10.1176/ps.62.6.pss6206_0666] [Medline: 21632737]

39.   Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: A randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]

40.   Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: A randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]

41.   Greaney ML, Sprunck-Harrild K, Bennett GG, Puleo E, Haines J, Viswanath KV, et al. Use of email and telephone prompts to increase self-monitoring in a Web-based intervention: Randomized controlled trial. J Med Internet Res 2012;14(4):e96 [FREE Full text] [doi: 10.2196/jmir.1981] [Medline: 22842775]

42.   Muñoz RF, Barrera A, Delucchi K, Penilla C, Torres L, Pérez-Stable EJ. International Spanish/English Internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 Sep;11(9):1025-1034 [FREE Full text] [doi: 10.1093/ntr/ntp090] [Medline: 19640833]

43.   Proudfoot J, Parker G, Manicavasagar V, Hadzi-Pavlovic D, Whitton A, Nicholas J, et al. Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: A randomised controlled trial. J Affect Disord 2012 Dec 15;142(1-3):98-105. [doi: 10.1016/j.jad.2012.04.007] [Medline: 22858215]

44. Schneider F, de Vries H, Candel M, van de Kar A, van Osch L. Periodic email prompts to re-use an Internet-delivered computer-tailored lifestyle program: Influence of prompt content and timing. J Med Internet Res 2013;15(1):e23 [FREE Full text] [doi: 10.2196/jmir.2151] [Medline: 23363466]

45. Santucci LC, McHugh RK, Elkins RM, Schechter B, Ross MS, Landa CE, et al. Pilot implementation of computerized cognitive behavioral therapy in a university health setting [First online: 17 April 2013]. Adm Policy Ment Health 2014 Jul;41(4):514-521. [doi: 10.1007/s10488-013-0488-2] [Medline: 23592231]

46. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an Internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]

47. Titov N, Andrews G, Choi I, Schwencke G, Johnston L. Randomized controlled trial of Web-based treatment of social phobia without clinician guidance. Aust N Z J Psychiatry 2009;43(10):913-919.

48. Tullis T, Albert B. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Amsterdam, the Netherlands: Elsevier/Morgan Kaufmann; 2008.

49. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]

50. Christensen H, Griffiths K, 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]

51. 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. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]

52. Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. J Clin Epidemiol 2013 Nov;66(11):1244-1250. [doi: 10.1016/j.jclinepi.2013.06.013] [Medline: 23953079]

53. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: A review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [Medline: 15667704]

54. Lalmas M, O'Brien H, Yom-Tov E. In: Marchionini G, editor. Measuring User Engagement: Synthesis Lectures on Information Concepts, Retrieval, and Services. San Rafael, CA: Morgan and Claypool Publishers; 2014.

55. 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]

Abbreviations

BCT: behavior change technique
CENTRAL: Cochrane Central Register of Controlled Trials
CINAHL: Cumulative Index to Nursing and Allied Health Literature
DI: digital intervention
ERIC: Education Resources Information Center
JMIR: Journal of Medical Internet Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: relative risk
SMD: standardized mean difference
UCL: University College London

<u>Original Paper</u>

A Guided Online and Mobile Self-Help Program for Individuals With Eating Disorders: An Iterative Engagement and Usability Study

Martina Nitsch[1,2], Dr Phil; Christina N Dimopoulos[3], BA; Edith Flaschberger[2,4], Dr Phil; Kristina Saffran[3], BA; Jenna F Kruger[3], MPH; Lindsay Garlock[5], MA; Denise E Wilfley[6], PhD; Craig B Taylor[3,7], MD; Megan Jones[3,5], PsyD

[1]Ferdinand Porsche Distance Learning University of Applied Sciences, Vienna, Austria

[2]Ludwig Boltzmann Institut Health Promotion Research, Vienna, Austria

[3]Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, United States

[4]Health Promotion Fund Austria, Austrian Public Health Institute, Vienna, Austria

[5]Lantern, San Francisco, CA, United States

[6]Department of Psychiatry, School of Medicine, Washington University, St. Louis, MO, United States

[7]mHealth Institute, Palo Alto University, Palo Alto, CA, United States

Corresponding Author:
Martina Nitsch, Dr Phil
Ferdinand Porsche Distance Learning University of Applied Sciences
Lothringerstraße 4-8
Vienna, 1040
Austria
Phone: 43 1 505 77 78 ext 152
Fax: 43 1505 32 28
Email: martina.nitsch@fernfh.ac.at

Abstract

Background: Numerous digital health interventions have been developed for mental health promotion and intervention, including eating disorders. Efficacy of many interventions has been evaluated, yet knowledge about reasons for dropout and poor adherence is scarce. Most digital health intervention studies lack appropriate research design and methods to investigate individual engagement issues. User engagement and program usability are inextricably linked, making usability studies vital in understanding and improving engagement.

Objective: The aim of this study was to explore engagement and corresponding usability issues of the Healthy Body Image Program—a guided online intervention for individuals with body image concerns or eating disorders. The secondary aim was to demonstrate the value of usability research in order to investigate engagement.

Methods: We conducted an iterative usability study based on a mixed-methods approach, combining cognitive and semistructured interviews as well as questionnaires, prior to program launch. Two separate rounds of usability studies were completed, testing a total of 9 potential users. Thematic analysis and descriptive statistics were used to analyze the think-aloud tasks, interviews, and questionnaires.

Results: Participants were satisfied with the overall usability of the program. The average usability score was 77.5/100 for the first test round and improved to 83.1/100 after applying modifications for the second iteration. The analysis of the qualitative data revealed five central themes: layout, navigation, content, support, and engagement conditions. The first three themes highlight usability aspects of the program, while the latter two highlight engagement issues. An easy-to-use format, clear wording, the nature of guidance, and opportunity for interactivity were important issues related to usability. The coach support, time investment, and severity of users' symptoms, the program's features and effectiveness, trust, anonymity, and affordability were relevant to engagement.

Conclusions: This study identified salient usability and engagement features associated with participant motivation to use the Healthy Body Image Program and ultimately helped improve the program prior to its implementation. This research demonstrates that improvements in usability and engagement can be achieved by testing and adjusting intervention design and content prior to program launch. The results are consistent with related research and reinforce the need for further research to identify usage

XSL•FO
RenderX

patterns and effective means for reducing dropout. Digital health research should include usability studies prior to efficacy trials to help create more user-friendly programs that have a higher likelihood of "real-world" adoption.

KEYWORDS

usability study; engagement; adherence; dropout; digital health intervention; online program; self-help; eating disorder; mobile application

Introduction

Digital health technologies are increasingly common and are used in the prevention, diagnosis, and treatment of mental health problems. In the eating disorders field, several online programs have been developed and demonstrate promising results [1-3]. Available evidence for these programs suggest significant improvements in preventing and treating eating disorders [2,4,5], yet poor adherence and high dropout rates remain common and challenging problems in most studies [4,6,7]. Inconsistent measures of program usage and dropout across studies contribute to high variability in interpretation of adherence and findings [2,7-10].

To date, there is scant research examining the specific reasons for dropout and poor user engagement in online programs. The multidimensional nature of user engagement complicates research design because engagement includes individual users' thoughts and feelings, degree of activity, and attitudes towards technical aspects of the program including aspects of usability and appeal [11]. User engagement is also inextricably linked to the usability of a program [11], which refers to aspects of effectiveness, efficiency, and satisfaction [12]. Fortunately, established methods to examine program usability exist, which can also be harnessed to evaluate user engagement issues of digital health interventions.

This study aimed to explore and reveal different usability and engagement issues in the course of the redesign of the Healthy Body Image Program (HBI) [13-17], which is an evidence-based, guided online intervention for individuals with body image concerns or disordered eating symptoms, prior to its implementation. We conducted a usability study including qualitative interview elements focusing specifically on engagement. We applied a mixed-methods approach to investigate the first phase of interaction because research on this early stage of engagement is rare [18] and critical to outcome. This interaction time-point is important for determining future usage patterns and possible dropout reasons. The results of an iterative usability study exploring potential users' initial phase of interaction with a prototype of a guided online self-help program are presented. This study also seeks to describe the method and demonstrate the value of investigating engagement issues within a usability study prior to program launch.

Methods

Participants and Recruitment

Participants were recruited through Web-based and print advertisements (eg, flyers, email listserve announcements) in the San Francisco Bay Area and on the campus of a large private university. Participants were offered free access to the online program and gift cards of US $10-30 depending on how much time they spent in the usability testing. The HBI program was originally designed for college-aged females, so this study also included women aged 18-25 years with an interest in improving body image and reducing disordered eating behaviors. Interested individuals were first contacted via phone in order to explain the study procedure and to conduct a short telephone screening, for which we used the SCOFF questionnaire [19]. SCOFF (Sick, Control, One stone, Fat, Food; the acronym comprises the questionnaire's 5 items) is a widely used and well-validated eating disorder symptom screen. Consistent with prior research using the SCOFF as a screening tool, the indication of a possible eating disorder diagnosis, as measured by a positive response to 2 or more of 5 questions, was an additional inclusion criterion for this study. Thus, we ensured that the participants were representative of the individuals who are usually directed to use this particular HBI track (described below). As this study was part of a larger intervention study, the following exclusion criteria applied: lack of English language fluency, hearing impairments, and participation in any depression or anxiety intervention research study.

The complete sample of this study consisted of 9 participants. Based on an iterative usability study design approach [6,20], we aimed to conduct tests in two rounds with no more than 5 participants per round, since usability testing with 5 users reveals 85% of usability problems and more than 5 users would produce repetitive information [21]. In the first round, 4 participants tested the prototype on the computer, after which major issues were addressed. In the second round, another 5 participants tested the revised and improved program, based on the results of the first round, as a mobile app on a smartphone. The transition from the prototype on the computer to mobile app was planned as a further step in the development cycle of the program and was thus directly factored into the research design. The intervention was ultimately intended to be used on mobile and Web. Of note, the intervention was designed mobile first, even though the intervention was first accessible on the Web.

Each participant used a prototype of the program in a usability testing session, which lasted from 45-110 minutes. The time it takes users to engage with the program and to answer questions naturally varies due to the nature of usability evaluation. Thus, we followed the timing and pace of each participant to better understand individual differences. Each test had participants use the program while performing the think-aloud technique, followed by a semistructured interview and a short questionnaire. The testing sessions were conducted from August through October 2014 and took place at Stanford University

School of Medicine, except three tests, which were held in other public places (eg, separate room in a library, backyard of a cafe) chosen by the participants. The privacy of the participants and a possible impact regarding their answers was taken into consideration by making sure that there were no people in close vicinity. We obtained human subjects approval from the Institutional Review Board at Stanford University.

The Intervention

HBI begins with a Web-based assessment to determine the severity of disordered eating symptoms and risk for developing an eating disorder. Based on assessment results, participants are directed to one of several tracks of a tailored online evidence-based intervention, referred to by variations of "Student Bodies," which is for individuals across the eating disorder risk and diagnostic spectrum [22]. Tailoring is done at the level of program assignment. This involves using the initial assessment to determine whether the intervention is suitable for the individual participant. The personalization occurs in several additional ways. First, participants select preferences, set goals, and receive dynamic feedback and recommendations based on interaction sequences. Second, each participant interacts with a personal coach and receives unique messages and feedback to support engagement and personal relevance.

In the current study, which is related to the redesign of the program, we used an offline-prototype of the guided self-help program Student Bodies–Eating Disorders (SB-ED), which is specifically designed for individuals who screen positive for a clinical or subclinical eating disorder as defined in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition (DSM-5), excluding full-syndrome anorexia nervosa. As part of a dissemination partnership and technology transfer, the Healthy Body Image Program and variations of "Student Bodies" were licensed to a private company, Lantern, which now provides the programs under its name. SB-ED aims to reduce disordered eating behaviors (eg, restrictive eating, binge eating, compensatory behaviors), improve body image, and support the development of effective coping skills. It includes daily sessions based on motivational principles of Cognitive Behavioral Therapy (CBT), Motivational Interviewing, and the Fogg Behavior Model for Persuasive Design (FBM) [23]. Additionally, we used the Supportive Accountability Model of guidance in Internet interventions [24] for coach-related motivational design.

SB-ED includes personal one-on-one in-app and phone-based coaching and is accessible via a mobile app (see Figure 1) and Web-based program. The user can connect with a personal coach through an introductory phone call and unlimited in-app messaging (see Figure 2). The platform includes clinical management, risk management, and quality assurance tools to support effective coaching. SB-ED includes 40 sessions lasting approximately 10 minutes each which, in the context of research, are accessible for 8 weeks. The sessions consist of a daily check-in to track eating habits and compensatory behaviors over the last 24 hours, other self-monitoring tools, psychoeducational learnings, interactive multimedia tools (eg, audioguided exercises, interactive tools), and CBT techniques.

Figure 1. Mobile app of the Student Bodies–Eating Disorders program.
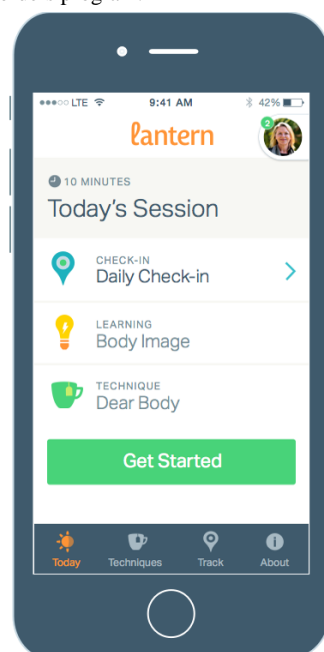
Figure 2. Integrated messaging function of the Student Bodies–Eating Disorders program.
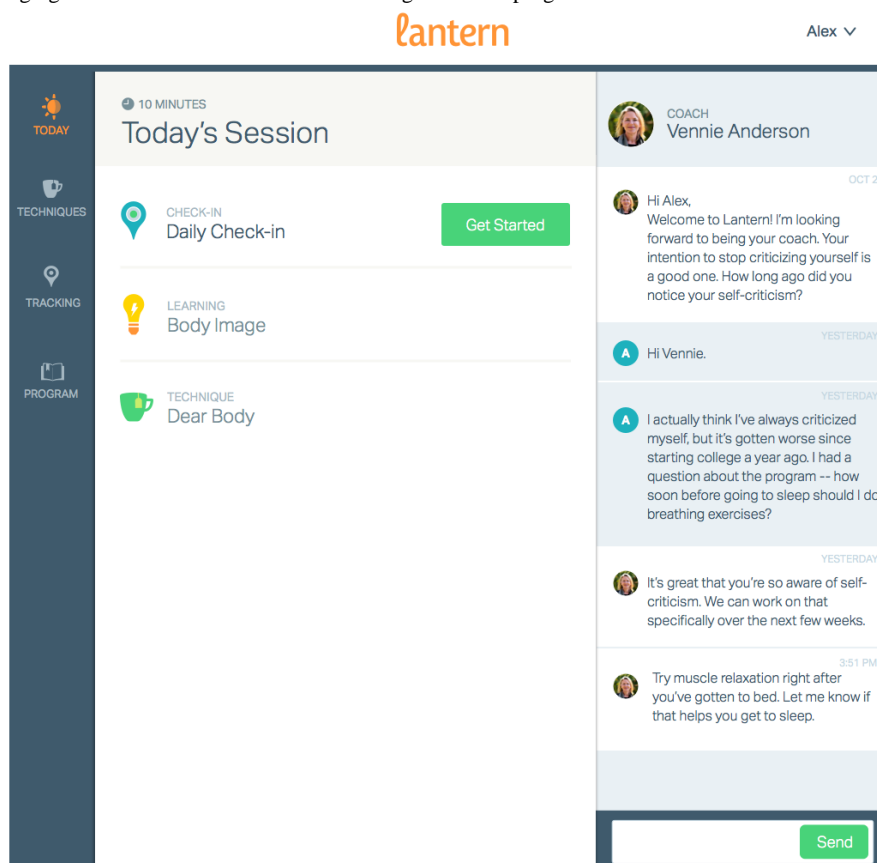
Data Collection and Analysis

Think-Aloud Task and Semistructured Interview

In each individual usability testing session, following informed consent, participants were first asked to think aloud while using the program. The think-aloud procedure involved two tasks: (1) completion of the online assessment and (2) completion of the first session of the SB-ED program. The procedure was pretested on an individual, who is not included in the 9 potential users. Participants were given an opportunity to practice the think-aloud technique by completing an Internet-based task, which was unrelated to the SB-ED program, before the usability testing started. The moderator guided the participant through the testing session by presenting the tasks and interrupted the process only if the interviewee appeared to be having difficulties thinking aloud, such as by prompting, "Tell me what you're thinking," "What are you looking at?," or "What's on your mind?" An observer recorded all comments and problems that the participants encountered. Additionally, video-analytic software (ScreenFlow) was used to capture the computer and mobile phone screen display as well as the verbal and non-verbal reactions of participants.

Next, a semistructured interview was conducted and audiotaped in order to explore important issues regarding usability and engagement. The research team developed a semistructured interview guide that included questions about the participants' experience with the transition from the assessment to the program, general navigation issues, interaction with the coach, motivation for completing the program, and whether the user would recommend the program to others.

The think-aloud tasks and interviews were transcribed verbatim including the non-verbal reactions of the participants according to the video recordings and the notes of the observers. We coded and organized the transcripts using Atlas.ti and Excel software. Since our main aim was to identify all emerging issues and the relations between the themes, we applied thematic analysis [25]. Two researchers coded the transcripts independently by identifying themes and their relevant characteristics (categories). The themes and categories were discussed and a coding framework was created, reviewed, and interpreted by the research team. In this research context, thematic analysis seemed to be the most suitable method to combine and analyze both the think-aloud tasks and the semistructured interview.

Questionnaire

Finally, participants were asked to fill out a questionnaire, which included the System Usability Scale (SUS) [26], which is a standardized 10-item Likert scale questionnaire to assess a system's usability with 5 response options ranging from "5=strongly agree" to "1=strongly disagree" (see Multimedia Appendix 1). The participant's scores for each item need to be converted to a new number (for items 1, 3, 5, 7, and 9, the score contribution is the scale position minus 1, and for items 2, 4, 6, 8, and 10, the contribution is 5 minus the scale position), summed, and then multiplied by 2.5 to convert the scores to 0-100 [26]. A SUS score of above 68 points would be considered as above average [27]. Additionally, participants were asked questions on sociodemographic and Internet use characteristics as well as psychological treatment questions.

Following a mixed-methods approach, the SUS questionnaire data were used to validate and complement the qualitative results. Questionnaire data were analyzed in Microsoft Excel using descriptive statistics. Due to missing values, the questionnaire data of one participant had to be excluded from the analysis.

Results

Participants

According to our inclusion criteria, all 9 participants were women aged 18-25 years. Due the prescreening phase, all participants gave a positive response to 2 or more of 5 questions from the SCOFF questionnaire [19], thus indicating a possible eating disorder diagnosis. One participant stated she had been given an eating disorder diagnosis and currently received psychological treatment. The sample consisted of 3 high school students, 3 participants with a Bachelor's degree, one participant with a Master's degree, one with a college degree, and one participant did not fill out this particular section of the questionnaire.

Questionnaire

On average, participants were satisfied with the overall usability of the program, which resulted in an average SUS score of 77.5

for the first test round and improved to 83.1 out of 100 points for the second test round. Both mean values are above the general average SUS score of 68 points [27]. The SUS is not designed to interpret individual items [26], thus, only the aggregate scores presented in Multimedia Appendix 1 were considered in the analysis.

Think-Aloud Task and Semistructured Interview

Despite the fact that not all improvements were implemented between the test runs, the rating of the interview question about recommending the program to others improved, from 6.5 for the first round to 7.9 out of 10 points for the second round (means based on verbal rating: 0=not at all likely, 10=extremely likely).

The following analysis of the think-aloud task and the semistructured interview revealed more nuanced results. We identified five central themes consisting of several categories, which are relevant in terms of engagement and usability, across all tested program stages (assessment, transition, Session 1) and participants. Whereas the themes layout, navigation, and content point to the usability aspect of the program (see Table 1), the themes support and engagement conditions mainly focus on user engagement issues.

Table 1. Major usability issues and resulting changes.

Central theme	Problem description	Resulting changes
Navigation	There was confusion about how to contact the coach.	Information on how to contact the coach and the possibility of contacting the coach before program start was added to the introduction of Session 1
	Symptom self-report scales with values were not clear enough and selection buttons did not function.	Values were made more apparent and technical problems were fixed.
	Confusion about how to revisit content or check past entries.	Icons were changed to make functions more obvious.
	Difference between program and assessment tool was not clear.	Information regarding the procedure and the provider was added in the recruiting email and the introduction of the assessment tool.
	Participants felt that typing longer texts on the mobile phone is not convenient.	One exercise including typing longer texts was replaced with an interactive motivational enhancement exercise.
Content	Participants raised doubts regarding the wording of the assessment results section.	More sensitive and tentative wording was used.
	Participants did not like the exercise "Dear Thighs."	Title was changed to "Dear Body," which had a general impact on engaging in the exercise.
	Questions were too long and/or hard to answer.	No changes possible since assessment is based on standardized instruments.
Layout	Answering format did not suit participants' needs.	No changes possible since assessment is based on standardized instruments.

Layout

The majority of participants liked the layout of the program and described it as "friendly," "youthful," and "pretty," similarly emphasizing the "nice colors and graphics" as well as the "easy format." A few users mentioned that the interface seemed familiar to them and that it looked like a start-up, which was interpreted positively by some, and negatively by others, as it seemed to "be just another algorithm." Additionally, a few

skeptical comments were made, most of which concerned the questions included in the assessment, which did not suit the needs of some participants. For example, one participant suggested providing an open-ended answering format instead of a closed one.

Navigation

The navigation of the program was described positively for its simplicity, intuitivism, interactivity, and guidance. A few

XSL•FO

RenderX

negative comments were made concerning select technical issues, for example, symptom self-report scales with values that were not clear enough or selection buttons that did not function. There was also some confusion about certain functions, such as how to revisit content, check past entries, or contact the coach in the first round of testing. The critical issue about how to contact the coach was improved before the second round by introducing the coach and the possibility of contact before the program start, after which the remaining participants had no problems finding and contacting the coach.

Comments concerning the transition from the assessment tool to the actual program were mixed. Some participants found the transition easy and clear, while others thought that the difference between the program and the assessment tool was not clear enough. This issue was improved before the second round by adding an explanation regarding the process and the provider in the recruiting email and in the introduction of the assessment tool.

In contrast to the first test round with the computer, some specific problems regarding typing longer text passages emerged during the second test round with the mobile app. Some participants mentioned that typing on the mobile phone is not as convenient and would take longer. One participant even "felt like the language was stifled on the phone," which might influence users' motivation for the program. As a consequence, an exercise that included typing longer texts was removed and replaced with an interactive motivational enhancement exercise.

Content

Participants had a general positive impression of the program and liked the repetitive encouragement provided, the "holistic" approach referring to the CBT approach, and the focus on "positive psychology" and "self-awareness." In terms of specific content, participants had concerns about the questions on the assessment. Some participants criticized them for being too long or for being unclear regarding reference points or definitions. In spite of this, participants seemed to recognize the questions as important.

Participants' views were especially mixed about information on the topic of "body image." Some participants described it as "educating" and "promising" and identified with the program content related to body image. Others mentioned that it "reads like any other eating disorder website" and that they would not read it since they felt they knew everything about body image already. The journal feature to track daily eating behavior was mainly seen positively, since "there is no calorie counting" and it "teaches [users] to track in a healthy way." Additionally, the use of a CBT approach was seen as positive and participants described it as "fancy" and "fascinating."

Two major issues resulted from the first test round and were changed before the second test round. Some participants raised doubts regarding the text passage showing the results of the assessment tool. They felt "shocked," "concerned," or "scared," and they expressed "moments of unease." They suggested avoiding strong or serious language and instead giving the user more personal and tentative feedback. As a consequence, we used more sensitive language for the assessment results. After

the wording had been changed and information had been added that the survey was not meant as a diagnostic tool, most participants in the second round agreed with their results and stated that they found it "useful" and "helpful to hear that you need help." However, some participants noted that there was "too much" text on the assessment results. Second, unfavorable wording seemed to be the issue with the exercise "Dear Thighs," which prompted participants to write a letter to a particular body part. Some participants were turned off by the exercise, saying that it felt "weird," "awkward," "crazy," and "cheesy", "like they were in middle school" and that they just wanted it to be over. However, when the exercise was renamed to "Dear Body," there seemed to be a dramatic change in participant reaction and most participants found the exercise "extremely helpful," "powerful," and "clever." Other minor wording issues occurred throughout the test: participants mentioned that some information was missing or that some words would benefit from an additional definition (eg, diet, meal restriction, peer).

Support

During the test, support in general, and more specifically, the support of the coach turned out to be major issues. All participants liked the idea of having a coach and perceived that the coach would be there to advise, help, and motivate them. This impression was made without participants' engaging in an active conversation with the coach (as this was not possible in the prototype they tested). There was confusion about the method of user-coach communication. With this knowledge, after the first test round, the concept and role of the coach was further clarified (see the details on navigation below).

Although participants liked the idea of the coach in general, they expressed mixed feelings about having the support of an online coach. One person mentioned that "some people prefer interaction with the computer" because they might feel embarrassed when talking to a real person about their problems. Other participants mentioned that it is more "convenient," "accessible for everyone because technology is omnipresent," and also "important, because other resources on campus are scarce." On the other hand, the majority of the participants mentioned that they would prefer to work with a coach in person rather than online or on the phone and that "online coaching can just be a support for personal counseling" indicating a belief that online coaching would not be sufficient on its own.

Besides the importance of the coach, some participants also highlighted the value of community support. Some participants mentioned interaction with other users who had already finished the program as being a great motivator, which may be integrated in the future.

Engagement Conditions

For engagement, intrinsic motivation and external motivators (eg, program features) were highly relevant. In terms of intrinsic motivators, the severity of disease seemed important. Some participants raised doubts that the "program is not for people with serious eating disorder issues," the "matter needs to be treated more seriously," and that the program would need "to be more interactive to tackle the complex issue of eating disorders." On the other hand, the program was seen "as a good

start for people without an official diagnosis." Other intrinsic motivators were curiosity, fun, or doing it for their family.

Time was identified as another major factor influencing likelihood of use. Generally, most participants were skeptical about the time they had to invest, stating that a daily commitment of 10 minutes was too long. They felt that it would be "too intensive for students" because they perceived that students really just "want a quick fix for their problem." Moreover, some participants mentioned that they felt like stopping during the assessment or the exercise "Dear Thighs" since it felt too long or it made them feel uncomfortable. One person who was undergoing therapy for an eating disorder stated that 10 minutes would not be enough to tackle her problem. Conversely, some participants found the procedure and the exercises engaging and the overall time commitment was fine.

Another important factor in terms of finishing the program was the question of trust, which was challenging in an online format where there was initial confusion about the nature and role of the remote coach. However, credibility in terms of having already heard about the program or its developers had a positive impact. Additionally, information about the program developers was added in the recruiting email before the second test round.

Effectiveness or success was seen as a central motivator because participants mentioned that they would finish the program if they were to see "improvements," "results," or a "gain in health." For instance, it was expected that the program or the coach would provide external motivation by integrating daily reminders, motivational pop-ups, affirmations, tips and advice, and detailed and customized feedback. Some participants also seemed to be impressed by the number of students enrolled in the program, the rate of symptom reduction (50%) in disordered eating behavior after completing the program, and the research background, which was outlined on the registration page. Anonymity and privacy were other factors that were highlighted positively and mentioned as relevant for program completion. Hence, no issues were raised regarding the confidentiality of the data. In addition to the program's affordability, additional incentives were discussed, such as "a gift card together with personal commitment would seal the deal" and suggestions that program use should be compensated with class credits.

In some cases, it was also misread as a fitness or weight management tool. More information was needed about how the assessment questions relate to the program personalization and other program abilities, such as customization of questions or if the program "learns" from the users' answers and "how the program gets to know the people." Some participants also indicated that they had not read or paid attention to the introduction text.

Participants reported mixed feelings about using the program on their mobile phone or via the computer, yet the majority of the participants said that they would rather use it on their mobile phone as a mobile app or a widget. However, participants also mentioned some possible pitfalls to this, such as "it is tiring to read long texts on the phone," that "texts are rather skimmed than read," and that "less scrolling would be better." Whereas the majority of participants described the mobile app as convenient, some participants also said that they would probably forget to use the app and others mentioned that they would prefer to use it on both the computer and mobile phone.

Discussion

Principal Findings

This study aimed to assess and improve usability and engagement aspects of a guided online self-help program for improving body image and reducing disordered eating symptoms. The difference between the first and the second round of testing shown in the SUS as well as in the think-aloud and interview section clearly highlights the value and reliability of performing an iterative usability study using a mixed-methods approach. This study's findings also further support the importance of incorporating usability and feasibility studies as part of the digital health intervention design process [28-30]. Usability testing is a valuable and effective method for executing a user-centered design process, illuminating end-user needs and perceptions, and facilitating intervention adaptation prior to a broad implementation.

This study found a need for intervention improvement in five major areas: layout, navigation, content, and support and engagement conditions. Regarding content, wording and language used was an important issue as it was found to trigger negative emotions in the first iteration (eg, when reading the results of the assessment and the title of the "Dear Thighs" technique). Choosing the right design, wording, and developing language in a user-centered and participatory design process is critical and may have a significant impact on engagement [31].

In terms of assessment, a number of participants had concerns about the assessment logic and assessment items. Because the items were derived from standardized instruments, it was not possible to change question text. However, other simple changes in design resulted in immediate improvement. Assessment results were presented to highlight "strengths" and "challenges" (and not diagnoses) and were intended to help participants gain perspective about their need for help and the apparent urgency and severity of their needs. The feedback provided to participants aimed to help them make an informed and empowered decision about whether the online intervention was appropriate or if they would be better served by seeking in-person evaluation and therapy (information with referral information was provided). The participant feedback on the standardized questions reveals a difficult challenge of using evidence-based and psychometric sound instruments for online assessments. Few research assessments are constructed on the basis of being user friendly, and many are developed using highly educated populations (eg, college students). In the future, assessments should be developed with consideration of how they might be used with digital programs, and text should be written at an average reading level to improve accessibility.

Concerning navigation, a major issue was that it was not obvious to participants how to contact the coach. This finding was important due to the central role of the coach and prompted addition of information and design change to help participants understand the coach's role and how to contact their coach.

For engagement, the identified main themes center on motivation, ability (simplicity), and triggers as outlined in the FBM [23]. The anticipation that the coach provides external motivation by setting reminders or affirmations refers to the Supportive Accountability Model [24]. The coach also played an important role in promoting engagement as indicated by participants, who mentioned that the support of the coach would be a primary motivator to finish the program. This is consistent with research confirming that the inclusion of professional therapist or "technician" support improves program adherence [5,8]. In this context, the participant's assumption that the coach would be a virtual coach was crucial for the research team, especially since the majority of the participants mentioned that they would prefer to work with a coach in person.

In terms of motivators, the users' symptom severity or specific diagnosis seemed important, as it refers to the core motivator "pleasure/pain" in the FBM [23], indicating that the need for seeking help rises according to the personal pain or severity level of the disease. In this study, some participants doubted the program's effectiveness in relation to their specific needs and expressed a desire for professional face-to-face therapist support. The preference for specialist treatment is in line with other recent research findings suggesting that individuals treated for anorexia nervosa prefer health professionals with high professional communication skills and an adequate knowledge of eating disorders [32]. Interestingly, Dölemeyer et al [2] found that studies exclusively enrolling participants with binge eating behaviors showed relatively low dropout rates, assuming that the motivation of this patient group is relatively high due to high psychological impairment and other related health problems. Consistent with Dölemeyer et al [2], previous studies using previous versions of the Healthy Body Image Program have also shown low dropout among participants with subclinical eating disorder symptoms and strong effects for participants with binge eating [13-17].

Some participants also seemed to be impressed by the success of other users, for example, the number of students enrolled in previous versions of the program so far and the rate of symptom reduction (50%), which was outlined on the registration page. "Seeing results" such as a reduction of disordered eating symptoms in their own lives was seen as another core motivator, which relates to the dimensions "hope/fear" in the FBM [23], which is characterized by the anticipation of an outcome. The Motivational Interviewing and Cognitive Behavioral theories also posit that expectancies influence engagement and outcome. In CBT, this originates from Bandura's Social Learning Theory [33], which, applied to this study, could explain why testimonials improve positive expectancies. Belief in the program effectiveness and success is important for enhancing engagement and preventing dropout [10]. In the future, information from the assessment data should be customized better to the user by noting how the program has worked with individuals with similar scores.

Time, which is described as an important element of simplicity in the FBM [23], also emerged as a potential factor in engagement, since some participants viewed 10 minutes of daily use as too much time. However, the possibility of using the mobile app was seen as more convenient, and previous research has highlighted convenience as an important criterion for the use of digital health interventions [34].

Overall, the results of this study are consistent with general characteristics of digital health interventions for behavior change and self-management suggested by Murray [35]. These interventions need a strong theoretical foundation, perceived personal relevance to the user, perceived effectiveness, tailoring, persuasive technologies, credibility, social networking, and regular "push factors" (including human support or periodic prompts) in order to increase adherence.

Strengths and Limitations

Although the concepts of usability and engagement are inextricably linked, adequate and standardized methods to investigate issues of engagement are scarce. Thus, the strength of this study is that it demonstrates that engagement issues can be investigated within the scope of a usability study. This research design proved to be effective in identifying a range of issues for improvement and facilitating measurable program improvement prior to implementation. Ultimately, the conduction of usability studies fits with the general demand for alternative methods to evaluate behavioral intervention technologies. Traditional evaluation methodologies such as randomized controlled trials are not compatible with fast-changing customer expectations and rapid technology advancement, which demand less time-intensive methods [36].

One major limitation of this study is the possibility that participants were influenced by the study situation itself, since the task of thinking aloud and simultaneously being observed might have provoked unintended reactions or statements. However, we tried to reduce this possible bias by practicing the think-aloud method with each participant prior to beginning the actual study. In three cases, the study situation might have been influenced because the tests were held in public places chosen by the participants. We tried to reduce possible influences by ensuring that there were no people in close vicinity, so that participants felt comfortable speaking openly. Another limitation was that the participants did not get individualized results based on their answers to the assessment questions since this was not possible at this stage of the development process of the program but instead a standardized results page according to their initial SCOFF results. This might have caused discrepancies regarding their expected results and thus led to negative statements about the wording of the results page. Another limitation is that participants tested a prototype rather than the actual program, so they could not use or test many of the program functionalities.

Conclusions

Despite the limitations, this usability study allowed us to improve and refine our guided online SB-ED program prior to its launch by making changes based on our target group's concerns and preferences. Our main findings regarding usability and engagement issues of online health programs are fairly consistent with prior research findings of similar studies, suggesting that this was a reliable and effective research method. The true advantage of conducting small-scale usability studies is evident in their ability to reveal specific program issues from the perspective of the target population in the implementation

phase and at the same time contribute to larger research-based insights. Usability studies of programs incorporating online assessments or questionnaires need to pay attention to standardized question items, which cannot easily be adapted to user needs and thus can highly interfere with usability and engagement aspects. Accordingly, future usability and engagement research for different stages of digital health program use is needed in order to identify general usage and adherence patterns, which can ultimately help improve program adherence and reduce dropout.

Acknowledgments

Conflicts of Interest

MJ and LG are employed at Lantern, which is a for-profit digital health company. However, they did not have any contact with participants, data, and were not involved in analysis and thus could not in any way bias results.

Multimedia Appendix 1

Individual SUS scores.

[XLSX File (Microsoft Excel File), 12KB - jmir_v18i1e7_app1.xlsx]

References

1. Serdar K, Kelly NR, Palmberg AA, Lydecker JA, Thornton L, Tully CE, et al. Comparing online and face-to-face dissonance-based eating disorder prevention. Eat Disord 2014;22(3):244-260. [doi: 10.1080/10640266.2013.874824] [Medline: 24456277]

2. Dölemeyer R, Tietjen A, Kersting A, Wagner B. Internet-based interventions for eating disorders in adults: a systematic review. BMC Psychiatry 2013;13:207 [FREE Full text] [doi: 10.1186/1471-244X-13-207] [Medline: 23919625]

3. Shingleton RM, Richards LK, Thompson-Brenner H. Using technology within the treatment of eating disorders: a clinical practice review. Psychotherapy (Chic) 2013 Dec;50(4):576-582 [FREE Full text] [doi: 10.1037/a0031815] [Medline: 23527906]

4. Schlegl S, Bürger C, Schmidt L, Herbst N, Voderholzer U. The potential of technology-based psychological interventions for anorexia and bulimia nervosa: a systematic review and recommendations for future research. J Med Internet Res 2015;17(3):e85 [FREE Full text] [doi: 10.2196/jmir.3554] [Medline: 25840591]

5. Aardoom JJ, Dingemans AE, Spinhoven P, Van Furth Eric F. Treating eating disorders over the internet: a systematic review and future research directions. Int J Eat Disord 2013 Sep;46(6):539-552. [doi: 10.1002/eat.22135] [Medline: 23674367]

6. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de VH, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]

7. 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]

8. Beintner I, Jacobi C, Schmidt UH. Participation and outcome in manualized self-help for bulimia nervosa and binge eating disorder - a systematic review and metaregression analysis. Clin Psychol Rev 2014 Mar;34(2):158-176. [doi: 10.1016/j.cpr.2014.01.003] [Medline: 24508686]

9. von Brachel R, Hötzel K, Hirschfeld G, Rieger E, Schmidt U, Kosfelder J, et al. Internet-based motivation program for women with eating disorders: eating disorder pathology and depressive mood predict dropout. J Med Internet Res 2014;16(3):e92 [FREE Full text] [doi: 10.2196/jmir.3104] [Medline: 24686856]

10. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]

11. O'Brien HL, Toms EG. Measuring engagement with search systems using the User Engagement Scale (UES). Inf Process Manage Sep 2013;49(5):1092. [doi: 10.1016/j.ipm.2012.08.005]

12. International Standards Organization (ISO). Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability. Geneva: International Organization for Standardization; 1998.

13. Jacobi C, Völker U, Trockel MT, Taylor CB. Effects of an Internet-based intervention for subthreshold eating disorders: a randomized controlled trial. Behav Res Ther 2012 Feb;50(2):93-99. [doi: 10.1016/j.brat.2011.09.013] [Medline: 22137366]

14. Beintner I, Jacobi C, Taylor CB. Effects of an Internet-based prevention programme for eating disorders in the USA and Germany--a meta-analytic review. Eur Eat Disord Rev 2012 Jan;20(1):1-8. [doi: 10.1002/erv.1130] [Medline: 21796737]

XSL•FO

RenderX

15. Völker U, Jacobi C, Barr TC. Adaptation and evaluation of an Internet-based prevention program for eating disorders in a sample of women with subclinical eating disorder symptoms: a pilot study. Eat Weight Disord 2011 Dec;16(4):e270-e273. [Medline: 22526132]

16. Jones M, Luce KH, Osborne MI, Taylor K, Cunning D, Doyle AC, et al. Randomized, controlled trial of an internet-facilitated intervention for reducing binge eating and overweight in adolescents. Pediatrics 2008 Mar;121(3):453-462. [doi: 10.1542/peds.2007-1173] [Medline: 18310192]

17. Taylor CB, Bryson S, Luce KH, Cunning D, Doyle AC, Abascal LB, et al. Prevention of eating disorders in at-risk college-age women. Arch Gen Psychiatry 2006 Aug;63(8):881-888 [FREE Full text] [doi: 10.1001/archpsyc.63.8.881] [Medline: 16894064]

18. Sánchez-Ortiz VC, House J, Munro C, Treasure J, Startup H, Williams C, et al. "A computer isn't gonna judge you": a qualitative study of users' views of an internet-based cognitive behavioural guided self-care treatment package for bulimia nervosa and related disorders. Eat Weight Disord 2011 Jun;16(2):e93-e101. [Medline: 21989103]

19. Morgan JF, Reid F, Lacey JH. The SCOFF questionnaire: assessment of a new screening tool for eating disorders. BMJ 1999 Dec 4;319(7223):1467-1468 [FREE Full text] [Medline: 10582927]

20. Nielsen J. Iterative user-interface design. Computer 1993 Nov;26(11):32-41. [doi: 10.1109/2.241424]

21. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. 1993 Presented at: ACM INTERCHI'93 Conference; 1993; Amsterdam, Netherlands p. 24-29. [doi: 10.1145/169059.169166]

22. Jones M, Kass AE, Trockel M, Glass AI, Wilfley DE, Taylor CB. A population-wide screening and tailored intervention platform for eating disorders on college campuses: the healthy body image program. J Am Coll Health 2014;62(5):351-356 [FREE Full text] [doi: 10.1080/07448481.2014.901330] [Medline: 24621000]

23. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. New York, NY: ACM; 2009 Presented at: International Conference on Persuasive Technology; 2009; Claremont, CA p. 26-29.

24. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

25. Froschauer U, Lueger M. Das qualitative Interview. 1st edition. Vienna: Facultas; 1993.

26. Brooke J. System usability scale (SUS): a quick-and-dirty method of system evaluation user information. Reading, UK: Digital Equipment Co Ltd; 1986.

27. Sauro J. A practical guide to the System Usability Scale: Background, benchmarks, & best practices. Denver, CO: Measuring Usability LLC; 2011.

28. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]

29. Stinson J, Gupta A, Dupuis F, Dick B, Laverdière C, LeMay S, et al. Usability testing of an online self-management program for adolescents with cancer. J Pediatr Oncol Nurs 2015;32(2):70-82. [doi: 10.1177/1043454214543021] [Medline: 25037173]

30. Breakey VR, Warias AV, Ignas DM, White M, Blanchette VS, Stinson JN. The value of usability testing for Internet-based adolescent self-management interventions: "Managing Hemophilia Online". BMC Med Inform Decis Mak 2013;13:113 [FREE Full text] [doi: 10.1186/1472-6947-13-113] [Medline: 24094082]

31. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]

32. Gulliksen KS, Nordbø RS, Espeset ES, Skårderud F, Holte A. The process of help-seeking in anorexia nervosa: patients' perspective of first contact with health services. Eat Disord 2015;23(3):206-222. [doi: 10.1080/10640266.2014.981429] [Medline: 25412397]

33. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.

34. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]

35. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2.0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]

36. Mohr DC, Cheung K, Schueller SM, Hendricks BC, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]

Abbreviations

CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
FBM: Fogg Behavior Model for Persuasive Design

HBI: Healthy Body Image Program
SB-ED: Student Bodies–Eating Disorders
SCOFF: Sick, Control, One stone, Fat, Food
SUS: System Usability Scale

XSL•FO
RenderX

Original Paper

Impact of a Mobile Phone Intervention to Reduce Sedentary Behavior in a Community Sample of Adults: A Quasi-Experimental Evaluation

Darla E Kendzor[1,2*], PhD; Kerem Shuval[3*], PhD; Kelley Pettee Gabriel[4], PhD; Michael S Businelle[1,2], PhD; Ping Ma[5], PhD; Robin R High[6], MA, MBA; Erica L Cuate[7], MPH; Insiya B Poonawalla[7], PhD; Debra M Rios[7], MS; Wendy Demark-Wahnefried[8], PhD, RD; Michael D Swartz[9], PhD; David W Wetter[10], PhD

[1]Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[2]Oklahoma Tobacco Research Center, Stephenson Cancer Center, Oklahoma City, OK, United States

[3]Intramural Research Department, Economic and Health Policy Research Program, American Cancer Society, Atlanta, GA, United States

[4]School of Public Health, Department of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Science Center, Austin, TX, United States

[5]Department of Clinical Sciences, The University of Texas Southwestern Medical Center, Dallas, TX, United States

[6]College of Public Health, Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, United States

[7]School of Public Health, University of Texas Health Science Center, Dallas, TX, United States

[8]Department of Nutrition Sciences, The University of Alabama at Birmingham, Birmingham, AL, United States

[9]School of Public Health, Department of Biostatistics, The University of Texas Health Science Center, Houston, TX, United States

[10]Department of Psychology, Rice University, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Darla E Kendzor, PhD
Department of Family and Preventive Medicine
University of Oklahoma Health Sciences Center
655 Research Parkway, Suite 400
Oklahoma City, OK, 73104
United States
Phone: 1 405 271 8001 ext 50478
Fax: 1 405 271 2808
Email: Darla-Kendzor@ouhsc.edu

Abstract

Background: Greater time spent sedentary is linked with increased risk of breast, colorectal, ovarian, endometrial, and prostate cancers. Given steadily increasing rates of mobile phone ownership, mobile phone interventions may have the potential to broadly influence sedentary behavior across settings.

Objective: The purpose of this study was to examine the short-term impact of a mobile phone intervention that targeted sedentary time in a diverse community sample.

Methods: Adults participated in a quasi-experimental evaluation of a mobile phone intervention designed to reduce sedentary time through prompts to interrupt periods of sitting. Participants carried mobile phones and wore accelerometers for 7 consecutive days. Intervention participants additionally received mobile phone prompts during self-reported sitting and information about the negative health impact of prolonged sedentariness. The study was conducted from December 2012 to November 2013 in Dallas, Texas. Linear mixed model regression analyses were conducted to evaluate the influence of the intervention on daily accelerometer-determined estimates of sedentary and active time.

Results: Participants (N=215) were predominantly female (67.9%, 146/215) and nonwhite (black: 50.7%, 109/215; Latino: 12.1%, 26/215; other: 5.6%, 12/215). Analyses revealed that participants who received the mobile phone intervention had significantly fewer daily minutes of sedentary time (B=–22.09, P=.045) and more daily active minutes (B=23.01, P=.04) than control participants.

XSL•FO
RenderX

Conclusions: A simple mobile phone intervention was associated with engaging in less sedentary time and more physical activity. Findings underscore the potential impact of mobile phone interventions to positively influence sedentary behavior and physical activity.

Introduction

Sedentary behavior has been defined as any activity that requires an energy expenditure no greater than 1.5 metabolic equivalents (METs) that is performed while sitting or reclining [1]. Exploring the health effects of sedentary behavior, independent of physical activity, has been a relatively new scientific pursuit, with a proliferation of studies published in the past decade. Evidence to date suggests that prolonged sedentary time is associated with increased risk for a variety of adverse health outcomes [2-12], including cancers of the breast, colon/rectum, ovaries, endometrium, and prostate [11-14]. In addition, greater sedentary time among adults is associated with weight gain, higher body mass index (BMI), and obesity [15-20], which is a known risk factor for cancer [21]. Nevertheless, adults in the United States are excessively sedentary with an average of approximately 8 hours per day spent sedentary during waking hours [22]. This high level of sedentary time has been observed in both men and women and across several racial/ethnic groups [22]. Emerging research has indicated that inactive-to-active transitions (henceforth "sedentary breaks") are linked with lower waist circumference, BMI, triglycerides, 2-hour glucose levels, and blood pressure [23-25]. Thus, interventions designed to reduce total sedentary behavior by interrupting prolonged sedentary bouts may have a substantial impact on health.

Despite the accumulating evidence supporting the deleterious health effects of prolonged sedentariness, few interventions have specifically focused on decreasing and interrupting sedentary time. Because engaging in habitual physical activity often requires significant effort and planning, it seems likely that modifying sedentary behavior through periodic interruptions during waking hours may be more achievable and sustainable over time. It is possible that modifying sedentary behavior represents a less complex behavior change, especially for inactive individuals [26,27]. Notably, Bond et al [28] recently reported promising initial findings demonstrating that a mobile phone intervention was associated with reductions in sedentary time among overweight/obese adults. The findings of a recent meta-analysis provide initial evidence that sedentary behavior interventions significantly reduce sedentary time [29], although most intervention studies have focused specifically on reducing occupational sitting time or screen time rather than total daily sitting time.

Recent research indicates that 64% of US adults owned mobile phones in 2015 [30], suggesting that mobile phone interventions have the potential to broadly influence sedentary behavior among adults across diverse settings. Thus, the purpose of the current study was to characterize the impact of a mobile phone-based sedentary behavior intervention that incorporated education, self-monitoring, and prompting in a community sample of adults. It was hypothesized that participants who received mobile phone prompts to decrease sedentary time would have significantly less sedentary time, more active time, and more sedentary breaks than those who did not receive mobile phone prompts over a 7-day period. In addition, it was anticipated that participants who received prompts to increase activity would be more acutely active following self-reported sitting than those who did not receive prompts.

Methods

Participants

A total of 248 adults were recruited from the Dallas metropolitan area through flyers posted on the University of Texas Southwestern campus (Dallas, TX), local advertising circulars, and word of mouth. Individuals were eligible to participate in the study if they were at least 18 years of age, possessed a valid home address and a functioning telephone number, and demonstrated greater than 6th grade English literacy level on the Rapid Estimate of Adult Literacy in Medicine (REALM) [31,32]. Of those screened, 10 were excluded because they were not able to demonstrate the minimum reading level, leaving a total study sample of 238 participants. Data collection began in December 2012 and concluded in November 2013.

Measures

Socioeconomic Status/Demographic Variables

Race/ethnicity, sex, age (in years), and educational attainment were assessed.

Body Mass Index

Participant's BMI was calculated based on objective measurements of height and weight using the standard formula (kg/m^2).

Smoking

Expired carbon monoxide (CO) levels were measured with a portable Vitalograph ecolyzer, which provided an objective indicator of current smoking status and level of smoking. CO levels of ≥8-10 parts per million (ppm) suggest recent cigarette smoking with a sensitivity and specificity of approximately 90% [33].

Sedentary Behavior and Physical Activity

Sedentary time at baseline was based on responses to 2 items from the International Physical Activity Questionnaire (IPAQ). The IPAQ assessed usual time spent sitting on a weekday and on a weekend day during the past week [34]. Weekday estimates were multiplied by 5, weekend day estimates were multiplied

by 2, and the resulting values were summed and divided by 7 to calculate the mean daily time spent sitting during the past week.

Physical activity and sedentary time were directly assessed using Actigraph GT3X (Pensacola, FL) triaxial accelerometers. Accelerometers were initialized via ActiLife6 software to begin data collection at midnight on the day of the baseline visit. Participants were instructed to wear the accelerometer on the waist and in line with their right hip, secured using an elastic belt during all waking hours for 7 consecutive days. Participants were asked to remove the accelerometer when sleeping, bathing or showering, and during all water activities. After the 7-day data collection period was completed, the monitors were returned at a second in-person visit and data were downloaded via ActiLife6 software. Before data reduction and processing, the downloaded data files were reintegrated and expressed as 60-second epochs. A 60-second epoch was used for consistency with previous research in nationally representative samples [22]. Research has shown that associations of activity estimates with key outcomes are not markedly different when shorter or longer epochs are used [35].

During the data reduction and processing stage, data were screened for periods of nonwear using established methods [36,37]. Nonwear periods were removed from further analysis. Total activity counts per day were calculated using summed daily counts detected over wear periods. Minutes spent in sedentary activity, as well as light and moderate lifestyle intensity activity were estimated using Matthews cut-points for all days with 10 hours or more per wear time [38]. Specifically, activity count ranges were 0 to 99 counts per minute for sedentary activity, 100 to 759 counts per minute for light-intensity activity, and 760 to 1951 counts per minute for moderate lifestyle intensity activity. An estimate reflecting total time spent active was also created using accumulated time ≥100 counts per minute. Sedentary breaks were defined as any period of sedentary time (ie, <100 counts/minute) that was immediately followed by a minute or more of active time (≥100 counts) [24] and sedentary breaks were summed across all waking hours.

Mobile Phone Assessments

All participants were provided with an Android mobile phone on which they were prompted to complete daily diary assessments and random assessments of health behavior and psychosocial variables (as part of a parent study; see Procedure section) over a 7-day observation period. Participants completed daily diary assessments once daily, 30 minutes after their self-reported usual wake time. In addition, participants were randomly prompted to complete assessments 4 times per day during self-reported waking hours. Participants were required to complete mobile phone assessments within 15 minutes,

although they were allowed to postpone assessments by 5 minutes for a total of 3 times. Of relevance to the current analyses, participants responded to the following daily diary and random assessment items, respectively: (1) "How many hours did you spend sitting yesterday?" and (2) "What were you doing right before your phone rang/vibrated?" Response options included sitting, talking, standing, walking/exercising, sleeping/resting, or other.

Procedure

The current study was approved by the Institutional Review Boards of the University of Texas Southwestern Medical Center and the University of Texas Houston Health Science Center. The sedentary behavior intervention described here was a post hoc addition to an observational prospective 7-day study. The parent study was designed to characterize proximal predictors of health behavior using mobile phone–based ecological momentary assessment. Thus, the current study had a quasi-experimental (nonrandomized) design whereby the first 131 consecutive participants who did not receive the intervention served as the control group. Control participants completed mobile phone assessments and wore an accelerometer to measure sedentary and active time over 7 consecutive days. The subsequent 107 participants who enrolled in the study additionally received the sedentary behavior intervention on the mobile phone. Participant recruitment and group allocation are depicted in Figure 1.

Potential participants were provided with the details of the study over the telephone and their interest in participating was assessed. Interested individuals were briefly screened by phone for eligibility and those eligible were scheduled to attend the initial study visit. The details of the study were reviewed at the first visit and informed consent was obtained from all participants. Reading level was assessed and participants who were unable to demonstrate greater than 6th grade reading level on the REALM were excluded from the study and compensated for their time with a US $20 gift card and a parking token.

Eligible participants completed study questionnaires on laptop computers. Height, weight, and CO were measured by trained staff. Participants were provided with a mobile phone and an accelerometer, instructed in their use, and asked to wear/carry the devices for 7 days. Participants received a US $50 gift card and a parking token for the completion of the baseline visit. Participants returned for a final visit and received up to US $80 in gift card compensation depending on the percentage of mobile phone assessments completed. A mobile phone assessment completion rate of 80% and the return of study mobile phones and accelerometers were required to earn the maximum compensation.

Figure 1. Participant flow for nonrandomized intervention trial.
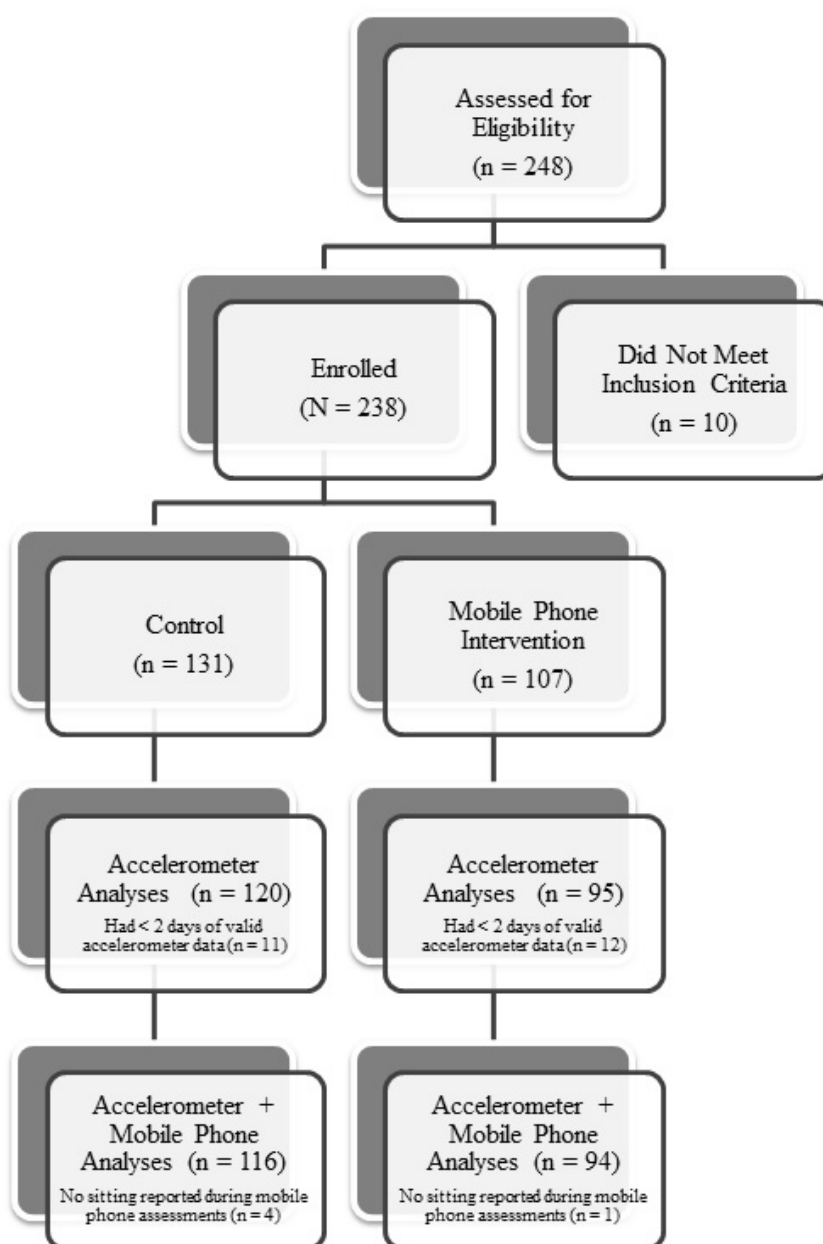
Intervention

Intervention information and prompts used in the current study were based on previously developed messages and materials [39,40]. The intervention group received a 1-page printed handout at the first visit describing the health-related importance of limiting sedentary time and increasing activity. The handout included suggestions about ways to reduce sedentary time and increase light-intensity activity throughout the day (eg, by moving around in the office). In addition, a daily message used in previous research [39] appeared on each participant's mobile phone at the end of daily diary assessment for 7 days: "Remember to STAND UP, SIT LESS, and MOVE MORE today!"

During the 7-day intervention period, participants who reported more than 2 hours of sitting during the previous day via the morning daily diary assessment received the following message: "Medical research has shown that long periods of uninterrupted sitting increase the likelihood of several health problems, including obesity and Type 2 diabetes. Make an effort to Stand Up more, Sit Less, and Move About more. This can be achieved by taking frequent standing and walking breaks (at least one break for every half hour of sitting), standing up when talking on the phone (at work or home), checking emails, etc, and replacing blocks of sitting time with standing time, such as doing household chores while watching TV." This message was adapted from previous research [39].

Participants who reported that they were sitting during any random mobile phone assessment received the following message: "Medical research has shown that long periods of uninterrupted sitting increase the likelihood of several health problems, including obesity and Type 2 diabetes. Please consider standing up *now* and moving about for 5 minutes. Make an effort to improve your health by standing up and moving around your home or office every half hour during periods of sitting." Note that both intervention and control participants completed mobile phone, questionnaire, anthropometric, and accelerometer assessments. However, only intervention participants received the education and mobile phone messages prompts.

Statistical Analyses

A series of linear mixed model (LMM) regression analyses were conducted to evaluate the influence of the sedentary behavior intervention (relative to the control group) on accelerometer-measured daily active minutes as well as daily minutes of sedentary, light, and moderate lifestyle activity over 7 days. Model 1 adjusted for daily minutes of wear time and study day. Model 2 adjusted for race (white vs nonwhite), education (≤high school vs >high school), CO level (ppm), age (in years), daily minutes of wear time, and study day (day 1-7). Total number of daily sedentary breaks was also examined as an outcome, with daily minutes of sedentary time additionally included in the models. Participants who did not have at least 2 days of accelerometer wear time of at least 10 hours per day were excluded from these analyses (n=23), leaving an analytic sample of 215 participants.

Additional analyses were conducted to compare active minutes and accelerometer counts between the groups during the 10 minutes following random mobile phone assessments where sitting was reported. A total of 5 participants did not endorse sitting during any random assessments and, therefore, the sample size was reduced to 210 participants in these analyses only. Model 1 adjusted for total random assessments completed, daily minutes of wear time, time of random assessments when sitting

was endorsed, and study day. Model 2 adjusted for race, education, CO level (smoking), age, daily minutes of wear time, time of random assessments when sitting was endorsed, and study day.

Results

Participant Characteristics

Participants (N=215) were predominantly female (67.9%, 146/215) and nonwhite (black: 50.7%, 109/215; Latino: 12.1%, 26/215; other: 5.6%, 12/215). See Table 1 for participant characteristics overall and by intervention group. Participants in the intervention group (n=95) were older, had higher CO levels, were more likely to be nonwhite, and were less likely to have completed greater than a high school education than those in the control group (n=120). Participants did not differ significantly by intervention group on self-reported mean daily sitting at baseline (mean 6.90 hours per day, SD 3.71), mean daily accelerometer wear time during the study (mean 843.63 minutes per day, SD 99.38), or days of accelerometer wear (mean 5.87 days, SD 1.45). Overall, participants completed 87.2% (mean 24.42, SD 4.15) of 28 possible random assessments via mobile phone over the 7-day study period, although control group participants had a slightly higher completion rate than those in the intervention group (mean 24.93, SD 3.35 vs mean 23.80, SD 4.90 completed assessments, *P*=.049).

Descriptive analyses of accelerometer estimates overall and by intervention group are presented in Table 2. Note that unadjusted comparisons indicate that intervention participants had significantly fewer daily accelerometer-measured sedentary minutes, spent less of their total accelerometer wear time in sedentary activity, and spent more of their daily accelerometer wear time active and engaged in light-intensity activity. Active minutes and total accelerometer counts in the 10-minute postprompt period were also greater in the intervention group.

Table 1. Baseline participant characteristics (N=215).

Participant characteristics	Total sample (N=215)	Intervention group (n=95)	Control group (n=120)	*P* [a]
Race (nonwhite), n (%)	147 (68.4)	77 (81)	70 (58.3)	<.001
Gender (female), n (%)	146 (67.9)	66 (70)	80 (66.7)	.66
Age (years), mean (SD)	43.90 (12.85)	46.75 (11)	41.65 (13.62)	.004
Education (>high school), n (%)	160 (74.4)	55 (58)	105 (87.5)	<.001
BMI (kg/m^2), mean (SD)	30.72 (7.81)	31.61 (8)	30.02 (7.90)	.14
CO level (ppm), mean (SD)	6.49 (10.18)	9.2 (11)	4.35 (8.46)	<.001

[a]Variables that were found to differ significantly between the intervention and control groups were included as covariates in adjusted analyses.

Intervention

The LMM regression analyses indicated that participants who received the sedentary behavior intervention had significantly fewer accelerometer-measured daily minutes of sedentary time and more daily active minutes over the 7-day study period than participants who did not receive the intervention in adjusted models 1 and 2 (see Table 3). Analyses indicated that those

included in the mobile phone intervention group engaged in significantly more minutes of light-intensity activity than control group participants in model 1 only. Daily minutes of moderate lifestyle intensity activity and total daily sedentary breaks did not differ significantly between groups in either model. Additional analyses indicated that intervention participants had significantly more active minutes (B=0.33, *P*=.01) and accelerometer counts (B=350.67, *P*=.01) than control

participants in the 10 minutes following random assessments where sitting was endorsed after adjustment for total random assessments completed, study day, time of random assessment when sitting, and daily minutes of accelerometer wear time.

However, when race, smoking (CO level), age, and education were added to the model, results were no longer significant for active minutes (B=.18, P=.21) or accelerometer counts (B=283.31, P=.06).

Table 2. Daily accelerometer estimates across valid wear days overall and by intervention group (N=215).

Accelerometer variables[a]	Total sample (N=215)	Intervention group (n=95)	Control group (n=120)	P
Mean daily wear time (minutes/day), mean (SD)	843.63 (99.38)	829.96 (96.28)	854.46 (100.86)	.07
Total days of observation (out of 7 possible), mean (SD)	5.87 (1.45)	5.74 (1.52)	5.98 (1.39)	.23
Sedentary, daily minutes, mean (SD)	531.64 (100.96)	507.20 (101.01)	550.99 (97.04)	.001
Sedentary, % of daily wear time, mean (SD)	62.92 (8.89)	61.08 (9.72)	64.38 (7.91)	.007
Active, daily minutes, mean (SD)	310.94 (80.59)	322.37 (88.01)	301.88 (73.30)	.06
Active, % of daily wear time, mean (SD)	36.96 (8.88)	38.87 (9.71)	35.44 (7.88)	.005
Light intensity, daily minutes, mean (SD)	215.12 (55.52)	222.83 (56.50)	209.01 (54.19)	.07
Light intensity, % of daily wear time, mean (SD)	25.54 (5.95)	26.87 (5.93)	24.49 (5.78)	.003
Moderate lifestyle intensity, daily minutes, mean (SD)	71.54 (31.07)	74.80 (35.90)	68.96 (26.51)	.17
Moderate lifestyle intensity, % of daily wear time, mean (SD)	8.53 (3.70)	9.03 (4.27)	8.14 (3.13)	.08
Inactive-to-active transitions, daily total, mean (SD)	94.17 (17.03)	94.53 (16.25)	93.88 (17.70)	.78
Total active minutes (10 minutes postprompt), mean (SD)[b]	2.44 (1.02)	2.59 (1.15)	2.32 (0.88)	.05
Total accelerometer counts (10 minutes postprompt), mean (SD)[b]	1782.07 (1096.64)	1970.61 (1322.76)	1629.29 (847.40)	.03

[a]Accelerometer estimates were defined as follows: sedentary activity was defined as less than 100 counts per minute, active time was defined as 100 or more counts per minute, light-intensity activity was defined as 100-759 counts per minute, and moderate lifestyle intensity activity was defined as 760-1951 counts per minute. An inactive-active transition (ie, sedentary break) was defined as a transition from less than 100 counts to 100 or more counts/minute.

[b] Sample size slightly reduced (N=210) because 5 participants had no reports of sitting during random mobile phone assessments.

Table 3. Effects of a mobile phone intervention on accelerometer-measured activity over 7 days (N=215).[a]

Accelerometer variables[b]	Model 1[c]		Model 2[d]	
	Unstandardized coefficient	P	Unstandardized coefficient	P
Sedentary, daily minutes	−27.33	.007	-22.09	.045
Active, daily minutes	28.52	.005	23.01	.04
Light intensity, daily minutes	18.94	.005	11.73	.10
Moderate lifestyle intensity, daily minutes	7.85	.06	7.14	.12
Inactive-to-active transitions, daily total[e]	3.06	.09	1.15	.56

[a]In the analyses, no intervention=0 and intervention=1.

[b] Accelerometer estimates were defined as follows: sedentary activity was defined as <100 counts per minute, active time was defined as ≥100 counts per minute, light-intensity activity was defined as 100-759 counts per minute, and moderate lifestyle intensity activity was defined as 760-1951 counts per minute. An inactive-active transition (ie, sedentary break) was defined as a transition from <100 counts to ≥100 counts/minute.

[c] Adjusted for daily minutes of accelerometer wear time and time/day.

[d] Adjusted for race, education, CO level, age, daily minutes of accelerometer wear time, and time/day.

[e] Daily minutes of sedentary time was additionally included in the models.

Discussion

The current study was among the first to evaluate a mobile phone intervention aimed at reducing sedentary behavior among

adults of diverse racial/ethnic backgrounds. Findings indicated that intervention participants had significantly fewer minutes of daily sedentary time and more daily minutes of active time than controls over the 7-day study period. Daily minutes of light-intensity activity was significantly higher among

intervention participants than those assigned to the control group in the partially adjusted model, although differences did not reach statistical significance in the fully adjusted model. Additionally, supplementary analyses indicated that activity was greater in the 10 minutes following self-reported sitting among intervention participants who received activity prompts than among control participants who did not receive prompts, although differences did not reach significance in the fully adjusted models. Overall, simple mobile phone prompts appear to be a promising strategy for reducing sedentary behavior and increasing activity, although adequately powered and well-designed studies will be needed to confirm these preliminary findings.

Intervention participants evidenced 3% less objectively measured sedentary time (of total accelerometer wear time) than control participants. To illustrate, 3% of 14 hours of mean wear time equals 25 minutes of time spent engaged in activity rather than in sedentary behavior. Differences in sedentary time noted in the current study are similar to the reductions reported with other types of sedentary intervention strategies [41,42]. Bond et al [28] specifically evaluated a mobile phone-based intervention using a within-subjects design and showed 3.3% to 5.9% decreases in sedentary time and 1.9% to 3.9% increases in light physical activity across 3 variations of the intervention. Although it is not certain whether these reductions in sedentary time (and increases in activity) have a significant impact on health, it is notable that differences in the current study were found using a very simple intervention which entailed (1) a printed handout, (2) a mobile phone reminder to "stand up, sit less, and move more" each morning, and (3) mobile phone prompts triggered by self-reported sitting several times daily. It is unclear why the intervention did not seem to impact sedentary breaks, although one possible explanation may be that the intervention was not intensive enough. Mobile phone interventions that are more intensive and those that use prompts based on real-time activity monitoring may have a greater impact on behavior.

Utilizing mobile phones to modify sedentary behavior is advantageous because phones can be used in most settings where individuals are sedentary, such as in the home or workplace [43]. In addition, rates of mobile phone ownership are steadily increasing, with the majority of US adults reporting that they owned a mobile phone in 2015 [30]. Although mobile phone interventions have the potential to broadly influence lifestyle behaviors, they have not been widely employed to modify sedentary behavior among adults. To date, most sedentary behavior interventions have focused primarily on reducing occupational sitting through the introduction of sit-stand desks and encouragement to use them throughout the day [39,44]. Although the workplace is an important place to target sedentary behavior, mobile phones offer the potential to influence sedentary behavior across settings where sedentary behavior is likely to take place. Plausibly, substantial reductions in sedentary behavior may reduce the likelihood of developing cancer and other diseases.

Study findings complement the findings of Bond et al [28], who tested 3 versions of a mobile phone intervention that provided feedback about time spent in objectively measured sedentary

behavior and prompted activity breaks following periods of continuous sedentariness in a sample of overweight/obese adults who were predominantly white and female. Notably, all versions of the mobile phone intervention were associated with within-subjects decreases in sedentary time, although mobile phone prompts to engage in 3 minutes of activity after 30 minutes of sedentary time were associated with greater reductions in sedentary time than prompts to engage in 12 minutes of activity after 120 minutes of sedentary time. Future research will be needed to determine whether mobile phone interventions have a sustained impact on sedentary behavior and to determine the optimal scheduling of prompts in longer-term interventions.

This study has notable strengths and limitations. Strengths include the application of novel technology that is scalable and has the potential to modify sedentary behavior across settings. In addition, although previous intervention research has primarily focused on reducing occupational sitting [39,44,45], the current intervention emphasized reducing sitting and promoting activity among adults across settings. Additionally, accelerometers were used to provide an objective measure of sedentary and active time; however, the accelerometers used in the study are not able to differentiate between sitting and standing. Because the mobile phone intervention messages encouraged both standing and moving around, it is noteworthy that the accelerometers used in this study did not have the capability of capturing increased time spent standing unless individuals were also moving around. Thus, it is possible that the impact of the mobile phone intervention on sedentary behavior was underestimated. It is also notable that mobile phone prompts to interrupt sedentary behavior were triggered based on self-reported sitting during random assessments throughout the day rather than objective activity monitoring.

A major limitation of the study was the quasi-experimental design (ie, nonrandom assignment). Sequential assignment of participants to the groups resulted in differences in participant characteristics including race, age, education, and smoking level. Although we attempted to control for differences in participant characteristics, randomization will be required to confirm study findings. Nevertheless, it is noteworthy that intervention participants seemed to be more vulnerable in many ways than the control group (eg, less education, more nonwhite, more smoking) and it seems promising that the intervention appeared to have a positive impact. Another limitation was that the intervention started during the first study visit (because it was embedded in the larger parent study) and, as a result, there was no baseline accelerometer measurement period. As such, we are unable to determine whether there were differences in objectively measured sedentary time between groups before the initiation of the intervention, although preintervention self-reports of daily sitting did not differ between groups. Finally, the study is limited by the short duration of the intervention (7 days). Randomized controlled trials will be needed to confirm these pilot study findings and determine the longer-term effectiveness of using mobile phone interventions to modify sedentary time.

In summary, although evidence indicates the importance of reducing and breaking up sedentary time throughout the day, it

remains unclear how to most effectively reduce sedentary behavior. This study evaluated the impact of an intervention that used mobile phone technology to prompt adults to reduce and break up their sedentary time and thereby increase activity. Intervention participants had less sedentary and more active time than control participants did during the 7-day study period. Findings also suggest that simple mobile phone messages may acutely increase activity in the 10 minutes following the prompt. These findings, although preliminary, underscore the potential impact of mobile phone interventions to modify sedentary behavior and positively influence health. Effective mobile phone interventions for sedentary behavior could be a practical and wide-reaching tool for cancer and disease prevention.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Sedentary Behaviour 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]
2. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
3. Ford ES, Caspersen CJ. Sedentary behaviour and cardiovascular disease: a review of prospective studies. Int J Epidemiol 2012 Oct;41(5):1338-1353 [FREE Full text] [doi: 10.1093/ije/dys078] [Medline: 22634869]
4. Matthews CE, George SM, Moore SC, Bowles HR, Blair A, Park Y, et al. Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. Am J Clin Nutr 2012 Feb;95(2):437-445 [FREE Full text] [doi: 10.3945/ajcn.111.019620] [Medline: 22218159]
5. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too little exercise and too much sitting: inactivity physiology and the need for new recommendations on sedentary behavior. Curr Cardiovasc Risk Rep 2008 Jul;2(4):292-298 [FREE Full text] [doi: 10.1007/s12170-008-0054-8] [Medline: 22905272]
6. Shuval K, Finley CE, Barlow CE, Gabriel KP, Leonard D, Kohl HW. Sedentary behavior, cardiorespiratory fitness, physical activity, and cardiometabolic risk in men: the cooper center longitudinal study. Mayo Clin Proc 2014 Aug;89(8):1052-1062. [doi: 10.1016/j.mayocp.2014.04.026] [Medline: 25034308]
7. Healy GN, Matthews CE, Dunstan DW, Winkler EA, Owen N. Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06. Eur Heart J 2011 Mar;32(5):590-597 [FREE Full text] [doi: 10.1093/eurheartj/ehq451] [Medline: 21224291]
8. Grøntved A, Hu FB. Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis. JAMA 2011 Jun 15;305(23):2448-2455 [FREE Full text] [doi: 10.1001/jama.2011.812] [Medline: 21673296]
9. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]
10. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996-2011. Am J Prev Med 2011 Aug;41(2):207-215. [doi: 10.1016/j.amepre.2011.05.004] [Medline: 21767729]
11. de Rezende LF, Rodrigues LM, Rey-López JP, Matsudo VK, Luiz OC. Sedentary behavior and health outcomes: an overview of systematic reviews. PLoS One 2014;9(8):e105620 [FREE Full text] [doi: 10.1371/journal.pone.0105620] [Medline: 25144686]
12. Proper KI, Singh AS, van Mechelen W, Chinapaw MJ. Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. Am J Prev Med 2011 Feb;40(2):174-182. [doi: 10.1016/j.amepre.2010.10.015] [Medline: 21238866]
13. Lynch BM. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. Cancer Epidemiol Biomarkers Prev 2010 Nov;19(11):2691-2709 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0815] [Medline: 20833969]
14. Schmid D, Leitzmann MF. Television viewing and time spent sedentary in relation to cancer risk: a meta-analysis. J Natl Cancer Inst 2014 Jul;106(7):dju098 [FREE Full text] [doi: 10.1093/jnci/dju098] [Medline: 24935969]

XSL·FO
RenderX

15. McCormack GR, Virk JS. Driving towards obesity: a systematized literature review on the association between motor vehicle travel time and distance and weight status in adults. Prev Med 2014 Sep;66:49-55. [doi: 10.1016/j.ypmed.2014.06.002] [Medline: 24929196]

16. Ding D, Gebel K, Phongsavan P, Bauman AE, Merom D. Driving: a road to unhealthy lifestyles and poor health outcomes. PLoS One 2014;9(6):e94602 [FREE Full text] [doi: 10.1371/journal.pone.0094602] [Medline: 24911017]

17. Heinonen I, Helajärvi H, Pahkala K, Heinonen OJ, Hirvensalo M, Pälve K, et al. Sedentary behaviours and obesity in adults: the Cardiovascular Risk in Young Finns Study. BMJ Open 2013;3(6) [FREE Full text] [doi: 10.1136/bmjopen-2013-002901] [Medline: 23794543]

18. Du H, Bennett D, Li L, Whitlock G, Guo Y, Collins R, China Kadoorie Biobank Collaborative Group. Physical activity and sedentary leisure time and their associations with BMI, waist circumference, and percentage body fat in 0.5 million adults: the China Kadoorie Biobank study. Am J Clin Nutr 2013 Mar;97(3):487-496 [FREE Full text] [doi: 10.3945/ajcn.112.046854] [Medline: 23364014]

19. Eriksen D, Rosthøj S, Burr H, Holtermann A. Sedentary work--associations between five-year changes in occupational sitting time and body mass index. Prev Med 2015 Apr;73:1-5. [doi: 10.1016/j.ypmed.2014.12.038] [Medline: 25575798]

20. Yang L, Hipp JA, Marx CM, Brownson RC. Occupational sitting and weight status in a diverse sample of employees in Midwest metropolitan cities, 2012-2013. Prev Chronic Dis 2014;11:E203 [FREE Full text] [doi: 10.5888/pcd11.140286] [Medline: 25412026]

21. American Cancer Society. Cancer Facts & Figures 2015. Atlanta, GA: American Cancer Society; 2015. URL: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf [accessed 2015-07-29] [WebCite Cache ID 6aO4bfxW5]

22. 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]

23. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983 [FREE Full text] [doi: 10.2337/dc11-1931] [Medline: 22374636]

24. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care 2008 Apr;31(4):661-666. [doi: 10.2337/dc07-2046] [Medline: 18252901]

25. Larsen RN, Kingwell BA, Sethi P, Cerin E, Owen N, Dunstan DW. Breaking up prolonged sitting reduces resting blood pressure in overweight/obese adults. Nutr Metab Cardiovasc Dis 2014 Sep;24(9):976-982. [doi: 10.1016/j.numecd.2014.04.011] [Medline: 24875670]

26. Shuval K, DiPietro L, Skinner CS, Barlow CE, Morrow J, Goldsteen R, et al. 'Sedentary behaviour counselling': the next step in lifestyle counselling in primary care; pilot findings from the Rapid Assessment Disuse Index (RADI) study. Br J Sports Med 2014 Oct;48(19):1451-1455. [doi: 10.1136/bjsports-2012-091357] [Medline: 22976910]

27. Shuval K, Hébert ET, Siddiqi Z, Leonard T, Lee SC, Tiro JA, et al. Impediments and facilitators to physical activity and perceptions of sedentary behavior among urban community residents: the Fair Park Study. Prev Chronic Dis 2013;10:E177 [FREE Full text] [doi: 10.5888/pcd10.130125] [Medline: 24176081]

28. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]

29. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. Obes Rev 2014 Nov;15(11):905-919 [FREE Full text] [doi: 10.1111/obr.12215] [Medline: 25112481]

30. Smith A. Pew Research Center. 2015 Apr 01. US smartphone use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2015-07-29] [WebCite Cache ID 6aO2y5TR2]

31. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]

32. Murphy P, Davis T, Long S, Jackson R, Decker B. Rapid Estimate of Adult Literacy in Medicine (REALM): a quick reading test for patients. J Reading 1993;37(2):124-130.

33. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]

34. 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]

35. Gabriel KP, McClain JJ, Schmid KK, Storti KL, High RR, Underwood DA, et al. Issues in accelerometer methodology: the role of epoch length on estimates of physical activity and relationships with health outcomes in overweight, post-menopausal women. Int J Behav Nutr Phys Act 2010;7:53 [FREE Full text] [doi: 10.1186/1479-5868-7-53] [Medline: 20550691]

XSL•FO
RenderX

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.   The International Physical Activity Questionnaire. URL: https://sites.google.com/site/theipaq/home [accessed 2015-07-29] [WebCite Cache ID 6aO57wfmJ]

38.   Matthew CE. Calibration of accelerometer output for adults. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S512-S522. [Medline: 16294114]

39.   Healy GN, Eakin EG, Lamontagne AD, Owen N, Winkler EA, Wiesner G, et al. Reducing sitting time in office workers: short-term efficacy of a multicomponent intervention. Prev Med 2013 Jul;57(1):43-48. [doi: 10.1016/j.ypmed.2013.04.004] [Medline: 23597658]

40.   Kravitz L. ACSM Information on...Reducing Sedentary Behaviors: Sitting Less and Moving More. Indianapolis, IN: American College of Sports Medicine; 2011. URL: http://www.acsm.org/docs/brochures/reducing-sedentary-behaviors-sitting-less-and-moving-more.pdf [accessed 2015-07-29] [WebCite Cache ID 6aO3xmcpg]

41.   Otten JJ, Jones KE, Littenberg B, Harvey-Berino J. Effects of television viewing reduction on energy intake and expenditure in overweight and obese adults: a randomized controlled trial. Arch Intern Med 2009 Dec 14;169(22):2109-2115. [doi: 10.1001/archinternmed.2009.430] [Medline: 20008695]

42.   Carr LJ, Karvinen K, Peavler M, Smith R, Cangelosi K. Multicomponent intervention to reduce daily sedentary time: a randomised controlled trial. BMJ Open 2013;3(10):e003261 [FREE Full text] [doi: 10.1136/bmjopen-2013-003261] [Medline: 24141969]

43.   Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]

44.   Alkhajah TA, Reeves MM, Eakin EG, Winkler EA, Owen N, Healy GN. Sit-stand workstations: a pilot intervention to reduce office sitting time. Am J Prev Med 2012 Sep;43(3):298-303. [doi: 10.1016/j.amepre.2012.05.027] [Medline: 22898123]

45.   Neuhaus M, Healy GN, Dunstan DW, Owen N, Eakin EG. Workplace sitting and height-adjustable workstations: a randomized controlled trial. Am J Prev Med 2014 Jan;46(1):30-40. [doi: 10.1016/j.amepre.2013.09.009] [Medline: 24355669]

Abbreviations

BMI: body mass index
IPAQ: International Physical Activity Questionnaire
LMM: linear mixed model
REALM: Rapid Estimate of Adult Literacy in Medicine

XSL•FO
RenderX

Original Paper

Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial

Elena Heber[1], MSc; Dirk Lehr[1,2], PhD; David Daniel Ebert[1,3], PhD; Matthias Berking[1,3], PhD; Heleen Riper[1,4,5], PhD

[1]Leuphana University Lueneburg, Division of Online Health Training, Innovation Incubator, Lueneburg, Germany

[2]Leuphana University Lueneburg, Institute of Psychology, Lueneburg, Germany

[3]Friedrich-Alexander-University Erlangen-Nuremberg, Department of Clinical Psychology and Psychotherapy, Erlangen, Germany

[4]VU University Amsterdam, Department of Clinical Psychology, Amsterdam, Netherlands

[5]Telepsychiatric Centre, University of Southern Denmark, Odense, Denmark

Corresponding Author:
Elena Heber, MSc
Leuphana University Lueneburg
Division of Online Health Training, Innovation Incubator
Scharnhorststrasse 1
Lueneburg, 21335
Germany
Phone: 49 4131 677 1708
Fax: 49 4131 677 1717
Email: heber@leuphana.de

Abstract

Background: Work-related stress is highly prevalent among employees and is associated with adverse mental health consequences. Web-based interventions offer the opportunity to deliver effective solutions on a large scale; however, the evidence is limited and the results conflicting.

Objective: This randomized controlled trial evaluated the efficacy of guided Web- and mobile-based stress management training for employees.

Methods: A total of 264 employees with elevated symptoms of stress (Perceived Stress Scale-10, PSS-10≥22) were recruited from the general working population and randomly assigned to an Internet-based stress management intervention (iSMI) or waitlist control group. The intervention (GET.ON Stress) was based on Lazarus's transactional model of stress, consisted of seven sessions, and applied both well-established problem solving and more recently developed emotion regulation strategies. Participants also had the opportunity to request automatic text messages on their mobile phone along with the iSMI. Participants received written feedback on every completed session from an e-coach. The primary outcome was perceived stress (PSS-10). Web-based self-report assessments for both groups were scheduled at baseline, 7 weeks, and 6 months. At 12 months, an extended follow-up was carried out for the iSMI group only.

Results: An intention-to-treat analysis of covariance revealed significantly large effect differences between iSMI and waitlist control groups for perceived stress at posttest ($F_{1,261}$=58.08, P<.001; Cohen's d=0.83) and at the 6-month follow-up ($F_{1,261}$=80.17, P<.001; Cohen's d=1.02). The effects in the iSMI group were maintained at 12-month follow-up.

Conclusions: This Web- and mobile-based intervention has proven effective in reducing stress in employees in the long term. Internet-based stress management interventions should be further pursued as a valuable alternative to face-to-face interventions.

Trial Registration: German Clinical Trials Register (DRKS): 00004749; http://drks-neu.uniklinik-freiburg.de/drks_web/setLocale_EN.do (Archived by WebCite at http://www.webcitation.org/6e8rl98nl)

XSL•FO
RenderX

Introduction

Stress and related adverse outcomes for physical and mental health are highly prevalent and pose a major threat to public health. Individuals with high stress levels face various negative consequences of stress including sleeping problems [1], burnout [2], an increased risk of depression, anxiety [3], and coronary heart disease [4,5]. According to a recent survey [6], 31% of US employees feel tense or stressed on a daily basis. Meanwhile, 64% report receiving insufficient stress management resources from their employers.

In recent years, Web-based and mobile-based interventions for coping with work-related stress have emerged. The advantages attributed to Web-based interventions include the potential for large-scale delivery, 24/7 availability, low costs and a low-access threshold [7]. Moreover, a recent meta-analysis reveals an equivalence between face-to-face and Internet-based guided cognitive behavioral therapy [8]. For populations experiencing high levels of work-related stress, Web-based interventions can be an appealing method for flexibly integrating stress management exercises into daily life. In particular, mobile behavioral intervention technologies for mental health offer the potential to deliver training components in real time and the real world [9]. Internet-based interventions may also reach those who are unwilling to participate in traditional face-to-face interventions [10].

Face-to-face training on stress management has been proven to be effective [11-13]. However, the evidence base for Internet-based stress management interventions (iSMIs) remains inconclusive, as only a limited number of randomized controlled trials (RCTs) have been conducted. Some of these studies showed a significant moderate reduction of stress for Web-based interventions compared with a waitlist group [14-17], a no-treatment group [18], and an attention control group [19]. Other studies did not find beneficial between-group effects for stress at posttest [20-22]. For instance, Wolever et al [18] found an effect size of d=0.74 for reduced stress for a guided mindfulness at work intervention, whereas Wiegand et al [21] did not find significant between-group effects for an unguided comprehensive Web program that included olfactory care products for women. This lack of conclusiveness of iSMIs also applies to other mental health indicators, such as depression. The differences in effectiveness may result from variations in the type and length of interventions studied, the usage of guidance, the outcomes, the measurements, or the setting. Likewise, little is known about the long-term efficacy of iSMIs. Two RCTs have investigated the efficacy of iSMIs at the 6-month follow-up relative to a control group finding a non-significant effect for stress reduction in students [23] and a small to moderate effect for the general population [14]. An extended follow-up conducted by Ruwaard et al [16] over a 3-year period yielded beneficial results for reducing stress. However, no RCTs investigating an intervention combining Web-based and mobile components with a focus on stress reduction have addressed employees as a target group relative to a control group in long-term follow-ups (eg, 6 months). With regard to the content of such interventions, currently available iSMIs do not base their theoretical foundation on a specific

stress model such as the effort-reward imbalance model [24] or the job-demand control model [25]. Likewise, more generic, established models of stress, such as Lazarus's transactional model of stress [26] are not applied. Lazarus's transactional model of stress specifies two coping strategies. Problem-focused coping is used to actively influence a stress situation in a positive way through the use of cognitive or behavioral efforts. Emotion-focused coping primarily serves the function of managing difficult emotions such as anger, disappointment, and sadness in relation to the specific situation. On the one hand, employees are often faced with problems that theoretically can be solved. Problem solving [27] is an established therapeutic method in dealing with such problems and has been proven to be successful in reducing mental and physical health problems [28]. This method has also been effectively used in Web-based interventions to manage depression, anxiety, and stress [29], although mixed results have been observed in studies targeting employees with depressive symptoms [30,31]. On the other hand, the working context also frequently requires dealing with problems that are unsolvable; such situations are commonly accompanied by strong negative emotions and require effective strategies on how to regulate these emotions. Emotion regulation skills have been shown to be relevant and successful in a broad range of mental disorders including depression and anxiety [32]; nevertheless, they remain largely untargeted in research on stress management interventions. From a theoretical perspective, promoting problem- and emotion-focused coping skills according to Lazarus's model as two major intervention components within the same intervention appears promising; however, this approach has not yet been introduced. This study aimed to fill this gap in the research by investigating an iSMI based on the combination of problem solving and emotion regulation.

This paper presents the results of a waitlist-controlled randomized trial to investigate the efficacy of a newly developed iSMI that includes mobile components for reducing stress in employees with elevated stress levels. We assessed whether the participants in the intervention group (iSMI) reported significantly lower scores on the primary outcome of perceived stress on the Perceived Stress Scale-10 (PSS-10) at posttest and at 6-month follow-up as compared to those in a waitlist control (WLC) group. Among the secondary outcomes, selected mental- and work-related health indicators often perceived to arise due to chronic stress, such as depression, anxiety, and emotional exhaustion, were also considered.

Methods

Trial Design

Using a 2-arm randomized controlled design, 264 participants were randomly allocated (at a ratio of 1:1 and a block size of 2) to the iSMI or to a WLC group. Both groups had full access to treatment as usual.

Participants

Participants 18 years and older were included if they were currently employed and scored 22 or above on the PSS-10. Due to the fact that the PSS is not a diagnostic measurement and there is no official cut-off available, we decided to use one

standard deviation (SD 6.2) above the mean (PSS-10=15.3) in a large working population [33] as a cut-off value to choose participants with an elevated level of stress. We excluded any applicants who were at risk of suicide (Beck Suicide Item >1; [34]) or self-reported to have been previously diagnosed with dissociative or psychotic symptoms. Participants were recruited in Germany from January to October 2013 in the general working population through newspaper articles and announcements by the ministry of education. Primarily, they were recruited through a large German health insurance company. The intervention addressed employees who were frequently stressed or exhausted, who felt that problems were increasingly difficult to handle, and who struggled to cope with difficult emotions. The intervention was advertised in the print membership magazine of a large German health insurance company. It was stated that, in cooperation with a university, the health insurance company offered this online training to employees, provided they were suitable for study inclusion. Moreover, the advantages of the intervention such as 24/7 availability, personal e-coaching, and participation free of charge were delineated. Those interested in participating had to provide an email address and a first and last name that could be pseudonyms if desired. Individuals received a link to the online screening questionnaire via email. Provided they were eligible, applicants had to submit their signed informed consent via regular post or scanned via email. Upon receipt of the informed consent, participants had to complete all baseline questionnaires. Subsequently, they were randomized into either the intervention or the waitlist control group.

Intervention

The iSMI GET.ON Stress is based on Lazarus's transactional model of stress [26]. This intervention applied both well-established problem solving and more recently developed emotion regulation strategies. Important principles for health behavior change such as goal setting, action planning, and coping planning were followed. The iSMI consisted of seven sessions and a booster session provided 4 weeks after training completion. Following psycho-education (Session 1), the participants learned a 6-step procedure to systematically solve problems (Sessions 2-3). In Sessions 4-6, the participants were introduced to emotion regulation techniques (muscle- and breathing relaxation, acceptance of negative emotions, and self-support in difficult situations). Session 7 included a plan for the future. The iSMI was specifically tailored to employees; this was reflected in the wording of the intervention, the example characters provided throughout the training, as well as in optional short informational material related to typical stress-related topics (eg, psychological detachment from work, time management, sleep hygiene, worrying, and organization of breaks during work) provided alongside the intervention. The application of exercises was strongly recommended. The participants were advised to complete 1-2 sessions per week. The program included interactive exercises, audio/video files, and downloadable material and was presented on a secured Web-based platform. Upon activation of the account through the research team, participants used their email address and a self-chosen password to log on. Within 48 hours after session completion, an e-coach provided approximately three quarters

of a page of written, non-therapeutic feedback intended to increase adherence and motivation. The e-coaches reported that the average time spent per feedback was 30 minutes. In the event of non-completion of a session, they also sent reminders. Each e-coach had a degree in psychology and followed a standardized manual on feedback writing. Fidelity and adherence to the feedback manual was ensured by providing extensive coaching on feedback writing and by employing a psychotherapist who provided supervision for the e-coaches. The participants could receive automatic text messages on their mobile phone along with the iSMI (eg, short relaxation exercises: "Relax your muscles in your hands and arms for 3 seconds now. Follow your breathing and each time you breathe out, relax a little more.") and were given the choice of either light (1 text message every other day) or intensive support (2-3 text messages per day) according to personal preferences. The text message coach was part of the intervention, aimed at reminding participants to practice and increasing the adherence to the intervention [9]. A more detailed description of the iSMI can be found in the protocol of the trial [35]. Screenshots of the intervention are available in the Multimedia Appendix 1.

Measurements

All questionnaires were self-assessed online at baseline (T1), 7 weeks (T2, post-treatment), 6 months (T3), and 12 months (T4, iSMI group only) after randomization. The WLC group received access to the intervention following T3.

Primary Outcome Measure

The primary outcome was the level of perceived stress as measured by the PSS-10 [36]. As this scale is based on Lazarus's transactional model of stress, it fits well with the theoretical basis of the intervention. We further decided to employ a general stress scale as previous research in a similar intervention for employees showed that work-related and non-work-related problems are equally often indicated and addressed [30]. The items were answered using a 5-point Likert scale (0=never; 1=almost never; 2=sometimes; 3=fairly often; 4=very often; range 0-40) referring to the past week. Cronbach alphas for this scale have been reported to range from .78 to .91 [37] and was .70 at T1, .90 at T2, .90 at T3, and .91 at T4 in this study.

Secondary Outcome Measures

Mental Health

Among the secondary outcomes concerning mental health, the following outcomes were measured using the specified scales: depression, using the Center for Epidemiological Studies' Depression Scale (CES-D) [38] (20 items; range 0-60; α=.91); insomnia severity, using the Insomnia Severity Index (ISI) [39] (7 items; range 0-28; α=.90); anxiety, using the subscale of the Hospital Anxiety and Depression Scales (HADS-A) [40] (7 items; range 0-21; α=.83); worrying, using the Penn State Worry Questionnaire, Ultra Brief Version-past week (PSWQ-PW) [41] (3 items; range 0-18; α=.87); and quality of life, using the Short Form 12 (SF-12) PH (physical health) and MH (mental health) [42].

Work-Related Health

Within the area of work-related health, we assessed emotional exhaustion, using the subscale of the Maslach Burnout Inventory (MBI-EE) [43] (5 items; range 1-6; α=.89); work engagement, using the Utrecht Work Engagement Scale (UWES) [44] (9 items; range 0-6; α=.94); and psychological detachment, using the subscale of the Recovery Experience Questionnaire (REQ-PD) [45] (4 items; range 1-5; α=.93). Moreover, mean days of absenteeism and presenteeism within the previous 3 months were assessed using the respective items of the German Version of the Trimbos and Institute of Medical Technology Assessment Cost Questionnaire for Psychiatry (TiC-P-G) [46].

Skills/Competencies

Emotion regulation in terms of comprehension (-C), acceptance (-A), and self-support (-SS) of the Emotion Regulation Skills Questionnaire (ERSQ), using the ERSQ-27 [47] (9 items; range 0-4; α=.87, .86, .85), and general distress, using the Emotion Specific Version, ERSQ-ES-GD [48] (12 items; range 0-4; α=.88) were assessed as measures of skills/competencies.

Other Measures

Client satisfaction, using the Client Satisfaction Questionnaire (CSQ-8) [49], demographic variables, and reasons for dropout will also be reported.

Sample Size

As the current evidence base for Web-based stress-management is limited and the effect sizes vary considerably between trials (from non-significant to moderate-large), we decided to use a conservative approach in order to also detect small effect sizes. We relied on a meta-analysis on traditional stress management interventions [13] and expected an effect size of d=0.35. Therefore, based on an alpha of .05 (two-tailed test), and a power of 80%, a sample size of 132 participants per group was necessary.

Randomization

The applied random integer list was created by an independent researcher using a Web-based randomization program (Randlist). The participants were informed about the randomization outcome via email.

Statistical Analyses

All analyses are reported according to the Consolidated Standards of Reporting Trials (CONSORT) statement regarding eHealth [50] using intention-to-treat (ITT) procedures (see Multimedia Appendix 2). Additionally, per-protocol and study completers-only analyses are reported. A significance level of .05 (two-sided) was used for all analyses. Analyses were performed using IBM SPSS version 22.

Missing Data

Multiple imputation was used to handle missing data [51]. Ten single imputations of the missing values were calculated based on the valid data for all outcome measures at all assessment points (T1, T2, T3, and T4) as well as age and gender and were aggregated into a single overall estimate of the effects of the intervention.

Intervention Effect

The iSMI and WLC groups were compared at 7 weeks (T2) and 6 months (T3) using analysis of covariance (ANCOVA) with baseline levels as covariates. Cohen's d with 95% confidence intervals (CIs) was calculated based on the imputed dataset by comparing the means and SDs of the iSMI and WLC groups at the respective time points (eg, post-means and post-SDs). According to Cohen [52], d=0.2 can be considered a small effect, d=0.5 a medium effect, and d=0.8 a large effect.

Reliable Change

The clinical significance in terms of reliable change was calculated according to the method of Jacobson and Truax [53] using the following formula: $1.96 \times SD1 \times sqrt(2) \times sqrt(1\text{-rel})$. Thereby, we used the standard deviation of the norm population (SD 6.2) and the reliability of the PSS-10 scale (α=.91) according to Cohen's and Janicki-Deverts' samples in 2006 and 2009 [37]. The participants were defined as having reliably changed if their PSS-10 score differed more than (+/-) 5.16 points from T1-T2 and T1-T3.

Symptom-Free Status

According to Jacobson and Truax [53], a cut-off point indicating symptom-free status was calculated and defined as scoring more than 2 SDs below the mean (T1) of the stressed population.

Number Needed to Treat

The number needed to treat (NNT) indicates the number of participants who must be treated to generate one additional clinically significant change. NNTs and their 95% confidence intervals [54,55] were calculated for reliable change and symptom-free status.

All procedures involved in the study were consistent with the generally accepted standards of ethical practice and were approved by the ethical committee of the University of Marburg (reference number AZ 2012-43K).

Results

Participants

A total of 450 individuals were screened for eligibility, and 186 were excluded primarily because they scored below the cut-off (136/450), because of a lack of informed consent/baseline (30/450), or other reasons (20/450). The study flow is illustrated in Figure 1.

Figure 1. Flow of participants.
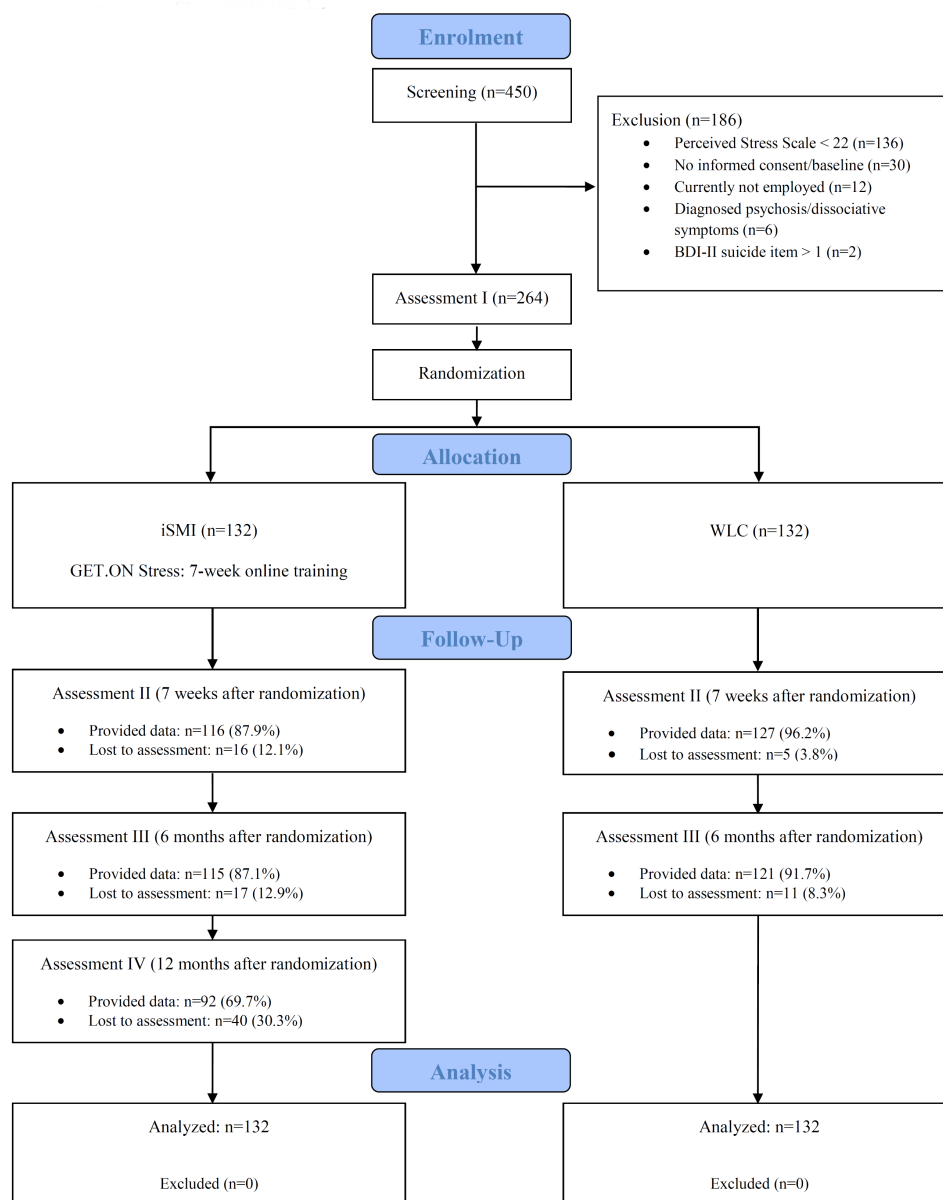
Baseline Characteristics

The demographic variables for all study participants are displayed in Table 1.

The average age of the participants was 43.3 years (SD 10.2). The sample was primarily female (193/264, 73.1%), married or in a relationship (160/264, 60.6%), and highly educated (203/264, 76.9%). Most participants were employed full-time (204/264, 77.3%); their average working experience was 18.1 (SD 11.1) years; and they were working in various sectors, most frequently in the social sector (97/264, 36.7%). Only a small percentage of participants had previously taken part in any health training (34/264, 12.9%). Having received psychotherapy was indicated by 95 (36.0%) of the 264 participants and currently being in psychotherapy by 16/264 (6.1%). Table 2 summarizes all means and SDs for the iSMI and WLC groups.

Table 1. Baseline characteristics.

Characteristics	All participants (n=264)	iSMI (n=132)	WLC (n=132)
Sociodemographic characteristics			
Age, mean (SD)	43.3 (10.2)	42.4 (10.7)	44.2 (9.6)
Gender, female, n (%)	193 (73.1)	97 (73.5)	96 (72.7)
Married or in a relationship, n (%)	160 (60.6)	80 (60.6)	80 (60.6)
Ethnicity, n (%)			
Caucasian/white	220 (83.3)	110 (83.3)	110 (83.3)
Prefer not to say	44 (16.7)	22 (16.7)	22 (16.7)
Educational level, n (%)			
Low	5 (1.9)	3 (2.3)	2 (1.5)
Middle	56 (21.2)	25 (18.9)	31 (23.5)
High	203 (76.9)	104 (78.8)	99 (75.0)
Work characteristics			
Full-time employed, n (%)	204 (77.3)	105 (79.5)	99 (75.0)
Part-time employed, n (%)	57 (21.6)	25 (18.9)	32 (24.2)
On sick leave, n (%)	3 (1.1)	2 (1.5)	1 (0.8)
Work experience in years, mean (SD)	18.1 (11.1)	17.2 (10.8)	18.9 (11.2)
Work sectors, n (%)			
Social	97 (36.7)	48 (36.4)	49 (37.1)
Service	43 (16.3)	21 (15.9)	22 (16.7)
Health	36 (13.6)	22 (16.7)	14 (10.6)
Economy	31 (11.7)	14 (10.6)	17 (12.9)
IT	15 (5.7)	8 (6.1)	7 (5.3)
Others	42 (16.0)	19 (14.3)	23 (17.4)
Income in Euro, per year, n (%)			
<10,000	10 (3.8)	7 (5.3)	3 (2.3)
10,000-30,000	48 (18.2)	28 (21.2)	20 (15.2)
30,000-40,000	63 (23.9)	26 (19.7)	37 (28.0)
40,000-50,000	47 (17.8)	26 (19.7)	21 (15.9)
50,000-60,000	29 (11.0)	15 (11.4)	14 (10.6)
60,000-100,00	32 (12.1)	17 (12.9)	15 (11.4)
>100,000	4 (1.5)	1 (0.8)	3 (2.3)
Prefer not to say	31 (11.7)	12 (9.1)	19 (14.4)
Experience, n (%)			
Previous health training	34 (12.9)	17 (12.9)	17 (12.9)
Previous psychotherapy	95 (36.0)	52 (39.4)	43 (32.6)
Current psychotherapy	16 (6.1)	5 (3.8)	11 (8.3)

Table 2. Means and standard deviations for the iSMI group and the WLC group (ITT sample).

| | T1 | | | | T2[a] | | | | T3[a] | | | | T4[a,b] | |
| | iSMI | | WLC | | iSMI | | WLC | | iSMI | | WLC | | iSMI | |
Outcome	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Primary outcome														
PSS-10	25.89	3.85	25.15	3.96	17.88	6.17	22.96	6.07	16.08	6.03	22.10	5.81	16.25	6.35
Mental health														
CES-D	23.34	8.47	23.77	7.59	15.61	9.09	21.35	8.82	13.83	7.71	21.49	8.48	14.56	8.71
ISI	15.53	6.62	14.82	6.48	8.79	5.81	11.20	6.15	8.18	5.58	11.00	5.35	7.82	4.95
HADS-A	11.23	3.34	10.67	3.36	7.79	3.98	10.32	3.46	6.73	3.38	9.65	3.58	6.51	3.46
PSWQ-PW	10.17	3.98	10.23	3.74	6.82	3.99	8.91	4.21	6.08	3.78	8.65	4.36	5.83	3.82
SF-12 MH[c]	32.29	8.44	32.55	8.08	N/A	N/A	N/A	N/A	43.38	10.56	36.54	9.50	44.16	9.69
SF-12 PH[c]	48.24	10.05	48.44	9.43	N/A	N/A	N/A	N/A	49.23	8.17	47.57	8.75	49.27	6.39
Work-related health														
MBI-EE	4.73	0.68	4.77	0.67	3.95	1.03	4.64	0.80	3.70	1.01	4.54	0.95	3.67	0.95
UWES[c]	3.18	1.26	3.31	1.15	3.38	1.29	3.10	1.13	3.46	1.17	3.16	1.14	3.46	1.13
REQ-PD[c]	2.11	0.81	2.16	0.85	2.75	0.91	2.26	0.92	3.06	0.95	2.38	0.87	2.97	0.88
Absenteeism[d]	4.93	8.70	4.40	9.62	N/A	N/A	N/A	N/A	3.64	6.70	5.23	12.10	5.88	10.57
Presenteeism[d]	15.98	14.27	17.29	16.51	N/A	N/A	N/A	N/A	11.32	12.88	11.47	11.93	8.22	9.59
Skills/Competencies														
ERSQ-C[c]	2.48	0.88	2.47	0.86	3.01	0.70	2.57	0.90	3.15	0.64	2.78	0.80	3.16	0.70
ERSQ-A[c]	1.95	0.93	1.98	0.79	2.64	0.76	2.16	0.85	2.84	0.78	2.31	0.88	2.82	0.78
ERSQ-SS[c]	2.02	0.91	2.15	0.82	2.70	0.73	2.32	0.88	2.79	0.82	2.38	0.86	2.73	0.86
ERSQ-ES-GD[c]	1.79	0.54	1.81	0.52	2.43	0.56	2.04	0.59	2.53	0.58	2.04	0.55	2.58	0.52

[a]Missing data imputed by multiple imputation.

[b]Extended follow-up for intervention group only.

[c]Higher scores indicate better outcomes.

[d]In relation to the previous 3 months.

Dropout Attrition and Handling of Missing Data

Overall, 8.0% (21/264) of participants at T2, 10.6% (28/264) of participants at T3, and 30.3% (40/132; iSMI only) of participants at T4 did not provide follow-up data for the primary outcome. A somewhat higher dropout rate was observed for the iSMI group (T2: 16/132, T3: 17/132) compared with the WLC (T2: 5/132, T3: 11/132). Thereby, groups significantly differed at T2 (χ^2_1=6.26; P<.05), but not at T3. Participants who did not provide follow-up data did not differ in a meaningful way from those who provided data, neither on baseline stress scores or any other baseline outcomes, with the exception of worrying (P<.05). Little's overall test of randomness indicated that data were missing completely at random. Thus, multiple imputations to estimate missing values could be performed [56].

Non-Usage Attrition

Intervention

Of the 132 individuals participating in the iSMI, Session 1 was completed by 122 of the participants (92.4%), Session 2 by 117 (88.6%), Session 3 by 112 (84.8%), Session 4 by 108 (81.8%), Session 5 by 103 (78.0%), Session 6 by 97 (73.5 %), Session 7 by 93 (70.5%), and the booster session by 72 (54.5%) of the participants. Because of a lack of time and changes in personal circumstances, 10 participants (7.6%) did not start the iSMI. Nine participants (6.8%) reported reasons for discontinuing the iSMI; these included lack of time (4/9), lack of motivation (3/9), technical problems (1/9), and dissatisfaction with the intervention (1/9). On average, the participants in the iSMI group completed 5.70 (SD 2.32) of the 7 sessions (81.4% of the intervention) and used the iSMI for 8.27 weeks (SD 8.54, range 0-56).

Text Message Support

Among the iSMI group, three-quarters (101/132, 76.5%) requested text message support via mobile phone. Of those, 43.6% (44/101) preferred light coaching and 56.4% (57/101) preferred intensive coaching at the start of the intervention.

Other Treatment During the Trial

In the WLC condition, 37 participants (28.0%) indicated at T2 that they had received other help within the previous 7 weeks (eg, psychotherapy, health training other than the iSMI) as opposed to 24 participants (18.2%) in the iSMI condition. No significant differences were found in stress levels between those participants who had received help and those who had not.

Primary Outcome Analyses

Intervention Effect

As shown in Table 3, a significant group effect in the ANCOVA indicated that lower scores on the PSS-10 (relative to the WLC) were present for the iSMI group at T2 ($F_{1,261}=58.08$, $P<.001$) and T3 ($F_{1,261}=80.17$, $P<.001$). Large between-group effect sizes were observed at T2 (Cohen's $d=0.83$; 95% CI 0.58-1.08) and T3 ($d=1.02$; 95% CI 0.76-1.27). In the intervention group, the within-group effect sizes were $d=1.54$ (95% CI 1.22-1.86) from pretest to post-test, $d=1.92$ (95% CI 1.55-2.29) from pre-test to 6-month follow-up, and $d=1.83$ (95% CI 1.45-2.21) from pre-test to 12-month follow-up. In the control group, within-group effect sizes of $d=0.41$ (95% CI 0.23-0.60) were observed from pre-test to post-test and $d=0.60$ (95% CI 0.39-0.81) from pre-test to 6-month follow-up. Figure 2 shows the PSS-10 scores for both groups at all assessment points.

Figure 2. Levels of perceived stress (means and SDs) according to the PSS-10 for the iSMI and WLC groups at all assessment points for the ITT sample at pre-test (T1), post-test (T2), 6 months (T3), and 12 months (T4, iSMI only) (asterisks indicate $P<.001$).
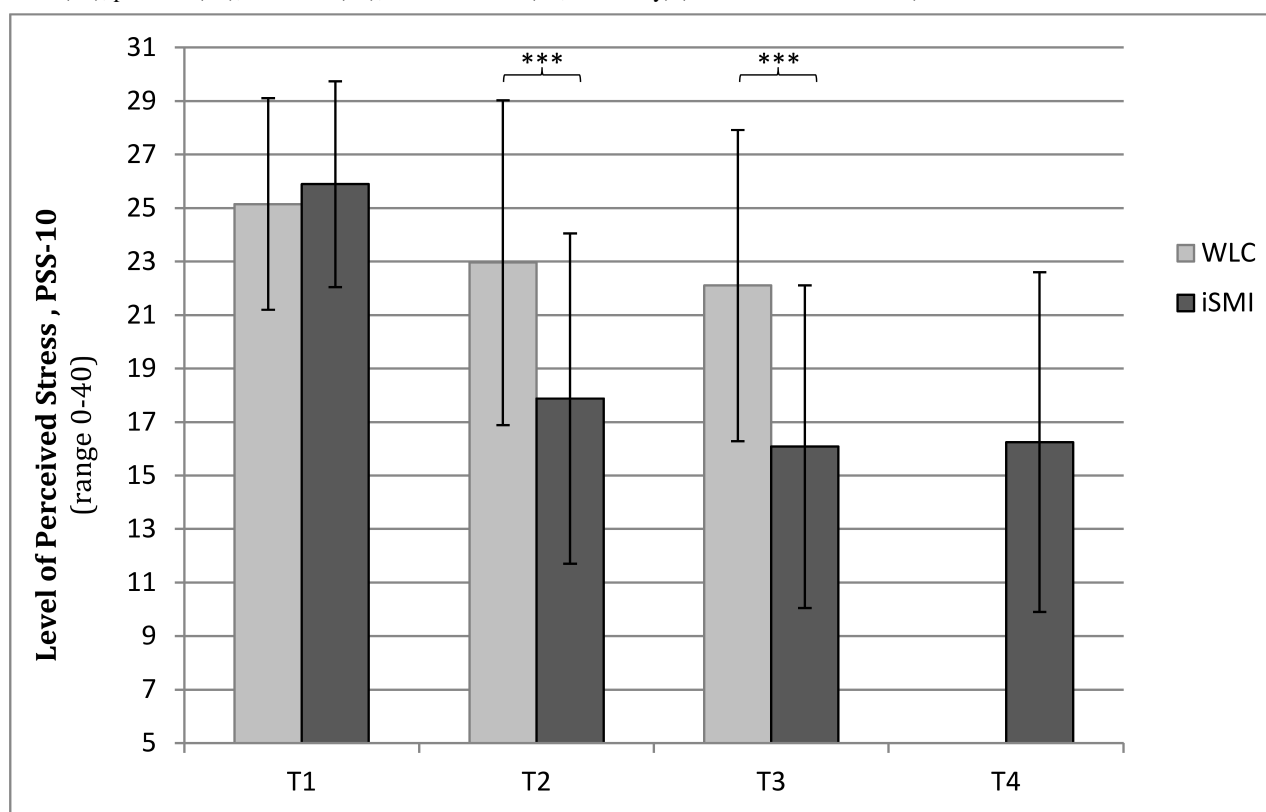
Reliable Change

At T2, more participants in the iSMI group (81/132, 61.4%) showed reliable improvement on the PSS-10 compared with the WLC (33/132, 25.0%). A reliable deterioration was present in 1.5% (2/132) of the iSMI and 8.3% (11/132) of the WLC, whereas 37.1% (49/132; iSMI) and 66.7% (88/132; WLC) were reliably unchanged. At T3, those showing reliable improvement numbered over three-quarters (102/132, 77.3%) in the iSMI and

nearly half (44/132, 33.3%) in the WLC group. Those showing reliable deterioration numbered 0.8% (1/132) in the iSMI and 6.1% (8/132) in the WLC group. No reliable change was present in 22.0% (29/132; iSMI) and 60.6% (80/132; WLC). The NNTs for reliable improvement were 2.75 (95% CI 2.11-3.96) at T2 and 2.28 (95% CI 1.83-3.01) at T3. The groups significantly differed from T1-T2 ($\chi^2_1=37.54$, $P<.001$) and from T1-T3 (Fisher's exact=53.53, $P<.001$).

Table 3. Results of the ANCOVAs and Cohen's *d* for the primary and secondary outcome measures (ITT sample) at posttest (T2) and at 6-month follow-up (T3).

Outcome	T2[a] Between-groups effect			T3[a] Between-groups effect		
	d (95% CI)	ANCOVA[b]		*d* (95% CI)	ANCOVA[b]	
		$F_{1,261}$	*P*		$F_{1,261}$	*P*
Primary outcome						
PSS-10	0.83 (0.58-1.08)	58.08	<.001	1.02 (0.76-1.27)	80.17	<.001
Mental health						
CES-D	0.64 (0.39-0.89)	34.92	<.001	0.95 (0.69-1.20)	68.29	<.001
ISI	0.40 (0.16-0.65)	20.43	<.001	0.52 (0.27-0.76)	28.82	<.001
HADS-A	0.68 (0.43-0.93)	49.57	<.001	0.84 (0.59-1.09)	78.94	<.001
PSWQ-PW	0.51 (0.26-0.75)	19.74	<.001	0.63 (0.38-0.88)	31.00	<.001
SF-12 MH	N/A	N/A		0.68 (0.43-0.93)	34.28	<.001
SF-12 PH	N/A	N/A		0.20 (-0.05 to 0.44)	4.27	.04
Work-related health						
MBI-EE	0.75 (0.50-1.00)	48.55	<.001	0.86 (0.60-1.11)	63.84	<.001
UWES	0.23 (-0.01 to 0.47)	13.28	<.001	0.26 (0.02-0.50)	10.09	.002
REQ-PD	0.54 (0.29-0.78)	27.45	<.001	0.75 (0.50-1.00)	52.82	<.001
Absenteeism	N/A	N/A		0.16 (-0.08 to 0.40)	2.94	
Presenteeism	N/A	N/A		0.01 (-0.23 to 0.25)	0.02	
Skills/Competencies						
ERSQ-27-C	0.55 (0.30-0.79)	31.56	<.001	0.51 (0.27-0.76)	22.03	<.001
ERSQ-27-A	0.60 (0.35-0.84)	30.94	<.001	0.64 (0.39-0.88)	34.96	<.001
ERSQ-27-SS	0.47 (0.23-0.71)	29.32	<.001	0.49 (0.24-0.73)	24.56	<.001
ERSQ-ES-GD	0.68 (0.43-0.93)	36.39	<.001	0.87 (0.61-1.12)	54.71	<.001

[a]Missing data imputed by multiple imputation.

[b]Controlling for pre-treatment scores (T1).

Symptom-Free Status

In this study, the cut-off score was 17.70 and below indicating a value of 2 SDs below the mean of the stressed population at T1 (mean 25.52, SD 3.91). More participants in the iSMI group met the criterion for full remission of stress symptoms compared with the WLC group at T2 (iSMI: 68/132, 51.5%; WLC: 26/132, 19.7%; χ^2_1=29.14, *P*<.001; NNT=3.14, 95% CI 2.34-4.78) and T3 (iSMI: 79/132, 59.8%; WLC: 31/132, 23.5%; χ^2_1=35.91, *P*<.001; NNT=2.75, 95% CI 2.11-3.95).

Completers-Only Analysis

Completers-only analyses on participants who completed all questionnaires revealed similar large effect sizes for the primary outcome at T2 (243/264, 92.0%, *d*=0.85; CI 0.59-1.11) and T3 (236/264, 89.4%, *d*=1.01; 95% CI 0.74-1.28).

Secondary Outcome Analyses

Table 3 also shows the results of the ITT analyses for secondary outcomes for mental health, work-related health, and skills/competencies. The ANCOVAs showed highly significant between-group effects for almost all outcomes at both assessment points; all significance levels were *P*<.001 apart from work engagement at T3 (*P*=.002) and the physical health component of quality of life at T3 (*P*=.04). Between-group effects were not significant for absenteeism and presenteeism.

At T2, the majority of effect sizes were in the range of moderate (eg, *d*=0.40 for insomnia) to large (eg, *d*=0.75 for emotional exhaustion) apart from work engagement for which a small effect size was obtained (*d*=0.23). At T3, almost all effect sizes became more pronounced apart from the comprehension subscale of the ERSQ-27, which only slightly decreased (from *d*=0.55 at T2 to *d*=0.51 at T3). Thereby, large effect sizes were found at T3 for depression (*d*=0.95), anxiety (*d*=0.84), emotional exhaustion (*d*=0.86), and emotion regulation skills regarding general distress (*d*=0.87). The additional measurements taken at T3 yielded effect sizes of *d*=0.68 for the mental health component and *d*=0.20 for the physical health component of quality of life, as well as *d*=0.16 for absenteeism and *d*=0.01 for presenteeism.

Extended Follow-Up at 12 Months

The within-group effect size (from T1-T4) for the primary outcome PSS-10 was *d*=1.83 (95% CI 1.45-2.21). At

XSL•FO
RenderX

12-month-follow-up, improvements in all other outcome measures were also maintained at the 6-month level in the iSMI group apart from absenteeism ($P<.05$). On a descriptive level, the gains in mean days for absenteeism in the iSMI group from T1 (mean 4.93, SD 8.70) to T3 (mean 3.64, SD 6.70) could not be maintained at T4 (mean 5.88, SD 10.57). In contrast, the mean days for presenteeism in the iSMI group were almost reduced by half from T1 (mean 15.98, SD 14.27) to T4 (mean 8.22, SD 9.59).

Client Satisfaction

Client satisfaction with the training was high, with 92.2% (107/116) being "satisfied in an overall, general sense" ("very satisfied" or "mostly satisfied"). The majority of the participants indicated that they have received the kind of training they wanted (92.2%, 107/116; "yes, definitely" or "yes, generally"), that the training met their needs (88.8%, 103/116; "almost all […]" or "most […]"), that they are satisfied with the amount of training they received (87.9%, 102/116; "very satisfied" or "mostly satisfied"), that the training has helped them to deal more effectively with their problems (92.2%, 107/116; "yes it helped a great deal" or "yes, it helped"), and that they would use the training again if they needed to (92.2%, 107/116; "yes, definitely" or "yes, I think so"). Moreover, 90.5% (105/116) stated that they would recommend the iSMI to a friend ("yes, definitely" or "yes, I think so").

Explorative Analyses

Intervention Completion

A separate per-protocol analysis was conducted for participants who completed the intervention (≥6 sessions), which was defined as working through all of the theoretical intervention content presented up to Session 6. The ANCOVA showed significant differences between the subsample of intervention completers (97/132) and the WLC (132) with regard to perceived stress in favor of the experimental condition at T2 ($F_{1,226}=66.85$, $P<.001$) and T3 ($F_{1,226}=74.70$, $P<.001$) with slightly higher effect sizes at T2 ($d=0.95$; 95% CI 0.69-1.20) and T3 ($d=1.05$; 95% CI 0.79-1.31) as compared to the total iSMI sample. Within the iSMI group, we further compared intervention completers to non-completers. The ANCOVA showed a significant difference for reduction of perceived stress at T2 ($F_{1,129}=7.76$, $P=.006$), but not at T3 or T4.

Text Message Support

There were no significant differences in the primary outcome between participants who received text messages and those who did not, nor was there any significant difference depending on the level of intensity of the individually chosen text message support.

Discussion

Principal Findings

The primary aim of this study was to evaluate the efficacy of a guided iSMI for employees. For this purpose, a two-arm, waitlist-controlled randomized trial was conducted. The results indicate that the training is highly effective in reducing employee stress levels in the short term ($d=0.83$) and long term ($d=1.02$) compared with the levels observed in a waitlist control group. Reduced stress levels in the iSMI group could be maintained up to 12 months. Significant medium to large between-group effects were also found for relevant secondary outcomes concerning mental health (eg, depression), work-related health (eg, emotional exhaustion), and stress-related skills (eg, emotion regulation competencies). High levels of client satisfaction and adherence were observed, and the study dropout rate was low.

The posttest effect size in stress reduction found in this study is larger than what has been found in other iSMI trials. Available RCTs on iSMI with employees show mixed between-group effects for stress at posttest, ranging from non-significant [22] to moderate effect sizes (eg, $d=0.74$; [18]). Several reasons could explain the large effect sizes found in this study. First, the intervention used a guided format. Guided iSMIs [16-18] appear to be more effective than unguided interventions [21,22,57,58], a result that is known from Web-based interventions for other mental health problems [59]. Second, the theoretical basis of the intervention was confined to two evidence-based components. Research on face-to-face interventions suggests that interventions with fewer treatment components are superior to those using more components [12]. Problem-solving training that has already been successfully introduced in other Web-based interventions to manage depressive symptoms [30] was combined here with evidence-based emotion regulation techniques including the acceptance of emotions and compassionate self-support based on the Affect Regulation Training (ART) [32]. Third, the use of mobile components to flexibly introduce training components into daily life in real time may have reinforced a regular application of the intervention exercises and therefore the efficacy of the training. The text messages were particularly popular among the participants, and given the choice of receiving the text messages, the vast majority requested this mobile component. However, we did not assess the actual engagement and future studies should compare the efficacy of the intervention with and without the mobile phone component. Fourth, the level of intervention adherence, which is regarded as leading to better treatment outcomes [60], was relatively high. Unfortunately, comparisons with adherence levels of other iSMIs are difficult as few studies have reported this information so far. Compared with the available intervention completion rates (eg, 38.5%, [14]; 44.0%, [61]; 88.2% [19]), the percentage of participants completing the intervention in this study was higher range (70.5%). Considerable efforts were undertaken to increase adherence through methods that are generally considered to be effective, including human support [62], interactive exercises [63], tailoring of the intervention [64], and reminders [65] via mobile phone. Finally, the effects may have further stabilized at T3 through the booster session, as booster sessions can be successful in maintaining treatment outcomes [66].

With regard to long-term follow-up, this work is the first study of an intervention combining Web-based and mobile components that focuses on stress reduction in employees to assess the effects compared to a control group over a longer time period (ie, 6 months). The results show that this type of

intervention can have large long-term effects. Moreover, the results of this study also compare favorably to the limited existing evidence on the long-term effectiveness of iSMIs in populations other than employees, including studies finding non-significant [23] and small to moderate effect sizes (d=0.37 [14]) at the 6-month follow-up point. In addition, the results found for stress show similar effect sizes to those found for face-to-face interventions (d=0.73 [12]). Thus, iSMIs may be a useful alternative to traditional interventions.

With regard to the relationship of treatment intensity and outcome, no clear conclusions can be drawn from this study. Although intervention completers showed significantly lower stress levels at posttest, this effect was not maintained at the later follow-up points. Participants receiving text message support were also not doing significantly better as compared to those who did not receive any messages. Future research may benefit from further information on the amount of time spent on the intervention exercises in between sessions and the actual engagement with the text messages.

Limitations

The following limitations of this trial must be acknowledged. First, for feasibility reasons, only self-report measures were assessed. Although the replacement of self-report measures with physiological measures is not recommended in occupational stress research [67], a combination of both could produce further valuable insights. Second, because this study was in the setting of indicated prevention, these results only account for participants showing relatively high baseline scores. The current sample was severely distressed and showed high baseline scores on all measures. Thus, no conclusions can be drawn regarding participants with lower stress levels (eg, in a universal prevention setting). Third, with regard to the generalizability of results, the fact that participants self-selected into the trial, the majority were female, and individuals working in the social sector were slightly overrepresented needs to be taken into consideration. Fourth, to determine the added value of the mobile component providing real-time support and encouragement, direct comparison studies would be needed comparing the intervention with and without mobile support. Fifth, the fact that the effects on physical health and work engagement were smaller than the effects on the other outcome measures demands an explanation. It is possible that Web-based interventions do not produce meaningful differences on these outcomes. Alternative explanations for the small effect sizes include that, for physical health, a rather global measure (SF-12) was applied and it may be more promising to use more specific, stress-related health measures. For work engagement, the original response categories were not adapted to the study period and the outcome measure may therefore not have been as sensitive to change. Finally, it is important to acknowledge that some improvements were also observed in the WLC group over time; in fact, this pattern has previously been found in other trials [17,31].

Future Directions

Future research should replicate the results of this trial and investigate the moderators of outcome and adherence. It is also of interest whether the coaching time spent on each individual (up to 4 hours) could be reduced without losses in treatment effects, thereby resulting in more economical versions of iSMI. Moreover, future research should test iSMIs against the gold standard in the field (ie, face-to-face interventions) and assess which training format works best for which type of participant and under what circumstances. Although both formats may be equally effective, they may work differently on participants with varying personal characteristics and Web-based interventions may be more advantageous in terms of efficiency and costs.

Conclusion

This trial contributes to the limited evidence base on the feasibility and efficacy of Web-based and mobile-supported stress management interventions and is among the first studies to include a longer follow-up period. The iSMI presented herein proved feasible and highly effective in improving perceived stress and other mental and work-related health indices in employees in the long term. These results indicate that this iSMI could be a valuable alternative to face-to-face trainings. Web-based interventions for coping with stress should be further evaluated as such interventions have the potential to improve the mental health of individuals on a large scale.

Authors' Contributions

MB obtained the funding for the study. All authors contributed to the design of this study. EH was responsible for the intervention development, study administration, and statistical analyses of the data. DL, HR and DE supervised this process. All authors contributed to the data interpretation. EH drafted the manuscript. All authors contributed to the further writing of the manuscript and approved the final version of the manuscript.

Conflicts of Interest

DL, DE, and MB are stakeholders of the Institute for Online Health Training, which aims to transfer scientific knowledge related to this research into routine health care. EH is an employee at this Institute. The Institute was founded in January 2015. At the time of planning, conducting, and evaluating the study, it did not yet exist.

Multimedia Appendix 1

Screenshots of the intervention.

[PDF File (Adobe PDF File), 458KB - jmir_v18i1e21_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.2 [67].

[PDF File (Adobe PDF File), 1MB - jmir_v18i1e21_app2.pdf]

References

1. Akerstedt T. Psychosocial stress and impaired sleep. Scand J Work Environ Health 2006 Dec;32(6):493-501 [FREE Full text] [Medline: 17173205]
2. Maslach C, Leiter M. The truth about burnout: how organizations cause personal stress and what to do about it. San Francisco, CA: Jossey-Bass; 1997.
3. Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE. Work stress precipitates depression and anxiety in young, working women and men. Psychol Med 2007 Aug;37(8):1119-1129 [FREE Full text] [doi: 10.1017/S0033291707000414] [Medline: 17407618]
4. Kivimäki M, Virtanen M, Elovainio M, Kouvonen A, Väänänen A, Vahtera J. Work stress in the etiology of coronary heart disease--a meta-analysis. Scand J Work Environ Health 2006 Dec;32(6):431-442 [FREE Full text] [Medline: 17173200]
5. Steptoe A, Kivimäki M. Stress and cardiovascular disease. Nat Rev Cardiol 2012 Jun;9(6):360-370. [doi: 10.1038/nrcardio.2012.45] [Medline: 22473079]
6. APA Center for Organizational Excellence. 2014 Work and well-being survey. 2014. URL: http://www.apaexcellence.org/assets/general/2014-work-and-wellbeing-survey-results.pdf [accessed 2015-09-08] [WebCite Cache ID 6bO5kHn8q]
7. 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]
8. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
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. Marks I, Cavanagh K, Gega L. Hands-on help: computer-aided psychotherapy. Hove, England: Psychology Press; 2007.
11. Bhui KS, Dinos S, Stansfeld SA, White PD. A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. J Environ Public Health 2012;2012:515874 [FREE Full text] [doi: 10.1155/2012/515874] [Medline: 22496705]
12. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
13. van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ. The benefits of interventions for work-related stress. Am J Public Health 2001 Feb;91(2):270-276. [Medline: 11211637]
14. Drozd F, Raeder S, Kraft P, Bjørkli CA. Multilevel growth curve analyses of treatment effects of a Web-based intervention for stress reduction: randomized controlled trial. J Med Internet Res 2013;15(4):e84 [FREE Full text] [doi: 10.2196/jmir.2570] [Medline: 23607962]
15. Hänggi Y. Kann web-basierte Stressprävention erfolgreich sein? [Can web-based parental training on coping with familiy stress be successful?]. Zeitschrift für Klinische Psychologie und Psychotherapie 2006 Jul 1;35(3):169-177. [doi: 10.1026/1616-3443.35.3.169]
16. Ruwaard J, Lange A, Bouwman M, Broeksteeg J, Schrieken B. E-mailed standardized cognitive behavioural treatment of work-related stress: a randomized controlled trial. Cogn Behav Ther 2007;36(3):179-192. [doi: 10.1080/16506070701381863] [Medline: 17852171]
17. Zetterqvist K, Maanmies J, Ström L, Andersson G. Randomized controlled trial of internet-based stress management. Cogn Behav Ther 2003;32(3):151-160. [doi: 10.1080/16506070302316] [Medline: 16291546]

18.   Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]

19.   Rose RD, Buckey JC, Zbozinek TD, Motivala SJ, Glenn DE, Cartreine JA, et al. A randomized controlled trial of a self-guided, multimedia, stress management and resilience training program. Behav Res Ther 2013 Feb;51(2):106-112. [doi: 10.1016/j.brat.2012.11.003] [Medline: 23262118]

20.   Glück TM, Maercker A. A randomized controlled pilot study of a brief web-based mindfulness training. BMC Psychiatry 2011;11:175 [FREE Full text] [doi: 10.1186/1471-244X-11-175] [Medline: 22067058]

21.   Wiegand B, Luedtke K, Friscia D, Nair M, Aleles M, McCloskey R. Efficacy of a comprehensive program for reducing stress in women: a prospective, randomized trial. Curr Med Res Opin 2010 Apr;26(4):991-1002. [doi: 10.1185/03007991003688193] [Medline: 20201624]

22.   Yamagishi M, Kobayashi T, Nakamura Y. Effects of web-based career identity training for stress management among Japanese nurses: a randomized control trial. J Occup Health 2008;50(2):191-193 [FREE Full text] [Medline: 18403870]

23.   Chiauzzi E, Brevard J, Thum C, Thurn C, Decembrele S, Lord S. MyStudentBody-Stress: an online stress management intervention for college students. J Health Commun 2008 Sep;13(6):555-572. [doi: 10.1080/10810730802281668] [Medline: 18726812]

24.   Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol 1996 Jan;1(1):27-41. [Medline: 9547031]

25.   Karasek R. Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. Administrative Science Quarterly 1979 Jun;24(2):285-308. [doi: 10.2307/2392498]

26.   Lazarus R, Folkman S. Stress, appraisal, and coping. New York: Springer Publishing Co; 1984.

27.   D'Zurilla T, Nezu A. Problem-solving therapy. In: Dobson KS, editor. Handbook of cognitive-behavioral therapies. 3rd edition. New York: Guilford Press; 2010.

28.   Malouff JM, Thorsteinsson EB, Schutte NS. The efficacy of problem solving therapy in reducing mental and physical health problems: a meta-analysis. Clin Psychol Rev 2007 Jan;27(1):46-57. [doi: 10.1016/j.cpr.2005.12.005] [Medline: 16480801]

29.   van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]

30.   Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. Scand J Work Environ Health 2014 Nov;40(6):582-596. [doi: 10.5271/sjweh.3449] [Medline: 25121986]

31.   Geraedts AS, Kleiboer AM, Wiezer NM, van MW, Cuijpers P. Short-term effects of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014;16(5):e121 [FREE Full text] [doi: 10.2196/jmir.3185] [Medline: 24800966]

32.   Berking M, Whitley B. Affect Regulation Training. New York, NY: Springer New York; 2014.

33.   Lesage F, Berjot S, Deschamps F. Psychometric properties of the French versions of the Perceived Stress Scale. Int J Occup Med Environ Health 2012 Jun;25(2):178-184 [FREE Full text] [doi: 10.2478/S13382-012-0024-8] [Medline: 22528542]

34.   Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.

35.   Heber E, Ebert DD, Lehr D, Nobis S, Berking M, Riper H. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. BMC Public Health 2013;13:655 [FREE Full text] [doi: 10.1186/1471-2458-13-655] [Medline: 23855376]

36.   Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

37.   Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]

38.   Hautzinger M, Bailer M, Hofmeister D, Keller F. Allgemeine Depressionsskala (ADS) Manual [Center for Epidemiological Studies Depression Scale (CES-D; Radloff LS, 1977)]. Boston, MA: Hogrefe Publishers; 2012.

39.   Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]

40.   Hermann-Lingen C, Buss U, Snaith R. Hospital Anxiety and Depression Scale - Deutsche Version (HADS-D). 3rd edition. Bern: Hans Huber; 2011.

41.   Stöber J. PSWQ-PW: Penn State Worry Questionnaire-Past Week. In: Brähler E, Schumacher J, Strauß B, editors. Diagnostische Verfahren in der Psychotherapie. Göttingen, Germany: Hogrefe; 2002:291-294.

42.   Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]

43.   Schaufeli W, Leiter M, Maslach C, Jackson S. Maslach Burnout Inventory - General Survey (MBI-GS). In: Maslach C, Jackson SE, Leiter MP, editors. Maslach Burnout Inventory Manual. 3rd edition. Palo Alto, CA: Consulting Psychologists Press; 1996.

44. Schaufeli W, Bakker A. UWES - Utrecht Work Engagement Scale: Test Manual. Utrecht, The Netherlands: Department of Psychology, Utrecht University; 2003.

45. Sonnentag S, Fritz C. The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work. J Occup Health Psychol 2007 Jul;12(3):204-221. [doi: 10.1037/1076-8998.12.3.204] [Medline: 17638488]

46. Hakkaart-van Roijen L. Manual Trimbos/iMTA Questionnaire for costs associated with psychiatric illness (in Dutch). Rotterdam: Institute for Medical Technology Assessment; 2002.

47. Berking M, Znoj H. Entwicklung und Validierung eines Fragebogens zur standardisierten Selbsteinschätzung emotionaler Kompetenzen (SEK-27) [Development and validation of a self-report instrument for the assessment of emotion-specific regulation skills (ERSQ-27)]. Zeitschrift für Psychiatrie, Psychologie und Psychotherapie 2008 Jan;56(2):141-153. [doi: 10.1024/1661-4747.56.2.141]

48. Ebert DD, Christ O, Berking M. Entwicklung und Validierung eines Fragebogens zur emotionsspezifischen Selbsteinschätzung emotionaler Kompetenzen (SEK-ES) [Development and validation of a self-report instrument for the assessment of emotion-specific regulation skills (ERSQ-ES)]. Diagnostica 2013 Jan;59(1):17-32. [doi: 10.1026/0012-1924/a000079]

49. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [Medline: 10259963]

50. Eysenbach G, Consort-EHEALTHGroup. 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]

51. Little R, Rubin D. Statistical analysis with missing data. Hoboken, NJ: Wiley; 2002.

52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

53. 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]

54. Altman DG. Confidence intervals for the number needed to treat. BMJ 1998 Nov 7;317(7168):1309-1312 [FREE Full text] [Medline: 9804726]

55. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995 Feb 18;310(6977):452-454 [FREE Full text] [Medline: 7873954]

56. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]

57. Billings DW, Cook RF, Hendrickson A, Dove DC. A web-based approach to managing stress and mood disorders in the workforce. J Occup Environ Med 2008 Aug;50(8):960-968 [FREE Full text] [doi: 10.1097/JOM.0b013e31816c435b] [Medline: 18695455]

58. Hinman M, Ezzo L, Hunt D, Mays A. Computerized exercise program does not affect stress levels of asymptomatic VDT users. J Occup Rehab 1997 Mar;7(1):45-51. [doi: 10.1007/BF02765875]

59. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]

60. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

61. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: A randomized controlled trial. Internet Interventions 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]

62. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

63. Hasson H, Brown C, Hasson D. Factors associated with high use of a workplace web-based stress management program in a randomized controlled intervention study. Health Educ Res 2010 Aug;25(4):596-607 [FREE Full text] [doi: 10.1093/her/cyq005] [Medline: 20150531]

64. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27 Suppl 3:S227-S232. [Medline: 14672383]

65. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

66. Whisman MA. The efficacy of booster maintenance sessions in behavior therapy: Review and methodological critique. Clinical Psychology Review 1990 Jan;10(2):155-170. [doi: 10.1016/0272-7358(90)90055-F]

67. Semmer N, Grebner S, Elfering A. Beyond self-report: using observational, physiological, situation-based measures in research on occupational stress. In: Perrewe P, Ganster D, editors. Emotional and Physiological Processes and Positive Intervention Strategies (Research in Occupational Stress and Well-being, Volume 3). Bingley, UK: Emerald Group Publishing Limited; 2003:205-263.

XSL•FO
RenderX

Abbreviations

ANCOVA: analysis of covariance
CES-D: Center for Epidemiological Studies' Depression Scale
CONSORT: Consolidated Standards of Reporting Trials
CSQ-8: Client Satisfaction Questionnaire
ERSQ-27-A: Emotion Regulation Skills Questionnaire – subscale acceptance
ERSQ-27-C: Emotion Regulation Skills Questionnaire – subscale comprehension,
ERSQ-ES-GD: Emotion Regulation Skills Questionnaire – Emotion specific version – subscale general distress
ERSQ-27-SS: Emotion Regulation Skills Questionnaire – subscale self-support
HADS-A: Hospital Anxiety and Depression Scales – subscale anxiety
ISI: Insomnia Severity Index
iSMI: Internet-based stress management intervention
ITT: intention-to-treat
MBI-EE: Maslach Burnout Inventory–subscale Emotional Exhaustion
NNT: number needed to treat
PSS-10: Perceived Stress Scale – 10 item version
PSWQ-PW: Penn State Worry Questionnaire-Past Week
RCT: randomized controlled trial
REQ-PD: Recovery Experience Questionnaire – subscale psychological detachment
SD: standard deviation
SF-12 MH: Short Form 12–Quality of life (mental health component)
SF-12 PH: Short Form 12–Quality of life (physical health component)
TIC-P-G: Trimbos and Institute of Medical Technology Assessment Cost Questionnaire for Psychiatry
UWES: Utrecht Work Engagement Scale
WLC: waitlist control

Original Paper

The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review

Yi-Ru Regina Chen[1], PhD; Peter J Schulz[2], PhD

[1]Hong Kong Baptist University, Kowloon Tong, China (Hong Kong)
[2]University of Lugano, Lugano, Switzerland

Corresponding Author:
Yi-Ru Regina Chen, PhD
Hong Kong Baptist University
Department of Communication Studies
Hong Kong Baptist University
Kowloon Tong,
China (Hong Kong)
Phone: 852 3411 5057
Fax: 852 3411 7375
Email: yrchen@hkbu.edu.hk

Abstract

Background: The aging of the population is an inexorable change that challenges governments and societies in every developed country. Based on clinical and empirical data, social isolation is found to be prevalent among elderly people, and it has negative consequences on the elderly's psychological and physical health. Targeting social isolation has become a focus area for policy and practice. Evidence indicates that contemporary information and communication technologies (ICT) have the potential to prevent or reduce the social isolation of elderly people via various mechanisms.

Objective: This systematic review explored the effects of ICT interventions on reducing social isolation of the elderly.

Methods: Relevant electronic databases (PsycINFO, PubMed, MEDLINE, EBSCO, SSCI, Communication Studies: a SAGE Full-Text Collection, Communication & Mass Media Complete, Association for Computing Machinery (ACM) Digital Library, and IEEE Xplore) were systematically searched using a unified strategy to identify quantitative and qualitative studies on the effectiveness of ICT-mediated social isolation interventions for elderly people published in English between 2002 and 2015. Narrative synthesis was performed to interpret the results of the identified studies, and their quality was also appraised.

Results: Twenty-five publications were included in the review. Four of them were evaluated as rigorous research. Most studies measured the effectiveness of ICT by measuring specific dimensions rather than social isolation in general. ICT use was consistently found to affect social support, social connectedness, and social isolation in general positively. The results for loneliness were inconclusive. Even though most were positive, some studies found a nonsignificant or negative impact. More importantly, the positive effect of ICT use on social connectedness and social support seemed to be short-term and did not last for more than six months after the intervention. The results for self-esteem and control over one's life were consistent but generally nonsignificant. ICT was found to alleviate the elderly's social isolation through four mechanisms: connecting to the outside world, gaining social support, engaging in activities of interests, and boosting self-confidence.

Conclusions: More well-designed studies that contain a minimum risk of research bias are needed to draw conclusions on the effectiveness of ICT interventions for elderly people in reducing their perceived social isolation as a multidimensional concept. The results of this review suggest that ICT could be an effective tool to tackle social isolation among the elderly. However, it is not suitable for every senior alike. Future research should identify who among elderly people can most benefit from ICT use in reducing social isolation. Research on other types of ICT (eg, mobile phone–based instant messaging apps) should be conducted to promote understanding and practice of ICT-based social-isolation interventions for elderly people.

KEYWORDS

social isolation; elderly; ICT intervention

Introduction

It is estimated that the proportion of the world population aged 60 years and older will reach 22% by 2050 [1]. Social isolation among the elderly is therefore a growing concern. Depending on the definition and measure, the prevalence of social isolation among people aged 60 years and older is 7% to 24% [2-7] compared to 7% in the general population [6]. In addition, perceived social isolation is more severe among the older old people (aged 75-85 years) than the younger old (aged 57-65 years) [8]. Most importantly, social isolation is a real threat to the mental and physical health of the elderly population [7-11], leading to depression [3,12], self-harming (eg, drug abuse, alcoholism, suicide) [13-15] or self-neglecting behavior [16], a higher level of cognitive and/or physical disability [17], and increased mortality [8,18]. Consequently, preventing or ameliorating social isolation in that age group is becoming a top social topic and a priority in policy-making in many countries [19-20].

Social isolation is a multidimensional concept that lacks a clear and consistent definition in the literature [21-22]. Some scholars see it as directly equivalent to loneliness and use the terms interchangeably [20]; others perceive the two concepts as related yet distinct. For example, social isolation has been defined as the absence of contact with people who provide social support [23]. Others have defined it as a 2-dimensional concept that contains an objective absence of contacts or interactions with the contacts and a subjective feeling of limited or lost companionship or social support (ie, loneliness) resulting from having limited contacts or interactions [8,21]. No matter which definition one adopts, social isolation is considered a result of the elderly population's reduced social interactions—particularly with family, friends, and community networks—caused by their retirement, physical changes (cognitive and physical disabilities), inevitable loss of spouse or friends (shrinking network size), and/or living alone or in institutions [8]. Information and communication technology (ICT) may overcome the social and spatial barriers of social interaction by enabling easy, affordable communication and activities of multiple forms (ie, textual, audio, and/or visual) between the elderly (often with limited mobilization) and others anytime and anywhere. Many researchers have therefore been investigating its potential for alleviating social isolation in the elderly.

A search of the literature identified 4 systematic reviews [20-21,24-25] that synthesized the effects of social isolation interventions. These reviews examined studies of various designs, including randomized controlled trials (RCTs), experiments, quasi-experimental studies, and before-and-after (cohort) studies, published in the periods 1970 to 2002 [21,24], 1976 to 2009 [20], and 2000 to 2013 [25]. While 3 reviews examined all forms of interventions for social isolation [20-21,24], Morris and colleagues [25] focused only on the interventions using smart technologies to synthesize the effect of interventions on social connectedness of the elderly living at home and found conflicting results.

The objective of our systematic review is to gain a synthesis of the evident effects of ICT interventions on social isolation in the elderly. Our review is timely and valuable for the following reasons: (1) it reviews the effect of ICT interventions on the elderly with various characteristics (eg, demographics, health status, and living arrangements); (2) it covers the most recent research, published between 2002 and 2015; and (3) in addition to quantitative research, it includes studies that used qualitative methods (ie, observations, in-depth interviews, and focus group interviews) to offer insights into the mechanisms underpinning the observed variations in ICT effectiveness.

Methods

Searching Strategy, Inclusion Criteria, and Study Selection

Electronic searches for this systematic review were conducted in July 2015 using PsycINFO, PubMed, MEDLINE, EBSCO, SSCI, Communication Studies: a SAGE Full-Text Collection, Communication & Mass Media Complete, Association for Computing Machinery Digital Library, and IEEE Xplore. These databases were used because they include research on subjects such as health, aging, social science, digital technologies, computer-mediated communication, and communication science. A unified search term using Boolean operators was applied for all databases: ((social isolation OR loneliness) AND elderly AND (Internet OR social media OR information and communication technology)). Next, to ensure a broad inclusion of published studies relevant to our review topic, we adopted the following criteria to select studies for the review: (1) publications must be in English; (2) studies must empirically investigate the effects of ICTs on one or more attributes of social isolation among the elderly; and (3) study participants must be aged 55 years or older.

The search yielded 424 publications, of which 51 duplicates were removed. The first author then checked the remaining titles and abstracts to determine their relevance. If the information provided by a title or abstract was insufficient for determination, the full paper was screened by 2 researchers who documented the reasons leading to the exclusion of full texts. An additional 2 studies were found in the systematic review of studies on the elderly population's social connectedness and smart technologies by Morris et al [25]. A total of 30 articles met the inclusion criteria outlined in Figure 1 and were retained for this systematic review. After carefully reading the full texts of the articles, researchers excluded 5 more studies because of a lack of a complete text (1 article), no examination of social isolation as the outcome of ICT use (3 articles), and the participants being aged younger than 55 years (1 article).

XSL•FO
RenderX

Figure 1. An overview of the inclusion process.
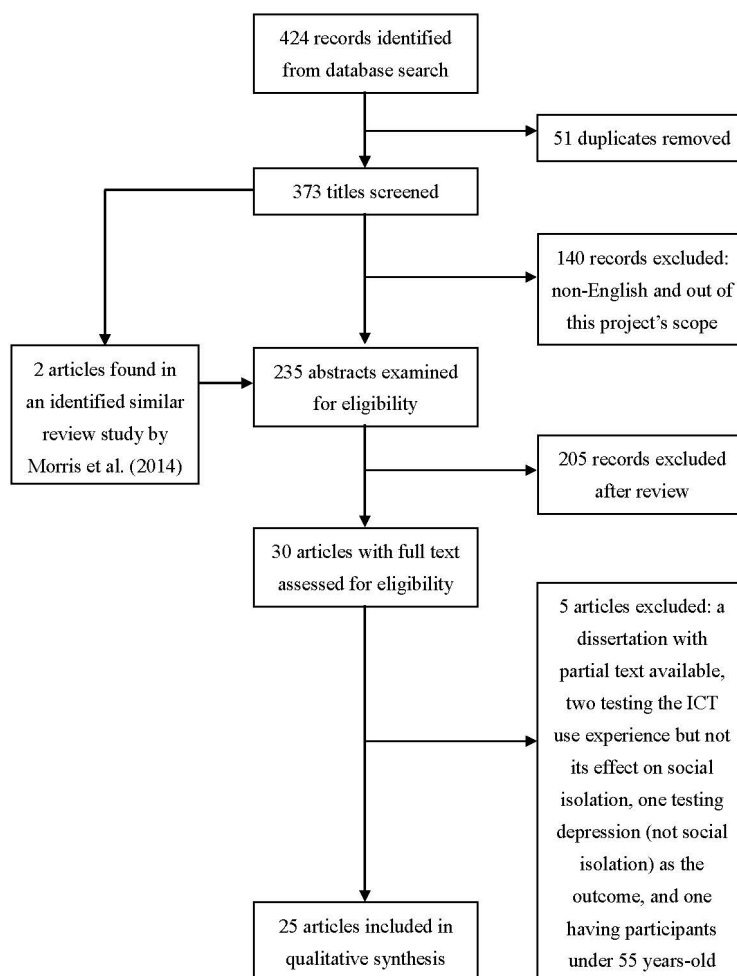
Data Analysis and Synthesis

Data on study design, sample size and characteristics, types of ICT applications, targets of elderly interaction via ICT applications, comparison groups, and outcomes were extracted from the selected studies and analyzed using a coding scheme. For research quality assessment of quantitative research, the Effective Public Health Practice Project (EPHPP) tool [26] was used because of its suitability for assessing such research with various study designs. The EPHPP tool evaluates 6 components of a quantitative study: selection bias, study design, confounders, blinding, data collection method, and withdrawals and dropouts. Based on performance in each component, an overall rating (ie, strong, moderate, or weak) of each study can be determined. The criteria proposed by Salmon [27] were used to evaluate the qualitative research: theoretical framework, value of study, data collection, participant description, data analysis, and data interpretations. For publications reporting more than one study, each study was independently analyzed. Data coding and quality appraisal were conducted by the first author and a research assistant, reaching an intercoder reliability of .91. Any inconsistencies between the reviewers were discussed between the 2 authors to achieve agreement.

The included studies differed in their research designs, research locations, participant characteristics, types and usage of interventions, and outcome measures. In view of the studies' heterogeneity, a narrative synthesis (instead of a meta-analysis) was performed [28].

Results

Characteristics of Examined Studies

All projects were published between 2002 and 2015, with 11 dated before 2010 [29-39] and 14 dated in or after that year [40-53]. They were conducted in 12 countries (Austria, Canada, Finland, Israel, Netherlands, New Zealand, Norway, Slovenia, Sweden, Taiwan, United Kingdom, and United States) with the highest number coming from the United States (n=9). In the 25 projects, 30 studies were reported (5 projects reported 2 studies: 1 quantitative and 1 qualitative). RCTs comprised 6 studies [36,39,43,47,51-52]; another 6 were cohort studies (2 with a control group [30,35] and 4 without [31,38,41,44]). Of the remaining studies, 4 were cross-sectional studies (surveys) [32,37,40,46] and 14 were qualitative studies: 9 employing in-depth interviews [29-31,34-35,42,45,48-49], 3 conducting focus group interviews [33,38,53], and 2 applying participant observations [39,50]. See Multimedia Appendix 1 for a complete description of the characteristics of the 25 reviewed publications.

Most research used some form of Internet or Web-based apps (eg, search, email, online chat rooms, videoconferencing, social

networking apps, and Web-based telehealth systems) on computers. Among those that did not, 1 study employed a telephone befriending intervention, 1 used mobile phones (smartphones), 1 focused on iPad use, 1 applied Nintendo Wii (a video game system), and 1 used a visual pet companion app that allowed the senior users to interact with a pet avatar in real time through an android tablet. The ICT intervention in all but 2 studies was implemented in the regular living environments of the participants, including private housing (n=13), assisted and independent living communities (n=2), congregate housing sites (n=1), retirement villages (n=2), nursing homes (n=4), day care centers (n=1), and no specifics on where they resided (n=2). The intervention was implemented in both settings for 3 projects [35,39,41] that worked with participants from 2 selected residential settings. The visual pet intervention [50] was mobile, and the participants used it in their familiar surroundings not bound by their living environment. The Finland study by Blažun et al [41] had the intervention set up at a community college. The ICT intervention in all but 2 studies aimed to facilitate interaction with other people in general (the participants most frequently contacted their family members, friends, significant others, doctors, and acquaintances made through online chat rooms). The other 2 studies [51-52] designed the intervention for the older person's interaction with family members only.

Characteristics of Participants

Sample size of the studies varied from 8 to 5203. The number of participants in the RCT studies ranged from 22 to 205. The sampling strategy of most studies (n=25) was convenience sampling; 3 [36,40,46] used random sampling and 2 [43,48] did not specify the sampling strategy.

Participants' average age ranged from 66 years (SD not given) to 83 years (SD 1.4) with heterogeneity across demographics, including age, gender, education, income, health status, mental status, living arrangements, and nationalities. Of the chosen studies, 1 [45] had equal numbers of male and female participants and 1 [38] had more males (n=17) than females (n=15). For the remaining studies, female participants often far outnumbered males even though ICT use among the elderly is highly associated with males [54]. There are two possible explanations for this phenomenon: (1) women have longer life expectancy across nations than men and (2) women are more likely to feel lonely and communicate with others and are therefore more likely to participate in such studies. Of the 25 projects, 2 [30,42] recruited living-alone elders only. In terms of participant health characteristics, 4 projects [29,36,41,47] targeted elderly people in generally good health and 5 [30-31,34,38,42,50] looked at those at high risk (lonely, frail, or chronically ill or physically handicapped seniors, those having dementia, and carers of spouses with dementia or after stroke). The other studies did not use health status as a filtering criterion for sampling. For other participant characteristics, 1 study [49] examined elderly former Soviet Union immigrants in Israel with financial difficulties and 1 [33] targeted elderly people interested in computer use. See Multimedia Appendix 2 for details on participant characteristics.

According to the EPHPP quality assessment tool, an attrition rate of 40% or above indicates weak data collection for a study

[26]. Based on this standard, 3 studies [30-31,52] were assessed to have a large number of dropouts. Fokkema and Kinpscheer [30] targeted solitary, lonely seniors with chronic illness or physical disability. The participants' physical and psychological conditions might account for the high attrition rate (43%) even considering the addition of 6 participants from the waiting list to replace the first 8 dropouts. The other 2 studies [31,52] were longitudinal projects lasting 12 months. The study duration contributed to the high rate of participant attrition, especially in a study targeted at the elderly in nursing homes [52]. Of the participants in Mellor's study [31], 60% were lost in the follow-up. In the study by Tsai et al [52], 44% of the participants in the control group did not complete the study. Of particular concern was a study by Machesney el al [50] in which the number of dropouts was unfortunately not specified but referred to as "several."

Dependent Variables and Outcome Measures

The outcome of ICT use was examined in 4 studies [42,43,48,49] by exploring its effect on social isolation in general, while the remaining studies assessed specific aspects of social isolation only. Social isolation as an outcome indicator was only quantitatively measured by Cotton et al [43] using a self-developed scale that contained 3 items, asking how frequently the participant was bothered by (1) not having a close companion, (2) not having enough friends, and (3) not seeing enough of people they feel close to. The other 3 qualitative studies did not clearly define the term. Catton and colleagues [42] seemed to regard social isolation as being forgotten and not belonging. Kahlbaough et al [47] and Karimi and Neustaedter [48] linked the concept to "not being connected with family, friends, and existing contacts." It should be noted that researchers in the 3 qualitative studies perceived social isolation and loneliness as highly interrelated, if not interchangeable, while Cotton et al [43] analyzed social isolation and loneliness as two separate outcomes of ICT use.

Studies that examined ICT impact on social isolation did so by looking at its effect on 1 or more of the 7 single attributes of social isolation: loneliness, social support, social contact, number of confidants, social connectedness/social connectivity, social networks, and social well-being. Among these, loneliness was the most tested dependent variable (n=18). It was measured by the University of California Los Angeles Loneliness Scale in 20 of the 25 projects. Fokkema and Knipscheer [30] used de Jong-Gierveld and Kamphuis' loneliness scale [55] whereas Aarts et al [40] used the scale's short version of 6 items [56]. Heo et al [46] employed the social support scale by Schuster et al [57] to assess loneliness while Sum et al [37] adopted the Social and Emotional Loneliness Scale. Rather than using a standardized scale, Blažun et al [41] used self-reported items of loneliness by the elderly participants in their pre-intervention survey to evaluate outcomes.

Social support was assessed by Tsai and colleagues [51-52] using Hsiung's Social Support Behaviors Scale, which includes subscales regarding (1) number of social networks, (2) quantity of social support behavior (emotional, informational, instrumental, and appraisal support), and (3) satisfaction with social support. The social support instrument used by Torp et

al [38] was adopted from the scale developed by Russel and colleagues [58]. Torp et al also examined social contact as another outcome indicator, applying Andersson's [59] Family and Friendship Contacts Scale. Social well-being was conceptualized as a multidimensional variable by Slegers et al [36] and was measured using de Jong-Gierveld and Kamphuis' loneliness scale and the number of social networks. Social connectedness/social connectivity and social networks examined in the reviewed qualitative studies were not clearly defined but were related to the number of connections with others and/or with society at large. Mellor et al [31], however, measured social connectedness using Lee and Robin's Social Connectedness Scale [60] in their cohort study.

It is worth noting that even though depression is not a dimension of social isolation, it is a related concept that attracts much academic attention. Of the reviewed studies, 6 also examined depression as an outcome variable [35,39,42,50-52]. This research tendency reflects the previous findings that social isolation leads to depression (a negative indicator of psychological well-being) among the elderly. Self-esteem, self-control, and quality of life were the other related outcomes of ICT intervention tested in the studies.

Effects of ICT Interventions on Alleviating Social Isolation

Of the studies addressing the relationship between ICT usage and social isolation in general, 4 demonstrated a positive result: the use of telephone befriending programs [42], computer and Internet [43,49], and ICT in general [48] lessened social isolation. The reported effect of ICT use on the individual dimensions of social isolation was consistent across studies, except for that on loneliness. ICT interventions significantly fostered social support, social contacts, social connectedness/social connectivity, and social networks among the participants, but no effect was found on number of confidants [39] or social well-being [36].

Of the studies examining loneliness, 15 of 18 revealed a significant reduction of loneliness among the elderly using ICT. Studies using communication programs (using landline phones, smartphones, iPads, emailing, and online chat rooms or forums) and high-technology apps (Wii, the TV gaming system, and Gerijoy, a virtual pet companion) consistently reported a positive effect on alleviating loneliness. The general use of computer and Internet in an RCT design was assessed in 2 nonsignificant-result studies [36,39], with 1 [36] targeting healthy elderly people living at home and the other [39] targeting elderly people living in subsidized housing or nursing facilities. The remaining non-significant study [40] examined the use of social networking sites in particular. Considering that other studies reporting a significant effect of such interventions also used the RCT and survey design targeting the elderly with different levels of health status and in various living situations, it is evident that the effect of the computer and Internet and of social networking sites on improving loneliness among the elderly was inconclusive. Another inconclusive finding concerns the effect of videoconferencing on loneliness reduction among the elderly. Blažun and colleagues [41] found that Slovene participants at nursing homes reported no change of loneliness

level after their use of Skype, while loneliness of Taiwanese nursing home participants was significantly lessened after their videoconferencing via Skype or Windows Live Messenger [51-52].

Furthermore, Sum et al [37] found that computer and Internet use functioned differently for various types of loneliness: social loneliness, family loneliness, and romantic loneliness. Using computers and the Internet to communicate with acquaintances alleviated elderly people's social loneliness, but heavy usage (of long duration) was positively associated with social loneliness. In addition, using the computer and Internet to make new contacts resulted in family loneliness. The impact of computer and Internet use on romantic loneliness was not determined.

Internet use increased social support among the elderly in general [46] and among those who were the main carers of their spouses with dementia or after a stroke in particular [38]. In a similar vein, Nahm's [32] survey data revealed a positive function of the elderly population's Internet use in building computer-mediated social networks, which led to social support. Interview data from the study by Dhillon et al [45] suggested that ICT (such as Facebook or networking games) fostered social interaction and social support that further alleviated loneliness among the elderly. Tsai et al [51] found that videoconferencing chats between elderly people at nursing homes and family members significantly increased emotional (ie, caring, empathy, love, and trust) and appraisal (ie, communicating information relevant to self-evaluation) support but not informational (ie, communicating information for problem-solving assistance) or instrumental (ie, tangible goods, services, and aid) support. However, this positive effect on social support was not found at the 6-month or 12-month stages of the intervention [52]. In addition, videoconferencing chats gave lower perceived instrumental support at the 6-month or 12-month stages while the frequency of in-person visits was not changed. The instrumental support finding may, as claimed by the researchers, imply that video chats assisted the elderly in better adapting to the living environment in the nursing home. Thus, their need for tangible goods, services, or aid dropped as their length of residence increased.

The relationship between ICT use and social connectivity/social connectedness or social networks was tested in 6 projects, which reported a generally consistent pattern. ICT in general (Internet, mobile/smartphones, iPads, social networking sites, and audio/video chat apps) served as an effective means for the elderly to remain connected with others [31,35,44,48] and expand their social networks [32,53]. It is important to note that Mellor and colleagues [31] reported that elderly people's use of computer and Internet at home increased their social connectedness at the 3-month stage of intervention but not at the 6-month or 9-month stage.

In addition to the social-isolation dimensions, a few studies explored the impact of ICT on related constructs, including depression, anxiety, negative affect, cognition, physical functioning (or daily activities), self-control (or perceived control), self-esteem, and quality of life (or life satisfaction). The results pertaining to the effect of ICT use on depression

XSL·FO

RenderX

were consistent and generally positive with only 1 study [50] reporting no clear information in the results section. While 2 of the 3 studies concluded that ICT led to positive effects, 1 [31] reported inconclusive results in this outcome. A favorable influence of ICT on life satisfaction was revealed in 4 studies [35,42,47,50], while 2 studies [31,39] found non-significant change in life satisfaction after using ICT. Neither self-esteem [31,39] nor control over life [36,39] was identified as a significant outcome of ICT use. Perception of self-control was, however, significantly increased after accessing ICT [35]. ICT use was found to improve physical health in the elderly [42]. Its effect on increasing the physical activities was inconclusive [36,47].

Lastly, a few studies assessed the effect of ICT use on the quantity and quality of communication of the elderly with others, and 3 studies [39,43,53] found a positive outcome. More elderly participants from Clark's study [29] stated that the Internet chat environment did not confine their messages for communication than said it did.

Quality Assessment of Examined Studies

While 4 quantitative studies were rated as moderate [35,43,46,52] and 4 as strong [36,39,47,51], 8 were rated as weak [30-32,37-38,40-41,44]. Among the 4 strong studies, Slegers et al [36] stood out with its rigorously controlled, randomized design. After randomizing twice, it compared the long-term effect of computer and Internet use on loneliness among the elderly in 4 conditions: training-intervention, training-no intervention, no training-no intervention (people in this condition had an interest in computer and Internet use), and control group (people here were not at all interested in ICT use).

The majority of the studies used a convenience sample that resulted in a high risk of selection bias. Consistent with the evaluation results of Morris et al [25], all but 6 of the chosen quantitative studies [36,40,44,49,51-52] failed to specify the proportion of the source population participating or the proportion of those who agreed to participate in the assigned group. The lack of such information makes it hard to determine the samples' representativeness. Of the 6 studies, 2 [40,44] had a participation proportion of less than 60%. None of the studies report any attempt to blind the participants from the intervention outcomes being examined. Information about whether the assessor (or caregiver) was aware of the intervention was very limited. There were 2 studies [30,43] that did not examine the possible differences between the experiment and control groups prior to the intervention. Furthermore, 3 studies [36,43,50] did not specify the duration of ICT intervention, and in 2 studies [35,43] the format of ICT training (ie, individual or group training) was not reported. Information about the training format is necessary because the literature suggests a relationship between the format and effectiveness of training for the elderly, who are likely to be slow learners of ICTs [31,39].

Of the 16 quantitative studies, 7 controlled for confounding factors in the analyses of effects of ICT interventions. Such factors included number of friends and family [43], physical/emotional/social limitations [43], number of children [35], positive life events [35], personal motivations (ie, learning new skills and gaining attention from others) [36], personality,

perceived psychological health [34], length of residency in nursing homes [51-52], educational level [40], sex [40], and age [40,51-52]. Of particular concern was the high percentage of participants who dropped out over the course of the trial in a few studies and the lack of power analysis conducted in all but 1 [36] of the RCT studies.

Among the qualitative studies, 4 [30-31,35,38] were conducted as a secondary analysis to provide further insights into the results of a (randomized) quantitative study. The quality of most qualitative studies was low because the authors failed to address several key areas, as proposed by Salmon [27]. First, most such studies (whether stand-alone or secondary) were descriptive or exploratory without examining specified propositions derived from the literature, while 3 studies [33,49,53] discussed the findings based on theories. Second, the interviewee recruitment processes were not clearly specified. Most studies reported the sampling frame and characteristics of the interviewees but some failed to provide information about the recruitment method (eg, randomly, purposively, or conveniently recruited) and others did not explain how a certain location, nursing home, or community was selected and why. Additionally, even though most studies reported the number of interviewees, none of them mentioned whether the number was a result of theoretical saturation. Without such information, readers are unable to determine the appropriateness and richness of the data. Furthermore, most studies [30-31,34-35,39,48,50] did not clearly report how the data were collected and analyzed. They often failed to report the interview protocol or the coding procedure even though some did state how they identified the emerging themes. This information is crucial because the researchers' approach to the data directly determines what the findings are. Lastly, a serious concern was that many authors reported the data superficially without interpretation or implications.

Discussion

This systematic review is, to our knowledge, the first to address the potential of ICT for preventing or reducing social isolation, a state that implies the risk of deteriorating physical and psychological well-being for the elderly. The results of this systematic review provide emerging quantitative and qualitative evidence to support the function of ICT in alleviating social isolation (in general or in particular dimensions) among elderly people. This review advances the mechanism of how ICT assists the elderly in combating social isolation and provides insights for policies and practices.

Social Isolation as an Untested Concept

Most studies of the review evaluated the effect of ICT use on single social-isolation dimensions, including loneliness, social support, and social connectedness. This pattern is consistent with that revealed in the review of studies by Dickens and colleagues [20] on social isolation interventions for the elderly, where only 2 of the 32 studies used social isolation as an outcome variable while the remaining studies mostly assessed loneliness, social network size, and social support. These findings suggest that social isolation of the elderly as a multidimensional concept is largely understudied [20,22]. Even though evidence shows that the use of ICT affects specific

aspects of social isolation, its effect on the overall perception of social isolation remains largely unknown. Therefore, more research is needed to unlock the relationship between ICT interventions and social isolation reduction.

The limited examination of the general concept of social isolation as a multidimensional construct might be a result of the lack of an appropriate scale. Sansoni et al [61] found the following 4 instruments of social isolation to be the leading ones in the literature: the Lubben Social Network Scale [62-64], the de Jong-Gierveld Loneliness Scale [55-56], the Medical Outcomes Study Social Support Survey [65], and the Multidimensional Scale of Perceived Social Support [66]. These instruments are clearly designed for measuring particular aspects rather than the overall concept of social isolation. Future research is required to develop a reliable scale of social isolation as a multidimensional variable by first discovering the concept's underpinnings from the perspective of the elderly. For instance, the comprehensive review of social isolation literature conducted by Nicholson [22] identified 5 key attributes of social isolation: (1) belonging, (2) social contacts, (3) quality of relationships, (4) fulfilling relationships, and/or (5) engagement. When investigating the complex relations between social isolation and health, Cornwell and Waite [8] operationalized social isolation as a variable of multiple components (ie, social contact frequency, social network size, social activity, loneliness, and social support) integrated into 2 forms: social disconnectedness and isolation. Further evaluation should be performed to validate the applicability of the instruments for measuring single aspects of social isolation versus that of the instruments tapping social isolation as a multidimensional construct. Even though some dimensions of social isolation were addressed by the studies included in this review, there are still some such as quality of relationships and engagement that remain untested in relation to the effect of ICT intervention. Researchers should explore these dimensions in future studies to advance our understanding of social isolation.

The Mechanism of ICT in Alleviating Social Isolation

ICT use consistently affected social isolation in general, social support, and social connectedness positively, but the positive ICT effect on social connectedness and social support rarely lasted for more than 6 months after the intervention. The results for loneliness were inconclusive. The results for self-esteem and control over life were consistently nonsignificant.

After triangulating the quantitative and qualitative data of the included studies in this review, it is suggested that the elderly's employment of ICT reduces their social isolation through the following mechanisms: connecting to the outside world, gaining social support, engaging in activities of interest, and boosting self-confidence. ICT helps the elderly stay connected with their family members (especially grandchildren), friends, former colleagues, acquaintances, and new contacts of shared interests or needs across temporal and geographical boundaries via digital interactions. Connections lead to social inclusion and foster social support. ICT also allows elderly people to renew their hobbies or competence and participate in enjoyable activities without the time constraint. Most importantly, ICT use boosts self-confidence among the elderly by making them "connected

to information," "feel young," "become one of the modern generation," "overcome challenges," "equip themselves with new skills," "stay socially active," and "help others online." It is worth noting that providing advice to the younger generation (acquainted or unknown) has a significant positive impact on the elderly population's self-confidence. The self-confidence gained leads to self-efficacy that goes beyond the use of ICT and participating in social activities. ICT use also empowers the elderly by engaging them in critical thinking and decision-making and providing access to information and resources. Self-confidence and empowerment further trigger their positive feelings toward themselves and their control over life and/or life satisfaction. Thus, a further examination of self-efficacy, mastery, and empowerment as outcomes should be promising for theory building in the field of social isolation.

ICT Use Among the Elderly

The findings of this review suggest that the elderly can benefit from ICT interventions and will use them (sometimes frequently) after proper training. At the same time, the high attrition rate of participants in the trials and the inconclusive results of ICT impact on loneliness reduction imply that ICT is not suitable for every senior. Spatial (eg, home-bound or institutionalized) and social (eg, immigrants or spousal carers) barriers to socialization, interest in ICT, motivations for ICT use, cognitive capability, sufficient eyesight, and basic physical ability to use the equipment (eg, figure or hand movement, skills of using the touch pad) are possible predictors of the suitability of ICT for the elderly. Furthermore, tailor-made training for the elderly (in terms of its setting, procedure, materials, timing, and instructor's style and attitude) is necessary for a maximum positive effect of the ICT on alleviating social isolation.

There are different mechanisms by which ICTs influence different kinds of loneliness and social support among the elderly [37,51-52]. The results reveal the interplay between the ICT-mediated activity and the effect of such behavior on particular types of loneliness and social support. Considering that there were only 2 studies addressing the types of loneliness and 2 examining types of social support, future research on these topics should advance the understanding of ICT's role in alleviating social isolation. Results of such research can provide insights into which individuals among the elderly can most benefit from ICT to reduce their loneliness or increase their social support in particular cases.

Future Development

The majority of the reviewed studies tested the ICT intervention as a one-time trial among a small number of participants. Thus, the generalizability of the results is limited. Further examination is needed to test their applicability.

Most ICT interventions examined in this systematic review involved the use of the computer and Internet. With the rapid development of ICT, other types of interventions should be explored. As stated by some interviewed participants of the reviewed studies, the use of ICT allowed them to adjust to their younger family members;#8217 communication style and preferences. As a result, it enhanced the quantity and quality of their intergenerational communication. Similarly, Clark [27]

observed that use of a particular platform, if one has a sufficient number of friends, lessens social isolation. ICT that is currently prevalent—instant messaging (eg, WhatsApp, Line, Snapchat), YouTube videos, and social networking sites (eg, Facebook, Instagram)—should be further investigated for the potential in reducing social isolation among the elderly. For example, Harley and Fitzpatrick [67] found that YouTube allowed a senior user to engage in communication beyond the family context with younger YouTubers who shared his interests using self-made videos (ie, videoblogging). Such behavior further fulfilled the senior's social and emotional needs and increased his self-confidence. Also of interest for further research are mobile phone apps, because elderly people demonstrate a fast-growing rate of mobile phone-based ICT adoption across their age groups in wealthy countries [68].

Additionally, the results of this review suggest that ICT use does not guarantee quality of communication. For example, when the ICT-mediated communication is not reciprocal, the ICT use could increase social isolation among the elderly [48]. Consequently, examining how to use ICT for generating quality communication between the elderly and others (eg, using videoconferencing for the elderly to virtually join family activities) can be a promising subject for future research on social isolation and/or intergenerational communication.

Strengths and Weaknesses of the Review

This systematic review tackled an emerging trend of social isolation research: ICT interventions for reducing social isolation in the elderly. The comprehensive search strategy and the inclusion of studies of all designs increased the likelihood of including all relevant studies in the field. Presenting the results of both randomized and nonrandomized research might be a limitation of this study. However, this review decision broadened our exploration of the available social-isolation interventions and their effectiveness and helped to better achieve the objective of this study.

The heterogeneity of studies included in this review limits the comparability and generalizability of our results. Although restricting the scope to studies published in English might introduce bias, the reviewed studies were conducted in America, Europe, and the Asia-Pacific region.

Conclusion

This systematic review has suggested a need for more well-designed studies on the effect of ICT interventions on the social isolation of elderly people. ICT in general is a promising tool for tackling social isolation of the elderly, but it is not for every senior. Research identifying who among the elderly can most benefit from ICT use and how the training and implementation of such intervention should be tailored to maximize its effect offers great value for clinical practice. In addition, with the rapid development of ICT, the effectiveness of other types of interventions (eg, mobile phone-based instant messaging apps and YouTube videos) in reducing social isolation should be empirically examined. Results of such research can facilitate innovative and effective practice of ICT-based social isolation interventions for elderly people.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of the selected studies.

[PDF File (Adobe PDF File), 247KB - jmir_v18i1e18_app1.pdf]

Multimedia Appendix 2

Participant characteristics of the selected studies.

[PDF File (Adobe PDF File), 160KB - jmir_v18i1e18_app2.pdf]

References

1. World Population Ageing 2007. New York: United Nations; 2007. Magnitude and speed of population ageing URL: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeingReport2007.pdf [accessed 2015-11-13] [WebCite Cache ID 6d0OGYWLn]
2. Victor C, Bowling A, Bond J, Scambler S. Loneliness, social isolation and living alone in later life. Sheffield, UK: University of Sheffield: Economic and Social Research Council Growing Older Programme; 2003.
3. Iliffe S, Kharicha K, Harari D, Swift C, Gillmann G, Stuck AE. Health risk appraisal in older people 2: the implications for clinicians and commissioners of social isolation risk in older people. Br J Gen Pract 2007 Apr;57(537):277-282 [FREE Full text] [Medline: 17394730]
4. Tomaszewski W, Barnes M. Investigating the dynamics of social detachment in old age. In: Banks J, Breeze E, Lessof C, Nazroo J. editors. Living in the 21st century: older people in England. The 2006 English Longitudinal Study of Ageing (wave 3). London: Institute of Fiscal Studies; 2008:150-185.
5. Hawthorne G. Measuring Social Isolation in Older Adults: Development and Initial Validation of the Friendship Scale. Soc Indic Res 2006 Feb 13;77(3):521-548. [doi: 10.1007/s11205-005-7746-y]

XSL•FO

RenderX

6.   Hawthorne G. Perceived social isolation in a community sample: its prevalence and correlates with aspects of peoples' lives. Soc Psychiatry Psychiatr Epidemiol 2008 Feb;43(2):140-150. [doi: 10.1007/s00127-007-0279-8] [Medline: 17994175]

7.   Saito M, Fujiwara Y, Kobayashi E, Fukaya T, Nishi M, Shinkai S. [Prevalence and characteristics of social isolation in the elderly in a dormitory suburb according to household composition]. Nihon Koshu Eisei Zasshi 2010 Sep;57(9):785-795. [Medline: 21061557]

8.   Cornwell EY, Waite LJ. Social disconnectedness, perceived isolation, and health among older adults. J Health Soc Behav 2009 Mar;50(1):31-48 [FREE Full text] [Medline: 19413133]

9.   Mulvaney-Day NE, Alegría M, Sribney W. Social cohesion, social support, and health among Latinos in the United States. Soc Sci Med 2007 Jan;64(2):477-495 [FREE Full text] [doi: 10.1016/j.socscimed.2006.08.030] [Medline: 17049701]

10.   Fiori KL, Smith J, Antonucci TC. Social network types among older adults: a multidimensional approach. J Gerontol B Psychol Sci Soc Sci 2007 Nov;62(6):P322-P330. [Medline: 18079416]

11.   Kharicha K, Iliffe S, Harari D, Swift C, Gillmann G, Stuck AE. Health risk appraisal in older people 1: are older people living alone an "at-risk" group? Br J Gen Pract 2007 Apr;57(537):271-276 [FREE Full text] [Medline: 17394729]

12.   Dorfman RA, Lubben JE, Mayer-Oakes A, Atchison K, Schweitzer SO, De Jong FJ, et al. Screening for depression among a well elderly population. Soc Work 1995 May;40(3):295-304. [Medline: 7761915]

13.   Hanson BS. Social network, social support and heavy drinking in elderly men--a population study of men born in 1914, Malmö, Sweden. Addiction 1994 Jun;89(6):725-732. [Medline: 8069173]

14.   Eng PM, Rimm EB, Fitzmaurice G, Kawachi I. Social ties and change in social ties in relation to subsequent total and cause-specific mortality and coronary heart disease incidence in men. Am J Epidemiol 2002 Apr 15;155(8):700-709 [FREE Full text] [Medline: 11943687]

15.   Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000 Sep;51(6):843-857. [Medline: 10972429]

16.   Locher JL, Ritchie CS, Roth DL, Baker PS, Bodner EV, Allman RM. Social isolation, support, and capital and nutritional risk in an older sample: ethnic and gender differences. Soc Sci Med 2005 Feb;60(4):747-761 [FREE Full text] [doi: 10.1016/j.socscimed.2004.06.023] [Medline: 15571893]

17.   Havens B, Hall M. Social isolation, loneliness, and the health of older adults in Manitoba, Canada. Indian J Gerontol 2001;1(15):126-144.

18.   Boden-Albala B, Litwak E, Elkind MSV, Rundek T, Sacco RL. Social isolation and outcomes post stroke. Neurology 2005 Jun 14;64(11):1888-1892. [doi: 10.1212/01.WNL.0000163510.79351.AF] [Medline: 15955939]

19.   Sum S, Mathews MR, Pourghasem M, Hughes I. Internet technology and social capital: How the Internet affects seniors' social capital and wellbing. J Comput Mediat Commun 2008;14(1):202-220. [doi: 10.1111/j.1083-6101.2008.01437.x]

20.   Dickens AP, Richards SH, Greaves CJ, Campbell JL. Interventions targeting social isolation in older people: a systematic review. BMC Public Health 2011;11:647 [FREE Full text] [doi: 10.1186/1471-2458-11-647] [Medline: 21843337]

21.   Cattan M, White M, Bond J, Learmouth A. Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. Ageing and Society 1999;25(1):41-67. [doi: 10.1017/S0144686X04002594]

22.   Nicholson NR. Social isolation in older adults: an evolutionary concept analysis. J Adv Nurs 2009 Jun;65(6):1342-1352. [doi: 10.1111/j.1365-2648.2008.04959.x] [Medline: 19291185]

23.   Wenger GC, Burholt V. Changes in levels of social isolation and loneliness among older people in a rural area: a twenty-year longitudinal study. Can J Aging 2004;23(2):115-127. [Medline: 15334812]

24.   Findley RA. Interventions to reduce social isolation amongst older people: where is the evidence? Ageing and Society 2003;23(5):647-658. [doi: 10.1017/S0144686X03001296]

25.   Morris ME, Adair B, Ozanne E, Kurowski W, Miller KJ, Pearce AJ, et al. Smart technologies to enhance social connectedness in older people who live at home. Australas J Ageing 2014 Sep;33(3):142-152. [doi: 10.1111/ajag.12154] [Medline: 24730370]

26.   Quality Assessment Tool for Quantitative Studies: Effective Public Health Practice Project. 2009. URL: http://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf [accessed 2015-05-01] [WebCite Cache ID 6YCW3AIuz]

27.   Salmon P. Assessing the quality of qualitative research. Patient Educ Couns 2013 Jan;90(1):1-3. [doi: 10.1016/j.pec.2012.11.018] [Medline: 23237574]

28.   Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks, Calif: Sage; 2001.

29.   Clark DJ. Older adults living through and with their computers. Comput Inform Nurs 2002;20(3):117-124. [Medline: 12021610]

30.   Fokkema T, Knipscheer K. Escape loneliness by going digital: a quantitative and qualitative evaluation of a Dutch experiment in using ECT to overcome loneliness among older adults. Aging Ment Health 2007 Sep;11(5):496-504. [doi: 10.1080/13607860701366129] [Medline: 17882587]

31.   Mellor D, Firth L, Moore K. Can the Internet improve the well-being of the elderly? Ageing Int 2008 Feb 9;32(1):25-42. [doi: 10.1007/s12126-008-9006-3]

32.   Nahm ES, Resnick B, Mills ME. A model of computer-mediated social support among older adults. AMIA Annu Symp Proc 2003:948 [FREE Full text] [Medline: 14728453]

33.   Richardson M, Weaver CK, Zorn Jr TE. 'Getting on': older New Zealanders' perceptions of computing. New Media & Society 2005 Apr 01;7(2):219-245. [doi: 10.1177/1461444805050763]

34.   Savolainen L, Hanson E, Magnusson L, Gustavsson T. An Internet-based videoconferencing system for supporting frail elderly people and their carers. J Telemed Telecare 2008;14(2):79-82. [doi: 10.1258/jtt.2007.070601] [Medline: 18348753]

35.   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]

36.   Slegers K, van Boxtel MPJ, Jolles J. Effects of computer training and Internet usage on the well-being and quality of life of older adults: a randomized, controlled study. J Gerontol B Psychol Sci Soc Sci 2008 May;63(3):P176-P184. [Medline: 18559683]

37.   Sum S, Mathews RM, Hughes I, Campbell A. Internet use and loneliness in older adults. Cyberpsychol Behav 2008 Apr;11(2):208-211. [doi: 10.1089/cpb.2007.0010] [Medline: 18422415]

38.   Torp S, Hanson E, Hauge S, Ulstein I, Magnusson L. A pilot study of how information and communication technology may contribute to health promotion among elderly spousal carers in Norway. Health Soc Care Community 2008 Jan;16(1):75-85. [doi: 10.1111/j.1365-2524.2007.00725.x] [Medline: 18181817]

39.   White H, McConnell E, Clipp E, Branch LG, Sloane R, Pieper C, et al. A randomized controlled trial of the psychosocial impact of providing internet training and access to older adults. Aging Ment Health 2002 Aug;6(3):213-221. [doi: 10.1080/13607860220142422] [Medline: 12217089]

40.   Aarts S, Peek STM, Wouters EJM. The relation between social network site usage and loneliness and mental health in community-dwelling older adults. Int J Geriatr Psychiatry 2015 Sep;30(9):942-949. [doi: 10.1002/gps.4241] [Medline: 25503540]

41.   Blažun H, Saranto K, Rissanen S. Impact of computer training courses on reduction of loneliness of older people in Finland and Slovenia. Computers Hum Behav 2012 Jul;28(4):1202-1212. [doi: 10.1016/j.chb.2012.02.004]

42.   Cattan M, Kime N, Bagnall A. The use of telephone befriending in low level support for socially isolated older people--an evaluation. Health Soc Care Community 2011 Mar;19(2):198-206. [doi: 10.1111/j.1365-2524.2010.00967.x] [Medline: 21114564]

43.   Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]

44.   Delello JA, McWhorter RR. Reducing the digital divide: Connecting older adults to iPad technology. J Appl Gerontol 2015 Jun 17:1-26. [doi: 10.1177/0733464815589985] [Medline: 26084479]

45.   Dhillon JS, Ramos CR, Wunsche BC, Lutteroth C. Designing a Web-based telehealth system for elderly people: An interview study in New Zealand. In: Proceedings of the 24th International Symposium on Computer-Based Medical Systems -- CBMS 2011 Bristol.: IEEE; 2011 Jun Presented at: the 24th International Symposium on Computer-Based Medical Systems; 27-30 June 2011; Bristol p. 1-6 URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5999157 [doi: 10.1109/CBMS.2011.5999157]

46.   Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]

47.   Kahlbaugh PE, Sperandio AJ, Carlson AL, Hauselt J. Effects of playing Wii on well-being in the elderly: Physical activity, loneliness, and mood. Activities, Adaptation & Aging 2011 Oct;35(4):331-344. [doi: 10.1080/01924788.2011.625218]

48.   Karimi A, Neustaedter C. From high connectivity to social isolation: communication practices of older adults in the digital age. : Association for Computing Machinery; 2012 Presented at: Computer Supported Cooperative Work; February 11-15, 2012; Seattle WA.

49.   Khvorostianov N, Elias N, Nimrod G. 'Without it I am nothing': The internet in the lives of older immigrants. New Media & Society 2011 Nov 11;14(4):583-599. [doi: 10.1177/1461444811421599]

50.   Machesney D, Wexler SS, Chen T, Coppola JF. Gerontechnology Companion: Virutal pets for dementia patients. : IEEE; 2014 Presented at: Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island; 2-2 May 2014; New York p. 1-3. [doi: 10.1109/LISAT.2014.6845226]

51.   Tsai HH, Tsai YF, Wang HH, Chang YC, Chu HH. Videoconference program enhances social support, loneliness, and depressive status of elderly nursing home residents. Aging Ment Health 2010 Nov;14(8):947-954. [doi: 10.1080/13607863.2010.501057] [Medline: 21069600]

52.   Tsai HH, Tsai YF. Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents. J Med Internet Res 2011;13(4):e93 [FREE Full text] [doi: 10.2196/jmir.1678] [Medline: 22086660]

53.   Winstead V, Anderson WA, Yost EA, Cotten SR, Warr A, Berkowsky RW. You can teach an old dog new tricks: a qualitative analysis of how residents of senior living communities may use the web to overcome spatial and social barriers. J Appl Gerontol 2013 Aug;32(5):540-560. [doi: 10.1177/0733464811431824] [Medline: 25474761]

54.   Helsper EJ, Eynon R. Distinct skill pathways to digital engagement. European Journal of Communication 2013 Sep 16;28(6):696-713. [doi: 10.1177/0267323113499113]

55.   de Jong-Gierveld J, Kamphuls F. The development of a Rasch-type loneliness scale. Applied Psychological Measurement 1985 Sep 01;9(3):289-299. [doi: 10.1177/014662168500900307]

56.   de Jong-Gierveld J, van Tilburg T. A 6-item scale for overall, emotional, and social loneliness confirmatory tests on survey data. Res Aging 2006;28(5):582-598.

57.   Schuster TL, Kessler RC, Aseltine RH. Supportive interactions, negative interactions, and depressed mood. Am J Community Psychol 1990 Jun;18(3):423-438. [Medline: 2264558]

58.   Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [Medline: 7431205]

59.   Andersson L. Intervention against loneliness in a group of elderly women: A process evaluation. Human Relations 1984 Apr 01;37(4):295-310. [doi: 10.1177/001872678403700402]

60.   Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist 2006 Aug;46(4):503-513. [Medline: 16921004]

61.   Sansoni J, Marosszeky N, Sansoni E, Fleming G. Final report: Effective assessment of social isolation.: Centre for Health Service Development, University of Wollongong; 2010 Jan 1. URL: http://www.adhc.nsw.gov.au/__data/assets/file/0007/236329/24_Social_Isolation_Report.pdf [accessed 2015-05-01] [WebCite Cache ID 6YCWVY6DJ]

62.   Lubben JE, Gironda M. Centrality of social ties to the health and well-being of older adults. In: Berkman B, Harooytoan LK, editors. Social Work and Health Care in an Aging World. New York: Springer; 2003:319-350.

63.   Lubben J, Gironda M. Measuring social networks and assessing their benefits. In: Phillipson C, Allan G, Morgan D, editors. Social Networks and Social Exclusion: Sociological and Policy Perspectives. Burlington, VT: Ashgate Publishing Company; 2004:20-34.

64.   Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist 2006 Aug;46(4):503-513. [Medline: 16921004]

65.   Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [Medline: 2035047]

66.   Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess 1990;55(3-4):610-617. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]

67.   Harley D, Fitzpatrick G. YouTube and intergenerational communication: the case of Geriatric1927. Univ Access Inf Soc 2008 May 29;8(1):5-20. [doi: 10.1007/s10209-008-0127-y]

68.   Nielsen J. Seniors as Web users.: Nielsen Norman Group; 2013. URL: http://www.nngroup.com/articles/usability-for-senior-citizens/ [accessed 2015-05-01] [WebCite Cache ID 6YCW89Kk4]

Abbreviations

EPHPP: Effective Public Health Practice Project
ICT: information and communication technology
RCT: randomized controlled trial

Tutorial

Adapting Behavioral Interventions for Social Media Delivery

Sherry Pagoto[1], PhD; Molly E Waring[2], PhD; Christine N May[1], PhD; Eric Y Ding[1], MS; Werner H Kunz[3], PhD; Rashelle Hayes[1], PhD; Jessica L Oleski[1], MA

[1]Department of Medicine, Division of Preventive and Behavioral Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[2]Departments of Quantitative Health Sciences and Obstetrics & Gynecology, University of Massachusetts Medical School, Worcester, MA, United States

[3]Digital Media Lab, University of Massachusetts Boston, Boston, MA, United States

Corresponding Author:
Sherry Pagoto, PhD
Department of Medicine
Division of Preventive and Behavioral Medicine
University of Massachusetts Medical School
S7-751
55 Lake Ave N
Worcester, MA, 01655
United States
Phone: 1 508 856 2092
Fax: 1 508 856 3840
Email: Sherry.Pagoto@umassmed.edu

Abstract

Patients are increasingly using online social networks (ie, social media) to connect with other patients and health care professionals—a trend called peer-to-peer health care. Because online social networks provide a means for health care professionals to communicate with patients, and for patients to communicate with each other, an opportunity exists to use social media as a modality to deliver behavioral interventions. Social media-delivered behavioral interventions have the potential to reduce the expense of behavioral interventions by eliminating visits, as well as increase our access to patients by becoming embedded in their social media feeds. Trials of online social network-delivered behavioral interventions have shown promise, but much is unknown about intervention development and methodology. In this paper, we discuss the process by which investigators can translate behavioral interventions for social media delivery. We present a model that describes the steps and decision points in this process, including the necessary training and reporting requirements. We also discuss issues pertinent to social media-delivered interventions, including cost, scalability, and privacy. Finally, we identify areas of research that are needed to optimize this emerging behavioral intervention modality.

Introduction

The term Web 2.0, coined in 1999, refers to Web technology that allows users to interact and create content in virtual communities, which represents a divergence from static websites that only allow users a passive role as consumers of information. Social media is a broad example of Web 2.0 and refers to online social networking sites such as Facebook, Twitter, Reddit, Pinterest, and Instagram, as well as blogs and message boards, all of which are tools that allow users to engage with one another and generate their own content. Social media usage has exploded in recent years such that it is nearly ubiquitous, with 89% of US adults now using the Internet and the majority of those (74%) having at least one social network account [1,2]. Prevalence of social media use is highest among younger adults. The Pew Internet Project reported in January 2014 that 89% of 18-29-year-olds use online social networking sites compared to 82% of 30-49-year-olds, 65% of 50-64-year-olds, and 49% of adults aged 65 years or older [2]. Similar rates of use of social media have been reported for men (74%) and women (76%), and among blacks (75%), Hispanics (80%), and whites (70%) [3]. The vast majority of users log into their preferred networks daily [4]. In fact, Facebook recently reported that US adult users spend, on average, 40 minutes a day on Facebook [5].

Online social network use is no longer limited to keeping in touch with friends and family; many users now seek and exchange information about health [6], parenting [7], and a wide variety of other topics. The 2011 Pew Internet Survey found that 34% of Internet users have read a commentary or experience about health or medical issues on a website or blog [8]. People are not just in search of health information on the Internet, but are also in search of other patients [9]. The US Department of Health and Human Services Chief Technology Officer,

Susannah Fox, labeled this emerging trend as "peer-to-peer health care" and explains, "Patients know things—about themselves, about treatments—and they want to share what they know to help other people." Fox refers to "peer-to-peer health care" as "the most exciting innovation in health care today" [10]. That social media provides a means for health care professionals to communicate with patients and for patients to communicate with each other presents an opportunity to use this modality to deliver behavior change programs.

Figure 1. The process of adapting a behavioral intervention for social media delivery.
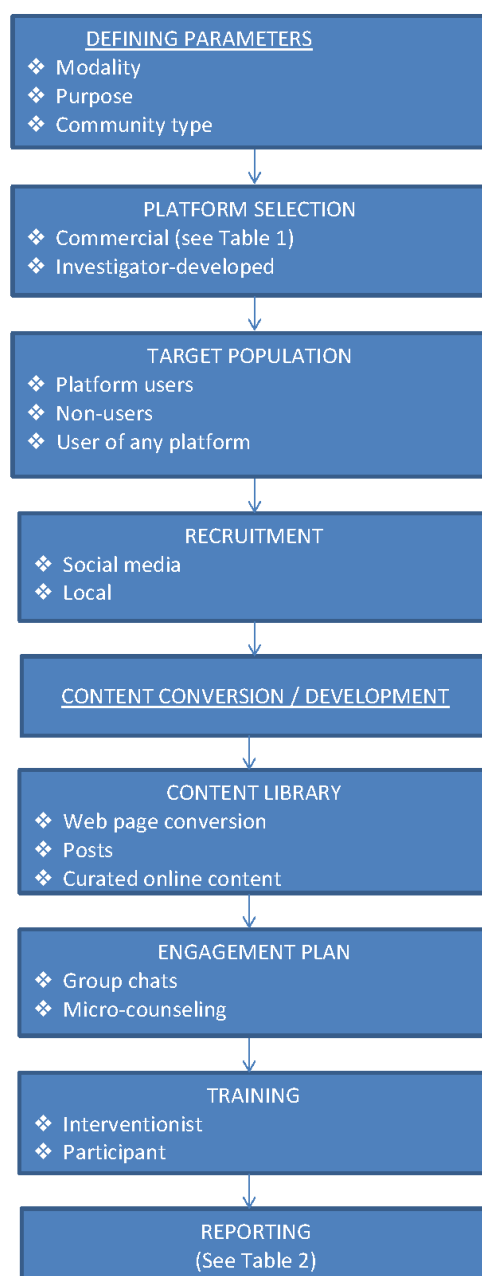
Scheduling constraints, family obligations, and lack of transportation negatively impact attendance in traditional in-person behavior change interventions [11]. The burden of the traditional intervention modality centers around the clinic visits (eg, high frequency and long duration) [12]. Delivering behavioral interventions via an online social network circumvents many of these barriers by reducing or eliminating visits. Using online social networks as a behavioral intervention

modality allows us to take advantage of how embedded online social networking is in people's lives. People can participate as part of their usual social media routine rather than carve out large chunks of time from their usual activities to participate. The absence of scheduled meetings allows people to engage flexibly each day, when the time is best for them. Advice, support, and cues can be provided in the moment, when participants need it the most. The ability to be "in the

participant's pocket" and deliver real-time counseling and support has great potential for changing behavior.

Using online social networks to deliver behavioral interventions is a relatively new idea, thus much work with regard to design, content, implementation, and reporting of trial outcomes and engagement is needed. In this paper, we discuss the process by which investigators can translate behavioral interventions for social media delivery. We present a model (see Figure 1) that describes the steps and decision points in this process, including the necessary training and reporting requirements. We also discuss issues pertinent to social media-delivered interventions, including cost, scalability, and privacy. Finally, we identify areas of research that are needed to optimize this new behavioral intervention modality.

Defining the Parameters of the Social Media-Delivered Intervention

Online Social Network as Primary Delivery Modality Versus Supportive Adjunct

In the context of behavioral interventions, an online social network may be (1) the primary modality of intervention, (2) one modality in a multimodal intervention, or (3) a supportive adjunct to an intervention that is primarily utilizing another modality (eg, visits, telephone, mobile app, and text messaging). When an online social network is the primary intervention modality, all intervention content is delivered via the online social network. This means all intervention content must be in a format that is consistent with formats typically used in that online social network. For example, on Twitter, posts are limited to 140 characters and often include links to webpages with more information. On Facebook, no character limit is in place and use of links, videos, and pictures is common. On Instagram, all content is in the form of photos or videos, and on YouTube, all content is video.

When an online social network is one modality in a multimodal intervention, some, but not necessarily all, key active ingredients of the behavioral intervention are delivered via the online social network. In this type of intervention, the content meant for social media delivery would need to be converted into a format that is consistent with content typically circulated in the target online social network.

When the online social network is an adjunct to an intervention that is delivered primarily via another modality, it might provide resources and information to users but not active ingredients of the intervention. Still, in this scenario, all resources (eg, links and tips) and information need to be converted into an appropriate format for the target online social network. For successful translation, the research team must be familiar with not only the features and capabilities of the online social network, but also with the norms of their target audience on that social network. Clearly reporting the role of the online social network in the intervention and the origin and purpose of the content disseminated via the network is essential.

Purpose of the Online Social Network

The purpose of the online social network may be to provide a means to deliver behavioral counseling; to provide supplemental information about behavioral strategies originally delivered via a different intervention modality (eg, clinic visits); to maximize attendance and retention in the intervention (eg, post reminders for meetings and follow-ups); to provide resources (eg, a list of local gyms for a physical activity study); to provide a place for participants to communicate, connect, and support each other; or any combination of the above. A major decision in intervention planning is whether the purpose involves facilitating participant engagement, information dissemination, or both [13].

When engagement is a purpose of the online social network, an engagement plan is needed. When information dissemination is the sole purpose of the online social network, efforts to elicit engagement are less important. In this case, a static website or email distribution list might also serve this purpose. The critical difference between a website and an online social network is the ability to facilitate user engagement. An advantage of an online social network over a static website is the ability to insert an intervention into participants' pre-existing social media feeds. This will only be the case, though, if the participants are regular users of the target social media platform. The combination of engagement and information dissemination may be the most effective approach, given studies showing that engagement in an online social network is associated with better outcomes on clinical end points [14-16].

Host- Versus User-Generated Content

In designing a social media-delivered intervention, matching the type of intervention to the purpose (eg, behavioral counseling platform, public health campaign, or information only) is a necessary consideration. Content in social media-delivered interventions may be host generated [17], such that the host generates all content and pushes it to users; user generated [18], such that users generate all content; or host and user generated [19,20], such that content is created by both the host and users. Host-generated feeds are unidirectional, such as in the case of public health campaigns, which involve a feed of information relating to a specific topic. Examples include the US Centers for Disease Control and Prevention (CDC) Twitter feeds (eg, @CDC_cancer) and the Skin Cancer Foundation Facebook page [21]. Users can comment on host posts and reply to each other's comments. Communities characterized primarily by user-generated content are typically started by a host who builds a forum for users to interact with one another, but does not provide content and does not typically interact with users (eg, PatientsLikeMe.com). Alternatively, a community may be initiated by a user for similar users to exchange information (eg, Fitbloggin Facebook page by a weight-loss blogger for weight-loss bloggers). Communities that include content that is both host and user generated are bidirectional such that both hosts and users generate content and interact with one another. Bidirectional communication is important when the purpose of the intervention is to be able to generate conversations, answer participant questions, help them solve problems, and/or provide behavioral counseling. Participants should be clear on the type

of community to which they have been invited (ie, host and/or user generated) and the expectations for engagement.

Social Network Platform Selection

Overview

Investigators can choose a commercial social networking platform, a commercial app that has a social networking function, or they may create their own social network platform. In terms of commercial platforms, many are freely available and have large user bases, for example, Facebook and Twitter. These two platforms have been in existence for 11 and 9 years, respectively, and consequently have had the benefit of long periods of refinement, user buy-in, and proven usability. Selecting the commercial platform that is most used by the target population will result in more openness to use and more experienced users, which may reduce engagement barriers. Investigators are referred to the Pew Survey of Internet and American Life report on the demographics of users of the most popular social media sites [22]. Alternatively, investigators may elect to utilize the social network capabilities of commercial mHealth apps. Many mobile apps (eg, Endomondo and MyFitnessPal) have social networking features that can be used to engage with participants individually and see data they have recorded using the mobile app. Finally, investigator-designed social network platforms, as in Brindal et al [23], can also be used when commercial platforms do not have the functionality to meet the intervention specifications.

Several considerations go into choosing the best platform. An investigator-designed social network requires technical expertise and overhead costs for development and maintenance. Developing an engaging user interface requires a high level of

technical expertise, cost, and development time. In a recent systematic review of weight-loss studies using online social networks, the authors explained that discussion boards and chat rooms used in investigator-designed websites lack the "sophisticated, user-friendly, vibrant platforms that incorporate a rich, pleasing, graphical environment" that are characteristic of mainstream online social network platforms [24]. Such an approach might be best matched for interventions that cannot be feasibly implemented using commercially developed platforms. An advantage of commercial platforms is that they are free and the technical skills required for use are becoming ubiquitous among many populations. A commercial platform that is heavily used by the target population (eg, targeted according to age, sex, and ethnicity) may not require extensive participant training. Software for managing, collecting, and analyzing social media data on many commercial platforms is also widely available and often free. However, researchers need to be aware of the limitations and challenges of using existing commercial online social networks, including privacy concerns, changing use policies, and changing data access policies. For example, Facebook at any time can change their application programming interface (API), which is the protocol that describes how different software programs can interact with Facebook. This may affect how data is extracted from Facebook and software compatibility. Data that was once accessible via certain means can become instantly inaccessible via those means. Consideration of the advantages and disadvantages of different platform options is an important step early in intervention development. The best platform for a particular behavioral target, intervention content, and target population may be inappropriate for other purposes. Table 1 describes the top nine commercial social media platforms [25-37].

Table 1. Characteristics of existing online social media platforms.

Platform	Year founded	Number of users	Medium of posts	Character limit	Private messages (yes/no)	Privacy functions allow creation of private groups (yes/no)	Chat function (yes/no)
Facebook	2004	1.44 billion [25]	Text, video, images	63,206 [26]	Yes	Yes	Yes
Twitter	2006	302 million [27]	Text, video, images	140	Yes	Yes	No
Pinterest	2010	72.8 million [28]	Text, video, images	500 [29]	Yes	Yes (private pin boards)	No
Snapchat	2011	100 million [30]	Video, images	31 for captions	Yes	No	Yes
LinkedIn	2002	364 million [31]	Text, images	600 [32]	Yes	Yes	Yes
Instagram	2010	300 million [33]	Video (<15 sec), images with captions	2200 for captions	Yes	No, but can send content to specific recipients	No
Google+	2011	300 million [34]	Text, video, images	100,000 [35]	Yes	Yes	Yes
Tumblr	2007	420 million users, 246 million blogs [36]	Text, video, images	No limit in blog posts; 250 in replies; 500 in "ask messages"	Yes	Yes (private group blog)	No
Vine	2013	40 million [37]	Video	N/A[a] (clips limited to <6 sec)	Yes	No	No

[a]N/A: not applicable.

Facebook

Facebook has two formats—fan pages and groups—that can be used for behavioral intervention delivery. Fan pages are public, where feeds can be followed when a user hits the "like" button of the fan page. These are ideal for public health campaign approaches. Groups, on the other hand, can be *public*, *private* (ie, only invited members can see content), or *secret* (ie, only invited members can see content, and existence of page is hidden) and can be used to generate conversations among a smaller group of people. A *secret* group is ideal when the investigators wish to keep all aspects of the intervention completely private.

An advantage of Facebook for behavioral intervention delivery is that it is the most popular social networking site, with 74% of US adults currently having an account. Participants may already be logging in regularly, which will bring them in regular contact with the intervention. Facebook also has settings in which communities can be created that allow users to participate privately even if their main account is public. Users can participate without their online friends being aware they are a member of the group. This might result in fewer privacy concerns and greater intervention receipt and participation. A drawback of Facebook is that it is not currently possible to change the organization of content. All exchanges appear on the "wall," or main feed of the group page, including exchanges between two users. This can result in a busier program feed or deter one-on-one interactions when participants do not feel comfortable having a side conversation appearing on the group wall. Another drawback is that the feed in a private group is organized in order of posts that have received the most recent interactions, as opposed to the chronological order of posts. This setting is not modifiable and, as a result, important content may fall to the bottom of the feed on days with heavy posting, making it difficult for participants to locate intervention content. Facebook allows the moderator to "pin" a single post to the top of the wall to circumvent this issue. Another disadvantage is that laypersons may not trust or understand Facebook's privacy settings, which can affect their interest in participating in the study [38].

A challenge specific to using fan pages on Facebook is that of intervention receipt. Facebook adjusts the number of posts from a fan page occurring in a user's feed to the level of interaction on that page by the user using a proprietary formula. This means that a participant with low initial engagement on the page will receive fewer updates from the page in their newsfeed and this may continue to decline as their engagement declines. Investigators can still boost receipt and engagement by analyzing audience engagement frequently and tailoring the social media strategy according to what the audience engages with most, as discussed elsewhere [39]. This does not occur in a *secret* group. Facebook settings are subject to change, thus investigators are encouraged to review current settings at the beginning of the study, test the setting before engaging study participants, and keep track of any setting changes during the study. We refer investigators to Facebook Help Center, which provides a wealth of information about settings and privacy.

An advantage of Facebook is that it is a way to leverage participants' close social ties given that Facebook is used to connect to existing friend networks. This may be useful when doing family-based interventions or interventions targeting groups of people who are connected socially. While secret

groups on Facebook do not leverage social support from participants' Facebook friends, engaging close social ties may not always be helpful. Our previous work showed that for weight loss, social media users felt their Facebook friends were less of a source of weight-loss social support and more of a source of negativity relative to their weight-loss community on Twitter [40]. Thus, the ability to separate a health-focused online community from users' personal communities may be an advantage when it comes to certain health behaviors. An alternative approach is to allow participants to invite their trusted Facebook friends into the intervention. As long as such individuals have consented, this approach could take advantage of social support from both close and loose social ties [41].

Twitter

Twitter, an online social network that allows users to communicate via 140-character-or-less posts, is used by 23% of Americans [2]. Twitter has been used to deliver behavioral interventions in several studies [42,43]. The vast majority of Twitter users (88%) communicate publicly [44], meaning their tweets are viewable by anyone on the Internet. However, Twitter has a *protected account* setting that allows users to communicate privately. When an account is *protected*, the only users who can view the content are those who have been approved by the user. Private groups can be created on Twitter by having users set their accounts to *protected,* and by only following each other. This feature makes it highly conducive to facilitating confidential discussions in behavioral intervention groups. Avatars and pseudonyms can also be used to anonymize participants' bios to fully protect participants' privacy. Unlike Facebook, creating a private group on Twitter requires the user to use *protected* privacy settings on their accounts, and in most cases, the user may prefer not to use their main account to participate in the intervention. The latter approach may be ideal because the intervention feed will then not get lost in the highly populated newsfeeds of experienced users. The Twitter app makes switching between accounts possible with a single click, as opposed to requiring logging out of one account and then into the other, as is the case with Facebook.

Users' newsfeeds on Twitter are organized chronologically, as opposed to by most recently engaged posts as in Facebook. Twitter allows users to send private direct messages (DMs) to other users, including group facilitators. This allows users to connect on a one-on-one basis and is an advantage over Facebook interventions, where private messages can only be passed between users who are "friends"; however, being a member of a secret Facebook group does not require a user to be "friends" with the other members. Group facilitators should be cautious, though, that too much private messaging with participants could end up moving instructive conversations out of the group, which can increase the time needed to manage the group.

Other Social Media Platforms

Other popular social media platforms include Instagram, Pinterest, YouTube, Vine, and Snapchat. Instagram involves sharing of images and videos, while YouTube and Vine involve sharing of videos. Pinterest involves sharing of links that are organized on "boards" that are decorated by images extracted from links. Clicking on the image takes the user to the link. Snapchat is a platform in which users can take photos and videos, add text and drawings, and send to selected users who can then view it for a time limit between 1 and 10 seconds. Multiple platforms can also be used in the same intervention. For example, videos in a Vine or YouTube feed can be shared on Facebook and/or Twitter. An alternative to commercial platforms is an investigator-designed platform that meets the specific needs of the intervention. In this case, content can be pushed to participants' main social media feeds via share buttons to take advantage of participants' social media presence on commercial platforms.

The Target Population

In designing an online social network-delivered intervention, it is important to consider the social media experience of the target population. If the online social network is providing active intervention ingredients, ensuring that all participants have access to the online social network is necessary to maximize intervention receipt. Different subgroups, as defined by age, sex, cultural group, or other characteristics, may contain more frequent users of certain commercial media platforms. Recruiting individuals who are not active users or who have no experience engaging in a social media platform may result in their not receiving the full dose of the intervention. Some studies recruit both current users of the target platform and also individuals who are willing to open an account to participate in the study [45,46]. Even though participants might agree to sign up for an account on the target platform to participate in the study, the opportunity to embed the intervention into a pre-existing habit is lost with this target population. Instead, the intervention requires the development of a social media habit, which may or may not occur. Their log-ins may be solely for research participation and so this may mean fewer opportunities for them to be exposed to intervention content relative to regular users who will see it during their usual social media activities, even when their interest in the intervention is waning.

Other studies take a more conservative approach and limit study entry criteria to current users of the target social media platform [17,18,47-53]. Given that 61% of Facebook users surveyed reported they take breaks from using Facebook up to several weeks at a time [54], some studies have even more stringent entry criteria regarding use patterns [20,55]. For example, in one study, participants were required to be daily users of Facebook [20], and in another, participants were required to use Facebook for at least 30 minutes per day [55]. When using novel platforms, restricting participants to those with social media experience may reduce usability issues. Our previous work revealed that participants who were recruited as nonusers engaged minimally, even when staff provided instruction on how to use the online social network platform [43]. To ensure that users are familiar with the social networking site, they may be recruited directly from the site. For example, Facebook ads can be used for study recruitment for a fee. On Twitter, recruitment ads can be disseminated via tweets for free. Advertisements can target specific subgroups by using keywords or hashtags frequently used by that community.

"Influencers"—people that represent a specific community and have a large following among that community—can also be engaged and/or incentivized to disseminate study advertisements [56].

Content Conversion

Delivery of intervention content via an online social network requires transforming communication into a format that is consistent with how people communicate in online social networks. For example, in traditional intervention delivery modalities, content is delivered via printed handouts and/or a manual that provides the foundation for counselor-led discussions. However, content in online social networks is typically in the form of brief posts that sometimes include graphics, videos, or links to articles. One study found that information about contraceptives shared over Facebook—in video, graphic, and game format—led to higher knowledge scores than when information was distributed via pamphlet [57]. This shows that information shared via social media has the potential to be even more effective than print formats.

Posts with a graphic, video, or link to an article typically include a brief headline describing the content to attract viewers to read and/or click on it. Individuals and/or concepts in graphics, videos, and other media should reflect the target population. For example, an intervention targeting Latino moms should utilize images and videos that feature Latino moms in the preferred language of the population. Most online discussions are not scheduled, but rather happen spontaneously in the form of comments/replies or hitting a "like" or "favorite" button to indicate agreement. Online discussions are asynchronous, meaning a question may be posted by a user at one time and then answers by other users may appear throughout the day and sometimes into the next day, but not typically longer. Prior to the intervention, behavioral content should be converted into a content library that aligns with how users interact on the target social media platform. To this end, we recommend that a study team member is a regular and active user who can educate the team about the norms of the target platform.

Content Library

Overview

A content library is organized in a similar fashion as a treatment manual, but the content itself is in a different format. Content can include articles written by investigators that are posted on a website, links to other online resources (eg, recipes), brief posts that introduce links or videos (eg, "Check out these 5 ways to squeeze exercise into your day! Which will you try this week?"), infographics, images, gifs (ie, images with animation), videos, status updates, conversation starters, polls, event invitations, and chat topics.

Documents-to-Webpage Conversion

Word processing documents (eg, .doc, .rtf, and .pdf) are rarely shared in online social networks and few networks even allow this capability. Instead, content can be broken down into a series of brief posts or into online articles. Online articles can be shared via links, a commonly shared format of information on social

media. Using blogging software (eg, WordPress), lessons in a treatment manual can be converted into online articles with photos and videos embedded. Online articles are typically brief (ie, 800 words) and include images. Essentially, the treatment manual can be converted into a study blog/website that may or may not allow comments and can be publicly available or completely private (ie, accessible only via links, not search engines). Images used can be developed by the team, purchased from stock image websites, or copied from free stock image websites. Investigators should be aware that using images found via search engines may violate copyright laws, which has consequences especially if the treatment manual is published or sold.

Creation of Posts

Once the treatment manual is converted into an online format, the next step is to create posts that introduce links in each post in a way that draws the users' attention to the link. The goal is to achieve a high engagement rate, which includes all activities that a user can do with a post (eg, click on a link, "like" or "favorite" it, share it, or comment on it). The content in the link can only be effective if clicked on and read. Some social media platforms have character limits for posts (eg, Twitter) and others will limit how many characters can be viewed without necessitating an extra click. Although Facebook does not have character limits, a study of 11,000 Facebook pages found the optimal length of a post was about 120 characters, with longer posts getting lower click-through rates [58].

Intervention posts should also accurately convey what is to be found in the link, being careful not to bait users with sensationalism (eg, "Emotional eating no more! How to get over it for GOOD!"). The term "click bait" is used colloquially on social networking sites to refer to posts that exaggerate or sensationalize content in the link for the purpose of "baiting" people to click. Images can also be used to accompany posts as a way to graphically illustrate a concept or generate emotion. One social media marketing study found that 87% of posts with Facebook engagement had a photo [59]. A study by the social media marketing company, HubSpot, found that posts with photos get 53% more likes, 104% more comments, and 84% higher click-through rates than text-only posts [60]. A study of a smoking cessation campaign found that the most common type of engagement was comments on photos, but while participants found many posts motivating, some triggered the desire to smoke, which suggests that certain images might cue unhealthy behavior [61]. Finally, a study of the National Cancer Institute Facebook page found that posts with images received the most engagement relative to videos, links, and status updates [62].

Infographics are increasingly being used as an alternative way to depict research findings or other information via social media. Free software can be used to make infographics or companies can be hired to design professional-quality infographics. Infographics are available online as well. For example, the US Centers for Disease Control and Prevention has a gallery of infographics available to include in websites and online publications [63]. Videos may be another way to deliver content, and are commonly shared on social media platforms [64]. The

majority of Internet users (78%) report watching videos online and 25% have uploaded videos [65]. Some platforms allow videos to be embedded into a post while in other platforms videos can be posted via links to their original source. However, investigators should avoid posting lengthy videos, as social media research reveals that the average length of time a user will watch a video is 2.7 minutes [66,67].

Curating Evidence-Based Resources

Other resources that can be linked to in a social networking feed include links to reputable online resources. For example, in a weight-loss intervention, links to healthy recipes can be curated and distributed to participants. In a smoking cessation intervention, links to information and resources posted on the American Cancer Society webpage might be leveraged. In general, nonprofit scientific organizations typically have a great deal of curated content on their websites and social media feeds, which could be rich sources of evidence-based information and tools to support an intervention. Leveraging existing evidence-based content is an excellent use of available resources and a way to connect and acquaint users with legitimate sources of health information on the Internet, given the tremendous amount of false information available online. Investigators are encouraged to confirm that the link is active before posting, given occasional changes to URLs or removal of content on external websites.

Engagement Plan

Overview

The engagement plan should describe group size, frequency of posting, whether posts are automated, and a guide for how and how often interventionists should engage with participants. Group size is a consideration given that it is likely to influence engagement. Very small groups may have low engagement due to size, but then very large groups might have so much engagement that intervention content gets buried in the newsfeed. In 19 studies we found using Facebook to deliver behavioral interventions, group size ranged from 3 to 7282 participants [17-19,45,46,48-50,52,53,55,61,68-74]. No data exists on the ideal size of an online social network group for a behavioral intervention.

In terms of post frequency, each social media platform has norms, and it would seem imperative to match the norms of the target platform. According to one social media marketing study, the ideal frequency is 1-2 times per day on Facebook and 3 times per day on Twitter [75]. Studies using social media for behavioral interventions report a posting frequency ranging from 1-2 posts per week [17,48-50], to daily [20,68,72,76], to 2 or more posts per day [19,42,46,73,77]. The frequency of posts is likely one factor in engagement; however, given the variability in engagement across studies, the nature of posts is likely an even more important factor. The ideal post frequency may also depend on the target population for the intervention. Investigators are encouraged to solicit feedback from participants during the design phase and/or during the pilot of their intervention. Frequency of posts should be reported in manuscripts so that its association with engagement can be examined across studies.

If using commercial social networking platforms like Facebook or Twitter, intervention content can be scheduled to post in advance at a predetermined timing and frequency using social media scheduling software (eg, Hootsuite and Buffer). Scheduling software also includes features that allow you to learn the times of day users are most likely to be logged in, which can increase the likelihood of posts being viewed. A social media marketing study found that engagement rates for Facebook are 18% higher on Thursdays and Fridays, while Twitter's highest click-through rates are on Saturday and Sunday [78]. Further, a study using Pinterest showed that articles on the topics of food and fitness are mainly posted on Sundays and Mondays [79]. These data represent average users, so ideal timing of posts may be highly dependent on the study population. Automating original posts is also helpful to keep the feed consistent and predictable, especially if new posts always appear at the same time of day. Even though posts are automated, interventionists can and should still engage with participants' comments on those posts and attempt to draw attention to those posts via their own comments. Automation can reduce the burden of posting on the interventionist, but one downside of automation is that it may lead to interventionists forgetting to log into the community regularly. Having a log-in schedule with reminders can help to keep interventionist log-ins regular. In addition to post frequency, interventionist reaction time to participant-initiated posts matters for user engagement. According to one study, 53% who tweet to a company expect a response within the hour. If the tweet is a complaint [80], 72% expect a response within the hour. Email notifications can be set up for interventionists to make them aware of participant posts and cue them to respond.

Group Chats

Group chats can be scheduled to conduct discussions in the same way that in-person group meetings are scheduled. Facebook has a function for conducting group chats. On Twitter, hashtags are often used to host group chats as a way to separate chat tweets from other tweets in the newsfeed and to allow people to easily follow the conversation. Chat tweets stay in the newsfeed, which allows the conversation to continue after the scheduled time of the chat. Moreover, people who missed the chat can view the chat later. Google Hangout can also be used to conduct video chats. While synchronous group chats may be more convenient than in-person meetings since they do not require transportation or childcare, they still require finding a time where everyone can attend, which may limit participation. On many platforms (eg, Twitter), group chat content can be viewed after the fact since it exists in the newsfeed. This allows participants who missed it to read the chat and comment on it later, and even allows them to reopen the discussion on a different day.

Microcounseling

An alternative to group chats is a form of interaction we refer to as microcounseling, which involves frequent, brief, asynchronous, yet timely exchanges between an interventionist and participants [43]. In microcounseling, the interventionist logs in at least once daily to initiate and engage in discussions. Although the goal is for informational posts to elicit

engagement, if they do not, the interventionist can stimulate engagement with a post that draws attention to the content by commenting on the post (eg, "I'm curious which of these strategies everyone wants to try this week?"). The interventionist can even specifically mention users in their comments to pull them into a discussion (eg, "@puppymama, you mentioned you were having trouble finding time to exercise, are any of these ideas helpful?"), similar to calling on someone in a traditional in-person group setting. Typical group dynamics emerge in online groups, such that some individuals are very talkative, while others are less so; some are advice givers, while others are advice seekers. Users of online social networks are used to fairly rapid (ie, same day) responses to their posts, thus, daily interventionist presence would seem essential to match such norms. A social media marketing study revealed that a Facebook fan page post will typically receive the majority of engagement within 3 hours [59]. Interventionists can take advantage of social media features that convey positive reinforcement (eg, "like" and "favorite" buttons) when users post and reply to increase the likelihood of such behavior occurring in the future. Another reason to address participants' comments and questions on a timely basis is the fast pace of a social media environment where posts can quickly get buried at the bottom of a newsfeed.

Peer Influence

The engagement plan can also involve strategies to facilitate peer influence on health behavior change, potentially impacting social norms. The influence of strong ties (ie, personal connections) may be particularly important given research showing that the social norms one perceives in their friend circle may influence outcomes in behavioral interventions [81]. Other research has shown that even weak ties can be influential [82,83]. Online social networks provide a unique opportunity to engage both strong and weak ties. Peer influence can be facilitated using team-based approaches involving strong and/or weak ties [84], allowing participants to engage their friend networks (ie, strong ties) into the intervention [70], recruiting groups of friends or family members into the intervention [85], or providing corrective feedback regarding perceived social norms that may be perpetuating unhealthy behavior [71].

Peer influence can also be leveraged to spread health messages throughout large online social networks. "Viral marketing" is a marketing technique in which messages are created by an entity, but then spread within online social networks by users [86]. A recent example is the Amyotrophic Lateral Sclerosis (ALS) Association ice bucket challenge, which went viral on Facebook in the summer of 2014 and resulted in unprecedented donations for the Association [87]. This technique could be leveraged by public health interventionists to spread health messages across large networks. While it is difficult to predict which messages have the capacity to spread virally, research studying viral messages may shed light on the characteristics of messages that are shared at high rates across online social networks.

Interventionist Training

Interventionists should ideally be experienced users of the target social media platform so that they are already acquainted with the norms of the platform. An interventionist with little or no experience with the platform would be equivalent to using an interventionist for telephone counseling who has never used or seen a phone. Training would need to be far more extensive (by an experienced user) and include review of how the device works, how people use it, and plenty of time to practice using it. Supervision should be provided throughout the intervention to flag issues. Regardless of the interventionist's experience level, the investigator should develop a written guide for how often interventionists should log in and expectations for engagement. Although exchanges are brief on social media, conversations are continuous and dynamic 7 days a week, including holidays, weekends, and evenings. The "off hours" (ie, evenings and weekends) are also times that people changing behavior are at high risk of encountering barriers. For example, people trying to lose weight [88] or quit smoking often encounter cues in the home and social environments. Having interventionists who respond daily takes advantage of the ability to insert intervention at these times when participants need it the most. Temporary absences from the interventionist could break the flow of the conversations and result in missed opportunities to intervene. Given the need for frequent, brief attention to the group, having multiple interventionists can be helpful to cover absences/vacations as well as to model interactive engagement in a group. Participants may also prefer one interventionist's engagement style over the other, thus, multiple interventionists reduces the possibility of disengagement from the study due to a nonpreferred interventionist.

To engage participants in discussion with interventionists and each other, informational posts alone might be insufficient, as this does not mimic typical group discussions where questions are posed and opinions are queried. Posts can be designed to engage participants into an interactive discussion by using open-ended questions, icebreakers, or conversation starters; otherwise inviting participants to respond can be used to generate discussion. In a Facebook intervention for weight loss in young adults, status updates, photos, and polls received the highest levels of engagement, with 75-97% receiving at least one interaction, while videos and links received much lower levels of engagement: 52-57% received at least one interaction [72]. Another study found that 64% of participant engagement on a study Facebook page occurred in response to the single post made by the interventionist [55]. The single post was an icebreaker, which asked participants to share experiences. In that study, the Facebook page was meant for participants to use to engage with each other, but the success of such an approach may be highly dependent on whether participants happen to feel comfortable engaging with strangers on a Facebook page. Ultimately, measuring engagement analytics throughout the course of a study will show which posts are most engaging, and this data can be used to refine the current intervention strategy in real time or in a future iteration.

Participant Training

Even the most experienced social media users may not be accustomed to using social media to engage in a behavioral intervention. In our study of an online social network-delivered

weight-loss intervention, one of the biggest barriers to engagement reported by participants was their not being sure what to post [43]. For this reason, some guidance at the outset of the intervention to inform participants on how to maximize their experience can be helpful. An orientation meeting can be held in person, by phone (eg, conference call), or online (eg, Google Hangout) to discuss the intervention and what is expected of participants. Participants can be encouraged to share their experiences, comment on posts, click the "like" button on posts they liked, and ask questions. Encouraging participants to post in the moment when they are struggling or have a question allows them to get help precisely when they need it. In addition to guidance on what to post, the orientation can give guidance on what not to post. For example, participants might be asked to refrain from posting anything for marketing purposes. On the other hand, having too many guidelines may cause participants to be inhibited from posting due to concerns about breaking rules.

When recruiting participants who are inexperienced with the social network platform, extensive training on the platform will be necessary. This would include help setting up an account, guidance regarding how to use the features and settings, familiarity with both the app and Web versions, and how to set up email notifications to cue the participant when something new has been posted or when they have been mentioned in a post. In our previous work, some participants felt that the group orientation meeting of 90 minutes was insufficient [52]. Individual meetings tailored to the participant's level of experience might be more suitable. A run-in period during which the participant gets used to engaging regularly on the platform before the intervention begins might be useful.

Cost and Scalability

To the extent that data prove online social network-delivered interventions efficacious, a major potential advantage could be cost-effectiveness given the elimination of clinic visits. However, implementation still requires time and effort. Potential costs to deliver the intervention include time spent setting up the community, scheduling social media posts, confirming the functionality of external links, setting up software tracking programs, interventionist/participant training, interventionist time to deliver the intervention, and other participant contact time by interventionists (eg, emails to participants who have low engagement). Studies that involve the development of a novel platform or translation of an existing behavioral intervention for social media delivery will incur additional costs related to intervention development. Investigators should take measures to track resources utilized and time spent by interventionists and participants so that cost can be accurately estimated.

The scalability of online social network-delivered interventions has not yet been explored, but would seem to have great potential given the lack of geographical barriers. Online social networks can be used to deliver interventions by health care organizations, public health organizations, and other entities that serve large numbers of people over large geographical areas. A key research question is how large can an online social

network be and still effectively deliver a behavioral intervention. Another factor affecting scalability will be how much an intervention can be automated given that automation will reduce cost. Complete automation of the intervention has the highest potential for scalability, although may come at the loss of personalization. The leveraging of artificial intelligence in intervention delivery may be one way to preserve personalization. Cost-effectiveness studies are needed to truly estimate the scalability of such interventions. Researchers are encouraged to consider how their social media-delivered behavioral intervention might be scaled up for widespread dissemination and impact.

Privacy and Human Subjects Issues

Privacy concerns can arise when using online social media platforms to deliver behavioral interventions. Privacy can be difficult to protect when using open or public settings and some people may not be comfortable engaging publicly or having others outside of the study know they are in a study [38,73]. Given that behavioral interventions are traditionally conducted in private and confidential environments, the use of private online social network communities is the best way to mimic this setting.

Investigators should make their local human studies committees and participants aware of the privacy policies of the social media platform. Because commercial platforms have access to data shared in their platform, it is not recommended that protected health information is collected over the platform, but instead via other more secure means, such as through Research Electronic Data Capture (REDCap [89]) [90]. Pretesting the group or page prior to the start of the intervention to review the privacy settings, functionality, and appearance of the content is a helpful way to identify and remedy problems before using it with study participants.

Privacy should be explained to participants at the outset of a social media-delivered intervention to ensure they understand who does and does not have access to their data. Given recent highly publicized online security breaches (eg, Target [91]), it is not only important for participants to understand the privacy settings and receive guidance on posting personal information, but it is also important for them to understand that the researchers cannot completely guarantee against a security breach.

Engagement Data

Although engagement data—in the form of views, likes, shares, comments, favorites, replies, retweets, posts, and tweets—can be obtained by manual abstraction from the newsfeed, this can be a tedious and time-consuming task, especially for interventions with a large number of participants or long duration. A more efficient approach is to work with a programmer to extract the desired data or to use social media analytics programs to analyze the metrics. Data extraction capabilities may differ by social media platform, thus, identifying what data can be extracted and in what format in advance is recommended. Whether extracting data manually,

via software, or by a programmer, proper budgeting will be important, as all require resources. Some investigators will want to analyze the content of posts made by participants, which will require capturing the text of posts so it can be analyzed. Content analyses can be conducted manually via coders or in an automated fashion using machine learning or natural language processing. Given the volume and nature of social network data, a team science approach that includes behavioral scientists, social media analysts, computer scientists, and biostatisticians is highly recommended.

Intervention Reporting

Because the literature on online social network-delivered interventions is sparse, no reporting standards exist. Without consistent reporting about the intervention, it is difficult to compare studies and elucidate which approaches are associated with the highest engagement and success rates. Consistent reporting is also essential for replication. Table 2 outlines reporting guidelines for the intervention and for participant engagement.

Table 2. Reporting guidelines for social media-delivered interventions.

Intervention and participant characteristics	Reporting guidelines
Intervention general	
Type (ie, host, user, or host and user generated)	Is the social network content intended to be host generated, user generated, or host and user generated?
Primary modality	Is the social network the primary intervention modality or adjunctive?
Purpose of SNS[a]	What is the purpose of the social network?
Participants	
Experience with social media	What is the social media experience level of participants? Current users? Nonusers? Expert users?
Intervention content	
Post frequency	How often will posts be made by the interventionists?
Content	What is the content of the posts?
Microcounseling	Will interventionists be providing counseling?
Automation	Will posts be automated? If so, how many? When?
Chats	Will moderated chats be held? If so, how often?
Participant engagement metrics	
Likes/favorites	How many likes did each post get? On average, what percentage of posts did each participant like?
Replies/comments	How many replies did each post get? On average, what percentage of posts did each participant reply to?
Original posts	How many original posts did participants make? On average, how many original posts did each participant make?
Intervention fidelity	
Page membership	What percentage of participants actually joined the group/page/community?
Posts	What percentage of planned posts were actually posted?
Views	How many views did each post get? On average, what percentage of posts did each participant view?
Interventionist log-in frequency	How often did the interventionist log in?
Interventionist likes	What percentage of participant posts/comments did the interventionist like?
Interventionist replies/comments	What percentage of participant posts/comments did the interventionist reply or comment on?
Retention	
Group membership termination	How many participants exited the group before the intervention ended?
View termination	How many participants stopped viewing posts before the end of the intervention? At what point in the intervention?
Dropout	How many participants did not attend follow-up visits?

[a]SNS: social networking site.

Future Research

Delivering behavioral interventions via online social networks is a relatively new endeavor; thus, many questions about best practices remain unanswered. We pose several questions to be explored in future research.

1. What is the optimal size for an online social network group for a behavioral intervention? The ideal size of a social network group for each purpose (eg, microcounseling, peer support, and information delivery) is unknown. In our previous work, we found that people who tweet about their weight-loss journey reported that their organically grown social network on Twitter was, on average, 494 followers (SD 635) [40]. However, it is unclear what percentage of their network a user interacts with in organically grown social networks. Studies should explore the impact of differently sized networks on both engagement and the behavioral outcome.

2. What is the ideal structure of a group intervention? Some studies use public groups and other private groups, and each approach has its merits. Public groups allow for growth and wider dissemination of content, while private groups allow for discussions that are more intimate. The ideal structure of the group likely depends on the goal of the intervention, but this has never been explored.

3. What is meaningful engagement? Engagement comes in many forms, including hitting a "like" button, voting in a poll, or posting original content. Not all engagement may be meaningful, in other words, it may not actually result in change in knowledge, behavior, or other key outcomes. Research is needed to discern which types of engagement are associated with better outcomes.

4. How can engagement be increased in an online social network? Studies have demonstrated links between engagement and outcomes in social network-delivered interventions, but what remains unclear is how to increase meaningful engagement. Research is needed to explore the effect of group size, interventionist engagement, post type, and participant characteristics on participant engagement. The identification of modifiable factors would be particularly helpful to inform future interventions.

5. For whom are social network-delivered interventions best suited? While it may be assumed that social network-delivered interventions are best suited for regular users, the ideal way to engage nonusers is unknown. The characteristics of users most likely to benefit are unknown. People who use social media heavily to socialize may not feel comfortable, or have interest in, using social media for the purposes of learning about a health condition or changing behavior. Evaluating the target population's interest in a social media-delivered intervention prior to attempting an intervention will likely be useful. For example, Waring and colleagues surveyed 63 overweight or obese women of childbearing age who were Twitter users to find out if they had an interest in participating in a weight-loss intervention delivered via Twitter. The majority (81%), but not all, were at least somewhat interested in such a program [92]. Further research is needed to explore which populations are most interested in this type of intervention. Replicating interventions in populations with different social media skill levels and personal characteristics, as well as using different online social network platforms, may shed light on which approaches work for whom and under what circumstances.

Conclusions

Social media has revolutionized interpersonal communication, which presents unique opportunities for communicating with patients and delivering behavioral interventions. The design of social network-delivered interventions requires an understanding of the target platform, its users, and the norms for communication on the platform. Such an understanding will inform how the platform can be used and what role it can play in the intervention. Content from traditional interventions will require translation into a format that is consistent with how content is exchanged on the target platform. The dawn of social network-delivered interventions has also introduced a science of engagement, which requires measurement of metrics unique to each platform. Although social media presents a new means of intervening on patient behavior, many challenges and unknowns exist in the process of translating traditional intervention models for social media delivery, including the translation of intervention content, privacy, requirements and cost, and identifying the target populations most likely to be responsive. Social media research requires a team science approach that includes experts in social media analysis, behavioral science, computer science, and big data analyses. Consistent reporting of intervention details and engagement data will be crucial to advancing this science.

Conflicts of Interest

None declared.

References

1.   Duggan M, Smith A. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Dec 30. Social media update 2013 URL: http://www.pewinternet.org/2013/12/30/social-media-update-2013/ [accessed 2015-08-27] [WebCite Cache ID 6b64FaV8V]

2.   Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2014. Social networking fact sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2015-06-24] [WebCite Cache ID 6ZVvgfc4q]

3.   Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2015 Jan 09. Frequency of social media use URL: http://www.pewinternet.org/2015/01/09/frequency-of-social-media-use-2/ [accessed 2015-08-27] [WebCite Cache ID 6b64Xzv2a]

4.   Hampton K, Sessions Goulet L, Rainie L, Purcell K. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2011 Jun 16. Part 2: Who are social networking site users? URL: http://www.pewinternet.org/2011/06/16/part-2-who-are-social-networking-site-users/ [accessed 2015-08-27] [WebCite Cache ID 6b64duHY5]

5.   Constine J. TechCrunch. 2014 Jul 23. American users spend an average of 40 minutes per day on Facebook URL: http://techcrunch.com/2014/07/23/facebook-usage-time/ [accessed 2015-08-27] [WebCite Cache ID 6b64rX8gD]

6.   Fox S. The Social Life of Health Information, 2011. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2015-08-27] [WebCite Cache ID 6b65CRzpl]

7.   Duggan M, Lenhart A, Lampe C, Ellison N. Pew Research Center. Washington DC: Pew Internet & American Life Project; 2015 Jul 16. Parents and social media URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2015-08-27] [WebCite Cache ID 6b65JivYD]

8.   Fox S. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2012. Health fact sheet URL: http://www.pewinternet.org/Commentary/2011/November/Pew-Internet-Health.aspx [accessed 2015-08-27] [WebCite Cache ID 6b65PHUkE]

9.   Fox S. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2011 Feb 28. Peer-to-peer health care URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [accessed 2015-08-27] [WebCite Cache ID 6b65r4Dy0]

10.   Fox S. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Aug 03. Peer-to-peer health care is a slow idea that will change the world URL: http://susannahfox.com/2013/08/03/peer-to-peer-health-care-is-a-slow-idea-that-will-change-the-world/ [accessed 2015-08-27] [WebCite Cache ID 6b65xvo8E]

11.   Montgomery KS, Bushee TD, Phillips JD, Kirkpatrick T, Catledge C, Braveboy K, et al. Women's challenges with postpartum weight loss. Matern Child Health J 2011 Nov;15(8):1176-1184. [doi: 10.1007/s10995-010-0681-9] [Medline: 20844941]

12.   Archer E, Groessl EJ, Sui X, McClain AC, Wilcox S, Hand GA, et al. An economic analysis of traditional and technology-based approaches to weight loss. Am J Prev Med 2012 Aug;43(2):176-182. [doi: 10.1016/j.amepre.2012.04.018] [Medline: 22813682]

13.   Syred J, Naidoo C, Woodhall SC, Baraitser P. Would you tell everyone this? Facebook conversations as health promotion interventions. J Med Internet Res 2014;16(4):e108 [FREE Full text] [doi: 10.2196/jmir.3231] [Medline: 24727742]

14.   Poncela-Casasnovas J, Spring B, McClary D, Moller AC, Mukogo R, Pellegrini CA, et al. Social embeddedness in an online weight management programme is linked to greater weight loss. J R Soc Interface 2015 Mar 6;12(104):20140686 [FREE Full text] [doi: 10.1098/rsif.2014.0686] [Medline: 25631561]

15.   Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]

16.   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]

17.   Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: A randomized pilot trial. BMC Womens Health 2015;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]

18.   Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: A qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015;17(1):e27 [FREE Full text] [doi: 10.2196/jmir.4061] [Medline: 25624064]

19.   Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: A cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]

20.   Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]

21.   Facebook. The Skin Cancer Foundation URL: https://www.facebook.com/skincancerfoundation [accessed 2015-12-27] [WebCite Cache ID 6e6GrGcY7]

22.   Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2015 Jan 09. Social media update 2014 URL: http://www.pewinternet.org/2015/01/09/social-media-update-2014/ [accessed 2015-08-27] [WebCite Cache ID 6b66RMgiH]

XSL•FO

RenderX

23. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a Web-based intervention: Randomized trial. J Med Internet Res 2012;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]

24. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: Systematic review. J Med Internet Res 2013;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]

25. PR Newswire. 2015 Apr 22. Facebook reports first quarter 2015 results URL: http://www.prnewswire.com/news-releases/facebook-reports-first-quarter-2015-results-300070539.html [accessed 2015-08-27] [WebCite Cache ID 6b66gYS3r]

26. Protalinski E. ZDNet. 2011 Nov 30. Facebook increases status update character limit to 63,206 URL: http://www.zdnet.com/article/facebook-increases-status-update-character-limit-to-63206/ [accessed 2015-08-27] [WebCite Cache ID 6b66pbnNx]

27. Quintaro P. Benzinga. 2015 Apr 28. Twitter MAU were 302M for Q1, up 18% YoY URL: http://www.benzinga.com/news/earnings/15/04/5452400/twitter-mau-were-302m-for-q1-up-18-yoy [accessed 2015-08-27] [WebCite Cache ID 6b67H0jH9]

28. Mangalindan JP. Mashable. 2015 Apr 29. Pinterest's Evan Sharp: Guys are on here, too URL: http://mashable.com/2015/04/29/pinterest-evan-sharp-users/ [accessed 2015-08-27] [WebCite Cache ID 6b67dU8Sz]

29. Guberti M. Marc's Blog: Digital Marketing & Entrepreneurship. 2015 Mar 23. 5 Pinterest limits URL: http://marcguberti.com/2014/03/5-pinterest-limits/ [accessed 2015-08-27] [WebCite Cache ID 6b67pxhUj]

30. Snapchat. 2015. 3V Advertising URL: https://www.snapchat.com/ads [accessed 2015-08-27] [WebCite Cache ID 6eJZz5wkE]

31. LinkedIn Newsroom. 2015. URL: https://press.linkedin.com/about-linkedin [accessed 2015-08-27] [WebCite Cache ID 6b68VCd6M]

32. LinkedIn: Pulse. 2015 Feb 06. LinkedIn maximum character counts for 2015 URL: https://www.linkedin.com/pulse/linkedin-maximum-character-counts-2015-pierson-linkedin-coach-%E3%83%84-?trk=pulse-det-nav_art [accessed 2016-01-05] [WebCite Cache ID 6eJaI72Gi]

33. Systrom K. Instagram Blog. 2014 Dec 10. 300 million: Sharing real moments URL: http://blog.instagram.com/post/104847837897/141210-300million [accessed 2015-08-28] [WebCite Cache ID 6b7S23p7V]

34. Barr A. USA Today. 2013 Oct 29. Google's social network sees 58% jump in users URL: http://www.usatoday.com/story/tech/2013/10/29/google-plus/3296017/ [accessed 2015-08-28] [WebCite Cache ID 6b7SS72TJ]

35. SMG Convonix Research Team. Convonix. 2012 Jan 05. Who has more "character"? Google+ or Facebook? URL: http://www.convonix.com/blog/social-media-marketing/googleplus-vs-facebook/ [accessed 2015-08-28] [WebCite Cache ID 6b7SbtMXw]

36. Boris C. Marketing Pilgrim. 2014 Oct 24. Tumblr announces 40 percent growth and a fancy new video player URL: http://www.marketingpilgrim.com/2014/10/tumblr-announces-40-percent-growth-and-a-fancy-new-video-player.html [accessed 2015-08-28] [WebCite Cache ID 6b7SpYaZ9]

37. Vine. Twitter. 2013 Aug 20. We've said this before and we'll say it again: This community - now more than 40 million of you - is amazing. Thank you for inspiring us URL: https://twitter.com/vine/status/369911739782946816 [accessed 2015-08-28] [WebCite Cache ID 6b7SzQuMo]

38. Ramo DE, Liu H, Prochaska JJ. A mixed-methods study of young adults' receptivity to using Facebook for smoking cessation: If you build it, will they come? Am J Health Promot 2015 Apr;29(4):e126-e135. [doi: 10.4278/ajhp.130326-QUAL-128] [Medline: 24575728]

39. DeMers J. Forbes. 2015 May 13. Why your organic Facebook reach is still falling -- And what to do about it URL: http://www.forbes.com/sites/jaysondemers/2015/05/13/why-your-organic-facebook-reach-is-still-falling-and-what-to-do-about-it/ [accessed 2015-08-28] [WebCite Cache ID 6eJeCVpqH]

40. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM, et al. Tweeting it off: Characteristics of adults who tweet about a weight loss attempt. J Am Med Inform Assoc 2014;21(6):1032-1037. [doi: 10.1136/amiajnl-2014-002652] [Medline: 24928175]

41. Kernot J, Olds T, Lewis LK, Maher C. Effectiveness of a Facebook-delivered physical activity intervention for post-partum women: A randomized controlled trial protocol. BMC Public Health 2013;13:518 [FREE Full text] [doi: 10.1186/1471-2458-13-518] [Medline: 23714411]

42. Turner-McGrievy G, Tate D. Tweets, apps, and pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]

43. Pagoto SL, Waring ME, Schneider KL, Oleski JL, Olendzki E, Hayes RB, et al. Twitter-delivered behavioral weight-loss interventions: A pilot series. JMIR Res Protoc 2015;4(4):e123 [FREE Full text] [doi: 10.2196/resprot.4864] [Medline: 26500186]

44. Beevolve. 2012 Oct 10. An exhaustive study of Twitter users across the world URL: http://www.beevolve.com/twitter-statistics/ [accessed 2015-10-08] [WebCite Cache ID 6c899nPPf]

45. Patrick K, Marshall SJ, Davila EP, Kolodziejczyk JK, Fowler JH, Calfas KJ, et al. Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). Contemp Clin Trials 2014 Jan;37(1):10-18 [FREE Full text] [doi: 10.1016/j.cct.2013.11.001] [Medline: 24215774]

46. Wojcicki JM, Geissler JD, Stokes CW, Heyman MB, Tran CT. The use of the RESPeRATE device to lower blood pressure in inner city obese adolescents and children: A pilot feasibility study. High Blood Press Cardiovasc Prev 2013 Jun;20(2):89-92 [FREE Full text] [doi: 10.1007/s40292-013-0014-3] [Medline: 23653175]

47. Kernot J, Olds T, Lewis LK, Maher C. Usability testing and piloting of the Mums Step It Up program--A team-based social networking physical activity intervention for women with young children. PLoS One 2014;9(10):e108842 [FREE Full text] [doi: 10.1371/journal.pone.0108842] [Medline: 25272039]

48. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]

49. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The efficacy of a walking intervention using social media to increase physical activity: A randomized trial. J Phys Act Health 2015 Jun;12 Suppl 1:S18-S25. [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]

50. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]

51. Prochaska JJ, Pechmann C, Kim R, Leonhardt JM. Twitter=quitter? An analysis of Twitter quit smoking social networks. Tob Control 2012 Jul;21(4):447-449 [FREE Full text] [doi: 10.1136/tc.2010.042507] [Medline: 21730101]

52. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: A pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615. [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]

53. 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 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]

54. Rainie L, Smith A, Duggan M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Feb 05. Coming and going on Facebook URL: http://www.pewinternet.org/2013/02/05/coming-and-going-on-facebook/ [accessed 2015-08-28] [WebCite Cache ID 6b7TZAx2I]

55. 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]

56. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]

57. Kofinas JD, Varrey A, Sapra KJ, Kanj RV, Chervenak FA, Asfaw T. Adjunctive social media for more effective contraceptive counseling: A randomized controlled trial. Obstet Gynecol 2014 Apr;123(4):763-770. [doi: 10.1097/AOG.0000000000000172] [Medline: 24785602]

58. What 120 Billion Impressions Tells Us About Successful Facebook Marketing. Boulder, CO: BlitzMetrics; 2013. URL: https://blitzmetrics.com/wp-content/uploads/2013/04/120BillionImpressions-V4.3b.pdf [accessed 2015-08-28] [WebCite Cache ID 6b7UEmwAP]

59. Wittman C. EdgeRank Checker. 2012 Jan 17. Post lifetime: How long does a Facebook post live? URL: https://www.socialbakers.com/edgerankchecker/blog/2012/01/post-lifetime-how-long-does-a-facebook-post-live/ [accessed 2015-08-28] [WebCite Cache ID 6b7UOduJd]

60. Corliss R. HubSpot Blogs. 2012 Nov 15. Photos on Facebook generate 53% more likes than the average post URL: http://blog.hubspot.com/blog/tabid/6307/bid/33800/Photos-on-Facebook-Generate-53-More-Likes-Than-the-Average-Post-NEW-DATA.aspx [accessed 2015-08-28] [WebCite Cache ID 6b7UfgH9C]

61. Struik LL, Baskerville NB. The role of Facebook in Crush the Crave, a mobile- and social media-based smoking cessation intervention: Qualitative framework analysis of posts. J Med Internet Res 2014;16(7):e170 [FREE Full text] [doi: 10.2196/jmir.3189] [Medline: 25016998]

62. Strekalova YA, Krieger JL. A picture really is worth a thousand words: Public engagement with the National Cancer Institute on social media. J Cancer Educ 2015 Sep 8:1-3. [doi: 10.1007/s13187-015-0901-5] [Medline: 26351003]

63. Centers for Disease Control and Prevention. Atlanta, GA: US Department of Health and Human Services; 2015. Infographics URL: http://www.cdc.gov/socialmedia/tools/infographics.html [accessed 2015-08-28] [WebCite Cache ID 6b7VpRC3U]

64. 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]

65. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Oct 10. Video: The rise of online video URL: http://www.pewinternet.org/2013/10/10/video-the-rise-of-online-video/ [accessed 2015-08-28] [WebCite Cache ID 6b7WSZ2YU]

66. Statistic Brain. 2015. Attention span statistics URL: http://www.statisticbrain.com/attention-span-statistics/ [accessed 2015-08-28] [WebCite Cache ID 6b7Wb7eEl]

67.   Weinreich H, Obendorf H, Herder E, Mayer M. Not quite the average: An empirical study of Web use. ACM Trans Web 2008 Feb 01;2(1):1-31. [doi: 10.1145/1326561.1326566]

68.   Edwards-Gaura A, Whitaker D, Self-Brown S. Can social networking be used to promote engagement in child maltreatment prevention programs? Two pilot studies. West J Emerg Med 2014 Aug;15(5):575-581 [FREE Full text] [doi: 10.5811/westjem.2014.4.21413] [Medline: 25157304]

69.   Lelutiu-Weinberger C, Gamarel KE, Golub SA, Parsons JT. Race-based differentials in the impact of mental health and stigma on HIV risk among young men who have sex with men. Health Psychol 2015 Aug;34(8):847-856. [doi: 10.1037/hea0000192] [Medline: 25545041]

70.   Rice E, Tulbert E, Cederbaum J, Barman AA, Milburn NG. Mobilizing homeless youth for HIV prevention: A social network analysis of the acceptability of a face-to-face and online social networking intervention. Health Educ Res 2012 Apr;27(2):226-236 [FREE Full text] [doi: 10.1093/her/cyr113] [Medline: 22247453]

71.   Ridout B, Campbell A. Using Facebook to deliver a social norm intervention to reduce problem drinking at university. Drug Alcohol Rev 2014 Nov;33(6):667-673. [doi: 10.1111/dar.12141] [Medline: 24689339]

72.   Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]

73.   Merchant G, Weibel N, Patrick K, Fowler JH, Norman GJ, Gupta A, et al. Click "like" to change your behavior: A mixed methods study of college students' exposure to and engagement with Facebook content designed for weight loss. J Med Internet Res 2014;16(6):e158 [FREE Full text] [doi: 10.2196/jmir.3267] [Medline: 24964294]

74.   Jones L, Saksvig BI, Grieser M, Young DR. Recruiting adolescent girls into a follow-up study: Benefits of using a social networking website. Contemp Clin Trials 2012 Mar;33(2):268-272 [FREE Full text] [doi: 10.1016/j.cct.2011.10.011] [Medline: 22101207]

75.   Ross P. Socialbakers. 2014 Apr 08. Photos are still king on Facebook URL: http://www.socialbakers.com/blog/2149-photos-are-still-king-on-facebook [accessed 2015-08-28] [WebCite Cache ID 6b7aDCnUp]

76.   Horodynski MA, Silk K, Hsieh G, Hoffman A, Robson M. Tools for teen moms to reduce infant obesity: A randomized clinical trial. BMC Public Health 2015;15:22 [FREE Full text] [doi: 10.1186/s12889-015-1345-x] [Medline: 25604090]

77.   Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: Use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]

78.   Cooper BB. Buffer Blog. 2013 Aug 29. A scientific guide to posting tweets, Facebook posts, emails, and blog posts at the best time URL: https://blog.bufferapp.com/best-time-to-tweet-post-to-facebook-send-emails-publish-blogposts [accessed 2015-08-28] [WebCite Cache ID 6b7aoClQu]

79.   Pinterest Blog. 2014 May 15. Pin trends of the week URL: https://blog.pinterest.com/en/pin-trends-week [accessed 2015-08-28] [WebCite Cache ID 6b7bjxeAH]

80.   Lithium. 2013. Consumers will punish brands that fail to respond on Twitter quickly URL: http://www.lithium.com/company/news-room/press-releases/2013/consumers-will-punish-brands-that-fail-to-respond-on-twitter-quickly [accessed 2016-01-10] [WebCite Cache ID 6eRaQnApq]

81.   Leahey TM, Doyle CY, Xu X, Bihuniak J, Wing RR. Social networks and social norms are associated with obesity treatment outcomes. Obesity (Silver Spring) 2015 Aug;23(8):1550-1554. [doi: 10.1002/oby.21074] [Medline: 26150394]

82.   Sandstrom GM, Dunn EW. Social interactions and well-being: The surprising power of weak ties. Pers Soc Psychol Bull 2014 Apr 25;40(7):910-922. [doi: 10.1177/0146167214529799] [Medline: 24769739]

83.   Van Voorhees BW, Gollan J, Fogel J. Pilot study of Internet-based early intervention for combat-related mental distress. J Rehabil Res Dev 2012;49(8):1175-1190 [FREE Full text] [Medline: 23341310]

84.   Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418 [FREE Full text] [doi: 10.1038/oby.2012.18] [Medline: 22310234]

85.   Wing RR, Jeffery RW. Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. J Consult Clin Psychol 1999 Feb;67(1):132-138. [Medline: 10028217]

86.   Subramani MR, Rajagopalan B. Knowledge-sharing and influence in online social networks via viral marketing. Commun ACM 2003 Dec 01;46(12):300. [doi: 10.1145/953460.953514]

87.   Braiker B. Digiday. 2014 Aug 14. The 'Ice Bucket Challenge': A case study in viral marketing gold URL: http://digiday.com/brands/ice-bucket-challenge-case-study-viral-marketing-success/ [accessed 2015-10-08] [WebCite Cache ID 6c8D76p5F]

88.   Orsama A, Mattila E, Ermes M, van Gils GM, Wansink B, Korhonen I. Weight rhythms: Weight increases during weekends and decreases during weekdays. Obes Facts 2014;7(1):36-47. [doi: 10.1159/000356147] [Medline: 24504358]

89.   REDCap. URL: http://project-redcap.org/ [accessed 2016-01-14] [WebCite Cache ID 6eXGGmMEv]

90.   Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

91.   Yang JL, Jayakumar A. The Washington Post. 2014 Jan 10. Target says up to 70 million more customers were hit by December data breach URL: http://www.washingtonpost.com/business/economy/target-says-70-million-customers-were-hit-by-dec-data-breach-more-than-first-reported/2014/01/10/0ada1026-79fe-11e3-8963-b4b654bcc9b2_story.html [accessed 2015-10-08] [WebCite Cache ID 6c8Dhwb08]

92.   Waring M, Evans M, Schneider K, Whited M, Appelhans B, Xiao R, et al. Interest in a Twitter-delivered weight loss program among women of childbearing age. In: Proceedings of the 36th Annual Meeting & Scientific Sessions of the Society of Behavioral Medicine. 2015 Presented at: The 36th Annual Meeting & Scientific Sessions of the Society of Behavioral Medicine; April 22-25, 2015; San Antonio, TX.

Abbreviations

ALS: amyotrophic lateral sclerosis
API: application programming interface
CDC: Centers for Disease Control and Prevention
DM: direct message
N/A: not applicable
NIH: National Institutes of Health
REDCap: Research Electronic Data Capture
SNS: social networking site

XSL•FO
RenderX

Original Paper

Use of the Internet for Sexual Health Among Sexually Experienced Persons Aged 16 to 44 Years: Evidence from a Nationally Representative Survey of the British Population

Catherine RH Aicken[1], BSc (Hons), MSc; Claudia S Estcourt[2], MD, MBBS, FRCP; Anne M Johnson[1], MD; Pam Sonnenberg[1], PhD; Kaye Wellings[3], FRCOG; Catherine H Mercer[1], PhD

[1]Research Department of Infection and Population Health, Institute of Epidemiology and Healthcare, University College London, London, United Kingdom

[2]Blizard Institute, Centre for Immunology and Infectious Diseases, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[3]Centre for Sexual and Reproductive Health Research, Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Catherine RH Aicken, BSc (Hons), MSc
Research Department of Infection and Population Health
Institute of Epidemiology and Healthcare
University College London
Centre for Sexual Health and HIV Research, Mortimer Market Centre
off Capper Street
London, WC1E 6JB
United Kingdom
Phone: 44 (0)20 3108 2067
Fax: 44 (0)20 3108 2079
Email: c.aicken@ucl.ac.uk

Abstract

Background: Those who go online regarding their sexual health are potential users of new Internet-based sexual health interventions. Understanding the size and characteristics of this population is important in informing intervention design and delivery.

Objective: We aimed to estimate the prevalence in Britain of recent use of the Internet for key sexual health reasons (for chlamydia testing, human immunodeficiency virus [HIV] testing, sexually transmitted infection [STI] treatment, condoms/contraceptives, and help/advice with one's sex life) and to identify associated sociodemographic and behavioral factors.

Methods: Complex survey analysis of data from 8926 sexually experienced persons aged 16-44 years in a 2010-2012 probability survey of Britain's resident population. Prevalence of recent (past year) use of Internet sources for key sexual health reasons was estimated. Factors associated with use of information/support websites were identified using logistic regression to calculate age-adjusted odds ratios (AORs).

Results: Recent Internet use for chlamydia/HIV testing or STI treatment (combined) was very low (men: 0.31%; women: 0.16%), whereas 2.35% of men and 0.51% of women reported obtaining condoms/contraceptives online. Additionally, 4.49% of men and 4.57% of women reported recent use of information/support websites for advice/help with their sex lives. Prevalence declined with age (men 16-24 years: 7.7%; 35-44 years: 1.84%, $P<.001$; women 16-24 years: 7.8%; 35-44 years: 1.84%, $P<.001$). Use of information/support websites was strongly associated with men's higher socioeconomic status (managerial/professional vs semiroutine/routine: AOR 1.93, 95% CI 1.27-2.93, $P<.001$). Despite no overall association with area-level deprivation, those in densely populated urban areas were more likely to report use of information/support websites than those living in rural areas (men: AOR 3.38, 95% CI 1.68-6.77, $P<.001$; women: AOR 2.51, 95% CI 1.34-4.70, $P<.001$). No statistically significant association was observed with number of sex partners reported after age adjustment, but use was more common among men reporting same-sex partners (last 5 years: AOR 2.44, 95% CI 1.27-4.70), women reporting sex with multiple partners without condoms (last year: AOR 1.90, 95% CI 1.11-3.26), and, among both sexes, reporting seeking sex online (last year, men: AOR 1.80, 95% CI 1.16-2.79;

women: AOR 3.00, 95% CI 1.76-5.13). No association was observed with reporting STI diagnosis/es (last 5 years) or (after age adjustment) recent use of any STI service or non-Internet sexual health seeking.

Conclusions: A minority in Britain used the Internet for the sexual health reasons examined. Use of information/support websites was reported by those at greater STI risk, including younger people, indicating that demand for online STI services, and Internet-based sexual health interventions in general, may increase over time in this and subsequent cohorts. However, the impact on health inequalities needs addressing during design and evaluation of online sexual health interventions so that they maximize public health benefit.

Introduction

Sexual health is increasingly recognized as encompassing physical, mental, and emotional well-being in relation to sexuality and sexual relationships, and freedom from coercion [1]. In Britain, and globally, there has been an expansion in online sexual health services [2-5]. As well as providing information, these services take advantage of the interactive potential of the Internet, such as for sexual health promotion [6], to aid contraceptive choices [7], or for individual counseling via Web chat [8,9]. Condoms and contraceptives are purchasable online from Internet vendors and pharmacies. Regarding sexually transmitted infections (STIs), England's National Chlamydia Screening Programme (NCSP) provides free, Internet-ordered home-sampling kits to those aged 16-24 years in many localities [5]. Privately provided Internet-ordered STI and human immunodeficiency virus (HIV) testing and STI treatment services are increasingly available, although they have been poorly regulated and of variable quality [3,5]. Recently, the British government legalized HIV home tests, which have been available for purchase online since 2015 [10].

Internet access is now nearly universal among people of reproductive age in the United Kingdom (98% aged 16-34 years, 93% aged 35-44 years in 2013) and more than one-third regularly uses the Internet to find information on health-related issues [11]. Although new Internet-based sexual health services continue to be developed [12-15], the number and characteristics of people who use currently available online sexual health services in Britain are unknown. To inform the design and delivery of new online sexual health interventions and services, we need to understand the demographic and behavioral characteristics of existing users. This will help inform whether Internet-based services could reach populations that underutilize conventional sexual health services relative to their need for sexual health care. This may include people at elevated risk of STI, such as young people (aged 16-24 years), people of black ethnic origins, men who have sex with men (MSM) [16], those who report multiple sexual partners, those living in deprived areas [17], and sexually active people who report no recent sexual health care use. This evidence is necessary for estimating the likely impact of online services which are currently being developed, and for informing the targeting of these services to maximize public health benefit. This study aims to fill this evidence gap by providing evidence of the British population's

use of existing Internet-based sexual health services and the population who report using them. We conjectured that those reporting use of the Internet for these reasons might represent a population likely to take up online sexual health services that are currently being developed. Our study's focus was on the year before the survey interview to provide a contemporary picture in a rapidly changing field.

Specific objectives were (1) to estimate the prevalence of reporting recent (in the previous year) use of the Internet as a source of chlamydia testing, HIV testing, STI treatment, condoms/contraceptive supplies, and help/advice with one's sex life from information/support websites among sexually experienced men and women; (2) to describe the population reporting this; and (3) to estimate the proportions reporting a preference for online sexual health care.

Methods

Natsal-3 Survey Design and Administration

Britain's third National Survey of Sexual Attitudes and Lifestyles (Natsal-3 [18,19]) is a probability sample survey conducted between 2010 and 2012 among the British resident population aged 16 to 74 years (N=15,162). Natsal-3 asked detailed demographic and behavioral questions and a number of questions about sources used for various types of sexual health care and advice/help with one's sex life (including the Internet). Detailed methods have been reported elsewhere; briefly, Natsal-3 used a multistage, clustered, and stratified probability sample design with a boost sample of those aged 16 to 34 years [18,19]. An interviewer visited each selected household and randomly selected one person in the eligible age range to participate, with oral informed consent. Participants completed the survey using a mixture of computer-assisted personal interview (CAPI) conducted face-to-face and computer-assisted self-interview (CASI) for the more sensitive questions [18,19]. Natsal-3 achieved an overall response rate of 57.7% and a cooperation rate (of eligible addresses contacted) of 65.8% [18,19].

The full survey is available online [20]. Variables used in this study were based on self-reported responses to closed-ended survey questions, except Index of Multiple Deprivation (IMD) quintile [21] and Output Area Classification (OAC) 2011 supergroup (OAC 2011 categorizes census output areas into 8 supergroups based on population characteristics) [22,23]. These

were added to the dataset according to participants' postcodes. National Statistics Socio-Economic Classification (NS-SEC) was derived from responses to standard questions [24].

Population of Interest: Sexually Experienced Persons Aged 16 to 44 Years

Several survey questions relevant to these analyses were not asked to participants aged 45 years and older. Therefore, the denominator for this study was limited to those aged 16 to 44 years, the age group in which most STI diagnoses occur [16], and which approximates women's reproductive age. We further limited the denominator to sexually experienced people, defined as those who reported ever having had any opposite- or same-sex sexual partners, because they are most likely to require sexual health services.

Outcome Variables

Outcome variables for this study included reported use of Internet services for key sexual health reasons (Table 1) and reporting the Internet as a preferred source of contraception, or for STI treatment/diagnosis if an STI was suspected (Table 2). The wording of these survey questions is described in Tables 1 and 2. Of specific relevance to the question about help/advice with one's sex life (first question in Table 1), shortly before this question, participants were presented with the following broad definition of *sex life*: "An individual's sex life includes their sexual thoughts, sexual feelings, sexual activity and sexual relationships."

For timeframe, the question on sources of contraceptive supplies referred to the past year. Questions on HIV testing, chlamydia testing, and STI treatment referred to the last occurrence. For comparability, only participants who indicated that this last occurrence was in the previous year (determined from responses to other survey questions) were included as reporting these behaviors.

Explanatory Variables

We had the following categories of explanatory variables: participants' sociodemographics, Internet access, area-level measures, sexual behavior (in the past year and past 5 years), sexual health care use, and STI diagnosis. Variables for sexual behavior and service use were selected to match the timeframe of the primary outcome variable (the year before the survey interview). Some variables corresponding to the 5 years before the interview were included (having had same-sex partners, number of sexual partners, sexual health clinic attendance, and STI diagnosis) to reflect greater variability in certain behaviors in the population over this longer period [25].

Statistical Methods

Data were analyzed using the complex survey functions of Stata 12 to take account of clustering, stratification, and weighting of the Natsal-3 sample. Weights were applied to adjust for unequal probabilities of selection for participation in the survey. All analyses were conducted separately by sex. Participants with missing data for a given variable were excluded from analyses using this variable because item nonresponse in Natsal-3 was low (typically less than 0.5% in the CAPI and 1%-3% in the CASI) [18].

Logistic regression was used to obtain crude odds ratios to compare estimates of the odds of reporting use of information/support websites for advice/help with one's sex life, by each explanatory variable. Multivariable logistic regression was used, adjusting only for age, as a potential confounder of associations with NS-SEC code, which contained a "full-time student" category; OAC 2011, which was based on population characteristics including age; recent STI diagnosis; and sexual behavior variables because young people report greater numbers of recent and new sexual partners than older adults [25].

The observed low prevalences of other outcome variables meant that it was not possible to explore their associated factors. Statistical significance was considered as $P<.05$ for all analyses.

Ethical Approval

The Natsal-3 study was approved by the Oxfordshire Research Ethics Committee A (Ref: 10/H0604/27).

Results

Prevalence of Reported Recent Use of the Internet for Selected Sexual Health Reasons

Among sexually experienced persons aged 16 to 44 years, Internet use for chlamydia testing, HIV testing, or STI treatment (combined) in the previous year was reported by 0.31% (12/3702) men and 0.16% (6/3716) women (Figure 1). (Note: numerators and denominators are weighted and rounded to the nearest integer so may be subject to rounding errors.) Mostly this was chlamydia testing. No one in the sample reported Internet treatment for STIs other than chlamydia. Also, no one aged 35 to 44 years reported using the Internet for chlamydia testing, HIV testing, or STI treatment. Use of Internet sources of contraception/condoms in the past year was a little more common, especially among men (men: 2.35%, 87/3702; women: 0.51%, 19/3716). (Participants were not asked which method they obtained online, but it is likely that this was mostly condoms: 114 of 122 men and women reporting obtaining contraceptive supplies online in the past year reported use of male [n=113] and/or female [n=2] condoms in this period.) Use of information and support websites for advice/help with one's sex life in the past year was more common still, reported by 4.49% (166/3702) men and 4.57% (170/3716) women. Overall, use of the Internet for any of these sexual health reasons in the past year was reported by 6.85% men (95% CI 6.02-7.78) and 5.15% women (95% CI 4.50-5.89). In contrast, 60.2% men (95% CI 58.2-62.1) and 71.7% women (95% CI 70.2-73.2) reported use of non-Internet sources of sexual health care or advice/help with their sex lives, in the past year. (We defined this as GUM clinic attendance; use of non-Internet sources of chlamydia/HIV testing, STI treatment, or condoms/contraceptive supplies; or non-Internet sources of advice/help with one's sex life, excluding self-help and friends/family, in the past year.)

Table 1. Details of the Natsal-3 survey questions used as outcome variables in these analyses of sexually experienced persons aged 16 to 44 years (unweighted N=8926, weighted N=7400).

Question wording	Timeframe; number of responses permitted	Response options	Respondents eligible for each survey question	Number eligible for each question, unweighted (weighted)
Have you sought help or advice regarding your sex life from any of the following sources in the last year?	During previous year; multiple responses	Information and support sites on the Internet;[a] family member/friend; self-help books/information leaflets; self-help groups; helpline; GP/family doctor; sexual health/GUM/STI clinic; psychiatrist or psychologist; relationship counsellor; other type of clinic or doctor; have not sought any help	Entire sample of the current study	8926 (7400)
Have you got contraception from any of these sources in the last year?	During previous year; multiple responses	Internet website;[a] a doctor or nurse at your GP's surgery; sexual health clinic (GUM clinic); family planning clinic / contraceptive clinic / reproductive health clinic; NHS antenatal clinic / midwife; private doctor or clinic; youth advisory clinic (eg, Brook clinic); pharmacy/chemist; supplies from school/college/university services; over the counter at a petrol station/supermarket/other shop; vending machine; mail order; hospital accident and emergency (A&E) department; any other type of place (please say where); I have not got contraception in the last year	Those reporting use of any contraceptive method[b] in the last year	7182 (5862)
When you were last tested for chlamydia, where were you offered the test?	Last occurrence; single response	Internet;[a] GP surgery; sexual health clinic (GUM clinic); NHS family planning clinic / contraceptive clinic / reproductive health clinic; antenatal clinic/midwife; private non-NHS clinics or doctor; youth advisory clinic (eg, Brook Clinic); School/college/university; termination of pregnancy (abortion) clinic; hospital accident and emergency (A&E) department; pharmacy/chemist; other non-health care place (eg, youth club, festival, bar); somewhere else	Those reporting chlamydia testing in the last year	2387 (1545)
Where were you tested? (the last HIV test if more than one)	Last occurrence; single response	Internet site offering postal kit;[a] GP surgery; sexual health clinic (GUM clinic); NHS family planning clinic / contraceptive clinic / reproductive health clinic; antenatal clinic / midwife; private non-NHS clinic or doctor; youth advisory clinic (eg, Brook clinic); termination of pregnancy (abortion) clinic; hospital accident and emergency (A&E) department; somewhere else	Those reporting HIV testing in the last year	802 (562)
Where were you last treated for [STI[c]]?	Last occurrence; single response	Internet site offering treatment;[a] GP surgery; sexual health clinic (GUM clinic); NHS family planning clinic / contraceptive clinic / reproductive health clinic; antenatal clinic / midwife; private non-NHS clinic or doctor; pharmacy/chemist; youth advisory clinic (eg, Brook clinic); termination of pregnancy (abortion) clinic; hospital accident and emergency (A&E) department; somewhere else	Those reporting having been told by a doctor / health professional that they had an STI in the last year	178 (117)

[a] Internet response options.

[b] Including condoms.

[c] Separate questions were asked about the following infections: chlamydia; gonorrhea; genital warts; syphilis; *Trichomonas vaginalis*; genital herpes; nonspecific urethritis (NSU) or nongonococcal urethritis (NGU).

Table 2. Natsal-3 survey questions about preferred sources of sexual health care.

Question wording[a]	Response options	Respondents eligible for each survey question	Number eligible for each question, un-weighted (weighted)
If you thought that you might have an infection that is transmitted by sex, where would you *first* go to seek diagnosis and/or treatment?	Internet site offering treatment;[b] GP surgery; sexual health clinic (GUM clinic); NHS Family planning clinic/contraceptive clinic/reproductive health clinic; NHS antenatal clinic/midwife; private non-NHS clinic or doctor; pharmacy/chemist; youth advisory clinic (eg, Brook clinic); hospital accident and emergency (A&E) department; somewhere else	Those reporting any life-time sexual partners	8858 (7338)
If all of these different types of service were available in your area and easy to get to, which one would *you* prefer to get contraception from?	NHS or Department of Health website;[b] a doctor or Nurse at your GP's surgery; sexual health clinic (GUM clinic); family planning clinic / contraceptive clinic / reproductive health clinic; youth advisory clinic (eg, Brook clinic); pharmacy/chemist; none of these; not needed	Those reporting use of any method in the last year	6909 (5524)

[a] Use of italics reflects emphasis given in the survey. One response could be selected at each question.

[b] Internet response options.

Figure 1. Percentage reporting seeking sexual health care and advice/help with one's sex life in the previous year, and specifically using the Internet to do so, among sexually experienced persons aged 16-44 years by gender and age group.
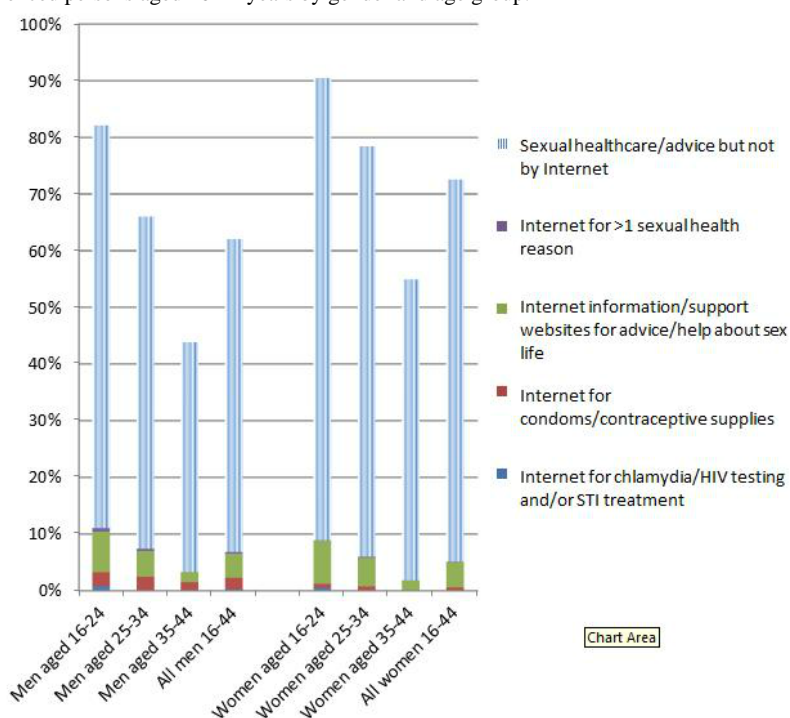
Associations with Reporting Use of Information and Support Websites for Advice/Help with One's Sex Life

Sociodemographic Factors

Mean age of men and women reporting use of Internet information/support websites for advice/help with their sex life (based on the first question described in Table 1 and hereon referred to as "Internet information/support" for brevity) was 25.9 years (SD 7.5) and 26.9 years (SD 8.8), respectively, in this sample aged 16 to 44 years. Those not reporting this were on average older (men: 31.0 years, SD 8.0; women: 31.3 years, SD 9.7). The prevalence of reporting use of Internet information/support declined steeply with increasing age among both sexes (7.7% men, 7.8% women aged 16-24 years to 1.84% men, 1.84% women aged 35-44, both *P*<.001). Tables 3 and 4 present univariate and age-adjusted analyses among men and women, respectively.

Table 3. Variation in the prevalence and odds of reporting recent (past year) use of information/support websites for advice/help with one's sex life (Internet information/support) among sexually experienced men aged 16 to 44 years.[a]

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
Sociodemographics						
Age (years)				<.001		—
16-24	1361 (994)	7.7% (6.3-9.4)	1		—	
25-34	1451 (1299)	4.93% (3.90-6.23)	0.62 (0.45-0.86)		—	
35-44	784 (1383)	1.84% (1.12-3.02)	0.22 (0.13-0.39)		—	
Ethnic group				.007		.004
White	3134 (3118)	4.01% (3.39-4.75)	1		1	
Asian/Asian British	190 (270)	6.9% (4.0-11.6)	1.77 (0.98-3.21)		2.11 (1.16-3.84)	
Black/black British	126 (140)	7.8% (3.7-15.4)	2.01 (0.92-4.42)		2.11 (0.93-4.81)	
Mixed/Chinese/other	108 (110)	9.4% (5.1-16.8)	2.49 (1.26-4.93)		2.2 (1.13-4.26)	
Education level [b]				<.001		<.001
No academic qualifications	252 (275)	0.8% (0.3-2.5)	0.60 (0.18-2.00)		0.65 (0.20-2.18)	
Academic qualifications typically gained at age 16	880 (912)	1.4% (0.8-2.3)	1		1	
Studying for/attained further academic qualifications	2354 (2419)	6.05% (5.13-7.13)	4.57 (2.68-7.78)		3.79 (2.20-6.51)	
Socioeconomic status [c]				<.001		.001
Managerial/professional	1060 (1262)	4.53% (3.42-5.98)	1.46 (0.97-2.19)		1.93 (1.27-2.93)	
Intermediate	509 (554)	3.0% (1.8-4.8)	0.94 (0.53-1.66)		1.16 (0.64-2.08)	
Semiroutine/routine	1321 (1300)	3.15% (2.40-4.11)	1		1	
No job	122 (99)	1.6% (0.4-6.4)	0.48 (0.11-2.08)		0.33 (0.08-1.42)	
Full-time student	574 (452)	11.1% (8.5-14.5)	3.85 (2.53-5.86)		1.95 (1.14-3.34)	
Internet access						
Access to Internet at home				.02		.02
Yes	3327 (3442)	4.73% (4.06-5.51)	1		1	
No	267 (232)	1.5% (0.6-3.9)	0.30 (0.11-0.82)		0.31 (0.11-0.84)	
Area-level measures						
Deprivation [d]				.51		.24
1 (least deprived)	642 (658)	5.7% (4.2-7.7)	1		1	
2	653 (699)	4.3% (3.1-6.0)	0.74 (0.46-1.20)		0.71 (0.44-1.14)	
3	690 (720)	4.6% (3.3-6.5)	0.81 (0.50-1.30)		0.76 (0.47-1.23)	
4	774 (823)	4.3% (2.9-6.4)	0.75 (0.45-1.26)		0.69 (0.41-1.15)	
5 (most deprived)	837 (776)	3.8% (2.7-5.3)	0.66 (0.41-1.06)		0.58 (0.36-0.93)	
Output Area Classification 2011				<.001		<.001
1: "Rural residents"	276 (294)	3.2% (1.8-5.6)	1		1	
2: "Cosmopolitans"	302 (329)	12.5% (9.0-17.2)	4.33 (2.17-8.63)		3.38 (1.68-6.77)	
3: "Ethnicity central"	181 (225)	5.4% (2.7-10.3)	1.71 (0.69-4.27)		1.58 (0.64-3.91)	
4: "Multicultural metropolitans"	516 (595)	3.7% (2.3-5.7)	1.15 (0.54-2.43)		1.04 (0.49-2.22)	
5: "Urbanites"	665 (667)	3.6% (2.4-5.3)	1.13 (0.55-2.30)		1.09 (0.53-2.24)	
6: "Suburbanites"	587 (597)	4.5% (3.2-6.3)	1.44 (0.72-2.85)		1.30 (0.65-2.59)	

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
7: "Constrained city dwellers"	331 (271)	4.1% (2.3-7.1)	1.28 (0.56-2.94)		1.06 (0.46-2.48)	
8: "Hard-pressed living"	738 (698)	2.8% (2.0-4.0)	0.87 (0.44-1.75)		0.76 (0.38-1.52)	
Sexual behavior, last year						
Number of sexual partners				.77		.29
0	191 (174)	4.6% (2.4-8.6)	1.06 (0.53-2.12)		0.95 (0.48-1.89)	
1	2320 (2612)	4.37% (3.63-5.26)	1		1	
2+	1051 (857)	5.0% (3.7-6.6)	1.14 (0.80-1.63)		0.72 (0.48-1.08)	
≥1 new sexual partners				<.001		.11
No	2129 (2503)	3.34% (2.71-4.12)	1		1	
Yes	1428 (1134)	7.14% (5.74-8.85)	2.22 (1.61-3.07)		1.39 (0.93-2.09)	
Number of sexual partners without a condom				.12		.30
0	862 (780)	5.9% (4.4-7.8)	1		1	
1	2139 (2412)	4.15% (3.40-5.05)	0.69 (0.48-0.98)		0.96 (0.66-1.38)	
≥2	523 (419)	4.5% (3.1-6.7)	0.75 (0.46-1.25)		0.69 (0.42-1.13)	
Seeking sexual partners online				.004		.009
No	3287 (3414)	4.28% (3.64-5.03)	1		1	
Yes	306 (257)	7.9% (5.4-11.6)	1.92 (1.24-3.00)		1.80 (1.16-2.79)	
Sexual behavior, last 5 years						
Number of sexual partners				.04		.96
0-1	1441 (1805)	3.63% (2.82-4.66)	1		1	
2-4	1106 (1012)	5.17% (3.98-6.70)	1.45 (0.99-2.13)		0.94 (0.63-1.41)	
≥5	1024 (837)	5.8% (4.4-7.6)	1.64 (1.11-2.42)		0.95 (0.60-1.49)	
≥1 same-sex partners				.002		.008
No	3459 (3561)	4.32% (3.68-5.06)	1		1	
Yes	137 (116)	10.9% (6.2-18.5)	2.71 (1.43-5.14)		2.44 (1.27-4.70)	
Sexual health care use and STI						
Non-Internet sexual health care or advice/help, last year [e]				.004		.42
Yes	2391 (2223)	5.46% (4.57-6.51)	1		1	
Not reported	1205 (1453)	3.10% (2.24-4.28)	0.55 (0.37-0.82)		0.84 (0.55-1.29)	
Attended STI clinic, last 5 years				.03		.89
Yes	861 (712)	5.9% (4.5-7.8)	1		1	
No	2670 (2902)	4.11% (3.41-4.95)	0.68 (0.48-0.97)		0.97 (0.67-1.41)	
STI service use, last year [f]				.27		.08
Yes	873 (703)	5.3% (3.9-7.0)	1		1	
Not reported	2723 (2974)	4.35% (3.64-5.19)	0.82 (0.57-1.17)		1.40 (0.96-2.02)	
STI [g] diagnosis, last 5 years				.68		.97
No	3300 (3408)	4.47% (3.81-5.24)	1		1	

XSL•FO
RenderX

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
Yes	257 (225)	5.0% (2.9-8.5)	1.13 (0.63-2.04)		0.99 (0.55-1.79)	

[a] Unweighted N=3614, weighted N=3697. Denominators vary due to item nonresponse.

[b] Denominator restricted to those aged 17 and older. No academic qualifications: left school at age 16 without passing any exams/gaining any qualifications (excludes qualifications gained at an older age); academic qualifications typically gained at age 16: left school at 16 having passed some exams/gained some qualifications (eg, English General Certificate of Secondary Education [GCSE] or equivalent); studying for or attained further academic qualifications: left school at age 17 or older.

[c] Based on National Statistics Socioeconomic Classification (NS-SEC) code. No job: no job of ≥10 hours per week in the last 10 years.

[d] Quintile of adjusted Index of Multiple Deprivation for Great Britain.

[e] Defined as reporting STI clinic attendance within the last year or responses other than "Internet" for questions listed in Table 1 within the last year. Exceptions (non-Internet responses which were ignored) were (1) where participants had indicated friend, parent/relative, or partner as sources of contraceptive supplies (free-text response) and (2) where participants had selected "family member/friend," "self-help books/information leaflets," "self-help groups," and "have not sought any help" as sources of advice/help about their sex life.

[f] Defined as reporting any of: STI clinic attendance, chlamydia testing, or HIV testing within this last year.

[g] Natsal definition of STIs excludes thrush.

A strong association was observed with education level; 1.4% of men and 2.0% of women who left school aged 16 years with General Certificates of Secondary Education (GCSEs), or equivalent qualifications, reported recent use of Internet information/support compared to 6.05% of men and 5.87% of women with, or studying for, further academic qualifications (both sexes: *P*<.001), an association which remained after age adjustment. Associations with socioeconomic status [24] followed similar trends. Men in high-status occupations were more likely to report Internet information/support than those in lower-status occupations, before and after age adjustment (managerial/professional men vs men in semiroutine/routine occupations: age-adjusted OR [AOR] 1.93, 95% CI 1.27-2.93, *P*<.001), whereas a similar finding among women reached borderline statistical significance after age adjustment. Full-time students of both genders were also more likely than those in lower-status occupations to report Internet information/support even after taking account of their younger age (men: AOR 1.95, 95% CI 1.14-3.34; women: AOR 1.93, 95% CI 1.24-3.00).

Despite associations with these individual measures of social status (education, socioeconomic status), no overall association was observed between recent use of Internet information/support and area-level deprivation [21]. Use of Internet information/support was associated with OAC 2011 supergroup.

Use was high among "cosmopolitans" (residents of densely populated urban areas characterized by relatively high proportions of single people, young adults, full-time students, and high ethnic integration) [23] (men: 12.5%, 95% CI 9.0-17.2; women 11.7%, 95% CI 8.3-16.3). There was little variation between other supergroups except, among women only, slightly lower use of Internet information/support in "hard-pressed living" areas (mostly urban areas in Northern England and Wales with higher unemployment and lower proportions with higher-level qualifications than the national average) [23]. Strong associations with OAC 2011 supergroup remained after age adjustment (see Tables 3 and 4).

No overall association with ethnicity was observed among women after age adjustment, but Asian/Asian British men were more likely to report use of Internet information/support than white men (AOR 2.11, 95% CI 1.16-3.84, *P*=.004). Notably, numbers in minority ethnic groups were relatively small.

Having home Internet access was reported by 93.5% (95% CI 92.9-94.0) of sexually experienced persons aged 16 to 44 years. The minority who did not have home Internet were less likely to report use of Internet information/support than those who had (men: OR 0.30, 95% CI 0.11-0.82, *P*=.02; women: OR 0.26, 95% CI 0.11-0.58, *P*<.001) with little change after adjusting for age.

Table 4. Variation in the prevalence and odds of reporting recent (past year) use of Internet information/support among sexually experienced women aged 16 to 44 years.[a]

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
Sociodemographics						
Age (years)				<.001		—
16-24	1713 (956)	7.8% (6.4-9.4)	1		—	
25-34	2386 (1317)	5.28% (4.32-6.45)	0.66 (0.49-0.89)		—	
35-44	1175 (1409)	1.84% (1.16-2.90)	0.22 (0.13-0.37)		—	
Ethnic group				.02		.07
White	4619 (3179)	4.39% (3.76-5.10)	1		1	
Asian/Asian British	258 (220)	3.8% (2.2-6.4)	0.86 (0.49-1.52)		0.96 (0.54-1.70)	
Black/black British	174 (136)	5.6% (3.0-10.2)	1.30 (0.67-2.52)		1.34 (0.70-2.59)	
Mixed/Chinese/other	176 (117)	11.1% (6.1-19.3)	2.71 (1.39-5.28)		2.32 (1.20-4.50)	
Education level [b]				<.001		<.001
No academic qualifications	372 (237)	0.6% (0.2-1.9)	0.29 (0.08-1.04)		0.28 (0.08-0.98)	
Academic qualifications typically gained at age 16	1186 (863)	2.0% (1.3-3.1)	1		1	
Studying for/attained further academic qualifications	3607 (2528)	5.87% (5.07-6.79)	3.05 (1.88-4.97)		2.49 (1.52-4.06)	
Socioeconomic status [c]				<.001		.06
Managerial/professional	1526 (1202)	4.08% (3.16-5.26)	1.21 (0.79-1.85)		1.56 (1.02-2.40)	
Intermediate	1006 (719)	3.9% (2.5-5.9)	1.14 (0.66-1.97)		1.32 (0.76-2.29)	
Semiroutine/routine	1582 (1028)	3.39% (2.50-4.60)	1		1	
No job	418 (285)	4.6% (2.9-7.3)	1.39 (0.78-2.46)		1.39 (0.79-2.46)	
Full-time student	717 (429)	10.2% (7.9-13.1)	3.23 (2.14-4.89)		1.93 (1.24-3.00)	
Internet access						
Access to Internet from home				.001		<.001
Yes	4828 (3444)	4.84% (4.21-5.56)	1		1	
No	443 (236)	1.3% (0.6-2.8)	0.26 (0.11-0.58)		0.23 (0.10-0.52)	
Area-level measures						
Deprivation [d]				.58		.35
1 (least deprived)	847 (632)	5.5% (4.0-7.4)	1		1	
2	952 (699)	4.4% (3.1-6.1)	0.79 (0.49-1.29)		0.78 (0.48-1.26)	
3	1031 (739)	4.8% (3.5-6.7)	0.88 (0.55-1.41)		0.83 (0.51-1.35)	
4	1183 (821)	4.8% (3.5-6.5)	0.87 (0.55-1.38)		0.82 (0.51-1.29)	
5 (most deprived)	1261 (792)	3.7% (2.7-5.1)	0.68 (0.42-1.08)		0.61 (0.38-0.97)	
Output Area Classification 2011				<.001		<.001
1: "Rural residents"	414 (313)	4.0% (2.5-6.4)	1		1	
2: "Cosmopolitans"	349 (266)	11.7% (8.3-16.3)	3.20 (1.72-5.96)		2.51 (1.34-4.70)	
3: "Ethnicity central"	307 (257)	5.7% (3.5-9.0)	1.45 (0.72-2.91)		1.32 (0.65-2.68)	
4: "Multicultural metropolitans"	772 (557)	5.5% (3.9-7.7)	1.40 (0.76-2.57)		1.27 (0.69-2.36)	
5: "Urbanites"	961 (667)	4.8% (3.4-6.6)	1.20 (0.65-2.22)		1.14 (0.61-2.14)	
6: "Suburbanites"	799 (608)	4.1% (2.8-5.8)	1.02 (0.55-1.90)		1.02 (0.55-1.92)	
7: "Constrained city dwellers"	488 (277)	3.3% (2.0-5.4)	0.83 (0.41-1.69)		0.70 (0.35-1.42)	

XSL•FO
RenderX

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
8: "Hard-pressed living"	1184 (736)	2.0% (1.3-3.1)	0.50 (0.26-0.94)		0.45 (0.24-0.86)	
Sexual behavior, last year						
Number of sexual partners				.008		.65
0	284 (187)	3.2% (1.7-6.0)	0.75 (0.38-1.48)		0.88 (0.45-1.73)	
1	3829 (2825)	4.22% (3.58-4.96)	1		1	
≥2	1113 (631)	6.9% (5.2-9.2)	1.69 (1.19-2.40)		1.18 (0.81-1.72)	
≥1 new sexual partners				<.001		.11
No	3670 (2748)	3.82% (3.19-4.56)	1		1	
Yes	1553 (892)	7.2% (5.7-8.9)	1.95 (1.43-2.65)		1.32 (0.94-1.85)	
Number of partners without a condom				<.001		.03
0	1007 (680)	4.3% (3.1-5.8)	1		1	
1	3620 (2635)	4.12% (3.47-4.89)	0.97 (0.67-1.40)		1.05 (0.73-1.50)	
≥2	575 (317)	10.0% (7.1-13.9)	2.51 (1.50-4.17)		1.90 (1.11-3.26)	
Seeking sexual partners online				<.001		<.001
No	5079 (3559)	4.38% (3.78-5.06)	1		1	
Yes	189 (116)	11.8% (7.5-18.1)	2.93 (1.74-4.94)		3.00 (1.76-5.13)	
Sexual behavior, last 5 years						
Number of sexual partners				<.001		.18
0-1	2649 (2116)	3.77% (3.05-4.65)	1		1	
2-4	1630 (995)	4.6% (3.6-5.8)	1.23 (0.88-1.71)		0.88 (0.63-1.23)	
≥5	958 (541)	8.1% (6.1-10.7)	2.25 (1.53-3.29)		1.31 (0.85-2.01)	
≥1 same-sex partners				.09		.24
No	4972 (3493)	4.47% (3.87-5.16)	1		1	
Yes	302 (189)	7.2% (4.3-11.9)	1.65 (0.93-2.93)		1.42 (0.80-2.52)	
Sexual health care use and STI						
Non-Internet sexual health care or advice/help, last year [e]				<.001		.11
Yes	4055 (2648)	5.42% (4.66-6.30)	1		1	
Not reported	1219 (1034)	2.53% (1.70-3.75)	0.45 (0.29-0.71)		0.68 (0.42-1.10)	
Attended STI clinic, last 5 years				<.001		.14
Yes	1342 (779)	7.4% (5.9-9.4)	1		1	
No	3865 (2855)	3.90% (3.27-4.63)	0.51 (0.37-0.69)		0.76 (0.53-1.09)	
STI service use, last year [f]				.02		.61
Yes	1908 (1130)	5.80% (4.65-7.22)	1		1	
Not reported	3366 (2552)	4.08% (3.39-4.90)	0.69 (0.51-0.94)		1.10 (0.77-1.58)	
STI [g] diagnosis, last 5 years				.75		.14
No	4830 (3419)	4.65% (4.03-5.36)	1		1	

Variable	N, unweighted (weighted)	Prevalence (95% CI)	OR (95% CI)	P	AOR (95% CI)	P
Yes	398 (225)	4.2% (2.4-7.3)	0.91 (0.50-1.64)		0.63 (0.35-1.16)	

[a] Unweighted N=5312, weighted N=3703. Denominators vary due to item nonresponse.

[b] Denominator restricted to those aged 17 and older. No academic qualifications: left school at age 16 without passing any exams/gaining any qualifications (excludes qualifications gained at an older age); academic qualifications typically gained at age 16: left school at 16 having passed some exams/gained some qualifications (eg, English General Certificate of Secondary Education [GCSE] or equivalent); studying for or attained further academic qualifications: left school at age 17 or older.

[c] Based on National Statistics Socioeconomic Classification (NS-SEC) code. No job: no job of ≥10 hours per week in the last 10 years.

[d] Quintile of adjusted Index of Multiple Deprivation for Great Britain.

[e] Defined as reporting STI clinic attendance within the last year or responses other than "Internet" for questions listed in Table 1 within the last year. Exceptions (non-Internet responses which were ignored) were (1) where participants had indicated friend, parent/relative, or partner as sources of contraceptive supplies (free-text response) and (2) where participants had selected "family member/friend," "self-help books/information leaflets," "self-help groups," and "have not sought any help" as sources of advice/help about their sex life.

[f] Defined as reporting any of: STI clinic attendance, chlamydia testing, or HIV testing within this last year.

[g] Natsal definition of STIs excludes thrush.

Sexual Behavioral Factors

Use of Internet information/support was more commonly reported by women (but not men) reporting multiple sexual partners in the last year and among both men and women reporting new sexual partners in the last year, but these associations disappeared after age adjustment. Among women (but not men), use of Internet information/support was more commonly reported by those who reported multiple sexual partners in the previous year with whom they had not used condoms (AOR 1.90, 95% CI 1.11-3.26, P=.03). Men reporting sex with another man in the previous 5 years were more likely to report use of Internet information/support (AOR 2.44, 95% CI 1.27-4.70, P=.008), whereas no association with same-sex sex in the previous 5 years was observed among women. Men and women reporting seeking sexual partners online within the previous year were more likely to report use of Internet information/support than those not reporting seeking partners in this way (men: AOR 1.80, 95% CI 1.16-2.79; women: AOR 3.00, 95% CI 1.76-5.13).

Sexual Health Care Use

No association was observed between reporting use of Internet information/support and reporting STI diagnosis or diagnoses in the past 5 years. Use of Internet information/support was more common among those reporting recent non-Internet sources of sexual health care and advice/help, and having attended an STI clinic in the last 5 years, but not after adjusting for age. No association was observed with having used STI services in the previous year.

Preference for Internet Sources of Diagnosis/Treatment of Sexually Transmitted Infections and Condoms/Contraception

Less than 2% of sexually experienced participants aged 16 to 44 years reported that the first place they would look for diagnosis/treatment if they suspected that they had an STI would be an Internet site offering treatment. Among sexually experienced persons aged 16 to 44 years reporting use of any contraceptive method in the previous year, 5.45% men and 1.14% women indicated they would prefer to obtain supplies from an NHS or Department of Health website (Table 5).

Table 5. Preference for Internet sources of diagnosis/treatment of sexually transmitted infections and condoms/contraception.

Header	Men		Women	
	N, unweighted (weighted)	% (95% CI)	N, unweighted (weighted)	% (95% CI)
Would first look on an Internet site offering treatment for diagnosis/treatment if STI suspected[a]	3589 (3668)	1.77% (1.27-2.46)	5269 (3670)	0.81% (0.57-1.14)
Preferred source of contraceptive supplies would be NHS/Dept of Health website[a]	2793 (2743)	5.45% (4.52-6.56)	4116 (2781)	1.14% (0.82-1.58)

[a] Question wording, response options, and eligible respondents are detailed in Table 2.

Discussion

Principal Findings

Although Internet access is nearly universal in Britain, data from a recent national probability sample survey show that use of the Internet for key sexual health reasons is rare in the British population. Specifically, prevalence of reported use of Internet STI services is very low and reported use of the Internet for condoms/contraceptive supplies is also uncommon, particularly among women. Reporting recent use of Internet information and support websites for help/advice about one's sex life was slightly higher, especially among younger people and among

XSL•FO

RenderX

those who reported higher sexual risk behavior, including MSM and people who sought sexual partners online. However, those using information/support websites for advice/help with their sex lives may be from populations typically considered to have better access to sexual health care: the better-educated, residents of certain urban areas, and (among men) those of higher socioeconomic status. Despite this potential social inequality, those who reported recent use of information/support websites were as likely to report at least one previous STI diagnosis as those who did not report this.

Findings in Relation to Other Studies

We know of no other studies that have estimated the prevalence of use of the Internet for sexual health reasons or identified associated factors in a nationally representative sample. The association we found between use of information/support websites for advice/help with one's sex life, and younger age, is unsurprising given young people's greater Internet use [26], smartphone ownership [27-29], and greater need for sexual health care indicated by levels of reported STI diagnoses and STI clinic use [16,17]. Research on the acceptability of using the Internet to deliver conventional sexual health services reveals similar findings with respect to age [30-32] and education [33].

Differences in men's and women's sexual behaviors [34,35] and health-seeking behaviors [36-38] are well-documented, but our study revealed little difference by sex in the prevalence of reported use of information/support websites for advice/help with one's sex life (although there were some differences in associations observed among men and women, and more men than women reported that they would first look online for diagnosis/treatment if they suspected that they had an STI). In the English chlamydia screening program, the NCSP, more tests are performed among young women than among young men [5], perhaps due to women's greater engagement with contraceptive and other health services where screening is offered. Women also account for a greater proportion of tests in the NCSP's Internet-ordered home-sampling services, but with less discrepancy by gender compared to other NCSP testing venues (with the exception of military settings) [5]. In our study, use of the Internet for condoms/contraceptive supplies was reported by more men than women, perhaps reflecting gendered norms about who obtains condoms [39].

Surveys of patients attending genitourinary medicine (GUM) clinics in England, conducted almost a decade before Natsal-3, found patients reporting Internet sex seeking were also more likely to report use of the Internet for sexual health information [40], similar to the association we observed between Internet sex seeking and use of information/support websites.

Echoing our study's finding, little difference was found by IMD quintile in the proportion of NCSP Internet-ordered chlamydia home-sampling kits returned (2010) [5]. However, we found no studies using NCSP data to compare demographic or behavioral characteristics of those using Internet-ordered kits with the wider population in the target age range. Although Internet-based sexual health services have been viewed as a promising way of reaching rural populations, we found relatively low use of information/support websites in these areas.

Strengths and Limitations

Use of Natsal-3 data has allowed our analyses to examine a wide range of sociodemographic, behavioral, and health service use variables, in a sample representative of the resident British population, in relation to use of information/support websites for advice/help with one's sex life. Despite survey data being self-reported and, therefore, subject to recall and social desirability biases, they are of high quality; use of CASI was demonstrated to facilitate reliable reporting of sensitive information [41] and cognitive testing of several survey modules maximized the likelihood of questions being interpreted as intended [42]. Furthermore, the survey's response rate was similar to that achieved for other major social surveys undertaken in Britain at that time [43,44] and item nonresponse was typically very low [18,19]. Importantly, in this rapidly evolving field, we focused on reported behavior in the year before the survey interview and Natsal-3 data are relatively recent (collected 2010-2012). However, there may have been changes in norms regarding Internet use for sexual health since data collection.

The very low prevalence of most outcome variables examined meant that their associations could not be explored. The exception was reported use of the Internet for advice/help with one's sex life, but even this was reported by less than 5% of the study population; therefore, rare behaviors could not be included as explanatory variables in the analysis. We adjusted only for age in the multivariable analysis. Due to small numbers in some subgroups, we had to treat some variables crudely (eg, ethnicity), creating categories large enough to obtain sufficient subgroup sizes. This limits explanatory potential; for example, we cannot explore differences between black Caribbean and black African ethnicities. The subgroup mixed/Chinese/other is not particularly meaningful, although creation of this category gave sufficient subgroup sizes to explore associations with Britain's major ethnic groups (Asian, black, white).

Natsal-3 survey questions (Tables 1 and 2) serve various purposes and were not designed for our particular study. We cannot be sure about how questions were interpreted. Our main outcome variable (use of information/support websites for advice/help with one's sex life) was based on responses to a question located in the survey module entitled "Sexual Function." However, we assumed that the question was interpreted more broadly than about sexual function alone because "sexual function" was neither mentioned in the question nor visible on the computer screen at the time, and sex life was defined broadly (see Methods). Supporting our assumption, we found that among sexually active persons aged 16 to 44 years who reported use of information/support websites at this question, more than half agreed that they felt satisfied with their sex life, more than half disagreed that they felt distressed or worried about it, and more than two-thirds disagreed that they had avoided sex because of sexual difficulties (their own or a partner's; data not shown). This suggests that many who reported use of Internet help/advice with their sex life were doing so for reasons other than sexual function problems, although we cannot be sure. In terms of applicability of our findings to sexual health broadly defined [1], our variable may not have captured use of the Internet in relation to all aspects

XSL•FO

RenderX

of sexual health, such as support and counseling following nonvolitional sex. It seems unlikely that participants would have considered this type of service use to be help/advice with their sex life, although perhaps they would if nonvolitional sex occurred in the context of a sexual relationship.

An advantage of our study is that we were able to consider those who had used the Internet for a range of sexual health reasons and also those who would prefer to use it for sexual health care, although we lack data on which particular websites were used/preferred. However, the low proportions who reported a preference for using the Internet for STI diagnosis/treatment, or a preference for accessing contraception from an NHS website, probably underestimate the proportions that might choose Internet-based services if they were well-regulated and based in the NHS. This is because relevant survey questions (Table 2) each allowed a single response and provided no description of the Internet services, which might be difficult for respondents to envisage or assumed to be costly because such services are not currently available through the NHS. The question also specified "if an STI was suspected": in this context, a consultation with a health care professional may seem most appropriate, whereas for a routine STI check-up, Internet services might hold greater appeal. Given how common it has become to look up symptoms and health information online before contacting a health professional, we believe that responses to the STI diagnosis/treatment question might poorly reflect the proportion that would use an Internet-ordered test if they found a reputable service offering this during their online search.

We acknowledge that even an ideal survey question cannot give us a definitive answer about who will use online sexual health interventions and services in the future. However, we feel our main outcome variable, which addresses use of information/support websites (as distinct from lay advice/help sought online) for sexual health broadly defined, reflects those who may take up online sexual health services and interventions, fitting with their existing sexual health-seeking behavior.

Implications for Policy and Practice

Low levels of use of the Internet for contraception and STI services may reflect the limited availability and quality of currently available online services—particularly at the time the data were collected (2010-12) and in relation to STI testing and treatment [3,5,45]. Also, many methods of contraception cannot feasibly or legally be provided online. Qualitative and quantitative research could explore awareness, expectations, and barriers to use of currently available online sexual health services.

Greater proportions reported use of information/support websites for advice/help with their sex lives, particularly among young people. This suggests scope for expansion of provision in the future, in this cohort and subsequent cohorts who have also grown up with the Internet, and as the range and quality of Internet sexual health care increase (as is likely given existing trends). An example of improved quality is the legalization and regulation of HIV home testing in the United Kingdom, available online [10]. However, our study suggests that if use of Internet sexual health care followed patterns of online help/advice seeking, health inequalities might increase, especially if expansion of online sexual health care was coupled with reduced provision of conventional sexual health care. "Digital divides" by socioeconomic status have been widely documented [11], with eHealth a specific area of concern [46,47]. This study's findings regarding education and socioeconomic status may reflect that Internet use is lower among those with less education and lower incomes [48]. Although home Internet access was high in the population of interest in Natsal-3, the survey did not ask about Internet use more generally, including via a personal device, which may vary across social strata. Having a smartphone or laptop/tablet might allow greater access to the Internet for sexual health than a household's shared personal computer if privacy from other household members is important. Since the data were collected for Natsal-3 between 2010 and 2012, there have been further increases in smartphone ownership [49,50] and Internet access [51], which may reduce differences in proportions using the Internet for sexual health by socioeconomic status and/or education. However, if these differences relate to differences in health care-seeking behavior, inequalities may be more persistent. Research should examine these associations further and evaluations of new Internet-based interventions and services should monitor and model impacts on both on STI transmission and on health inequalities. Interventions may also be required to promote eHealth should groups be identified that have good Internet access, yet are underserved by online and conventional health care.

Conflicts of Interest

None declared.

References

1. World Health Organization. Defining Sexual Health: Report of a Technical Consultation on Sexual Health, 28-31 January 2002, Geneva. Geneva: World Health Organization; 2006. URL: http://www.who.int/reproductivehealth/publications/sexual_health/defining_sexual_health.pdf [accessed 2015-02-09] [WebCite Cache ID 6WDV1Q9GJ]

2. Greacen T, Friboulet D, Fugon L, Hefez S, Lorente N, Spire B. Access to and use of unauthorised online HIV self-tests by internet-using French-speaking men who have sex with men. Sex Transm Infect 2012 Aug;88(5):368-374. [doi: 10.1136/sextrans-2011-050405] [Medline: 22436195]

3. Owens SL, Arora N, Quinn N, Peeling RW, Holmes KK, Gaydos CA. Utilising the internet to test for sexually transmitted infections: results of a survey and accuracy testing. Sex Transm Infect 2010 Apr;86(2):112-116. [doi: 10.1136/sti.2009.037226] [Medline: 19854701]

4. Peeling RW. Testing for sexually transmitted infections: a brave new world? Sex Transm Infect 2006 Dec;82(6):425-430 [FREE Full text] [doi: 10.1136/sti.2005.017251] [Medline: 17151028]

5. Woodhall SC, Sile B, Talebi A, Nardone A, Baraitser P. Internet testing for Chlamydia trachomatis in England, 2006 to 2010. BMC Public Health 2012;12:1095 [FREE Full text] [doi: 10.1186/1471-2458-12-1095] [Medline: 23253518]

6. 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]

7. French RS, Cowan FM, Wellings K, Dowie J. The development of a multi-criteria decision analysis aid to help with contraceptive choices: My Contraception Tool. J Fam Plann Reprod Health Care 2014 Apr;40(2):96-101. [doi: 10.1136/jfprhc-2013-100699] [Medline: 24265469]

8. Survivors UK. URL: http://www.survivorsuk.org/web-chat/ [accessed 2015-12-01] [WebCite Cache ID 6dRx2zjZd]

9. Relate. Relationship help: talk to someone URL: http://www.relate.org.uk/relationship-help/talk-someone [accessed 2015-02-09] [WebCite Cache ID 6WDYlJY9T]

10. Mundasad S. BBC News. 2015 Apr 27. HIV home test kit goes on sale in UK URL: http://www.bbc.co.uk/news/health-32453192 [accessed 2015-09-23] [WebCite Cache ID 6bkvrnaMp]

11. Ofcom. Adults' Media Use and Attitudes Report, 2014. 2014 Apr. URL: http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/adults-2014/2014_Adults_report.pdf [accessed 2015-02-09] [WebCite Cache ID 6WDZ4Wcj3]

12. Averbach S, Wendt JM, Levine DK, Philip SS, Klausner JD. Increasing access to emergency contraception through online prescription requests. J Reprod Med 2010 Mar;55(3-4):157-160. [Medline: 20506679]

13. Gibbs J, Sutcliffe L, Sadiq T, Sonnenberg P, Ashcroft R, Gkatzidou V, et al. TP 110 Development of an automated online clinical consultation with electronic antibiotic prescribing for the remote management of genital Chlamydia trachomatis infection within the eSTI2 Consortium. In: Sex Transm Dis. 2014 Jun Presented at: 2014 STD Prevention Conference; Jun 9-12, 2014; Atlanta, GA p. S69 URL: http://journals.lww.com/stdjournal/Citation/2014/06001/2014_STD_Prevention_Conference_.1.aspx

14. Fuller SS, Aicken C, Sutcliffe LJ, Estcourt CS, Gkatzidou V, Hone K, et al. O22.4 What are young people's perceptions of using electronic self-tests for STIs linked to mobile technology for diagnosis and care (eSTI2)? Sex Transm Infect 2013 Jul 13;89(Suppl 1):A69-A70. [doi: 10.1136/sextrans-2013-051184.0212]

15. Spielberg F, Levy V, Lensing S, Chattopadhyay I, Venkatasubramanian L, Acevedo N, et al. Fully integrated e-services for prevention, diagnosis, and treatment of sexually transmitted infections: results of a 4-county study in California. Am J Public Health 2014 Dec;104(12):2313-2320. [doi: 10.2105/AJPH.2014.302302] [Medline: 25320878]

16. Public Health England. Health Protection Report 9(22). 2015 Jun 23. Sexually transmitted infections and chlamydia screening in England, 2014 URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/437433/hpr2215_STI_NCSP_v6.pdf [accessed 2015-09-23] [WebCite Cache ID 6bl80mNzu]

17. Sonnenberg P, Clifton S, Beddows S, Field N, Soldan K, Tanton C, et al. Prevalence, risk factors, and uptake of interventions for sexually transmitted infections in Britain: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). Lancet 2013 Nov 30;382(9907):1795-1806 [FREE Full text] [doi: 10.1016/S0140-6736(13)61947-9] [Medline: 24286785]

18. Erens B, Phelps A, Clifton S, Hussey D, Mercer C, Tanton C, et al. Natsal. 2013. National survey of sexual attitudes and lifestyles 3: technical report URL: http://www.natsal.ac.uk/media/823219/natsal-3-technical-report.pdf [accessed 2015-02-13] [WebCite Cache ID 6WJKbenjB]

XSL•FO
RenderX

19. Erens B, Phelps A, Clifton S, Mercer CH, Tanton C, Hussey D, et al. Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). Sex Transm Infect 2014 Mar;90(2):84-89 [FREE Full text] [doi: 10.1136/sextrans-2013-051359] [Medline: 24277881]

20. Natsal. The National Survey of Sexual Attitudes and Lifestyles Questionnaire: Revised for Natsal 3. 2013. URL: http://natsal.ac.uk/media/2097/final-questionnaire_technical-report-appendix-b.pdf [WebCite Cache ID 6bl8TlDy3]

21. Payne R, Abel G. UK indices of multiple deprivation-a way to make comparisons across constituent countries easier. Health Stat Q 2012 Feb 21;53(Spring 2012):21-37 [FREE Full text]

22. Office for National Statistics. 2014. 2011 area classifications URL: http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/ns-2011-area-classifications/index.html [accessed 2015-09-23] [WebCite Cache ID 6bl9LN06J]

23. Office for National Statistics. Office for National Statistics. 2014. Pen portraits and radial plots URL: http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/ns-2011-area-classifications/pen-portraits-and-radial-plots/index.html [accessed 2015-02-10] [WebCite Cache ID 6WF3mh0Ws]

24. Office for National Statistics. The Standard Occupational Classification (SOC) 2010. Volumes 1-3. Basingstoke: Palgrave Macmillan; 2010.

25. Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, et al. Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). Lancet 2013 Nov 30;382(9907):1781-1794 [FREE Full text] [doi: 10.1016/S0140-6736(13)62035-8] [Medline: 24286784]

26. Pew Research Center. 2014. Internet user demographics URL: http://www.pewinternet.org/data-trend/internet-use/latest-stats/ [accessed 2015-02-10] [WebCite Cache ID 6WF4DPCsz]

27. Pew Research Center. 2014. Cell phone and smartphone ownership demographics URL: http://www.pewinternet.org/data-trend/mobile/cell-phone-and-smartphone-ownership-demographics/ [WebCite Cache ID 6WF4f4nKI]

28. International Telecommunication Union. The World in 2010: ICT Facts and Figures. Geneva: ITU; 2010. URL: http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF6iOuwL]

29. YouGov. YouGov Survey Results: Smartphones. 2011. URL: http://d25d2506sfb94s.cloudfront.net/today_uk_import/yg-archives-cons-firstsourcesolutions-smartphones-080811.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF6Hj365]

30. Miners A, Llewellyn C, Pollard A, Lagarde M, Richardson D, Cairns J, et al. Assessing user preferences for sexually transmitted infection testing services: a discrete choice experiment. Sex Transm Infect 2012 Nov;88(7):510-516 [FREE Full text] [doi: 10.1136/sextrans-2011-050215] [Medline: 22661632]

31. Challenor R, Deegan S. What service users want: a new clinic results service. Can we satisfy both patients' needs and wants? Int J STD AIDS 2009 Oct;20(10):701-703. [doi: 10.1258/ijsa.2009.009077] [Medline: 19815915]

32. Tripathi A, Duffus WA, Kissinger P, Brown TJ, Gibson JJ, Mena LA. Delivering laboratory results by text message and e-mail: a survey of factors associated with conceptual acceptability among STD clinic attendees. Telemed J E Health 2012 Sep;18(7):500-506. [doi: 10.1089/tmj.2011.0251] [Medline: 22827295]

33. Greenland KE, Op de Coul EL, van Bergen JE, Brouwers EE, Fennema HJ, Götz HM, et al. Acceptability of the internet-based Chlamydia screening implementation in the Netherlands and insights into nonresponse. Sex Transm Dis 2011 Jun;38(6):467-474. [doi: 10.1097/OLQ.0b013e318204546e] [Medline: 21217416]

34. Johnson AM, Mercer CH, Erens B, Copas AJ, McManus S, Wellings K, et al. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. Lancet 2001 Dec 1;358(9296):1835-1842. [doi: 10.1016/S0140-6736(01)06883-0] [Medline: 11741621]

35. Wellings K, Nanchahal K, Macdowall W, McManus S, Erens B, Mercer CH, et al. Sexual behaviour in Britain: early heterosexual experience. Lancet 2001 Dec 1;358(9296):1843-1850. [doi: 10.1016/S0140-6736(01)06885-4] [Medline: 11741623]

36. Adamson J, Ben-Shlomo Y, Chaturvedi N, Donovan J. Ethnicity, socio-economic position and gender--do they affect reported health-care seeking behaviour? Soc Sci Med 2003 Sep;57(5):895-904. [Medline: 12850114]

37. Moreira ED, Brock G, Glasser DB, Nicolosi A, Laumann EO, Paik A, GSSAB Investigators' Group. Help-seeking behaviour for sexual problems: the global study of sexual attitudes and behaviors. Int J Clin Pract 2005 Jan;59(1):6-16. [doi: 10.1111/j.1742-1241.2005.00382.x] [Medline: 15707457]

38. O'Brien R, Hunt K, Hart G. 'It's caveman stuff, but that is to a certain extent how guys still operate': men's accounts of masculinity and help seeking. Soc Sci Med 2005 Aug;61(3):503-516. [doi: 10.1016/j.socscimed.2004.12.008] [Medline: 15899311]

39. Fennell JL. Men bring condoms, women take pills: men's and women's roles in contraceptive decision making. Gender & Society 2011 Aug 04;25(4):496-521. [doi: 10.1177/0891243211416113]

40. Malu MK, Challenor R, Theobald N, Barton SE. Seeking and engaging in internet sex: a survey of patients attending genitourinary medicine clinics in Plymouth and in London. Int J STD AIDS 2004 Nov;15(11):720-724. [doi: 10.1258/0956462042395230] [Medline: 15537455]

41. Johnson AM, Copas AJ, Erens B, Mandalia S, Fenton K, Korovessis C, et al. Effect of computer-assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment. AIDS 2001 Jan 5;15(1):111-115. [Medline: 11192852]

XSL•FO
RenderX

42. Gray M, Nicholson S. National Survey of Sexual Attitudes and Lifestyles 2010: Findings and Recommendations from Cognitive Question Testing. London: National Centre for Social Research; 2009 Apr. URL: http://www.natsal.ac.uk/media/822297/natsal%202010_cognitive%20pilot%20report%20-%20final.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF73cHSq]

43. NatCen Social Research. In: Park A, editor. British Social Attitudes 28 (British Social Attitudes Survey series). London: SAGE Publications Ltd; 2012.

44. Craig R, Mindell J. Health and Social Care Information Centre. 2011 Dec 15. Health Survey for England 2010-volume 1: respiratory health URL: http://www.hscic.gov.uk/pubs/hse10report [accessed 2015-02-10] [WebCite Cache ID 6WF7dL4hW]

45. Vivancos R, Schelenz S, Loke YK. Internet treatment of sexually transmitted infections-a public health hazard? BMC Public Health 2007;7:333 [FREE Full text] [doi: 10.1186/1471-2458-7-333] [Medline: 18005431]

46. Hsu J, Huang J, Kinsman J, Fireman B, Miller R, Selby J, et al. Use of e-Health services between 1999 and 2002: a growing digital divide. J Am Med Inform Assoc 2005;12(2):164-171 [FREE Full text] [doi: 10.1197/jamia.M1672] [Medline: 15561786]

47. Lustria MLA, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]

48. Office for National Statistics Statistical Bulletin. 2010 Jul 27. Internet access 2010, households and individuals URL: http://www.ons.gov.uk/ons/publications/index.html?pageSize=50&sortBy=none&sortDirection=none&newquery=Internet+access+households+and+individuals&content-type=publicationContentTypes

49. eDigital Research. 2013. The explosion of mobile: infographic URL: https://www.edigitalresearch.com/news/item/nid/655462669 [accessed 2015-02-10] [WebCite Cache ID 6WF7xr5Q9]

50. The Deloitte Consumer Review. 5th ed. London: Deloitte; 2013. Beyond the hype: the true potential of mobile URL: http://www2.deloitte.com/content/dam/Deloitte/se/Documents/consumer-business/uk-cb-consumer-review-edition-5130620.pdf [accessed 2015-09-02] [WebCite Cache ID 6WDRujhFK]

51. Office for National Statistics Statistical Bulletin. 2015 Aug 06. Internet access-households and individuals, 2015 URL: http://www.ons.gov.uk/ons/dcp171778_412758.pdf [accessed 2015-09-23] [WebCite Cache ID 6blBFs3R2]

Abbreviations

AOR: age-adjusted odds ratio
CAPI: computer-assisted personal interview
CASI: computer-assisted self-interview
GCSE: General Certificate of Secondary Education
GUM: genitourinary medicine
HIV: human immunodeficiency virus
IMD: Index of Multiple Deprivation
MSM: men who have sex with men
NHS: National Health Service
NS-SEC: National Statistics Socio-Economic Classification
OAC: Output Area Classification
STI: sexually transmitted infection

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Features of Computer-Based Decision Aids: Systematic Review, Thematic Synthesis, and Meta-Analyses

Ania Syrowatka[1,2], PhD(c); Dörthe Krömker[3], PhD; Ari N Meguerditchian[1,4,5], MD, MSc, FRCS, FACS; Robyn Tamblyn[1,2,6], PhD

[1]Clinical and Health Informatics Research Group, McGill University, Montreal, QC, Canada

[2]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

[3]McGill Centre for the Convergence of Health and Economics, McGill University, Montreal, QC, Canada

[4]Department of Surgery, McGill University Health Centre, Montreal, QC, Canada

[5]Department of Oncology, McGill University Health Centre, Montreal, QC, Canada

[6]Department of Medicine, McGill University, Montreal, QC, Canada

Corresponding Author:
Ania Syrowatka, PhD(c)
Clinical and Health Informatics Research Group
McGill University
1140 Pine Avenue West
Montreal, QC, H3A1A3
Canada
Phone: 1 514 934 1934 ext 32999
Fax: 1 514 843 1551
Email: anna.syrowatka@mail.mcgill.ca

Abstract

Background: Patient information and education, such as decision aids, are gradually moving toward online, computer-based environments. Considerable research has been conducted to guide content and presentation of decision aids. However, given the relatively new shift to computer-based support, little attention has been given to how multimedia and interactivity can improve upon paper-based decision aids.

Objective: The first objective of this review was to summarize published literature into a proposed classification of features that have been integrated into computer-based decision aids. Building on this classification, the second objective was to assess whether integration of specific features was associated with higher-quality decision making.

Methods: Relevant studies were located by searching MEDLINE, Embase, CINAHL, and CENTRAL databases. The review identified studies that evaluated computer-based decision aids for adults faced with preference-sensitive medical decisions and reported quality of decision-making outcomes. A thematic synthesis was conducted to develop the classification of features. Subsequently, meta-analyses were conducted based on standardized mean differences (SMD) from randomized controlled trials (RCTs) that reported knowledge or decisional conflict. Further subgroup analyses compared pooled SMDs for decision aids that incorporated a specific feature to other computer-based decision aids that did not incorporate the feature, to assess whether specific features improved quality of decision making.

Results: Of 3541 unique publications, 58 studies met the target criteria and were included in the thematic synthesis. The synthesis identified six features: content control, tailoring, patient narratives, explicit values clarification, feedback, and social support. A subset of 26 RCTs from the thematic synthesis was used to conduct the meta-analyses. As expected, computer-based decision aids performed better than usual care or alternative aids; however, some features performed better than others. Integration of content control improved quality of decision making (SMD 0.59 vs 0.23 for knowledge; SMD 0.39 vs 0.29 for decisional conflict). In contrast, tailoring reduced quality of decision making (SMD 0.40 vs 0.71 for knowledge; SMD 0.25 vs 0.52 for decisional conflict). Similarly, patient narratives also reduced quality of decision making (SMD 0.43 vs 0.65 for knowledge; SMD 0.17 vs 0.46 for decisional conflict). Results were varied for different types of explicit values clarification, feedback, and social support.

Conclusions: Integration of media rich or interactive features into computer-based decision aids can improve quality of preference-sensitive decision making. However, this is an emerging field with limited evidence to guide use. The systematic review and thematic synthesis identified features that have been integrated into available computer-based decision aids, in an

XSL•FO
RenderX

effort to facilitate reporting of these features and to promote integration of such features into decision aids. The meta-analyses and associated subgroup analyses provide preliminary evidence to support integration of specific features into future decision aids. Further research can focus on clarifying independent contributions of specific features through experimental designs and refining the designs of features to improve effectiveness.

Introduction

Over the past decade, health care has shifted from paper-based practice to electronic health records [1]. Patient information and education, such as decision aids, are also gradually moving toward online, computer-based environments [2]. Decision aids are support tools intended to help patients engage in high-quality decision making for preference-sensitive decisions [3]. Preference-sensitive decisions are cases where there is more than one clinically appropriate option (eg, choosing between mastectomy and lumpectomy for treatment of breast cancer [4]), and as a result, depend on patients' values and preferences [3,5]. High-quality decision making occurs when well-informed patients consider evidence in the context of personal values and preferences to make a health decision. Decision aids have been shown to be effective for improving quality of decision making [3].

Considerable research has been conducted to guide content and presentation of decision aids [2,6-18]. In addition, the International Patient Decision Aid Standards Collaboration has established quality criteria for development and assessment of decision aids: (1) providing information in sufficient detail, (2) presenting probabilities in an unbiased manner, (3) including methods to clarify values and preferences, (4) providing structured guidance for deliberation and communication, (5) presenting information in a balanced manner, (6) using a systematic development process, (7) using up-to-date evidence, (8) disclosing conflicts of interest, (9) using plain language, and (10) ensuring that the decision is informed and values-based [6]. Additional quality criteria are recommended if the decision aid is delivered on the Internet or uses narratives. Further research has been conducted to expand upon these criteria. For example, risk communication literature builds on these standards by guiding how risk information (eg, probabilities) should be presented to facilitate high-quality decision making [19]. Similarly, research has been expanding around patient narratives [20,21] and explicit methods to clarify values and preferences [22]. However, given the relatively new shift to computer-based support, little attention has been given to how multimedia and interactivity can improve upon paper-based decision aids [2].

Theory suggests that integration of media rich or interactive features into computer-based decision aids can have a positive impact on quality of decision making by engaging patients in decision making beyond traditional static approaches [2]. Therefore, this review focuses on features that can be implemented only on electronic platforms and are not feasible in paper-based decision aids. The first objective of this review

was to summarize published literature into a proposed classification of features that have been integrated into computer-based decision aids. Building on this classification, the second objective was to assess whether integration of specific features was associated with higher-quality decision making.

Methods

Study Selection

Search Strategy

Four databases (MEDLINE, Embase, CINAHL, and CENTRAL) were searched for all relevant studies published from 1946-2013. Three main concepts of decision support, the patient, and computer were mapped to the most relevant controlled vocabulary using Medical Subject Headings (MeSH), and free-text terms were added where necessary. Full search strategies are outlined in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

The review identified studies that evaluated computer-based decision aids for adults faced with a preference-sensitive medical decision (ie, treatment, risk management, screening, or prevention) and that reported at least one quality of decision-making outcome (ie, knowledge, decisional conflict [23], decisional certainty, satisfaction with decision making, or decisional self-efficacy). All comparative study designs were considered; eligible controls included alternative aids and no decision support (eg, usual care or pre-post assessments). Studies were excluded if the article did not report original research, was not published in the English language, or if the decision aid was intended for proxy decision making.

Screening and Data Abstraction

Screening of articles was completed in two stages. Articles were first screened for relevance based on the information provided in the title and abstract and were then evaluated for inclusion based on the full text. Two reviewers independently screened articles at each stage (AS and DK). Disagreements were resolved by discussion and consensus between the 2 reviewers. Overall kappa score was calculated to assess interrater reliability [24].

One reviewer completed data abstraction (AS), which focused on citation information, study design, decision context, interventions, controls (eg, usual care or alternative aids), components being tested, and quality of decision-making outcomes. If an article included in the review cited a

development paper or webpage, then information from these sources was used to supplement data abstracted from the article.

Thematic Synthesis

All studies identified for the systematic review were included in the thematic analysis. Data abstracted from the articles was used to create a proposed classification of features that have been integrated into computer-based decision aids to date. The classification was developed based on logical groupings and informed by themes from decision aid literature [6,7].

Selection of groupings was completed by 1 reviewer (AS), guided by steps outlined by Thomas & Harden for conducting thematic analysis: (1) line-by-line coding of articles to record components, (2) development of descriptive themes, and (3) creation of analytical themes [25]. Components were grouped into types (ie, descriptive themes) if the function of the components was similar; for example, components that improved navigation were grouped together. Similarly, types of components were judged to be similar and grouped into features (ie, analytical themes) if they served a similar purpose; for example, components that improved access to information were grouped together within the same feature. The proposed features classification went through an iterative process of review by co-authors and colleagues and was revised accordingly.

Meta-Analyses

Only randomized controlled trials (RCTs) that reported knowledge or decisional conflict were included in the quantitative synthesis. Decisional certainty, satisfaction with decision making, and decisional self-efficacy were not included due to the lower number of studies that reported these outcomes. The standardized mean difference (SMD; ie, Cohen's *d* effect size) was calculated for each study [26]. For studies that reported multiple comparisons, only the comparison that tested the highest number of features was included in the analysis. If multiple measures were taken over time, only the first measure of knowledge or decisional conflict was included in the analysis. If studies reported more than one type of knowledge or more than one subscale of the decisional conflict scale (without reporting the total), then an average of the SMDs was calculated. For articles that did not provide sufficient information, study authors were contacted for additional information required to calculate the SMD.

The overall effect of computer-based decision aids was estimated by pooling the SMD of each study using Review Manager (version 5.3). Studies were pooled using inverse variance weighting and random effects models with 95%

confidence intervals. Heterogeneity of pooled SMDs was assessed based on I^2 statistics [27].

Tests for Subgroup Differences

Subgroup analyses were conducted to test whether specific features (or types of components) could explain some of the heterogeneity in the overall effect. Subgroup analyses compared pooled SMDs for decision aids that incorporated a specific feature to other computer-based decision aids that did not incorporate the feature to assess whether specific features were associated with improvements in quality of decision making. The Review Manager test for subgroup differences was used to assess statistical significance.

Results

Study Selection

The search identified 3541 eligible articles. The title and abstract screen retained 135 articles. Full text screening identified 58 studies that met the target criteria and were included in the thematic synthesis. The overall kappa score for screening was 0.60, reflecting moderate interrater agreement [28]. A subset of 26 RCTs from the thematic synthesis was used to conduct the meta-analyses with 18 of the articles reporting knowledge and 21 of the articles reporting decisional conflict. A modified Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart is presented in Figure 1 [29], and studies included in the meta-analyses are indicated in Multimedia Appendix 2.

Data abstracted from articles are presented in Multimedia Appendix 2, including citation information, study design, decision context, interventions, controls, components being tested, and SMDs for RCTs reporting knowledge or decisional conflict [30-101]. The studies were published between 1996 and 2013 and came primarily from the United States (37/58, 64%), Canada (7/58, 12%), and the United Kingdom (7/58, 12%). The majority (33/58, 57%) of decision aids were developed for treatment decisions. The remaining decision aids were created for risk management (6/58, 10%), screening (15/58, 26%), and prevention decisions (5/58, 9%).

Studies assessed quality of decision making by measuring knowledge (36/58, 62%), decisional conflict (30/58, 52%), decisional certainty (21/58, 36%), satisfaction with decision making (16/58, 28%), and decisional self-efficacy (7/58, 12%). Studies compared computer-based decision aid performance to usual care (18/58, 31%), alternative aids (29/58, 50%), or based on pre-assessments (14/58, 24%).

Figure 1. Modified PRISMA study selection flowchart.
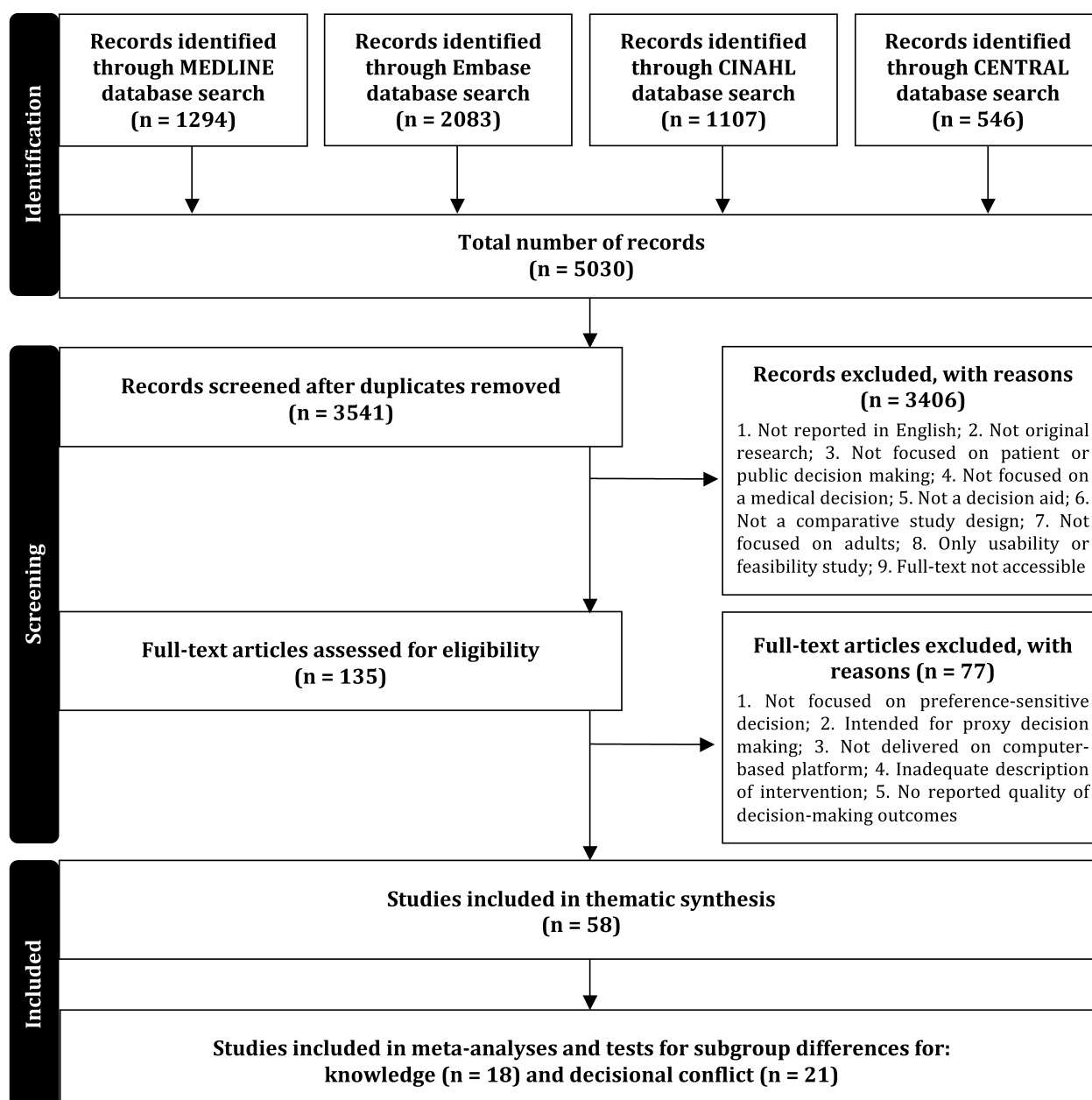
Thematic Synthesis

The thematic analysis identified six main features that have been integrated into computer-based decision aids: content control, tailoring, patient narratives, explicit values clarification, feedback, and social support. A proposed classification for these features and types of components is presented in Table 1. References to decision aids that implemented these features, as well as mode of presentation for each decision aid (eg, Web-based, videobooklet, CD-ROM) are provided in Multimedia Appendix 2, under columns titled "Features being tested" and "Interventions," respectively.

Table 1. Proposed classification of features that have been integrated into computer-based decision aids (58 studies).

Features	Types of components	Examples of components
Content control: Patient has control over access to information	Navigation	Menu bar, search function, television-like interface, touchscreen, help menu
	Clarity of information	Glossary, information summaries, supplementary risk diagrams, metaphors, narration
	Optional information	"Learn more" sections for detailed information about topics of interest
	Access to external resources	Reference lists, links to summaries of recent studies or clinical practice guidelines, developer contact information
Tailoring: Patient receives personalized information	Demographics	Patient age, gender, race/ethnicity, family history, health literacy
	Clinical condition	Specific diagnosis, stage of disease, comorbidities, current symptoms, current medications, past treatment experience, eligibility for interventions
	Values, preferences, and beliefs	Language, preferred role in decision making, stage of decision making, preference for colloquial vs technical terms, beliefs around efficacy of screening or treatment
	Knowledge deficits	Focus on information that is unclear or incorrect based on knowledge pre-tests
Patient narratives: Patient reflects on experiences of others	Patient stories (focus on personal experiences)	Video of patient sharing personal experience
	Behavior modeling (focus on process of deliberation)	Video of patient weighing options, video vignettes of common concerns around decision making
Explicit values clarification: Patient examines personal values and preferences	Decision points	Strategically placed questions to determine whether patient is prepared to move forward to next section of decision aid
	Notebook	Memory aid used to store issues of concern, "bookmarks" for important sections
	Weighting exercises	Simple yes/no questions, feeling thermometer, balance scale, selecting initial treatment decision
	Trade-off exercises	Simple rank order exercises, adaptive conjoint analysis-based tools
	Social matching	"Soap opera" episodes with questions to determine which character embodies patient's values and preferences
	Personal reflection	Patient considers perspectives of others affected by the decision (eg, partners, family members, or others)
Feedback: Patient receives important information around decision making based on interactions with aid	Decision aid progress	Program tracks information that has been covered, and suggests important information that has not been accessed
	Knowledge	Self-evaluations provide feedback on comprehension of evidence presented
	Summary of preferences	Bar graphs depicting relative importance of personal values and preferences
	Optimal choice	Patient values and preferences are incorporated into an algorithm to determine the most suitable option
	Decisional consistency	Alerts patient if initial treatment decision is not consistent with optimal choice
	Summary of decision aid activity (usually printed)	Plan of action based on initial treatment decision, personal risk summaries
Social support: Patient encouraged to involve others in decision-making	Community support	Celebrity endorsement, video of patient celebration after completing treatment, links to support groups
	Integration of family	Modules specific to others affected by the decision, information on how to communicate with partner
	Facilitation of shared decision making	Video of physician describing options and outcomes, video of physician encouraging patient to adhere to chosen option, video coaching to overcome physician communication barriers, recommended questions for physician consultations, copy of decision aid summary placed in patient chart, physician-specific modules

XSL•FO
RenderX

The majority of studies included in the thematic analysis provided content control (42/58, 72%). Two-thirds tailored information to the patient (38/58, 66%), and almost half incorporated patient narratives (28/58, 48%). Over half of the studies provided explicit values clarification (31/58, 53%), feedback (36/58, 62%), or social support (32/58, 55%). One third of the studies incorporated five (13/58, 22%) or all six (10/58, 17%) of these features.

Meta-Analyses

Eighteen studies were included in the meta-analysis to assess whether or not use of computer-based decision aids improved knowledge. The studies included were published between 2001 and 2013. Most computer-based decision aids performed significantly better than usual care or alternative aid controls (14/18, 78%); the performance of the remaining decision aids was not significantly different from controls. Overall, computer-based decision aids were associated with significant improvements in knowledge with a pooled SMD of 0.54 (95% CI 0.36-0.71; P<.001). A forest plot is presented in Figure 2.

We included 21 studies in the meta-analysis to assess whether or not use of computer-based decision aids improved decisional conflict. The studies included were published between 2002 and 2013. Most computer-based decision aids performed significantly better than usual care or alternative aid controls (13/21, 62%); the performance of the remaining decision aids was not significantly different from controls. Overall, computer-based decision aids were associated with significant improvements in decisional conflict with a pooled SMD of 0.35 (95% CI 0.23-0.48; P<.001). A forest plot is presented in Figure 3.

Although computer-based decision aids performed significantly better than usual care or alternative aids, there was a high level of heterogeneity in study-level SMDs. The I^2 statistics were 84% and 75% for knowledge and decisional conflict, respectively.

Figure 2. Forest plot of SMDs for improvements in knowledge (18 studies).

Study or Subgroup	Std. Mean Difference	SE	Weight	Std. Mean Difference IV, Random, 95% CI
Allen 2010	0.19	0.08	7.0%	0.19 [0.03, 0.35]
Ellison 2008	0.58	0.22	5.1%	0.58 [0.15, 1.01]
Frosch 2008	0.25	0.11	6.7%	0.25 [0.03, 0.47]
Huyghe 2009	0.51	0.45	2.6%	0.51 [−0.37, 1.39]
Jibaja–Weiss 2011	0.75	0.24	4.9%	0.75 [0.28, 1.22]
Manne 2010	1.09	0.16	6.0%	1.09 [0.78, 1.40]
Mathieu 2010	0.59	0.12	6.6%	0.59 [0.35, 0.83]
Montgomery 2007	0.57	0.1	6.8%	0.57 [0.37, 0.77]
Ozanne 2007	0.63	0.38	3.2%	0.63 [−0.11, 1.37]
Protheroe 2007	0.57	0.2	5.4%	0.57 [0.18, 0.96]
Rostom 2002	0.64	0.29	4.2%	0.64 [0.07, 1.21]
Saver 2007	0.49	0.22	5.1%	0.49 [0.06, 0.92]
Sawka 2012	1.88	0.28	4.3%	1.88 [1.33, 2.43]
Schapira 2007	−0.15	0.15	6.2%	−0.15 [−0.44, 0.14]
Schroy 2010	0.91	0.1	6.8%	0.91 [0.71, 1.11]
Simon 2012	0.23	0.08	7.0%	0.23 [0.07, 0.39]
Thomson 2007	0.06	0.19	5.6%	0.06 [−0.31, 0.43]
Wilkie 2013	0.43	0.13	6.4%	0.43 [0.18, 0.68]
Total (95% CI)			**100.0%**	**0.54 [0.36, 0.71]**

Heterogeneity: Tau^2 = 0.11; Chi^2 = 110.67, df = 17 (P < 0.00001); I^2 = 85%
Test for overall effect: Z = 5.97 (P < 0.00001)

Favors control Favors computer–based aid

Figure 3. Forest plot of SMDs for improvements in decisional conflict (21 studies).

Study or Subgroup	Std. Mean Difference	SE	Weight	Std. Mean Difference IV, Random, 95% CI
Allen 2010	0.17	0.08	6.5%	0.17 [0.01, 0.33]
Banegas 2013	0.56	0.09	6.3%	0.56 [0.38, 0.74]
Berry 2013	0.08	0.1	6.2%	0.08 [−0.12, 0.28]
Davison 2007	0.04	0.11	6.0%	0.04 [−0.18, 0.26]
de Achaval 2012	0.37	0.17	4.8%	0.37 [0.04, 0.70]
Diefenbach 2012	0.62	0.27	3.1%	0.62 [0.09, 1.15]
Feldman−Stewart 2012	0.1	0.17	4.8%	0.10 [−0.23, 0.43]
Frosch 2008	0.07	0.11	6.0%	0.07 [−0.15, 0.29]
Huyghe 2009	1.23	0.49	1.4%	1.23 [0.27, 2.19]
Jibaja−Weiss 2011	0.25	0.22	3.9%	0.25 [−0.18, 0.68]
Manne 2010	0.36	0.15	5.2%	0.36 [0.07, 0.65]
Montgomery 2007	0.28	0.1	6.2%	0.28 [0.08, 0.48]
Murray 2001	0.54	0.15	5.2%	0.54 [0.25, 0.83]
Ozanne 2007	0.78	0.38	2.0%	0.78 [0.04, 1.52]
Protheroe 2007	1.04	0.18	4.6%	1.04 [0.69, 1.39]
Saver 2007	−0.02	0.21	4.0%	−0.02 [−0.43, 0.39]
Sawka 2012	1.48	0.26	3.3%	1.48 [0.97, 1.99]
Schapira 2007	0.07	0.15	5.2%	0.07 [−0.22, 0.36]
Simon 2012	0.24	0.08	6.5%	0.24 [0.08, 0.40]
Thomson 2007	0.42	0.19	4.4%	0.42 [0.05, 0.79]
Volk 2008	0.33	0.18	4.6%	0.33 [−0.02, 0.68]
Total (95% CI)			100.0%	0.35 [0.23, 0.48]

Heterogeneity: $Tau^2 = 0.06$; $Chi^2 = 79.71$, $df = 20$ ($P < 0.00001$); $I^2 = 75\%$
Test for overall effect: $Z = 5.59$ ($P < 0.00001$)

Favors control   Favors computer−based aid

Tests for Subgroup Differences

The six features and associated types identified through the thematic analysis were used to inform subgroup analyses. The results are presented in Tables 2 (for knowledge) and 3 (for decisional conflict), which present and compare pooled SMDs for decision aids that include a specific feature (or type) to other computer-based decision aids that do not include the feature. The number of studies included in each subgroup and P values associated with the Review Manager test for subgroup differences are also reported.

Overall, integration of content control was positively associated with quality of decision making, although the association was only significant for knowledge ($P=.008$). However, specific types had differing effects. Knowledge and decisional conflict were negatively affected by offering navigation ($P=.56$, $P=.08$, respectively), whereas knowledge and decisional conflict both improved by providing clarity around evidence ($P=.03$, $P=.07$, respectively), optional in-depth information ($P=.05$, $P=.42$, respectively), or access to external resources ($P=.65$, $P=.15$, respectively).

Conversely, tailoring was negatively associated with knowledge and decisional conflict ($P=.08$, $P=.07$, respectively). This association was consistent across all types; both knowledge and decisional conflict were negatively affected by tailoring based on patient demographics ($P=.07$, $P=.31$, respectively), clinical condition ($P=.06$, $P=.14$, respectively), or values, preferences, and beliefs ($P=.14$, $P=.02$, respectively).

Similarly, patient narratives reduced quality of decision making; however, the association was significant only for decisional conflict ($P=.005$). Both knowledge and decisional conflict were negatively affected by presenting patient stories ($P=.54$, $P=.11$, respectively), or behavior modeling ($P=.32$, $P=.01$, respectively).

Explicit values clarification reduced knowledge ($P=.42$) but did not affect decisional conflict. However, effects varied for different types. Both knowledge and decisional conflict improved by integrating notebooks ($P=.68$, $P=.56$, respectively), or trade-off exercises ($P=.84$, $P=.60$, respectively). Knowledge was reduced by providing weighting exercises ($P=.18$) or social matching ($P=.47$), with no effect on decisional conflict. Likewise, personal reflection reduced knowledge ($P=.47$).

Overall, providing feedback was negatively associated with knowledge and decisional conflict ($P=.40$, $P=.63$, respectively). Again, effects varied by type. Reporting progress through the decision aid improved decisional conflict ($P=.32$). Likewise, both knowledge and decisional conflict improved by providing knowledge feedback ($P=.80$, $P=.07$, respectively). Providing a summary of preferences did not affect decisional conflict. Optimal choice feedback reduced knowledge ($P=.44$), but improved decisional conflict ($P=.54$). Both knowledge and decisional conflict were negatively affected by providing feedback around decisional consistency ($P<.001$, $P=.31$, respectively), or providing a summary of decision aid activity ($P=.35$, $P=.62$, respectively).

Table 2. Number of studies and pooled SMDs for improvements in knowledge comparing decision aids including each feature to decision aids not including the feature (18 studies).

Feature and types of components	Studies, n	Feature included, pooled SMD (95% CI)	Studies, n	Reference (no feature), pooled SMD (95% CI)	P value[a]
Overall: Any feature	18	0.54 (0.36-0.71)	0	—	—
Content control	15	0.59 (0.39-0.79)	3	0.23 (0.05-0.41)	.008
Navigation	7	0.47 (0.19-0.76)	11	0.59 (0.34-0.83)	.56
Clarity of information	13	0.65 (0.44-0.87)	5	0.24 (-0.05-0.54)	.03
Optional information	7	0.76 (0.42-1.09)	11	0.38 (0.21-0.54)	.05
Access to external resources	6	0.63 (0.15-1.10)	12	0.51 (0.32-0.70)	.65
Tailoring	10	0.40 (0.18-0.62)	8	0.71 (0.44-0.99)	.08
Demographics	9	0.38 (0.15-0.62)	9	0.71 (0.45-0.96)	.07
Clinical condition	8	0.36 (0.11-0.61)	10	0.69 (0.46-0.93)	.06
Values, preferences, and beliefs	3	0.31 (0.00-0.62)	15	0.59 (0.38-0.79)	.14
Knowledge deficits	0	—	18	0.54 (0.36-0.71)	—
Patient narratives	8	0.43 (0.19-0.68)	10	0.65 (0.37-0.93)	.26
Patient stories	7	0.47 (0.20-0.75)	11	0.59 (0.34-0.83)	.54
Behavior modeling	3	0.39 (0.11-0.67)	15	0.57 (0.36-0.78)	.32
Explicit values clarification	11	0.48 (0.30-0.65)	7	0.67 (0.23-1.12)	.42
Decision points	0	—	18	0.54 (0.36-0.71)	—
Notebook	3	0.59 (0.35-0.84)	15	0.53 (0.33-0.73)	.68
Weighting exercises	8	0.41 (0.24-0.58)	10	0.65 (0.34-0.95)	.18
Trade-off exercises	3	0.58 (0.12-1.04)	15	0.53 (0.33-0.72)	.84
Social matching	1	0.43 (0.18-0.68)	17	0.55 (0.36-0.73)	.47
Personal reflection	1	0.43 (0.18-0.68)	17	0.55 (0.36-0.73)	.47
Feedback	8	0.46 (0.27-0.64)	10	0.60 (0.31-0.89)	.40
Decision aid progress	0	—	18	0.54 (0.36-0.71)	—
Knowledge	2	0.60 (0.12-1.08)	16	0.53 (0.35-0.72)	.80
Summary of preferences	0	—	18	0.54 (0.36-0.71)	—
Optimal choice	3	0.42 (0.11-0.73)	15	0.57 (0.36-0.78)	.44
Decisional consistency	2	0.17 (0.03-0.31)	16	0.60 (0.40-0.79)	<.001
Summary of decision aid activity	6	0.44 (0.23-0.65)	12	0.60 (0.34-0.86)	.35
Social support	10	0.58 (0.32-0.84)	8	0.50 (0.23-0.76)	.67
Community support	4	0.91 (0.34-1.48)	14	0.45 (0.27-0.63)	.14
Integration of family	3	0.50 (0.29-0.72)	15	0.54 (0.34-0.74)	.82
Facilitation of shared decision making	6	0.44 (0.13-0.75)	12	0.59 (0.36-0.82)	.45

[a]Review Manager test for subgroup differences.

Table 3. Number of studies and pooled SMDs for improvements in decisional conflict comparing decision aids including each feature to decision aids not including the feature (21 studies).

Feature and types of components	Studies, n	Feature included, pooled SMD (95% CI)	Studies, n	Reference (no feature), pooled SMD (95% CI)	P value[a]
Overall: Any feature	21	0.35 (0.23-0.48)	0	—	—
Content control	14	0.39 (0.23-0.56)	7	0.29 (0.08-0.49)	.42
Navigation	8	0.22 (0.10-0.34)	13	0.42 (0.23-0.60)	.08
Clarity of information	12	0.46 (0.28-0.65)	9	0.23 (0.07-0.40)	.07
Optional information	6	0.44 (0.20-0.68)	15	0.32 (0.17-0.47)	.42
Access to external resources	5	0.72 (0.12-1.33)	16	0.28 (0.18-0.37)	.15
Tailoring	12	0.25 (0.13-0.37)	9	0.52 (0.26-0.79)	.07
Demographics	10	0.29 (0.16-0.42)	11	0.43 (0.20-0.65)	.31
Clinical condition	10	0.26 (0.12-0.40)	11	0.46 (0.23-0.68)	.14
Values, preferences, and beliefs	7	0.18 (0.07-0.30)	14	0.44 (0.27-0.61)	.02
Knowledge deficits	0	—	21	0.35 (0.23-0.48)	—
Patient narratives	8	0.17 (0.08-0.26)	13	0.46 (0.28-0.65)	.005
Patient stories	5	0.20 (0.03-0.38)	16	0.39 (0.24-0.54)	.11
Behavior modeling	4	0.16 (0.05-0.27)	17	0.41 (0.25-0.56)	.01
Explicit values clarification	13	0.36 (0.20-0.51)	8	0.36 (0.14-0.58)	.97
Decision points	0	—	21	0.35 (0.23-0.48)	—
Notebook	4	0.48 (-0.02 to 0.98)	17	0.32 (0.20-0.44)	.56
Weighting exercises	9	0.35 (0.16-0.53)	12	0.36 (0.19-0.54)	.89
Trade-off exercises	3	0.48 (-0.08 to 1.04)	18	0.33 (0.20-0.45)	.60
Social matching	1	0.33 (-0.02 to 0.68)	20	0.36 (0.23-0.49)	.89
Personal reflection	0	—	21	0.35 (0.23-0.48)	—
Feedback	11	0.32 (0.16-0.49)	10	0.39 (0.19-0.58)	.63
Decision aid progress	1	0.62 (0.09-1.15)	20	0.35 (0.22-0.47)	.32
Knowledge	1	1.23 (0.27-2.19)	20	0.34 (0.22-0.46)	.07
Summary of preferences	1	0.37 (0.04-0.70)	20	0.35 (0.23-0.48)	.93
Optimal choice	4	0.45 (0.09-0.81)	17	0.33 (0.19-0.46)	.54
Decisional consistency	2	0.24 (0.02-0.45)	19	0.37 (0.23-0.51)	.31
Summary of decision aid activity	9	0.32 (0.15-0.50)	12	0.39 (0.20-0.57)	.62
Social support	11	0.38 (0.19-0.57)	10	0.34 (0.17-0.51)	.75
Community support	4	0.50 (-0.08 to 1.07)	17	0.33 (0.21-0.45)	.58
Integration of family	2	0.64 (-0.30 to 1.58)	19	0.35 (0.22-0.47)	.54
Facilitation of shared decision making	8	0.29 (0.13-0.45)	13	0.38 (0.21-0.56)	.46

[a]Review Manager test for subgroup differences.

Social support improved knowledge (P=.67) but did not affect decisional conflict. Both knowledge and decisional conflict improved by providing community support (P=.14, P=.58, respectively). Integration of family support did not affect knowledge but improved decisional conflict (P=.54). Both knowledge and decisional conflict were negatively affected by facilitation of shared decision making (P=.45, P=.46, respectively).

Discussion

Principal Results

This review summarizes published literature into a proposed classification of features that have been integrated into computer-based decision aids. The thematic synthesis identified six main features of content control, tailoring, patient narratives, explicit values clarification, feedback, and social support. Building on this classification, meta-analyses with tests for

subgroup differences were conducted to evaluate whether specific features improved quality of decision making. Overall, decision aids that integrated these features performed significantly better than usual care or alternative aids. The exploratory subgroup analyses rank-ordered the features. Overall, content control performed better than other features. Conversely, tailoring and patient narratives performed worse compared to other features. Results were varied for different types of explicit values clarification, feedback, and social support.

Thematic Synthesis

The proposed features classification is the first of its kind for decision aids. It serves two purposes: to provide the first step towards improving reporting of features that are integrated into computer-based decision aids and to promote use of such features in future decision aids. Currently, reporting standards for interventions are specific about the overarching goal of replicability; however, they offer little guidance around how to reach this goal. For example, the Consolidated Standards of Reporting Trials (CONSORT) statement simply states that authors should report sufficient information to ensure replicability, including detail around how and when interventions were administered [102]. The CONSORT statement extension for nonpharmacologic interventions builds on the former by stating that authors should report different components of interventions and procedures for tailoring interventions [103]. As a result of such vague recommendations, published descriptions of interventions, including decision aids, often do not provide sufficient information to guide replication [104]. In addition, lack of consistent language in reporting means that similar interventions can be reported using different descriptions, which can complicate evaluations to identify features that make interventions successful. Consequently, there has been a call for better reporting guidance [104]. To address such limitations in the field of behavior change, Abraham & Michie developed a classification of active behavior change interventions [105]. Using the classification, a systematic review and meta-regression were conducted to evaluate behavior change interventions to promote healthy eating or physical activity [106]. The analysis highlighted "active ingredients" of successful interventions that may have otherwise been missed. Similarly, the classification proposed in this study can lead to better reporting of features that are integrated into computer-based decision aids, using consistent language. Improved reporting will build further evidence around the value of these features, which can guide integration of these features into future decision aids.

Meta-Analyses

As expected, computer-based decision aids were associated with significant improvements in knowledge and decisional conflict compared to usual care or alternative aids [3]. The subgroup analyses served as exploratory assessments of specific features and provided insight into which features perform better than others.

Tests for Subgroup Differences

Content Control

Overall, content control improved quality of decision making. All types of content control performed better than other features, with the exception of navigation. Content control is intended to provide patients with control over order, detail, and type of evidence presented [2]. The concept of content control is promising, considering that patients have different preferences for the amount of detail presented [107]. In addition, these preferences, often correlating with health literacy, can change over time, which highlights that a one-size-fits-all solution is not appropriate [107]. Ideally, content control should improve engagement, which has the potential to facilitate understanding and retention of evidence. In addition, providing the means to self-personalize content may promote autonomy and empower patients to take ownership over health care decisions.

Interestingly, navigation reduced quality of decision making compared to other features. Given that navigation is a foundational piece of computer-based interventions, this may represent a reporting bias. As a result of journal space limitations, navigation may have been underreported in exchange for reporting novel or impressive decision aid components. This relates back to the need for a classification to ensure that all features are reported; otherwise, important features may be overlooked as a result of biased evaluations.

Tailoring

Tailoring reduced quality of decision making, with all subgroups performing worse than other features. In general, tailoring is intended to translate evidence into patient-specific information to improve engagement. The effects of tailoring can be split into two categories: (1) effects on calculation of risk estimates, and (2) effects on presentation of information.

Tailoring can be used to frame evidence in terms of patient demographics or clinical condition to present only viable treatment options with more accurate estimates of associated risks and benefits. Ideally, this should provide a better understanding of personal situations and lead to high-quality decision making. However, evidence around the benefits of tailoring risk estimates is varied [13]. Tailoring may not be effective or may even be detrimental to decision making if personalized risk estimates are not considered appropriate by the patient [108]. Patients may distrust risk estimates if the estimates are not congruent with previous knowledge or if the risk calculations are not transparent or omit risk factors perceived as important [108]. Most of the tailoring incorporated in the decision aids included in this study was superficial and did not present risk information in the context of population risks. It is possible that the tailoring approaches were not developed appropriately and did not tailor sufficiently to be considered useful by the patients.

Tailoring can also be used to present evidence in terms of patient preferences or to address knowledge deficits, in an effort to facilitate understanding or to correct misinformation. However, this form of tailoring may limit the amount or type of evidence that is presented. For example, decision aids can be tailored to information-seeking style (ie, high or low levels of detail) [30]

or information preferences (ie, focus on topics of interest) [31]. Such tailoring approaches may result in an incomplete or unbalanced presentation of evidence and unintentionally reduce quality of decision making. In fact, this form of tailoring could be considered the reverse of content control. Specifically, content control may promote patient autonomy by providing control over evidence reviewed, whereas tailoring employs a paternalistic model of presenting evidence, and as a result, may reduce autonomy.

Patient Narratives

Similarly, patient narratives reduced quality of decision making. Patient narratives are intended to provide insight into patient experiences and bring attention to important evidence to consider throughout the decision-making process. In addition, information presented through patient narratives is processed differently than written information and can improve understanding and retention of evidence [15,20,109]. However, narratives can unintentionally present biased or unbalanced information, which may result in lower-quality decision making [2,20,110]. To date, there is no consensus around the value of including patient narratives in decision aids [2,15].

In this study, both patient stories and behavior modeling scenarios reduced quality of decision making. Shaffer and Zikmund-Fisher have developed a taxonomy for patient narratives outlining dimensions that are expected to impact decision making: (1) purpose of the narrative, (2) content of the narrative, and (3) evaluative valence (ie, tone of the narrative) [20]. This taxonomy highlights different types of content and proposes underlying mechanisms on how content might affect decision making for outcome, experience, and process narratives. Given that most articles simply reported inclusion of patient narratives, it is unclear whether the dimensions of purpose, content, and tone can explain the reduction in quality of decision making. However, based on this taxonomy, if the content of patient stories was disproportionally focused on adverse events, then risks associated with selecting a particular option may have been overrepresented. If anecdotal information presented through patient stories contradicts statistical data presented in the decision aid, then these conflicting messages may have reduced quality of decision making. Behavior modeling narratives (similar to process narratives described in the taxonomy) are intended to increase deliberation, by guiding the patient to consider or disregard specific factors related to decision making. However, if a narrative places more emphasis on factors not considered to be important or relevant to the patient, this could impede the decision-making process. This may even lead to lower-quality decision making by encouraging patients to focus on factors that are not in line with personal values or preferences. Further research can help determine the value of specific types of patient narratives, and what type of content and presentation facilitate higher-quality decision making.

The negative effects of tailoring and patient narratives on quality of decision making in decision aids were unexpected, considering the positive impact of tailoring and patient narratives when employed in behavior change interventions [111-113]. A potential reason for this may lie in the fundamental differences in purpose between decision aids and behavior change interventions. Specifically, behavior change interventions are intended to persuade the user to take up a specific health behavior, which is considered to be the best option [114]. However, this is not the purpose of decision aids, which are developed for preference-sensitive decisions where there is more than one clinically appropriate option [3]. The overarching goal of decision aids is to provide complete, balanced, and unbiased information to facilitate high-quality decision making. Tailoring and patient narratives, as implemented in the decision aids included in this systematic review, may have unintentionally presented superficial, unbalanced, or biased information, which may have reduced the quality of decision making. Further research needs to be conducted to refine the content and presentation of these two features to improve their effectiveness in decision aids.

Explicit Values Clarification

Specific types of explicit values clarification had a positive effect on quality of decision making. Explicit values clarification methods are intended to guide patients through specific tasks to identify personal values and preferences [14]. Few have been evaluated, and findings are mixed [14]. In this study, providing a "notebook" to record topics that were unclear or of concern was associated with higher-quality decision making. This device allowed patients to highlight areas where they needed more information, or where they needed to reflect more deeply on values and preferences. As a result, patients were more knowledgeable and had lower decisional conflict. Likewise, trade-off exercises (eg, rank-ordering all outcomes from most to least important) also improved quality of decision making. Such exercises provide a realistic approach to decision making, where the patient must consider and trade-off between risks and benefits of two or more options. In contrast, weighting exercises (eg, ranking each outcome on a scale of 1-10) can be inadequate if the patient ranks everything as equally important. In essence, trade-off exercises may help to better clarify issues around decision making.

Feedback

Specific types of feedback were also associated with improvements in quality of decision making. Feedback is intended to provide the patient with important information around decision making based on interactions with the decision aid. Progress through the decision aid and knowledge feedback both improved quality of decision making. Both are intended to ensure that the patient is well informed by confirming that all necessary information is reviewed by the patient and to correct misinformation, respectively. Summary of preferences, optimal choice, and decisional consistency are types of feedback that are specific to explicit values clarification methods. Summary of preferences provides feedback around how patients personally value risks and benefits integral to decision making. Optimal choice builds on summary of preferences, by suggesting which option is best based on patients' values and preferences, which had a positive effect on decisional conflict. Similarly, research has shown that providing implications of stated values (ie, optimal choice) may have a positive effect on decision making [22]. In contrast, decisional consistency reduced quality

of decision making. This form of feedback elicits patients' initial decisions and reports whether or not the initial decisions are consistent with patients' values and preferences. This may lead to increased decisional conflict when the decision aid reports a lack of decisional consistency. This gives the impression that the initial decision was wrong and may elicit negative emotional responses from the patient.

Social Support

Specific types of social support improved quality of decision making. This feature is intended to reinforce that the patient is not alone in their experiences or decision making. Social support is a recurring theme throughout patient needs assessments for medical care [115-117], and patient decision aids specifically [118]. However, little guidance has been offered around what type of social supports can be integrated into decision aids and how. This review identified three types: community (ie, support from others faced with the same decision), family (ie, support from others affected by the specific decision), and clinician support (ie, facilitation of shared decision making). Community support improved knowledge and decisional conflict. Patients who had access to this type of support were likely able to have knowledge questions answered, as well as discuss how personal values and preferences fit into decision making. Integration of family support had positive effects on decisional conflict. Patients who had access to this type of support were likely able to better engage their partners and family members in the decision-making process and discuss values and preferences. Social support, specifically community and family support, seems to be promising but was limited by sample size in this study.

Principles for Decision Aid Development and Future Directions

Based on the study findings, content control should be integrated into decision aids to allow patients to select the order, level of detail, and type of information presented. This approach allows the patient to directly access topics of interest, view alternative presentations of information for clarity, and access optional information or external resources. However, to ensure balanced representation of all options, it is important to integrate safeguards to ensure that the patient reviews all necessary evidence (ie, not "optional" information) prior to making a final decision.

Tailoring, as currently developed and presented, should be used with caution, as it may reduce quality of decision making. Ineffective tailoring may have resulted from superficial or non-transparent tailoring, which patients did not believe reflected their true risk. Allowing patients to "self-tailor" through content control may be a viable option until effective strategies for tailoring information are established.

Patient narratives should also be used with caution, as they may reduce quality of decision making. Patient narratives may unintentionally present unbalanced or biased information, which may undermine statistical data presented in the decision aid or encourage patients to focus on factors that are not in line with personal values or preferences. Further research should focus on identifying types of narrative content and presentation that facilitate quality decision making.

Further research is also needed in the areas of feedback, explicit values clarification, and social support to guide future integration. There was substantial heterogeneity in effects between types of components within each of these features, which may reflect artificial grouping of components. In addition, small sample sizes limited appropriate assessments, with many components having been tested only in one decision aid, which limited guidance for integration of these features into decision aids.

Limitations

Studies included in the meta-analyses had a high level of heterogeneity with regard to patient populations, decision context, characteristics of the interventions, and components being tested, as well as choice of usual care or alternative aid controls. Studies were selected for inclusion based on testing a computer-based decision aid intervention, evaluating quality of decision making by measuring either knowledge or decisional conflict, and using an RCT design. For each subgroup analysis, decision aids that incorporated a specific feature (or type of component) were compared to decision aids that did not incorporate the feature. Effectively, this approach compared groupings of studies that tested various complex decision aids against very different control groups. Therefore, results from this study should be interpreted as "hypothesis-generating" and should be considered preliminary evidence to guide future work in this area.

Small numbers of studies incorporated certain types of components, which reduced the power to detect significant subgroup differences but also increased the probability of false positives. Nevertheless, subgroup analyses were conducted for all features and types of components for completion, and the number of studies in each subgroup was considered when interpreting the results.

Similarly, the study did not adjust for numerous comparisons generated through the subgroup analyses, which also increased the probability of false positives. However, given the exploratory nature of the study, such adjustments may not be necessary, since findings will require further research to establish independent contributions of each feature [27].

Given that the majority of decision aids incorporated multiple features, conducting subgroup analyses limited the capacity to disentangle the effects of specific features or to assess whether specific bundles of features were more effective for improving quality of decision making. Ideally, conducting a meta-regression, similar to the analysis described by Michie et al, would address these shortcomings [106]. However, a meta-regression was not considered appropriate for this study. The regression coefficients would have been unstable given the low number of studies (18 for knowledge and 21 for decisional conflict), and it would have been impossible to control for the substantial heterogeneity of decision aids included in the analysis. Therefore, straightforward exploratory subgroup analyses were selected as a viable alternative.

Conclusions

Integration of media rich or interactive features into computer-based decision aids can improve quality of preference-sensitive decision making beyond traditional static approaches. However, this is an emerging field with limited evidence to guide implementation. The systematic review and thematic synthesis identified features used in available computer-based decision aids, in an effort to facilitate reporting of these features and to promote integration of such features into decision aids. The meta-analyses and associated subgroup analyses provide preliminary evidence to support integration of specific features into future decision aids. Further research can focus on clarifying independent contributions of specific features through experimental designs and refining the designs of features to improve effectiveness.

Authors' Contributions

AS and RT were responsible for study conception and design; AS and DK acquired data; AS, AM, and RT analyzed and interpreted data; AS drafted the paper; and DK, AM, RT made critical revisions. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Systematic review search strategies for MEDLINE, Embase, CINAHL, and CENTRAL.

[PDF File (Adobe PDF File), 360KB - jmir_v18i1e20_app1.pdf]

Multimedia Appendix 2

Information about studies included in the systematic review.

[PDF File (Adobe PDF File), 698KB - jmir_v18i1e20_app2.pdf]

References

1. Collier R. National Physician Survey: EMR use at 75%. CMAJ 2015 Jan 6;187(1):E17-E18 [FREE Full text] [doi: 10.1503/cmaj.109-4957] [Medline: 25487665]
2. Hoffman AS, Volk RJ, Saarimaki A, Stirling C, Li LC, Härter M, et al. Delivering patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. BMC Med Inform Decis Mak 2013;13 Suppl 2:S13 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S13] [Medline: 24625064]
3. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, Wu Julie H C. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014;1:CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
4. Hawley ST. Involving patients in the decision-making process regarding breast cancer treatment: implications for surgery utilization. Womens Health (Lond Engl) 2010 Mar;6(2):161-164. [doi: 10.2217/whe.09.87] [Medline: 20187720]
5. Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. Health Aff (Millwood) 2002;Suppl Web Exclusives:W96-114 [FREE Full text] [Medline: 12703563]
6. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
7. Volk RJ, Llewellyn-Thomas H, Stacey D, Elwyn G. Ten years of the International Patient Decision Aid Standards Collaboration: evolution of the core dimensions for assessing the quality of patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S1 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S1] [Medline: 24624947]
8. Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S2 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S2] [Medline: 24625093]

9.   Barry MJ, Chan E, Moulton B, Sah S, Simmons MB, Braddock C. Disclosing conflicts of interest in patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S3 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S3] [Medline: 24625114]

10. Feldman-Stewart D, O'Brien MA, Clayman ML, Davison BJ, Jimbo M, Labrecque M, et al. Providing information about options in patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S4 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S4] [Medline: 24625127]

11. Montori VM, LeBlanc A, Buchholz A, Stilwell DL, Tsapas A. Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: a quality dimension of the International Patient Decision Aid Standards. BMC Med Inform Decis Mak 2013;13 Suppl 2:S5 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S5] [Medline: 24625191]

12. Abhyankar P, Volk RJ, Blumenthal-Barby J, Bravo P, Buchholz A, Ozanne E, et al. Balancing the presentation of information and options in patient decision aids: an updated review. BMC Med Inform Decis Mak 2013;13 Suppl 2:S6 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S6] [Medline: 24625214]

13. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han P, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak 2013;13 Suppl 2:S7 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S7] [Medline: 24625237]

14. Fagerlin A, Pignone M, Abhyankar P, Col N, Feldman-Stewart D, Gavaruzzi T, et al. Clarifying values: an updated review. BMC Med Inform Decis Mak 2013;13 Suppl 2:S8 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S8] [Medline: 24625261]

15. Bekker HL, Winterbottom AE, Butow P, Dillard AJ, Feldman-Stewart D, Fowler FJ, et al. Do personal stories make patient decision aids more effective? A critical review of theory and evidence. BMC Med Inform Decis Mak 2013;13 Suppl 2:S9 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S9] [Medline: 24625283]

16. McCaffery KJ, Holmes-Rovner M, Smith SK, Rovner D, Nutbeam D, Clayman ML, et al. Addressing health literacy in patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S10 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S10] [Medline: 24624970]

17. Stacey D, Kryworuchko J, Belkora J, Davison BJ, Durand M, Eden KB, et al. Coaching and guidance with patient decision aids: A review of theoretical and empirical evidence. BMC Med Inform Decis Mak 2013;13 Suppl 2:S11 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S11] [Medline: 24624995]

18. Sepucha KR, Borkhoff CM, Lally J, Levin CA, Matlock DD, Ng CJ, et al. Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. BMC Med Inform Decis Mak 2013;13 Suppl 2:S12 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S12] [Medline: 24625035]

19. Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. Ann Intern Med 2014 Aug 19;161(4):270-280. [doi: 10.7326/M14-0295] [Medline: 25133362]

20. Shaffer VA, Zikmund-Fisher BJ. All stories are not alike: a purpose-, content-, and valence-based taxonomy of patient narratives in decision aids. Med Decis Making 2013 Jan;33(1):4-13. [doi: 10.1177/0272989X12463266] [Medline: 23065418]

21. Shaffer VA, Hulsey L, Zikmund-Fisher BJ. The effects of process-focused versus experience-focused narratives in a breast cancer treatment decision task. Patient Educ Couns 2013 Nov;93(2):255-264. [doi: 10.1016/j.pec.2013.07.013] [Medline: 23962673]

22. Witteman HO, Scherer LD, Gavaruzzi T, Pieterse AH, Fuhrel-Forbis A, Exe N, et al. Supporting patient values: a systematic review of values clarification exercises. 2014 Presented at: 35th Annual Meeting for the Society for Medical Decision Making; October 22, 2013; Baltimore, Maryland.

23. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making 1995;15(1):25-30. [Medline: 7898294]

24. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]

25. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]

26. Cohen J. Statistical power analysis for the behavioral sciences, revised edition. New York, NY: Academic Press; 1977.

27. The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions, version 5.1.0. 2011. URL: http://handbook.cochrane.org [WebCite Cache ID 6aXuMa5av]

28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]

29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269, W64. [Medline: 19622511]

30. Diefenbach MA, Mohamed NE, Butz BP, Bar-Chama N, Stock R, Cesaretti J, et al. Acceptability and preliminary feasibility of an internet/CD-ROM-based education and decision program for early-stage prostate cancer patients: randomized pilot study. J Med Internet Res 2012;14(1):e6 [FREE Full text] [doi: 10.2196/jmir.1891] [Medline: 22246148]

31. Davison BJ, Goldenberg SL, Wiens KP, Gleave ME. Comparing a generic and individualized information decision support intervention for men newly diagnosed with localized prostate cancer. Cancer Nurs 2007;30(5):E7-15. [doi: 10.1097/01.NCC.0000290819.22195.d6] [Medline: 17876177]

32. Akl EA, Grant B, Guyatt GH, Montori VM, Schünemann HJ. A decision aid for COPD patients considering inhaled steroid therapy: development and before and after pilot testing. BMC Med Inform Decis Mak 2007;7:12 [FREE Full text] [doi: 10.1186/1472-6947-7-12] [Medline: 17504536]

33.   Berman L, Curry L, Goldberg C, Gusberg R, Fraenkel L. Pilot testing of a decision support tool for patients with abdominal aortic aneurysms. J Vasc Surg 2011 Feb;53(2):285-292.e1 [FREE Full text] [doi: 10.1016/j.jvs.2010.08.075] [Medline: 21067886]

34.   Berry DL, Halpenny B, Hong F, Wolpin S, Lober WB, Russell KJ, et al. The Personal Patient Profile-Prostate decision support for men with localized prostate cancer: a multi-center randomized trial. Urol Oncol 2013 Oct;31(7):1012-1021 [FREE Full text] [doi: 10.1016/j.urolonc.2011.10.004] [Medline: 22153756]

35.   Berry DL, Halpenny B, Wolpin S, Davison BJ, Ellis WJ, Lober WB, et al. Development and evaluation of the personal patient profile-prostate (P3P), a Web-based decision support system for men newly diagnosed with localized prostate cancer. J Med Internet Res 2010;12(4):e67 [FREE Full text] [doi: 10.2196/jmir.1576] [Medline: 21169159]

36.   Brink SG, Birney AJ, McFarren AE. Charting your course: formative evaluation of a prostate cancer treatment decision aid. Int Electron J Health Educ 2000;3(1):44-54.

37.   de Achaval S, Fraenkel L, Volk RJ, Cox V, Suarez-Almazor ME. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. Arthritis Care Res (Hoboken) 2012 Feb;64(2):229-237 [FREE Full text] [doi: 10.1002/acr.20646] [Medline: 21954198]

38.   Deyo RA, Cherkin DC, Weinstein J, Howe J, Ciol M, Mulley AG. Involving patients in clinical decisions: impact of an interactive video program on use of back surgery. Med Care 2000 Sep;38(9):959-969. [Medline: 10982117]

39.   Spunt BS, Deyo RA, Taylor VM, Leek KM, Goldberg HI, Mulley AG. An interactive videodisc program for low back pain patients. Health Educ Res 1996 Dec;11(4):535-541 [FREE Full text] [Medline: 10163960]

40.   Phelan EA, Deyo RA, Cherkin DC, Weinstein JN, Ciol MA, Kreuter W, et al. Helping patients decide about back surgery: a randomized trial of an interactive video program. Spine (Phila Pa 1976) 2001 Jan 15;26(2):206-211;discussion 212. [Medline: 11154542]

41.   Diefenbach MA, Butz BP. A multimedia interactive education system for prostate cancer patients: development and preliminary evaluation. J Med Internet Res 2004 Jan 21;6(1):e3 [FREE Full text] [doi: 10.2196/jmir.6.1.e3] [Medline: 15111269]

42.   Feldman-Stewart D, Tong C, Siemens R, Alibhai S, Pickles T, Robinson J, et al. The impact of explicit values clarification exercises in a patient decision aid emerges after the decision is actually made: evidence from a randomized controlled trial. Med Decis Making 2012;32(4):616-626. [doi: 10.1177/0272989X11434601] [Medline: 22287534]

43.   Feldman-Stewart D, Brennenstuhl S, Brundage MD, Roques T. An explicit values clarification task: development and validation. Patient Educ Couns 2006 Nov;63(3):350-356. [doi: 10.1016/j.pec.2006.04.001] [Medline: 16860521]

44.   Fraenkel L, Rabidou N, Wittink D, Fried T. Improving informed decision-making for patients with knee pain. J Rheumatol 2007 Sep;34(9):1894-1898. [Medline: 17696268]

45.   Fraenkel L, Peters E, Charpentier P, Olsen B, Errante L, Schoen RT, et al. Decision tool to improve the quality of care in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2012 Jul;64(7):977-985 [FREE Full text] [doi: 10.1002/acr.21657] [Medline: 22392766]

46.   Heller L, Parker PA, Youssef A, Miller MJ. Interactive digital education aid in breast reconstruction. Plast Reconstr Surg 2008 Sep;122(3):717-724. [doi: 10.1097/PRS.0b013e318180ed06] [Medline: 18766034]

47.   Jibaja-Weiss ML, Volk RJ, Friedman LC, Granchi TS, Neff NE, Spann SJ, et al. Preliminary testing of a just-in-time, user-defined values clarification exercise to aid lower literate women in making informed breast cancer treatment decisions. Health Expect 2006 Sep;9(3):218-231. [doi: 10.1111/j.1369-7625.2006.00386.x] [Medline: 16911136]

48.   Jibaja-Weiss ML, Volk RJ, Granchi TS, Neff NE, Robinson EK, Spann SJ, et al. Entertainment education for breast cancer surgery decisions: a randomized trial among patients with low health literacy. Patient Educ Couns 2011 Jul;84(1):41-48. [doi: 10.1016/j.pec.2010.06.009] [Medline: 20609546]

49.   Liao L, Jollis JG, DeLong ER, Peterson ED, Morris KG, Mark DB. Impact of an interactive video on decision making of patients with ischemic heart disease. J Gen Intern Med 1996 Jun;11(6):373-376. [Medline: 8803746]

50.   Maslin AM, Baum M, Secker Walker J, A'Hern R, Prouse A. Shared decision-making using an interactive video disk system for women with early breast cancer. J Res Nurs 1998;3(6):444-454. [doi: 10.1177/174498719800300609]

51.   Molenaar S, Sprangers MA, Rutgers EJ, Luiten EJ, Mulder J, Bossuyt PM, et al. Decision support for patients with early-stage breast cancer: effects of an interactive breast cancer CDROM on treatment decision, satisfaction, and quality of life. J Clin Oncol 2001 Mar 15;19(6):1676-1687. [Medline: 11250997]

52.   Morgan MW, Deber RB, Llewellyn-Thomas HA, Gladstone P, Cusimano RJ, O'Rourke K, et al. Randomized, controlled trial of an interactive videodisc decision aid for patients with ischemic heart disease. J Gen Intern Med 2000 Oct;15(10):685-693 [FREE Full text] [Medline: 11089711]

53.   Murray E, Davis H, Tai SS, Coulter A, Gray A, Haines A. Randomised controlled trial of an interactive multimedia decision aid on hormone replacement therapy in primary care. BMJ 2001 Sep 1;323(7311):490-493 [FREE Full text] [Medline: 11532844]

54.   Piercy GB, Deber R, Trachtenberg J, Ramsey EW, Norman RW, Goldenberg SL, et al. Impact of a shared decision-making program on patients with benign prostatic hyperplasia. Urology 1999 May;53(5):913-920. [Medline: 10223483]

55.   Protheroe J, Bower P, Chew-Graham C, Peters TJ, Fahey T. Effectiveness of a computerized decision aid in primary care on decision making and quality of life in menorrhagia: results of the MENTIP randomized controlled trial. Med Decis Making 2007;27(5):575-584. [doi: 10.1177/0272989X07306785] [Medline: 17898242]

56.   Rostom A, O'Connor A, Tugwell P, Wells G. A randomized trial of a computerized versus an audio-booklet decision aid for women considering post-menopausal hormone replacement therapy. Patient Educ Couns 2002 Jan;46(1):67-74. [Medline: 11804772]

57.   Sawka AM, Straus S, Gafni A, Brierley JD, Tsang RW, Rotstein L, et al. How can we meet the information needs of patients with early stage papillary thyroid cancer considering radioactive iodine remnant ablation? Clin Endocrinol (Oxf) 2011 Apr;74(4):419-423 [FREE Full text] [doi: 10.1111/j.1365-2265.2010.03966.x] [Medline: 21198742]

58.   Sawka AM, Straus S, Gafni A, Meiyappan S, O'Brien MA, Brierley JD, et al. A usability study of a computerized decision aid to help patients with early stage papillary thyroid carcinoma in decision-making on adjuvant radioactive iodine treatment. Patient Educ Couns 2011 Aug;84(2):e24-e27. [doi: 10.1016/j.pec.2010.07.038] [Medline: 20732775]

59.   Sawka AM, Straus S, Rotstein L, Brierley JD, Tsang RW, Asa S, et al. Randomized controlled trial of a computerized decision aid on adjuvant radioactive iodine treatment for patients with early-stage papillary thyroid cancer. J Clin Oncol 2012 Aug 10;30(23):2906-2911 [FREE Full text] [doi: 10.1200/JCO.2011.41.2734] [Medline: 22753906]

60.   Schapira MM, Gilligan MA, McAuliffe T, Garmon G, Carnes M, Nattinger AB. Decision-making at menopause: a randomized controlled trial of a computer-based hormone therapy decision-aid. Patient Educ Couns 2007 Jul;67(1-2):100-107. [doi: 10.1016/j.pec.2007.02.007] [Medline: 17400413]

61.   Simon D, Kriston L, von Wolff A, Buchholz A, Vietor C, Hecke T, et al. Effectiveness of a web-based, individually tailored decision aid for depression or acute low back pain: a randomized controlled trial. Patient Educ Couns 2012 Jun;87(3):360-368. [doi: 10.1016/j.pec.2011.10.009] [Medline: 22154867]

62.   Sivell S, Edwards A, Manstead ASR, Reed MWR, Caldon L, Collins K, et al. Increasing readiness to decide and strengthening behavioral intentions: evaluating the impact of a web-based patient decision aid for breast cancer treatment options (BresDex: www.bresdex.com). Patient Educ Couns 2012 Aug;88(2):209-217. [doi: 10.1016/j.pec.2012.03.012] [Medline: 22541508]

63.   Sivell S, Marsh W, Edwards A, Manstead ASR, Clements A, Elwyn G, et al. Theory-based design and field-testing of an intervention to support women choosing surgery for breast cancer: BresDex. Patient Educ Couns 2012 Feb;86(2):179-188. [doi: 10.1016/j.pec.2011.04.014] [Medline: 21571485]

64.   Taylor KL, Davis KM, Lamond T, Williams RM, Schwartz MD, Lawrence W, et al. Use and evaluation of a CD-ROM-based decision aid for prostate cancer treatment decisions. Behav Med 2010;36(4):130-140. [doi: 10.1080/08964289.2010.525263] [Medline: 21186436]

65.   Thomson RG, Eccles MP, Steen IN, Greenaway J, Stobbart L, Murtagh MJ, et al. A patient decision aid to support shared decision-making on anti-thrombotic treatment of patients with atrial fibrillation: randomised controlled trial. Qual Saf Health Care 2007 Jun;16(3):216-223 [FREE Full text] [doi: 10.1136/qshc.2006.018481] [Medline: 17545350]

66.   Thomson R, Robinson A, Greenaway J, Lowe P. Development and description of a decision analysis based decision support tool for stroke prevention in atrial fibrillation. Qual Saf Health Care 2002 Mar;11(1):25-31 [FREE Full text] [Medline: 12078365]

67.   van Til JA, Drossaert CHC, Renzenbrink GJ, Snoek GJ, Dijkstra E, Stiggelbout AM, et al. Feasibility of web-based decision aids in neurological patients. J Telemed Telecare 2010;16(1):48-52. [doi: 10.1258/jtt.2009.001012] [Medline: 20086268]

68.   Hooker GW, Leventhal K, DeMarco T, Peshkin BN, Finch C, Wahl E, et al. Longitudinal changes in patient distress following interactive decision aid use among BRCA1/2 carriers: a randomized trial. Med Decis Making 2011;31(3):412-421 [FREE Full text] [doi: 10.1177/0272989X10381283] [Medline: 20876346]

69.   Kaufman EM, Peshkin BN, Lawrence WF, Shelby R, Isaacs C, Brown K, et al. Development of an Interactive Decision Aid for Female BRCA1/BRCA2 Carriers. J Genet Couns 2003 Apr;12(2):109-129. [doi: 10.1023/A:1022698112236] [Medline: 26140844]

70.   Huyghe E, Martinetti P, Sui D, Schover LR. Banking on Fatherhood: pilot studies of a computerized educational tool on sperm banking before cancer treatment. Psychooncology 2009 Sep;18(9):1011-1014 [FREE Full text] [doi: 10.1002/pon.1506] [Medline: 19061198]

71.   Montgomery AA, Emmett CL, Fahey T, Jones C, Ricketts I, Patel RR, et al. Two decision aids for mode of delivery among women with previous caesarean section: randomised controlled trial. BMJ 2007 Jun 23;334(7607):1305 [FREE Full text] [doi: 10.1136/bmj.39217.671019.55] [Medline: 17540908]

72.   Emmett CL, Murphy DJ, Patel RR, Fahey T, Jones C, Ricketts IW, et al. Decision-making about mode of delivery after previous caesarean section: development and piloting of two computer-based decision aids. Health Expect 2007 Jun;10(2):161-172. [doi: 10.1111/j.1369-7625.2006.00429.x] [Medline: 17524009]

73.   Saver BG, Gustafson D, Taylor TR, Hawkins RP, Woods NF, Dinauer S, et al. A tale of two studies: the importance of setting, subjects and context in two randomized, controlled trials of a web-based decision support for perimenopausal and postmenopausal health decisions. Patient Educ Couns 2007 May;66(2):211-222. [doi: 10.1016/j.pec.2006.12.004] [Medline: 17317080]

XSL•FO
RenderX

74.   Schwartz MD, Valdimarsdottir HB, DeMarco TA, Peshkin BN, Lawrence W, Rispoli J, et al. Randomized trial of a decision aid for BRCA1/BRCA2 mutation carriers: impact on measures of decision making and satisfaction. Health Psychol 2009 Jan;28(1):11-19 [FREE Full text] [doi: 10.1037/a0013147] [Medline: 19210013]

75.   Wilkie DJ, Gallo AM, Yao Y, Molokie RE, Stahl C, Hershberger PE, et al. Reproductive health choices for young adults with sickle cell disease or trait: randomized controlled trial immediate posttest effects. Nurs Res 2013;62(5):352-361 [FREE Full text] [doi: 10.1097/NNR.0b013e3182a0316b] [Medline: 23995469]

76.   Gallo AM, Wilkie DJ, Wang E, Labotka RJ, Molokie RE, Stahl C, et al. Evaluation of the SCKnowIQ tool and reproductive CHOICES intervention among young adults with sickle cell disease or sickle cell trait. Clin Nurs Res 2014 Aug;23(4):421-441 [FREE Full text] [doi: 10.1177/1054773813479377] [Medline: 23572406]

77.   Allen JD, Mohllajee AP, Shelton RC, Drake BF, Mars DR. A computer-tailored intervention to promote informed decision making for prostate cancer screening among African American men. Am J Mens Health 2009 Dec;3(4):340-351 [FREE Full text] [doi: 10.1177/1557988308325460] [Medline: 19477736]

78.   Allen JD, Othus MKD, Hart A, Tom L, Li Y, Berry D, et al. A randomized trial of a computer-tailored decision aid to improve prostate cancer screening decisions: results from the Take the Wheel trial. Cancer Epidemiol Biomarkers Prev 2010 Sep;19(9):2172-2186 [FREE Full text] [doi: 10.1158/1055-9965.EPI-09-0410] [Medline: 20716619]

79.   Ellison GL, Weinrich SP, Lou M, Xu H, Powell IJ, Baquet CR. A randomized trial comparing web-based decision aids on prostate cancer knowledge for African-American men. J Natl Med Assoc 2008 Oct;100(10):1139-1145 [FREE Full text] [Medline: 18942274]

80.   Evans R, Joseph-Williams N, Edwards A, Newcombe RG, Wright P, Kinnersley P, et al. Supporting informed decision making for prostate specific antigen (PSA) testing on the web: an online randomized controlled trial. J Med Internet Res 2010;12(3):e27 [FREE Full text] [doi: 10.2196/jmir.1305] [Medline: 20693148]

81.   Frosch DL, Bhatnagar V, Tally S, Hamori CJ, Kaplan RM. Internet patient decision support: a randomized controlled trial comparing alternative approaches for men considering prostate cancer screening. Arch Intern Med 2008 Feb 25;168(4):363-369. [doi: 10.1001/archinternmed.2007.111] [Medline: 18299490]

82.   Green MJ, Peterson SK, Baker MW, Harper GR, Friedman LC, Rubinstein WS, et al. Effect of a computer-based decision aid on knowledge, perceptions, and intentions about genetic testing for breast cancer susceptibility: a randomized controlled trial. JAMA 2004 Jul 28;292(4):442-452 [FREE Full text] [doi: 10.1001/jama.292.4.442] [Medline: 15280342]

83.   Kuppermann M, Norton ME, Gates E, Gregorich SE, Learman LA, Nakagawa S, et al. Computerized prenatal genetic testing decision-assisting tool: a randomized controlled trial. Obstet Gynecol 2009 Jan;113(1):53-63. [doi: 10.1097/AOG.0b013e31818e7ec4] [Medline: 19104360]

84.   Leung KY, Lee CP, Chan HY, Tang MHY, Lam YH, Lee A. Randomised trial comparing an interactive multimedia decision aid with a leaflet and a video to give information about prenatal screening for Down syndrome. Prenat Diagn 2004 Aug;24(8):613-618. [doi: 10.1002/pd.927] [Medline: 15305348]

85.   Lindblom K, Gregory T, Wilson C, Flight IHK, Zajac I. The impact of computer self-efficacy, computer anxiety, and perceived usability and acceptability on the efficacy of a decision support tool for colorectal cancer screening. J Am Med Inform Assoc 2012;19(3):407-412 [FREE Full text] [doi: 10.1136/amiajnl-2011-000225] [Medline: 21857024]

86.   Manne SL, Meropol NJ, Weinberg DS, Vig H, Ali-Khan CZ, Manning C, et al. Facilitating informed decisions regarding microsatellite instability testing among high-risk individuals diagnosed with colorectal cancer. J Clin Oncol 2010 Mar 10;28(8):1366-1372 [FREE Full text] [doi: 10.1200/JCO.2009.25.0399] [Medline: 20142594]

87.   Mathieu E, Barratt AL, McGeechan K, Davey HM, Howard K, Houssami N. Helping women make choices about mammography screening: an online randomized trial of a decision aid for 40-year-old women. Patient Educ Couns 2010 Oct;81(1):63-72. [doi: 10.1016/j.pec.2010.01.001] [Medline: 20149953]

88.   Australian screening mammography decision aid trial: a decision aid for women aged 40 thinking about starting mammography screening. 2003. Screening and Test Evaluation Program (STEP) at the University of Sydney URL: http://www.mammogram.med.usyd.edu.au/ [accessed 2015-07-29] [WebCite Cache ID 6aOFhtKW3]

89.   Ruffin MT, Fetters MD, Jimbo M. Preference-based electronic decision aid to promote colorectal cancer screening: results of a randomized controlled trial. Prev Med 2007 Oct;45(4):267-273. [doi: 10.1016/j.ypmed.2007.07.003] [Medline: 17689600]

90.   Rupert DJ, Squiers LB, Renaud JM, Whitehead NS, Osborn RJ, Furberg RD, et al. Communicating risk of hereditary breast and ovarian cancer with an interactive decision support tool. Patient Educ Couns 2013 Aug;92(2):188-196. [doi: 10.1016/j.pec.2013.04.008] [Medline: 23664232]

91.   Schroy PC, Emmons K, Peters E, Glick JT, Robinson PA, Lydotes MA, et al. The impact of a novel computer-based decision aid on shared decision making for colorectal cancer screening: a randomized trial. Med Decis Making 2011;31(1):93-107 [FREE Full text] [doi: 10.1177/0272989X10369007] [Medline: 20484090]

92.   Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine. Your disease risk: the source on prevention. Cancer - colon cancer. 2013. URL: http://www.yourdiseaserisk.wustl.edu/YDRDefault.aspx?ScreenControl=YDRGeneral&ScreenName=YDRcolon [WebCite Cache ID 6aOF8hynW]

XSL•FO
RenderX

93. Volk RJ, Jibaja-Weiss ML, Hawley ST, Kneuper S, Spann SJ, Miles BJ, et al. Entertainment education for prostate cancer screening: a randomized trial among primary care patients with low health literacy. Patient Educ Couns 2008 Dec;73(3):482-489 [FREE Full text] [doi: 10.1016/j.pec.2008.07.033] [Medline: 18760888]

94. Banegas MP, McClure JB, Barlow WE, Ubel PA, Smith DM, Zikmund-Fisher BJ, et al. Results from a randomized trial of a web-based, tailored decision aid for women at high risk for breast cancer. Patient Educ Couns 2013 Jun;91(3):364-371 [FREE Full text] [doi: 10.1016/j.pec.2012.12.014] [Medline: 23395006]

95. Fagerlin A, Dillard AJ, Smith DM, Zikmund-Fisher BJ, Pitsch R, McClure JB, et al. Women's interest in taking tamoxifen and raloxifene for breast cancer prevention: response to a tailored decision aid. Breast Cancer Res Treat 2011 Jun;127(3):681-688 [FREE Full text] [doi: 10.1007/s10549-011-1450-1] [Medline: 21442198]

96. Ozanne EM, Annis C, Adduci K, Showstack J, Esserman L. Pilot trial of a computerized decision aid for breast cancer prevention. Breast J 2007;13(2):147-154. [doi: 10.1111/j.1524-4741.2007.00395.x] [Medline: 17319855]

97. Sheridan SL, Shadle J, Simpson RJ, Pignone MP. The impact of a decision aid about heart disease prevention on patients' discussions with their doctor and their plans for prevention: a pilot randomized trial. BMC Health Serv Res 2006;6:121 [FREE Full text] [doi: 10.1186/1472-6963-6-121] [Medline: 17005051]

98. Pignone M, Sheridan SL, Lee YZ, Kuo J, Phillips C, Mulrow C, et al. Heart to Heart: a computerized decision aid for assessment of coronary heart disease risk and the impact of risk-reduction interventions for primary prevention. Prev Cardiol 2004;7(1):26-33. [Medline: 15010625]

99. Sheridan SL, Griffith JM, Behrend L, Gizlice Z, Jianwen C, Pignone MP. Effect of adding a values clarification exercise to a decision aid on heart disease prevention: a randomized trial. Med Decis Making 2010;30(4):E28-E39 [FREE Full text] [doi: 10.1177/0272989X10369008] [Medline: 20484089]

100. Pignone MP, Sheridan SL, Communication for Health Applications and Interventions (CHAI) Core. Heart to heart: a tool for improving communication & decision-making about heart disease prevention. URL: https://www.med-decisions.com/h2hv2/ [accessed 2015-07-29] [WebCite Cache ID 6aOFcHHK9]

101. Sheridan SL, Draeger LB, Pignone MP, Keyserling TC, Simpson RJ, Rimer B, et al. A randomized trial of an intervention to improve use and adherence to effective coronary heart disease prevention strategies. BMC Health Serv Res 2011;11:331 [FREE Full text] [doi: 10.1186/1472-6963-11-331] [Medline: 22141447]

102. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]

103. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008 Feb 19;148(4):295-309. [Medline: 18283207]

104. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? BMJ 2008 Jun 28;336(7659):1472-1474 [FREE Full text] [doi: 10.1136/bmj.39590.732037.47] [Medline: 18583680]

105. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]

106. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

107. Matsuyama RK, Kuhn LA, Molisani A, Wilson-Genderson MC. Cancer patients' information needs the first nine months after diagnosis. Patient Educ Couns 2013 Jan;90(1):96-102. [doi: 10.1016/j.pec.2012.09.009] [Medline: 23058682]

108. Scherer LD, Ubel PA, McClure J, Greene SM, Alford SH, Holtzman L, et al. Belief in numbers: When and why women disbelieve tailored breast cancer risk statistics. Patient Educ Couns 2013 Aug;92(2):253-259 [FREE Full text] [doi: 10.1016/j.pec.2013.03.016] [Medline: 23623330]

109. Young LM, Anderson RP. The use of personal narrative in classroom case study analysis to improve long-term knowledge retention and cultivate professional qualities in allied health students. J Microbiol Biol Educ 2010;11(2):107-112 [FREE Full text] [doi: 10.1128/jmbe.v11i2.204] [Medline: 23653708]

110. Winterbottom A, Bekker HL, Conner M, Mooney A. Does narrative information bias individual's decision making? A systematic review. Soc Sci Med 2008 Dec;67(12):2079-2088. [doi: 10.1016/j.socscimed.2008.09.037] [Medline: 18951673]

111. 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]

112. Lustria M, 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]

113. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]

114. Michie S, van Stralen M, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

115. Burg MA, Adorno G, Lopez E, Loerzel V, Stein K, Wallace C, et al. Current unmet needs of cancer survivors: analysis of open-ended responses to the American Cancer Society Study of Cancer Survivors II. Cancer 2015 Feb 15;121(4):623-630. [doi: 10.1002/cncr.28951] [Medline: 25581252]

116. van Walsem MR, Howe EI, Iversen K, Frich JC, Andelic N. Unmet needs for healthcare and social support services in patients with Huntington's disease: a cross-sectional population-based study. Orphanet J Rare Dis 2015;10:124 [FREE Full text] [doi: 10.1186/s13023-015-0324-8] [Medline: 26411462]

117. Song Y, Song H, Han H, Park S, Nam S, Kim MT. Unmet needs for social support and effects on diabetes self-care activities in Korean Americans with type 2 diabetes. Diabetes Educ 2012;38(1):77-85 [FREE Full text] [doi: 10.1177/0145721711432456] [Medline: 22222514]

118. Ng CJ, Mathers N, Bradley A, Colwell B. A 'combined framework' approach to developing a patient decision aid: the PANDAs model. BMC Health Serv Res 2014;14:503 [FREE Full text] [doi: 10.1186/s12913-014-0503-7] [Medline: 25341370]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

XSL•FO
RenderX

Original Paper

Primary Care Providers' Views of Patient Portals: Interview Study of Perceived Benefits and Consequences

David P Miller Jr[1], MD, MS; Celine Latulipe[2], PhD; Kathryn A Melius[3], MS; Sara A Quandt[4], PhD; Thomas A Arcury[3], PhD

[1]Wake Forest School of Medicine, Department of Internal Medicine, Winston-Salem, NC, United States

[2]College of Computing and Informatics, Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, United States

[3]Wake Forest School of Medicine, Department of Family and Community Medicine, Winston-Salem, NC, United States

[4]Wake Forest School of Medicine, Department of Epidemiology and Prevention, Winston-Salem, NC, United States

Corresponding Author:
David P Miller Jr, MD, MS
Wake Forest School of Medicine
Department of Internal Medicine
Medical Center Blvd
Winston-Salem, NC, 27157
United States
Phone: 1 3367134156
Fax: 1 3367167359
Email: dmiller@wakehealth.edu

Abstract

Background: The United States government is encouraging physicians to adopt patient portals—secure websites that allow patients to access their health information. For patient portals to recognize their full potential and improve patient care, health care providers' acceptance and encouragement of their use will be essential. However, little is known about provider concerns or views of patient portals.

Objective: We conducted this qualitative study to determine how administrators, clinic staff, and health care providers at practices serving a lower income adult population viewed patient portals in terms of their potential benefit, areas of concern, and hopes for the future.

Methods: We performed in-depth interviews between October 2013 and June 2014 with 20 clinic personnel recruited from health centers in four North Carolina counties. Trained study personnel conducted individual interviews following an interviewer guide to elicit perceptions of the benefits and disadvantages of patient portals. Interviews were recorded and transcribed. Research team members reviewed transcribed interviews for major themes to construct a coding dictionary. Two researchers then coded each transcript with any coding discrepancies resolved through discussion.

Results: The interviews revealed that clinic personnel viewed patient portals as a mandated product that had potential to improve communication and enhance information sharing. However, they expressed many concerns including portals' potential to generate more work, confuse patients, alienate non-users, and increase health disparities. Clinic personnel expected few older and disadvantaged patients to use a portal.

Conclusions: Given that clinic personnel have significant concerns about portals' unintended consequences, their uptake and impact on care may be limited. Future studies should examine ways portals can be implemented in practices to address providers' concerns and meet the needs of vulnerable populations.

(J Med Internet Res 2016;18(1):e8) doi:10.2196/jmir.4953

KEYWORDS

XSL•FO
RenderX

Introduction

Electronic health records can reduce medical error and enhance efficiency, particularly by facilitating the sharing of medical information [1]. Many electronic health records include patient portals—secure websites where patients can access their health information, request medication refills, and even communicate electronically with their health care provider. Recognizing the potential benefits of electronic health records and patient portals, the United States government is encouraging their adoption. In 2009, Congress passed the Health Information Technology for Economic and Clinical Health (HITECH) Act, which authorized incentive payments to physicians who demonstrated "meaningful use" of these new systems [2]. Beginning in 2015, providers who fail to adopt these new technologies will be penalized a small percentage of their Medicare reimbursements [3].

The Centers for Medicare and Medicaid Services is responsible for developing the "meaningful use" criteria, which they are releasing in stages. The initial meaningful use criteria included items entirely under health care providers' control, such as giving visit summaries to patients or electronically sending medication refills to pharmacies. However, the latest round of criteria, released in July 2014, included items that require patient engagement, such as specifying that at least 5% of patients access their health information through a patient portal [4].

In large health care systems that have implemented patient portals, the initial response from patients has been tepid. A 2-year study found that only 10% of veterans had authenticated their patient portal account within the Veterans Health Administration system [5]. Even in large commercial health systems, typically less than 30-40% of patients activate their online access [6-8]. In addition, older patients and those from vulnerable populations are even less likely to use a patient portal [7,9,10]. In clinics serving primarily disadvantaged populations, portal use has been less than 10% [11]. For patient portals to recognize their full potential and improve patient care, health care providers' acceptance and encouragement of their use will be essential [12-14]. However, little is known about providers' concerns or views of patient portals.

We conducted this study to learn from the early experiences of clinics that have implemented or are in the process of implementing a patient portal. All clinics served a lower income population, allowing us to specifically examine issues related to vulnerable groups. We particularly wanted to determine how administrators, clinic staff, and health care providers viewed patient portals in terms of their potential benefit, areas of concern, and hopes for the future. Knowing this information could help health care systems optimize their use of patient portals, leading to improvements in patient care and fulfillment of meaningful use criteria.

Methods

Data collection was completed between October 2013 and June 2014. We conducted this study as part of a larger multi-component investigation of factors that facilitate or hinder the use of patient portals among low-income older adults [15].

The project is a collaboration of a large academic medical center, a state university, and a network of 16 health centers located across rural North Carolina. The project protocol was approved by the Wake Forest Baptist Health Institutional Review Board, and all participants provided signed informed consent.

Participants

We recruited 20 participants from health centers in four North Carolina counties representing variation on the urban-rural continuum [16]. The health centers also represented diversity and included an urban academic health center and three rural federally qualified health centers. To identify the full spectrum of barriers and facilitators of patient portal implementation, we purposefully recruited participants who had varying experience with portals ranging from current experience to prior experience to no experience. Investigators selected potential participants from the health centers to reflect a desired diversity in job category and geographic distribution. While it is suggested that 12-15 interviews are generally adequate for qualitative research [17,18], the number is ultimately determined by the researchers. Based on this study teams' experience, 20 interviews would provide ample data to accurately describe health care providers' experiences with patient portals across various study sites.

Data Collection

Three trained interviewers completed one-on-one in-depth interviews with each participant. Interviewers met participants at a location of the participants' choice, usually their office or clinic. Following each interview, participants received a small incentive (US $20) to thank them for their time. All interviews were audio recorded and transcribed for later analysis. Project staff contacted potential participants until the desired number of participants was achieved. Saturation was reached as it was determined that variability within the dataset had been achieved and no novel information was being gathered.

Interview Content

The interview was designed to elicit use of technology and electronic patient information management systems by health care providers and to understand their perception of patient use of patient portals (see Multimedia Appendix 1). First, participants were asked about the use of patient portals in their practices, including if their practice had implemented a patient portal, the observed or anticipated impact of a portal on providers and patients, and the anticipated advantages and disadvantages of using a patient portal. Second, participants were asked about any privacy or security concerns regarding electronic personal health information and concerns their patients had about using the patient portal. Third, participants were asked about the environmental and community factors that impact use of patient portals, including facilitators and barriers for practitioners and patients.

Analysis

Data analysis was based on a systemic, computer-assisted approach [19]. Mechanics of data management were accomplished through the use of ATLAS.ti (Scientific Software Development GmbH). All interviews were transcribed verbatim, and each transcript was edited for accuracy. Data analysis began with the collection and ongoing reflection on interview content

through listening to interview recordings and reading the interview transcripts. Research team members reviewed each interview and recorded themes, patterns, and issues that arose in those narratives [20].

The entire research team discussed this information and developed a coding dictionary to reflect themes present in the interviews as structured according to Davis' Technology Acceptance Model (TAM) [21]. Based on Fishbein and Azjen's Theory of Reasoned Action and Theory of Planned Behavior [22], TAM posits the belief-attitude-intention-behavior causal relationship for predicting acceptance of information technology. Two beliefs are fundamental determinants of technology use: perceived usefulness and perceived usability. Perceived usefulness is "the degree to which a person believes that using a particular system would enhance his or her job performance," while perceived usability is "the degree to which a person believes that using a particular system would be free of effort" [21]. Further, perceived attributes are important: users are more likely to use technology if the applications involved are both easy to use and meet users' values and needs [23,24].

During an initial training period, the research team members practiced coding interviews to reach agreement on assignment of codes. Each transcript was coded by 2 research team members, and any differences were resolved through discussion.

This double coding throughout data processing was a check on completeness and drift from the original code definitions.

Results

Study team members contacted 30 health care providers to reach our target sample of 20 participants. Among the 10 non-participants, one refused to participate for lack of interest, 3 did not respond, and 6 were added to a waitlist. The 20 participants represented a range of positions from the 4 health centers (Table 1).

Approximately one-third of participants worked at clinics that were planning a new portal implementation to replace a prior portal that was discontinued for lack of use. The prior portal required an email address to register, and few of these rural clinics' patients had email accounts. The planned new portal does not require an email to register.

Interviews with each participant ranged from approximately 30 minutes to 2 hours in length. Following the structure of our interviewer guide, we organized our findings along four broad categories as displayed in Figure 1. Within the categories of "potential benefits" and "potential disadvantages," themes emerged detailing factors that primarily impacted the clinic, the patients, or the larger health care system.

Table 1. Characteristics of the study sample (N=20).

Characteristics	Value
Clinic position, n (%)	
Nurse	6 (30)
Physician/advanced practice provider	8 (40)
Other non-medical clinicians	2 (10)
Clinic manager/administration	4 (20)
Clinic location, n (%)	
Rural	10 (50)
Urban	10 (50)
Female, n (%)	15 (75)
Age, mean (range)	44 (24-70)
Race, n (%)	
White, non-Hispanic	9 (45)
African American	4 (20)
American Indian	3 (15)
Other	4 (20)
Availability of patient portal at clinic site, n (%)	
Active patient portal in place	6 (30)
Had portal previously; planning new portal implementation	7 (35)
Never had portal; planning first time portal implementation	7 (35)

Figure 1. Categories of major findings from interviews conducted with clinic personnel.
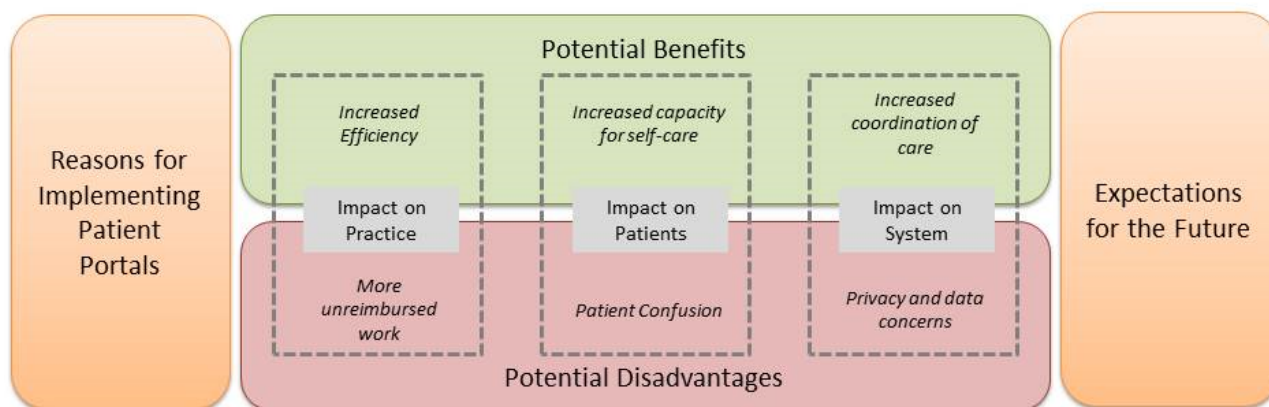
Reasons for Implementing Patient Portals

The main motivator for implementing patient portals was external pressure, and in particular, new federally mandated requirements for "meaningful use" of electronic health records. We heard concern about being "rated" on portal use in the future. As one urban physician stated, "So it's going to be an issue of 'How many of your patients are using the patient portal?' Two. 'Oh? Two? Well, you failed'." Similarly, a rural clinic administrator offered, "It's just one more thing that you have to do, and it's mandated. You've got to do it."

Potential Benefits of Patient Portals

While feeling pressured to adopt patient portals, staff and providers see several potential benefits to their use. These benefits are listed in Table 2 and described in further detail below.

Table 2. Potential benefits of patient portals identified by clinic personnel.

Potential benefit	Representative statements
Office efficiency	
Decreasing phone calls	Clinic staff: "If there's a lot of questions that can be answered through the portal that may cut down on even patients having to call in…"
	Nurse: "If somebody calls just for a prescription refill, they can do that online and save that phone call for somebody that really, really needs it."
Handling messages more quickly	Physician: "It's also a lot easier just to type a message, because I could reply to an email in a few seconds, versus having to sit down, open a chart, pick up the phone, call, hope they answer, and if they don't, having to call back later."
Eliminating need to inform patients of normal results	Physician: "There are a lot of steps in the process right now of notifying patients to let them know of their lab results. So if a patient were able to access that without us having to go through so many steps, that would be nice."
Patient/caregiver access to information	
Increasing patient ability to manage their health	Clinic staff: "they would get more information,...they would understand more, they would be more informed."
Increasing caregiver ability to assist with medical affairs	Nurse: "One patient was telling me that their family member was out in California, I think, a long way off, and she was her power of attorney but she couldn't get out there every time she needed to go to the doctor…but she was able to see what took place at the doctor's office, what exactly was going on with her, and she said that helped her make decisions to benefit the family member instead of make the wrong decision."
Increasing patient satisfaction	Nurse: "So I had talked to a couple of patients that had actually gotten on it. They was going crazy. 'I love it. I love being in this,' because they feel like they have some charge on their own health."
Information sharing with other health professionals	
Decreases duplicate tests	Physician: "I mean tests are so duplicated when people go to different providers. It's just ridiculous. And so if you had a patient who was going to their hematologist, and they were like, 'Okay, I already got my CBC done at my primary,' they could just go, 'Here it is,' and have a look at it."
Reduces medical error	Physician: "it's good because it will allow less miscommunication between providers and less medical errors and less medication errors."

Improving Office Efficiency

Many staff thought portals could improve office efficiency and save time, particularly by decreasing the volume of incoming phone calls for prescription refills and lab result requests. Nursing and clinic personnel can also triage and address electronic messages sent via a portal more quickly and efficiently than a telephone message. As one clinic administrator noted, "[nurses] get those messages on the computer where they're sitting anyways. So they can quickly see them, cipher

through what the patient said and actually try to get them an answer even before they call them back."

Portals may have other unique benefits for clinics where providers are not present on a daily basis, such as resident clinics. For these settings, portals give providers immediate access to patients' electronic requests, rather than requiring providers to physically return to the clinic to check a mailbox. Similar to nursing staff, physicians can also respond to electronic messages faster than phone messages. In addition, giving patients online access to their information reduced the need for staff to notify patients of routine results.

Improving Patient/Caregiver Access to Information

Similarly, portals can save patients time by giving them access to their health information: "Depending on their jobs or situation it may be easier for them to just quickly log on and check their information than having to make a phone call" (Clinic staff member). This feature was seen as particularly helpful for requesting prescription refills or making appointments, both of which can be done via a portal.

Additionally, staff expect portals to help patients better manage their care by providing them with easy access to their lab results, medication lists, and visit summaries. This feature could particularly benefit older adults who may have trouble understanding information during the medical encounter. One clinic administrator commented, "If [older adults] are having difficulty hearing or remembering what the doctor said, they

can look it up." Granting family members and caregivers access to patients' medical information was seen as another benefit of portals.

Increasing Patient Satisfaction

If these benefits are realized, staff expect patient satisfaction and trust with the practice would improve. Clinic personnel observed that early adopters of portals greatly enjoyed the increased access to their health information. As one clinic administrator observed, "there's something about that that gives you power and control. And everybody likes to know that they're in control."

Improving Information Sharing With Other Health Professionals

Improving patients' access to medical records can also improve information sharing with other physicians, potentially improving care and decreasing duplicate tests. This improved information sharing could potentially result in fewer medical errors caused by patients' not knowing their medication list, duplicate prescriptions, or incorrect therapies.

Potential Disadvantages of Patient Portals

Despite these potential benefits, staff expressed many more concerns about the negative impacts of portals on their practices (Table 3). In general, their concerns can be categorized into threats to the practice, threats to patients, and threats to the health care system.

Table 3. Potential disadvantages of patient portals identified by clinic personnel.

Potential disadvantage	Representative statements
Threats to practice	
Potential for high volume of messages	Nurse: "If the patients don't wanna wait, they'll email a hundred times or they'll call a hundred times."
New time pressures	Clinic Staff: "We're taking a phone call away but we just added an everyday procedure new thing that has to be done by the nurse that's already overwhelmed."
Decrease in office visits	Nurse: "I saw that a patient thought, 'Well, since I'm doing this now I don't need to come to you as much,' and that's fine, but don't miss your appointments. That part was not so good for us."
Liability concerns	Physician: "you have to be very careful about what you write, how you write, and what you're telling the patient."
Threats to patients	
Causing patient confusion/anxiety	Physician: "If a patient had all of that data right in front of them without understanding which values may or may not be important that could just lead to unnecessary confusion; whereas if you just…go over results in person with a patient or just send a result card saying your labs are normal you don't have to go into that level of confusion."
Alienating older patients	Nurse: "But as far as my older population – when I say "older" I mean 65 and up – I think that it's gonna be a challenge, because they don't understand...And a lot of them don't really care about that stuff [computers]. I mean, when they're 65 and older they come here, they want me to tell them, and that's it."
Widening health disparities	Physician: "This is actually going to create a gap between people that are educated and have private insurance, they can have easier access to health and health questions, and people that aren't – the barriers are just going to be bigger."
Threats to system	
Inaccurate data entry	Clinic Administrator: "And they understand that a human error could put some misinformation in there. And then they'll say, 'Well how're you gonna get it out?' And that's a good question you know, that does not happen easy."
System failures	Clinic Staff: "I guess like anything technology is not perfect so if there were to be any glitches – anything can happen…but of course with the Internet and with anything there's going to always be complications."
Privacy concerns	Physician: "People could potentially have medical information leaked through it. If they don't use a strong password on their account, it's certainly possible for someone to gain access to their information."

Threats to Practice

While some staff saw the potential for portals to improve office efficiency, many more comments were made about how portals could hamper workflows and increase stress. Several nurses and physicians feared some patients would inappropriately send repeated messages, overwhelming clinic staff. One physician stated, "I've heard of other colleagues who have had patients who maybe sort of abuse it, and write a little too many emails back and forth, and are just – you know, it's one question after the next after the next after the next."

The potential demands portals could place upon scarce time was a theme in several interviews. One physician worried that patients may expect immediate responses to their electronic requests. Nursing staff were also concerned about the extra tasks patient portals would introduce into their days. Just the task of informing patients of the portal was viewed as a burden. One nurse in a clinic explained,

> It's hard to take the time with them patients and show 'em—"This is what you have to do, this is what it is"—because we have like 5 other patients waiting for that provider in that room. Our goal is to get everybody out to lunch and then everybody off on time so there's no overtime.

Portals also may inadvertently decrease the perceived need for office visits. As one physician explained, "Like right now there's a problem in medicine that people want all their care over the phone, and this just adds another layer to 'I want all of my care for free'."

We observed some disagreement about whether patient portals would increase or decrease liability risk. One physician cited the risk of a privacy breach and a large monetary fine if a computer containing patient communications were lost. Another physician viewed electronic exchanges in the portal as being held to the same standard as an office visit and cautioned, "You have to also be careful about the information that you send because any information that you send is like seeing a patient." Conversely, another physician felt portals would protect physicians by saving "a perfect record of the entire conversation." If a patient were to later complain, this physician believed these electronic "records could just be given and everyone would know exactly what transpired" rather than having to "go to court."

Threats to Patients

Some staff and providers were concerned that patients would not fully understand the information present in the portal, triggering more phone calls and questions. Physicians remarked that lab results could be particularly troublesome because clinically insignificant abnormal results are common. Providers also felt their older patients would not want to use the portal and may feel alienated from the practice if they do not. As one administrator stated, "for younger more tech savvy patients, it's awesome and they think it's great." However, older patients could "feel a little left behind" as practices implement portals. One nurse especially saw this concern for patients who lacked literacy or technical skills:

> Every time they come they have the same question, the same problem, and they're just not understanding it and it makes them feel not wanting to come here if somebody's gonna be pushing something like that [the portal] on them. They feel like we're pushing it.

Two physicians expressed concerns that patients with insurance, higher education, or better access to technology would benefit from the additional services of patient portals while older or vulnerable patients who do not use portals would become further disadvantaged. For patients who fail to use a portal, "the barriers are just going to be bigger" (Physician).

Threats to System

Several staff and providers expressed concerns about the stability of new technology and security of information. Administrators and staff acknowledged the inevitability of user error and the potential for incorrect information to be entered into charts. New technology was seen as prone to technical bugs and breakdowns. Last, there was a general concern for the security of information on the Internet. Any new portal was seen as "a potential information leak" (Physician) that could occur through a number of means: a stolen password, a shared password, or hackers (Physician and Staff at three separate sites).

Expectations for the Future

Low Expectations for Immediate Use

We found general agreement among staff and providers that few patients would use a patient portal. For one physician, this low uptake of the portal was seen as a reality: "Let's put it this way, I saw a patient with a resident earlier last week, and it actually said that they have an active [patient portal] account, and I was surprised. That's how infrequently I see people that have it." An administrator in a different county described her experience with a prior attempt to launch a patient portal: "even with our effort, there was nobody who actually used it after we had about 100 sign up."

The low uptake of portals by patients discouraged providers from using the portals as well: "Because a lot of my patients haven't signed up for it, I don't use it to communicate systematically with them. I don't think of sending them letters or communicating with them on [the portal]" (Physician). Similarly, a nurse with prior experience with a portal stated, "they wanted us to check it every day and that type of thing. As I checked it, like I said, it was the same thing the whole time. I just stopped checking it."

Higher Expectations for Future Use

Although staff and providers viewed current use of patient portals as being very low, they had greater expectations for the future. Because patient portals are still a relatively new technology, some envisioned that usage would increase in the future as the population changes but predicted it would take several years to see uptake increase significantly (Clinic administrator). One clinic staff member compared the current use of portals to the early days of electronic banking: "Like ATMs and banking I think there would be a transition period where some people are still going to go inside and want that."

Rural/Urban Similarities and the Digital Divide

We found similar attitudes about the barriers to implementing patient portals in rural and urban clinics, as well as in clinics with a current active portal compared to those with no portal. Rather than a geographic digital divide, we observed a divide defined by age, education, and income. In particular, older adults were viewed as lacking the skills to operate a computer or smartphone, limiting their access to the Internet. This lack of computer literacy was attributed in part to a general anti-technology attitude among the elderly. As examples, one nurse stated "a computer scares them to death" and a physician remarked "[they] don't typically like computers, even if they have one."

All of the clinics in our sample served socioeconomically disadvantaged patient populations. In these clinics, providers believed a large number of their patients lacked the education to know how to use a computer or the income to afford home access. In some clinics, staff estimated that half the patients had no home Internet connection. The free Internet access provided in libraries was viewed as a poor substitute for home access: "If you are sitting in your house and you have a question about your medical record, if the practice is open, you're going to call. Are you really going to get in the car and drive over to the library?" (Clinic administrator).

In contrast, clinic providers felt "[patient portals] are a good idea in the private practice setting, because you have people who have smartphones, you have people who are knowledgeable, you have people who know how to navigate them."

Discussion

Principal Findings

By interviewing staff and providers from a variety of health centers, we were able to gather "front line" views of the early stages of implementing a patient portal. We purposively recruited clinics that serve a low-income population, allowing us to identify issues relevant to vulnerable populations. Clinic personnel in our study identified some benefits to portals, such as their potential to improve communication, give patients easy access to information, and enhance information sharing. However, we heard many more concerns about portals' potential to generate additional work, confuse patients, and perhaps alienate non-users. In general, staff saw patient portals as a mandated product that will rarely be used by older adults. This, in turn, discouraged providers from embracing this new technology. Perhaps because providers expected few patients to actually use the portal, we heard very few concerns about the potential for portals to shift reimbursable office visits to un-reimbursable electronic exchanges.

While this study is one of the first to investigate practices' early experiences with patient portals, a few other studies have explored physicians' thoughts about allowing patients to view health records or communicate electronically with clinicians. Similar to our findings, the majority of physicians believed implementing a patient portal would increase their workload [25,26]. Likewise, practice managers and physicians who use electronic communication with patients agree that it creates more work and adds pressure to their day [27]. In the interviews we conducted, staff and physicians worried that some patients may abuse a portal's easy access to providers. Because electronic patient-provider messaging is a recent development, practices should define clear expectations for appropriate use to guide patients and minimize misuse until new cultural norms emerge.

Many clinicians in our study feared the information in portals could confuse patients causing concern and more calls to the clinic. Some health systems have granted their patients access to view their health records online. In these systems, the majority of physicians shared these concerns, yet fewer than 20% of patients agreed [28]. Another small study of primary care residents and faculty at a single academic institution found that after a portal was implemented, only 13% felt their workload had increased [25]. Although low portal uptake may have contributed to the minimal change in workload observed, these and other studies still suggest that clinicians' fears of patient confusion and increased messages may not come to fruition [29,30].

Providers in our study believed few older adults would use a portal, a belief supported by a study in a large managed care organization reporting that the oldest adults were the least likely to log on to their portal [7]. An age-related digital divide may partially explain this finding. In national surveys, adults in the oldest age groups are the least likely to use the Internet or email [31,32]. In addition, those with functional impairments are also less likely to use the Internet [32,33]. Not surprisingly, patients without home computers and patients who do not use the Internet are less likely to register for a patient portal [6,34,35].

Compounding this issue, patients often rate portals as difficult to use and not user friendly [36]. A recent evaluation of three currently available personal health records found the majority of low socioeconomic status adults had difficulty navigating and using the systems, frequently requiring assistance [37]. Other analyses have found that members of vulnerable populations including those with less education, lower income, and low health literacy are the least likely to use the Internet or enroll in patient portals [7,8,10,13,14,32,38]. Despite this relatively low use of patient portals, interest in portals and electronic communication is often higher among racial/ethnic minorities and those with chronic medical conditions [13,39]. Still, if efforts to reach out to vulnerable populations fails to occur, health care disparities could increase as portal adopters reap the benefits of easier access to information leaving non-adopters behind [40]. Indeed, several studies have found that portal use is associated with improved patient self-management of disease, better patient-provider communication, and use of preventive health services [36,38,41].

Federal meaningful use criteria currently require at least 5% of a practice's patients to view, download, or transmit their health information electronically, and at least 5% of patients must send a secure electronic message [4]. The Centers for Medicare and Medicaid recently proposed softening these requirements to at least 1% of patients accessing their health information electronically and documenting that a secure electronic

messaging system was enabled [42]. Given clinics' concerns about their low-income patients' willingness and ability to use a patient portal, we are in favor of these relaxed requirements.

One of the most common reasons cited by patients for not using a portal is lack of knowledge or motivation [43]. This finding suggests that educating patients about the portal could help lessen the digital divide and prevent health disparities from increasing. A challenge for health systems will be identifying who can provide this training. In our interviews, we consistently heard that both clinicians and medical staff lack the time to take on extra tasks. Future research should focus on strategies for increasing portal adoption in vulnerable populations. One study found that showing a promotional video during a clinic visit had a small effect on increasing portal registrations [44]. Other potential strategies could include using non-clinical staff as trainers, holding workshops for interested patients, and creating user-friendly online tutorials. In general, clinics that have used a planned, systematic implementation strategy have seen higher rates of portal uptake than clinics that rely on clinicians to inform and enroll patients [45]. However, the importance of having clinician and provider buy-in before implementation has also been highlighted [29].

Limitations

Our study has limitations. Because we were primarily interested in learning how older and vulnerable adults use patient portals, we selected providers from clinics that serve a primarily disadvantaged population. The attitudes and barriers we identified may not be as prevalent in practices serving a higher socioeconomic patient base. Similarly, although we selected clinics from a mixture of urban and rural locations, all our study clinics are located in North Carolina. Different regions of the country may experience different barriers unique to their populations. Likewise, clinics that operate under different reimbursement structures, such as Health Maintenance Organizations, may view things differently, for example, a portal's potential to encourage more out-of-visit communication.

Conclusion

In conclusion, clinic staff from every health center in our sample recognized potential benefits to patient portals but were also concerned about the new work and confusion portals could bring. Uptake of portals was seen as very low, further discouraging providers from embracing them. Future studies should examine ways portals can be implemented efficiently in practices and strategies for increasing portal usage in vulnerable populations, including older adults. For portals to reach their full potential and meaningfully improve care, clinicians and patients will need to view them as a technology that adds value to care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Health Care Providers Interview Guide.

[PDF File (Adobe PDF File), 15KB - jmir_v18i1e8_app1.pdf]

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. 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]
3. Centers for Medicare & Medicaid Services. 2014 Aug. Payment Adjustments & Hardship Exceptions Tipsheet for Eligible Professionals URL: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/PaymentAdj_HardshipExcepTipSheetforEP.pdf [accessed 2014-12-19] [WebCite Cache ID 6a6177Hv0]
4. Centers for Medicare & Medicaid Services. 2014. Eligible Professional Meaningful Use Core Measures: Measure 7 of 17 Internet URL: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/Stage2_EPCore_7_PatientElectronicAccess.pdf [accessed 2015-02-02] [WebCite Cache ID 6a61HoQIY]
5. 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]

6. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]

7. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

8. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]

9. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]

10. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]

11. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]

12. National Learning Consortium. 2013 May. How to optimize patient portals for patient engagement and meet meaningful use requirements URL: http://www.healthit.gov/sites/default/files/nlc_how_to_optimizepatientportals_for_patientengagement.pdf [accessed 2014-12-01] [WebCite Cache ID 6a61M3d5D]

13. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

14. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]

15. Latulipe C, Gatto A, Nguyen H, Miller D, Quandt S, Bertoni A, et al. Design Considerations for Patient Portal Adoption by Low-Income, Older Adults. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, South Korea p. 3859-3868 URL: http://dl.acm.org/citation.cfm?doid=2702123.2702392

16. USDA Economic Research Service. Rural-Urban Continuum Codes URL: http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx [accessed 2015-02-05] [WebCite Cache ID 6a61PKrBa]

17. Guest G. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]

18. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [Medline: 7899572]

19. Arcury TA, Quandt SA. Qualitative methods in arthritis research: sampling and data analysis. Arthritis Care Res 1998 Feb;11(1):66-74 [FREE Full text] [Medline: 9534496]

20. Luborsky M. The identification and analysis of themes and patterns. In: Gubrium JF, Sankar A, editors. Qualitative methods in aging research. Thousand Oaks, CA: Sage; 1994:189-210.

21. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

22. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Reading, MA: Addison-Wesley Pub. Co; 1975.

23. Cain M, Mittman R. Diffusion of Innovation in Health Care. Oakland, CA: California HealthCare Foundation; 2002 May. URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/PDF%20D/PDF%20DiffusionofInnovation.pdf [accessed 2015-10-14] [WebCite Cache ID 6cH724EE1]

24. Rogers E. Diffusion of Innovations, 5th Edition. New York: Free Press; 2003.

25. Keplinger LE, Koopman RJ, Mehr DR, Kruse RL, Wakefield DS, Wakefield BJ, et al. Patient portal implementation: resident and attending physician attitudes. Fam Med 2013 May;45(5):335-340 [FREE Full text] [Medline: 23681685]

26. Ross SE, Todd J, Moore LA, Beaty BL, Wittevrongel L, Lin CT. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005;7(2):e13 [FREE Full text] [doi: 10.2196/jmir.7.2.e13] [Medline: 15914460]

27. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]

28. Walker J, Leveille SG, Ngo L, Vodicka E, Darer JD, Dhanireddy S, et al. Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. Ann Intern Med 2011 Dec 20;155(12):811-819 [FREE Full text] [doi: 10.7326/0003-4819-155-12-201112200-00003] [Medline: 22184688]

29. Wald JS. Variations in patient portal adoption in four primary care practices. AMIA Annu Symp Proc 2010;2010:837-841 [FREE Full text] [Medline: 21347096]

XSL•FO
RenderX

30. Kittler AF, Carlson GL, Harris C, Lippincott M, Pizziferri L, Volk LA, et al. Primary care physician attitudes towards using a secure web-based portal designed to facilitate electronic communication with patients. Inform Prim Care 2004;12(3):129-138. [Medline: 15606985]

31. Zickuhr K. Pew Research Center. Washington, DC; 2013 Sep. Who's not online and why URL: http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_Offline%20adults_092513_PDF.pdf [accessed 2013-10-21] [WebCite Cache ID 6a67eFkA0]

32. Greysen SR, Chin GC, Sudore RL, Cenzer IS, Covinsky KE. Functional impairment and Internet use among older adults: implications for meaningful use of patient portals. JAMA Intern Med 2014 Jul;174(7):1188-1190 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1864] [Medline: 24839165]

33. 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;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]

34. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]

35. van der Vaart R, Drossaert C, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar M. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]

36. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]

37. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May;57(3):491-506. [doi: 10.1177/0018720814549238] [Medline: 25875437]

38. Goldzweig CL, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, et al. Evidence-based Synthesis Program. Washington, DC: Department of Veterans Affairs; 2012 Jul. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes URL: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0048536/pdf/PubMedHealth_PMH0048536.pdf [accessed 2015-12-18] [WebCite Cache ID 6dsMJUy9H]

39. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]

40. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]

41. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]

42. Department of Health and Human Services, Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Modifications to Meaningful Use in 2015 Through 2017. Federal Register 2015 Apr 15;CFR 495:20345-20399.

43. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]

44. North F, Hanna B, Crane S, Smith S, Tulledge-Scheitel S, Stroebel R. Patient portal doldrums: does an exam room promotional video during an office visit increase patient portal registrations and portal use? J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i24-i27 [FREE Full text] [doi: 10.1136/amiajnl-2011-000381] [Medline: 21976028]

45. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]

Abbreviations

HITECH: Health Information Technology for Economic and Clinical Health
TAM: Technology Acceptance Model

XSL•FO
RenderX

Original Paper

Estimating Skin Cancer Risk: Evaluating Mobile Computer-Adaptive Testing

Ngadiman Djaja[1,2,3], MEd; Monika Janda[1,3], PhD; Catherine M Olsen[4], PhD; David C Whiteman[1,3,4], PhD; Tsair-Wei Chien[5,6], MBA

[1]School of Public Health and Social Work, Institute for Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

[2]Institute for Health and Biomedical Innovation, Institute for Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia

[3]National Health and Medical Research Council Centre for Research Excellence, Sun and Health (CRESH), Brisbane, Australia, Australia

[4]QIMR Berghofer Medical Research Institute, Brisbane, Australia

[5]Research Department, Chi-Mei Medical Center, Tainan, Taiwan

[6]Department of Hospital and Health Care Administration, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan

Corresponding Author:
Tsair-Wei Chien, MBA
Research Department
Chi-Mei Medical Center
No. 901, Chung Hwa Road, Yung Kung Dist., Tainan 710, Taiwan
Tainan, 710
Taiwan
Phone: 886 937399106
Fax: 886 62820534
Email: smile@mail.chimei.org.tw

Abstract

Background: Response burden is a major detriment to questionnaire completion rates. Computer adaptive testing may offer advantages over non-adaptive testing, including reduction of numbers of items required for precise measurement.

Objective: Our aim was to compare the efficiency of non-adaptive (NAT) and computer adaptive testing (CAT) facilitated by Partial Credit Model (PCM)-derived calibration to estimate skin cancer risk.

Methods: We used a random sample from a population-based Australian cohort study of skin cancer risk (N=43,794). All 30 items of the skin cancer risk scale were calibrated with the Rasch PCM. A total of 1000 cases generated following a normal distribution (mean [SD] 0 [1]) were simulated using three Rasch models with three fixed-item (dichotomous, rating scale, and partial credit) scenarios, respectively. We calculated the comparative efficiency and precision of CAT and NAT (shortening of questionnaire length and the count difference number ratio less than 5% using independent t tests).

Results: We found that use of CAT led to smaller person standard error of the estimated measure than NAT, with substantially higher efficiency but no loss of precision, reducing response burden by 48%, 66%, and 66% for dichotomous, Rating Scale Model, and PCM models, respectively.

Conclusions: CAT-based administrations of the skin cancer risk scale could substantially reduce participant burden without compromising measurement precision. A mobile computer adaptive test was developed to help people efficiently assess their skin cancer risk.

(J Med Internet Res 2016;18(1):e22)   doi:10.2196/jmir.4736

KEYWORDS

computer adaptive testing; skin cancer risk scale; non adaptive test; Rasch analysis; partial credit model

Introduction

In Australia, skin cancers account for approximately 80% of all newly diagnosed cancers [1]. There are three main types of skin cancer: (1) melanoma (the most dangerous form of skin cancer), (2) basal cell carcinoma (BCC), and (3) squamous cell carcinoma (SCC). BCC and SCC are often grouped together as nonmelanoma or keratinocyte skin cancers. Australia's incidence

XSL·FO
RenderX

of skin cancer is one of the highest in the world: two to three times the rates observed in Canada, the United States, and the United Kingdom [2], with age-standardized incidence rates for cutaneous melanoma at 65.3×10^{-5} and 1878×10^{-5} for keratinocyte cancer [1]. From a population of only 23 million, more than 434,000 people are treated for one or more nonmelanoma skin cancers in Australia each year [1].

Ultraviolet radiation exposure from sunlight is the major causal factor for skin cancer [2]. Personal behaviors to reduce excessive sunlight exposure are important modifiable factors for the prevention of skin cancers. The World Health Organization recommends several suitable behaviors such as appropriate use of sunscreens, staying in the shade, covering with sun protective clothing, giving up sunbathing, and abstaining from using sunbeds [3].

Requirement for Model-Data-Fit Detection

In practice, we do not know the real skin cancer risk for a person. Thus, assuming a person has characteristic attributes that correlate highly with the underlying construct of skin cancer, risk can be assessed through questions (ie, questionnaire items); for example, phenotypic measures such as freckles, hair color, eye color, tendency to burn, or behavioral factors such as attitudes to tanning and use of sunbeds. Using the responses to these items, it should be possible to create a unidimensional (ie, addable) scale to measure these attributes and calculate an overall skin cancer risk score. Ideally, such a score would be precise and characterized by a small standard error (SE).

Statistical validity is the correlation between each person's measures (or scores) on a questionnaire and those persons' unobservable true status [4]. Such unobservable variables (eg, true score or behaviors relating to sun protection and sun exposure) are considered latent traits (ie, exists but cannot be directly observed). The question is how to obtain optimal correlation (or validity) between the items when the true score is unknown. Rasch models [5] can be a gateway to assess how well the items measure the underlying latent trait [6-8]. That is, a unidimensional scale can be verified by Rasch analysis: when the data fit to the Rasch model, all items can be added.

Questionnaires that are built and tested using the Rasch model have become common in educational assessment for many years but are now also increasingly appreciated in health assessment, including measures of patient outcomes (quality of life, pain, depression) and other diverse latent traits such as perceptions of patient hospitalization and nurse bullying [9,10]. We previously applied the Rasch model to the assessment of the quality of an instrument to measure attitudes to skin self-examination [11]. Rasch analysis allows researchers to calculate a precise estimate of the latent trait by assessment of unidimensionality of the items, assessment of differential item functioning [12] (eg, probability of giving a certain response on an item by people from different groups with the same latent trait), and the possibility of transferring static questionnaires to computer adaptive testing (CAT) [13].

Multimedia Graphical Representations to Improve Patients' Health Literacy

Patients' health literacy is increasingly recognized as a critical factor affecting patient-physician communication and health outcomes [14], as a mediator for cancer screening behavior [15], and as a pathway between health literacy and cancer screening [16]. Adults with below basic or basic health literacy are less likely than adults with higher health literacy to get information about health issues from written sources (eg, newspapers, magazines, books, brochures, or the Internet) and more likely than adults with higher health literacy to get a lot of information about health issues from radio and television [17]. A mobile CAT with multimedia graphical representations (ie, similar to radio and television) could increase awareness of the risk of developing skin cancer (ie, health literacy) and motivate patient-physician communication and subsequently behavioral change. However, no mobile CAT app with graphical representations has been available until now.

Study Aims

Using data from a large cohort study of skin cancer from Queensland, Australia [18], we conducted a simulation study with a methodological focus to apply Rasch models to an existing skin cancer risk questionnaire. Further, we sought to compare static (nonadaptive) presentation as commonly used in paper and pencil questionnaires versus computer adaptive testing (CAT) for its precision in measurement. We hypothesized that compared to nonadaptive testing (NAT), CAT would result in greater precision (lower SE) for a similar item number or a shorter questionnaire of similar SE.

Methods

Data Source

De-identified data from the QSkin Sun and Health study baseline questionnaire were used [18]. This is a population-based cohort study of 43,794 men and women aged 40-69 years randomly sampled from the population of Queensland, Australia, in 2011 (Figure 1). We randomly partitioned the data into a calibration dataset (two-thirds, n=29,314) and a validation dataset (one-third, n=14,480). In the calibration dataset, 7213 participants had a history of skin cancer and 22,101 participants did not (Figure 2).

Approval for this study was obtained from the QIMR Berghofer Medical Research Institute Human Research Ethics Committee (approval #P1309). Participants joined the study by completing consent forms and the survey and returning them in a reply-paid envelope. Participants completed two consent forms. The first consent form covered the use of information provided in the survey, permission for data linkage to cancer registries, pathology laboratories, and public hospital databases. The second consent form gave permission for data linkage to Medicare Australia (Australia's universal national health insurance scheme) to ascertain whether or not participants had developed skin cancer.

The baseline questionnaire consisted of 46 items and was answered by all QSkin participants. All items were examined using the Rasch Partial Credit Model (PCM) [19] (Figure 2).

For optimal fit, the Rasch model requires a unidimensional measurement with criteria of Infit and Outfit mean square errors of each item 1.5 [20]. PCM allows for items to have a variable number of thresholds and step difficulties in contrast to the more commonly used Rating Scale Model (RSM) [8,9,21], which requires all items to use the same response categories.

For item invariance, the item estimation should be independent of the subgroups of individuals completing the questions and should work equally across populations [22]. Items not demonstrating invariance are commonly referred to as exhibiting differential item functioning (DIF) [23,24] or item bias. The chi-square test used for detecting DIF was computed from a comparison of the observed overall performance of each trait group on the item with its expected performance [25]. Its probability (eg, $P<.05$) reports the statistical probability of observing a chi-square value when the data fit the Rasch model. We used WINSTEPS [26] to detect items above the thresholds for DIF.

In addition, the category structure for each of the items in the skin cancer item bank should display monotonically increasing thresholds following the Linacre's guidelines [27] to improve the utility of the resulting measures.

Figure 1. Sample selection flowchart.
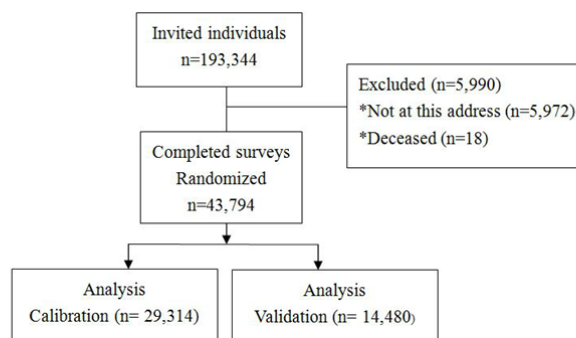
Determining a Cut-Off Point of Skin Cancer Risk

Traditionally in clinical practice, researchers use C-statistics, or area under the receiver operating characteristic (ROC) curve to plot the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings [28]. In this study, we plotted two sample normal distributions incorporated with ROC in Figure 3 when their means and standard deviations were known.

Much information such as cut point, area under ROC curve, and a graphical vertical bar showing cut points can be displayed on a plot. WINSTEPS software [26] was used to estimate means and standard deviations of cases with and without previous skin cancers to determine a cut-off point of skin cancer risk with maximal sensitivity and specificity in MS Excel (Figure 3). Providing the cut-off points in graphical form makes the results clear and easily understandable for readers or clinicians to interpret.

Mobile Computer Adaptive Testing Designed for Examining Personal Skin Cancer Risk

The CAT item bank (fitting to Rasch model's requirement regarding unidimensionality, local dependence, and monotonicity as well as DIF absence on gender) was constructed, consisting of all 31-item parameters obtained from the calibration using WINSTEPS [26].

To start the CAT, an initial item was selected randomly from the item bank. Using this initial item, a provisional person measure was estimated by the expected a posteriori (EAP) method [29] in an iterative Newton-Raphson procedure [9,30]. After each item was answered, EAP was recalculated, until the final score for the person was determined by the maximum of the log-likelihood function before terminating the CAT (Figure 2). The next item selection was based on the highest Fisher information (ie, item variance) of the remaining unanswered items interacting with the provisional person measure.

Two termination rules were set. The first was a minimum standard error of measurement (SEM) of 0.47 required for stopping the CAT. This SEM was set based on the internal consistency of the calibration sample (Cronbach alpha=.78). SE_i was the person SE of the estimated measure according to their item variances of the finished items on CAT, where $SEM=SD \times sqrt (1 - reliability)$ and $SE_i=1/sqrt(\Sigma information[i])$, where i refers to the CAT finished items responded to by a person [31], and SD is the person standard deviation of the derivation sample of 29,314 cases. The second termination rule was that each person must answer at least 10 items according to a simulation study on the data bank for attaining a minimal average personal reliability at a desired level (eg, 0.78) [32].

Simulation to Compare Efficiency and Precision of Computer Adaptive Testing and Nonadaptive Testing

Using the item parameters generated from the derivation cohort, 1000 cases following a normal distribution (mean logit 0, SD logit 1) were simulated [33-35] using three Rasch models (ie, dichotomous, 5-point RSM, and PCM) with three respective fixed-item scenarios (ie, 10, 20, and 30 items; see Tables 1-3).

Figure 2. Study simulation and CAT flowchart (interested readers can run a test of the mobile CAT through the QR code).
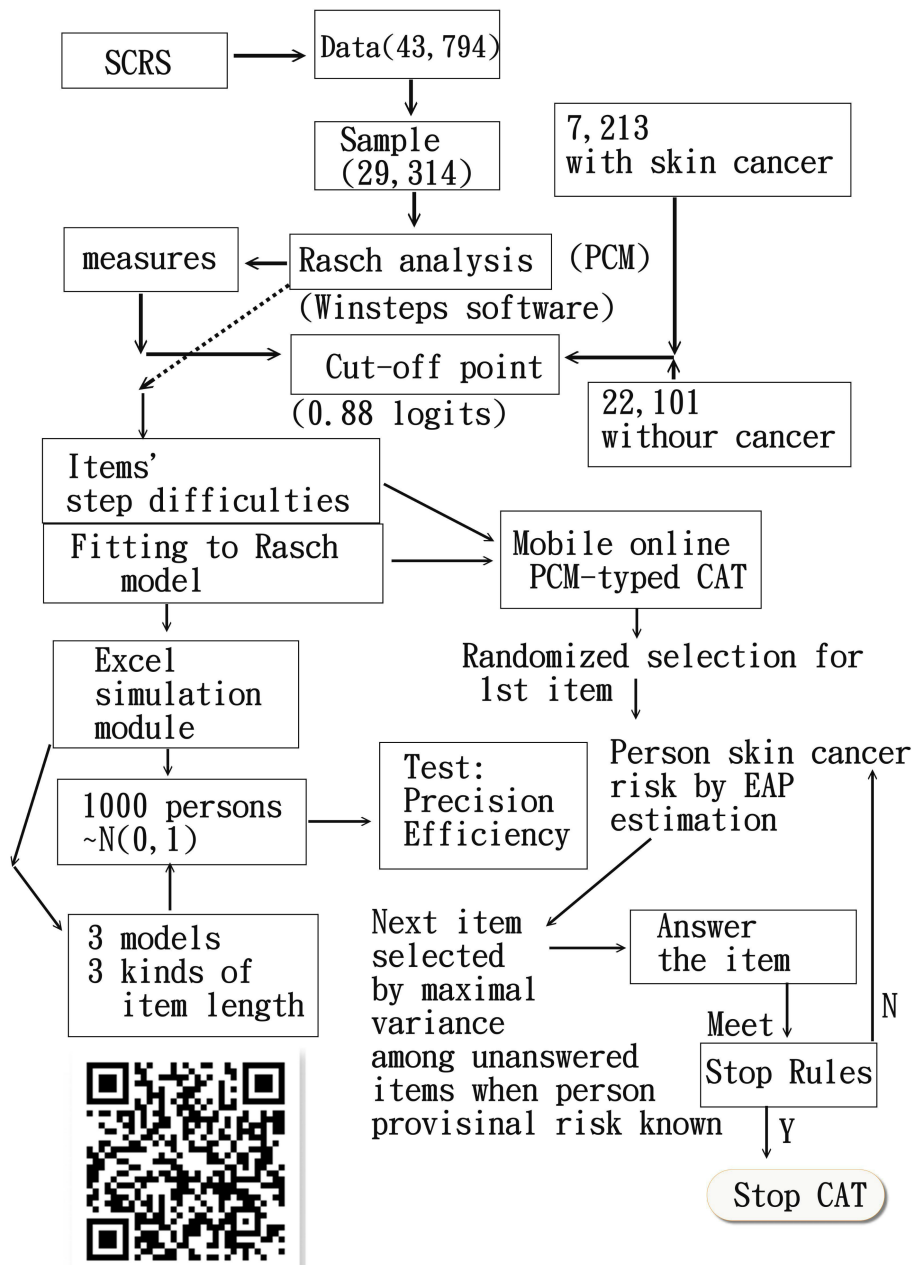
Figure 3. Cut-off point determined.
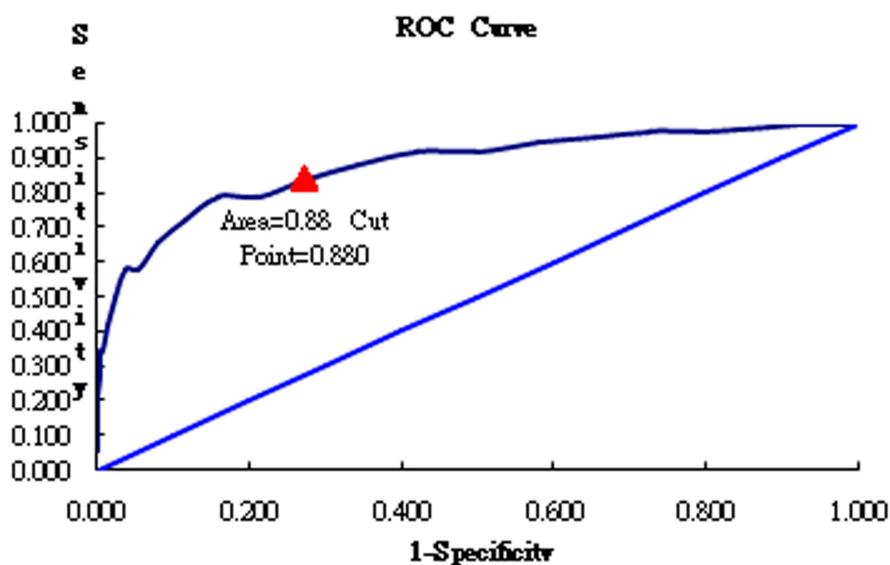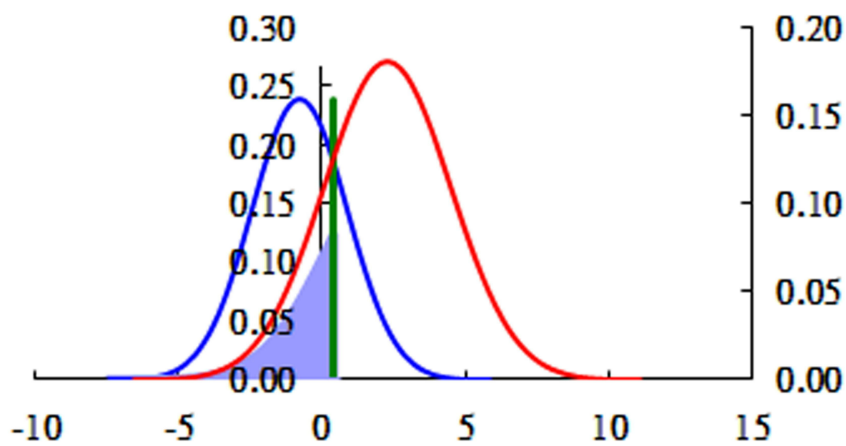
Table 1. 10, 20, or 30 items in static NAT format.

Datasets	Dichotomous		RSM		PCM	
	Mean	SE	Mean	SE	Mean	SE
10 items	-0.007	0.829	0.03	0.414	-0.179	0.398
20 items	-0.008	0.555	0.02	0.289	-0.19	0.272
30 items	0.045	0.439	-0.039	0.235	-0.084	0.224
CAT	-0.021	0.613	0.021	0.361	-0.154	0.32

Table 2. Precision of CAT.

Precision	Dichotomous		RSM		PCM	
	Diff. (%)[a]	Corr.[b]	Diff. (%)[a]	Corr.[b]	Diff. (%)[a]	Corr.[b]
10 items	0.40	0.863	0.30	0.952	0.00	0.931
20 items	0.00	0.957	0.00	0.988	0.00	0.986
CAT	0.13	0.925	0.05	0.958	0.10	0.946

[a]Diff. (%): Different number ratio compared to the 30-item dataset.

[b]Corr: Correlation coefficient of person theta to NAT.

Table 3. Efficiency of CAT.

Efficiency	Dichotomous		RSM		PCM	
	CAT item length	%[a]	CAT item length	%[a]	CAT item length	%[a]
CAT	15.55	48.20	10	66.70	10.13	67.32

[a]Efficiency=1 - CIL/30.

To allow testing of dichotomous and 5-point rating scale Rasch models, all item (or step) difficulties were converted from the calibrated results of the PCM. The overall difficulty for each item was designated to be the respective threshold of the dichotomous scale. In contrast, the step difficulties of the 5-point RSM [21] ranged from -2 to 2, with an advance 1.0 logit interval added to the overall difficulty of the respective item as to the PCM.

We calculated the comparative efficiency and precision for CAT and NAT by varying the number of items presented (10, 20, and 30 items) and by testing the difference in precision and efficiency compared to answering all available 31-items using independent *t* tests to count different number ratio less than 5% as shown in the following formula [36], respectively:

$$t = |\theta_{cat} - \theta_{30}|/\sqrt{SE^2_{cat} + SE^2_{30}}$$

In addition, a comparison of average person SEs achieved across all different conditions was made to verify precision for CAT and NAT. We ran an author-created Visual Basic for Applications module in MS Excel to conduct the simulation study (Figure 2) and mobile CAT.

Results

Determining a Cut-Off Point

The mean and SD of skin cancer risk for participants without skin cancer (mean -0.79, SE 1.67) or with skin cancer (mean 2.29, SE 2.21) were calculated and used to determine the optimal cut-off point at 0.88 logit with sensitivity at 0.79 and specificity at 0.74. Using this cut-off, the area under the ROC curve was 0.88 (see Figure 3).

Simulation to Compare Efficiency and Precision of Computer Adaptive Testing and Nonadaptive Testing

Using simulation data, we found that using more items yielded higher Cronbach alpha scores (Figure 4). Dichotomous scales had the lowest Cronbach alpha and dimension coefficient [37]. The PCM scales had the highest Cronbach alpha. The RSM scales gained the highest dimension coefficient.

As shown in Figure 4, CAT gained a relatively smaller SE corresponding to item length (ie, compared to NAT, shorter CATs result in larger SE). At equivalent precision, CAT reduces the response burden by 48.20%, 66.70%, and 66.20%, respectively for dichotomous, RSM, and PCM models (Figure 5).

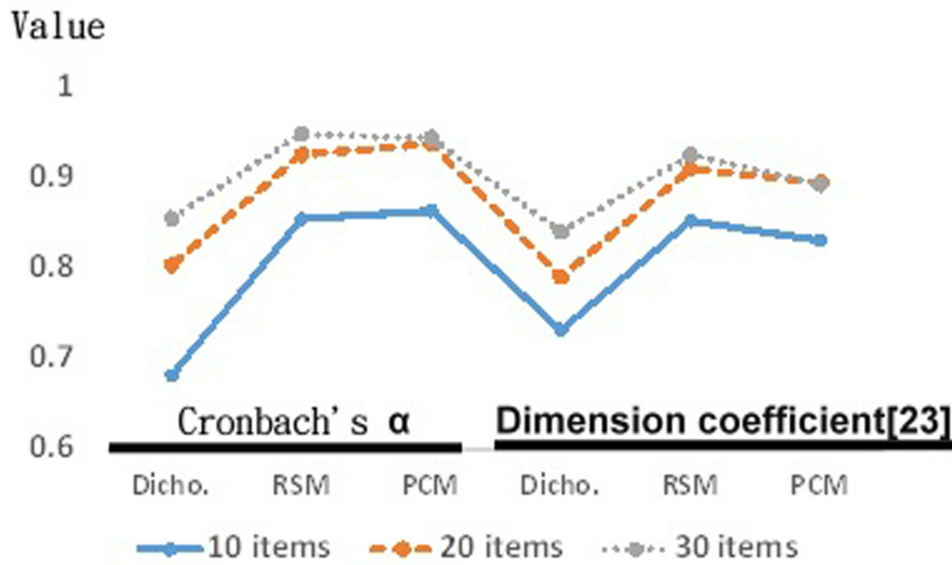**Figure 4.** Simulation data generated with 3 Rasch models.

Figure 5. Efficiency and precision of CAT, compared to using 10, 20, or 30 items in static NAT format.
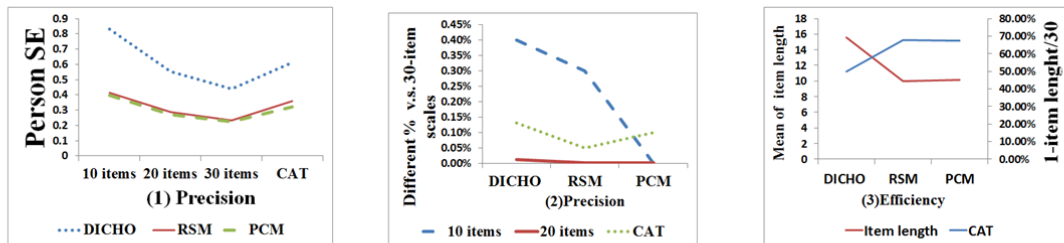
Mobile Computer Adaptive Testing Evaluating Skin Cancer Risk

We developed a mobile CAT survey procedure (see QR code in Figure 2 and Multimedia Appendix 1) to practically demonstrate the newly designed PCM-type CAT app in action. The CAT process was demonstrated item by item and is shown at the top of Figure 6. Person theta is the provisional ability estimated by the CAT module. The mean square error at the bottom of Figure 6 was generated by the formula of

$1/sqrt(\Sigma information[i])$, where i refers to the CAT presented items responded to by a person [31]. In addition, the residual at the top of Figure 6 was the average of the last five change differences between the pre-and-post estimated abilities on each CAT step. CAT will stop if residual value 0.05. The "corr" refers to the correlation coefficient between the CAT estimated measures and the step series numbers using the last 5 estimated theta values. The flatter of the theta trends means the higher probability of the person measure convergent to a final estimation.

Figure 6. A graphical CAT report shown after each response (top) and the more item length, the less standard errors in CAT process (bottom).
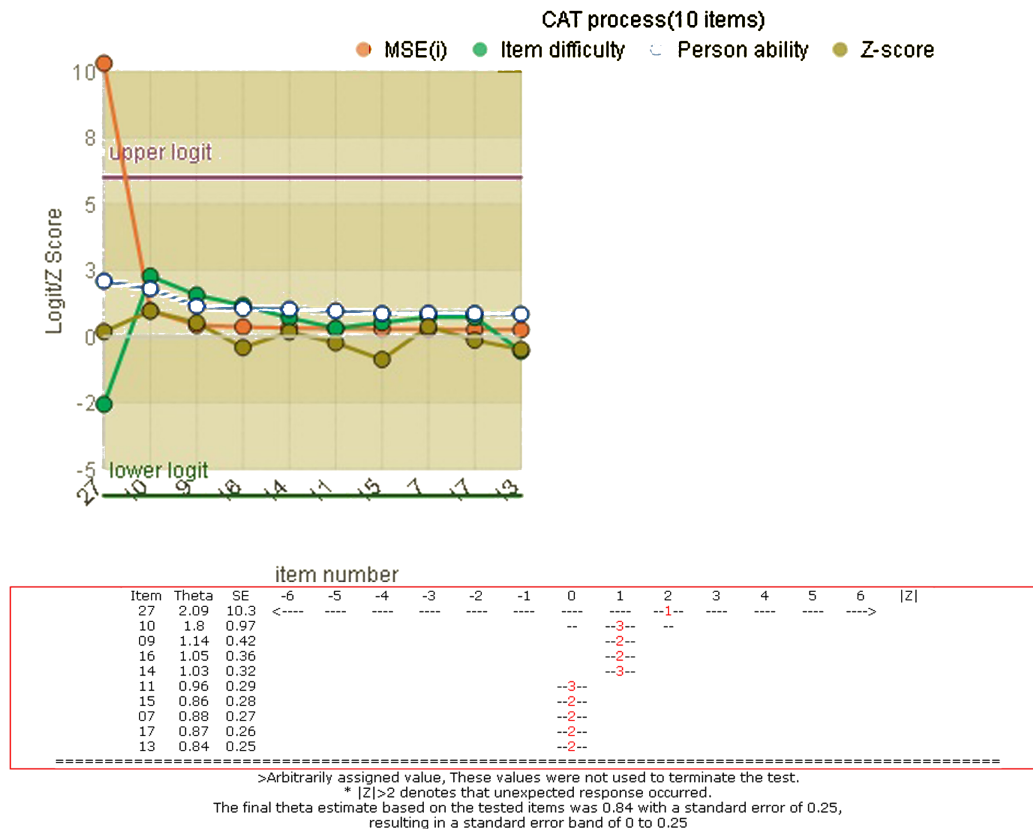
Discussion

Principal Findings

We used two different approaches to measure risk of skin cancer: nonadaptive testing and computer adaptive testing. Using data from a very large cohort of more than 43,000 people, we were able to show that our scale was able to accurately identify people at highest risk for skin cancer. On our risk scale, we identified a very high discriminatory accuracy of 0.88 (ie, the proportion of area under ROC curve) using a cut-off of 0.88 logits (the higher, the worse). Using CAT results in a smaller SE at high efficiency (fewer items answered), and therefore without compromising test precision, reduces response burden by 48.20%, 66.70%, and 66.20% for dichotomous, RSM, and PCM models, respectively. A prototype mobile online CAT for evaluating skin cancer risk has been developed and could be used to assess skin cancer risk at considerable reduction of respondent burden.

Consistent with the literature [8,9,30,34,35], the efficiency of CAT over NAT was supported for this skin cancer risk scale. We confirm the PCM-type CAT (ie, different from others by using simpler Rasch family models) requires significantly fewer items to measure a person's risk than NAT but does not compromise the precision of measurement. This mobile assessment could be used to quickly estimate a person's skin cancer risk and educate them about the need for skin protection on a personal level [38-40]. We confirm that participants with a history of skin cancer had a higher mean score of responses than those without a history of skin cancer.

Implications

Patients' health literacy (eg, understanding their own skin cancer risk) is increasingly recognized as a critical factor affecting patient-physician communication and health outcomes [14]. Adults with below basic or basic health literacy are more likely than adults with higher health literacy to get information about health issues from multimedia graphical representation [17], rather than the traditional newspapers, magazines, books, brochures, or *pamphlets*. A brief CAT such as the one we developed could be used to inform people quickly about their skin cancer risk and how to improve their sun protection behaviors.

This CAT module is a practical tool that can gather responses from patients efficiently and precisely. The tool offers diagnostics that can help practitioners assess whether responses are distorted or abnormal. For example, outfit mean-square values of 2.0 or greater suggest an unusual response. In instances where responses do not fit with the model's requirement, they can be highlighted for suspected cheating, careless responding, lucky guessing, creative responding, or random responding [41]; otherwise, one can take follow-up action [8,34,35] if the result shows a high cancer risk. For example, if a person's measure/risk is 1.0 logit (ie, log odds), their probability of developing skin cancer approaches $0.53(=\exp(1-0.88)/(1+\exp(1-0.88)))$. Interested readers can run

a test of the mobile CAT through the QR code shown in Figure 2.

A mobile online CAT could be used for evaluating skin cancer risk and might reduce the item length in clinical settings. The CAT can be improved in the future by expanding the item pool allowing use among more diverse samples. It must be noted that (1) item overall (ie, on average) and step (threshold) difficulties of the questionnaire must be calibrated in advance using Rasch analysis or other item response theory models before creating an item bank, (2) pictures used for the subject or response categories for each question should be well prepared with a Web link that can be shown simultaneously with the item appearing in the animation module of CAT, and (3) the model can be used for many kinds of models based on item response theory.

Strengths and Limitations

There are two major forms of standardized assessments in clinical settings [42]: (1) a traditional self-administered questionnaire, and (2) a rapid short-form scale [43,44]. Each has its advantages and drawbacks. Traditional pencil-and-paper questionnaires have a large respondent burden, often because they require patients to answer questions that do not provide additional information about their risk of disease in order to achieve adequate precision measurement [45]. CAT can target the optimal question for a specific person and therefore end at an appropriate number of items more economically according to the required SE (or say, criterion of person reliability). However, along with the advantages offered by CAT, there are some drawbacks as well, such as impossibility of estimating the ability in case of all extreme responses, CAT algorithms requiring serious item calibration, several items from the item bank being overexposed, and other test items not being used at all [46].

The strengths of this study include its very large sample size of more than 40,000 participants, permitting detailed analysis of the performance of questionnaire items and the ability to further test the performance of the items in a validation dataset. We simulated data by varying the types of models and item length to execute the CAT. (Interested readers who wish to see the video demonstration or use the MS Excel-type module can contact the corresponding author).

As with all forms of Web-based technology, advances in mobile health (mHealth) and health communication technology are rapidly emerging [47]. Use of mobile online CAT is promising and worth considering in many fields of health assessment, similar to its prominent role in education and staff selection testing. However, several issues should be considered more thoroughly in further studies. The scale's Cronbach alpha (=.78 yielded by studied 29,314 cases), sensitivity at 0.79, and specificity at 0.74 are slightly low. Second, the CAT module has a potential limitation for people using languages other than English because the interface may need to be modified for use in real world. A multiple language interface should be developed in the future. Third, the CAT graphical representation shown in Figure 6 might be confusing and difficult to interpret for people unfamiliar with CAT and may need to be improved to become a standard part of CAT routine.

Conclusions

The PCM-type CAT for skin cancer risk can reduce respondents' burden without compromising measurement precision and increases endorsement efficiency. The CAT module can be used for mobile phones and easy online assessment of patients' disease risks. This is a novel and promising way to capture information about skin cancer risk, for example while waiting outside physician consultation offices.

Authors' Contributions

All authors read and approved the final manuscript. ND and T-WC developed the study concept and design. MJ and CMO analyzed and interpreted the data. ND, T-WC, and DCW drafted the manuscript, and all authors provided critical revisions for important intellectual content. The study was supervised by T-WC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demonstration of an online CAT.

[WMV File (Windows Media Video), 2MB - jmir_v18i1e22_app1.wmv]

References

1.   Australian Institute of Health and Welfare & Australasian Association of Cancer Registries. Cancer in Australia: an overview. 2012. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129542353 [accessed 2015-11-20] [WebCite Cache ID 6dAmb82r4]
2.   Narayanan DL, Saladi RN, Fox JL. Ultraviolet radiation and skin cancer. Int J Dermatol 2010 Sep;49(9):978-986. [doi: 10.1111/j.1365-4632.2010.04474.x] [Medline: 20883261]
3.   Global Solar UV Index: A Practical Guide. Geneva: World Health Organization; 2002. URL: http://www.who.int/uv/publications/en/GlobalUVI.pdf [accessed 2015-11-20] [WebCite Cache ID 6dAp54PvL]
4.   Linacre J. True-score reliability or Rasch statistical validity? Rasch Measurement Transactions 1996;9(4):455.

XSL•FO
RenderX

5. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980.

6. Lerdal A, Kottorp A, Gay CL, Grov EK, Lee KA. Rasch analysis of the Beck Depression Inventory-II in stroke survivors: a cross-sectional study. J Affect Disord 2014 Apr;158:48-52 [FREE Full text] [doi: 10.1016/j.jad.2014.01.013] [Medline: 24655764]

7. Forkmann T, Boecker M, Wirtz M, Eberle N, Westhofen M, Schauerte P, et al. Development and validation of the Rasch-based Depression Screening (DESC) using Rasch analysis and structural equation modelling. J Behav Ther Exp Psychiatry 2009 Sep;40(3):468-478. [doi: 10.1016/j.jbtep.2009.06.003] [Medline: 19589499]

8. Sauer S, Ziegler M, Schmitt M. Rasch analysis of a simplified Beck Depression Inventory. Personality and Individual Differences 2013;54(4):530-535.

9. Chien T, Wang W, Huang S, Lai W, Chow C. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. J Med Internet Res 2011;13(3):e61 [FREE Full text] [doi: 10.2196/jmir.1785] [Medline: 21844001]

10. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]

11. Djaja N, Youl P, Aitken J, Janda M. Evaluation of a skin self examination attitude scale using an item response theory model approach. Health Qual Life Outcomes 2014;12:189 [FREE Full text] [doi: 10.1186/s12955-014-0189-x] [Medline: 25539671]

12. Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P. Differential item functioning in the Danish translation of the SF-36. J Clin Epidemiol 1998 Nov;51(11):1189-1202. [Medline: 9817137]

13. Ruo B, Choi SW, Baker DW, Grady KL, Cella D. Development and validation of a computer adaptive test for measuring dyspnea in heart failure. J Card Fail 2010 Aug;16(8):659-668 [FREE Full text] [doi: 10.1016/j.cardfail.2010.03.002] [Medline: 20670845]

14. Williams MV, Davis T, Parker RM, Weiss BD. The role of health literacy in patient-physician communication. Fam Med 2002 May;34(5):383-389. [Medline: 12038721]

15. Lee HY, Rhee TG, Kim NK. Cancer literacy as a mediator for cancer screening behaviour in Korean adults. Health Soc Care Community 2015 May 14:-. [doi: 10.1111/hsc.12243] [Medline: 25975449]

16. Kim K, Han HR. Potential links between health literacy and cervical cancer screening behaviors: a systematic review. Psychooncology 2015 Jun 18:-. [doi: 10.1002/pon.3883] [Medline: 26086119]

17. Cutilli CC, Bennett IM. Understanding the health literacy of America: results of the National Assessment of Adult Literacy. Orthop Nurs 2009;28(1):27-32; quiz 33 [FREE Full text] [doi: 10.1097/01.NOR.0000345852.22122.d6] [Medline: 19190475]

18. Olsen CM, Green AC, Neale RE, Webb PM, Cicero RA, Jackman LM, et al. Cohort profile: the QSkin Sun and Health Study. Int J Epidemiol 2012 Aug;41(4):929-929i [FREE Full text] [doi: 10.1093/ije/dys107] [Medline: 22933644]

19. Masters G. A rasch model for partial credit scoring. Psychometrika 1982 Jun;47(2):149-174. [doi: 10.1007/BF02296272]

20. Lai W, Chien T, Lin H, Su S, Chang C. A screening tool for dengue fever in children. Pediatr Infect Dis J 2013 Apr;32(4):320-324. [doi: 10.1097/INF.0b013e31827e111e] [Medline: 23249912]

21. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561-573.

22. Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of item parameter estimates. J Appl Meas 2003;4(2):153-163. [Medline: 12748407]

23. Holland P, Wainer H. Differential item functioning. Hillsdale: Lawrence Erlbaum Associates; 1993.

24. Tennant A, Pallant J. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. Rasch Mea Trans 2007;20(4):1082-1084.

25. Linacre J. RUMM2020 Item-Trait Chi-Square and Winsteps DIF Size. Rasch Meas Trans 2007;21(1):1096 [FREE Full text]

26. Linacre JM. WINSTEPS Software. URL: http://www.winsteps.com/index.htm [accessed 2015-11-20] [WebCite Cache ID 6dAqKCeRC]

27. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3(1):85-106. [Medline: 11997586]

28. Stephan C, Wesseling S, Schink T, Jung K. Comparison of eight computer programs for receiver-operating characteristic analysis. Clin Chem 2003 Mar;49(3):433-439 [FREE Full text] [Medline: 12600955]

29. Bock R, Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika 1981 Dec;46(4):443-459. [doi: 10.1007/BF02293801]

30. Embretson S, Reise S. Item response theory for psychologists. Mahwah, NJ: L. Erlbaum Associates; 2000.

31. Linacre J. Computer-Adaptive Tests (CAT), Standard Errors and Stopping Rules. Rasch Meas Trans 2006;20(2):1062.

32. Hsueh I, Chen J, Wang C, Hou W, Hsieh C. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. Phys Ther 2013 May;93(5):681-693 [FREE Full text] [doi: 10.2522/ptj.20120173] [Medline: 23329557]

33. Linacre J. How to Simulate Rasch Data. Rasch Meas Trans 2007;21(3):1125 [FREE Full text]

34. Chien T, Wu H, Wang W, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. Health Qual Life Outcomes 2009;7:39 [FREE Full text] [doi: 10.1186/1477-7525-7-39] [Medline: 19416521]

XSL•FO

RenderX

35.   Wainer H, Dorans N. Computerized adaptive testing: a primer. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

36.   Smith EJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3(2):205-231. [Medline: 12011501]

37.   Chien T. Cronbach's alpha with the dimension coefficient to jointly assess a scale's quality. Rasch Meas Trans 2012;26(3):1379 [FREE Full text]

38.   Robinson KJ, Gaber R, Hultgren B, Eilers S, Blatt H, Stapleton J, et al. Skin Self-Examination Education for Early Detection of Melanoma: A Randomized Controlled Trial of Internet, Workbook, and In-Person Interventions. J Med Internet Res 2014;16(1):e7 [FREE Full text] [doi: 10.2196/jmir.2883]

39.   Brady MS, Oliveria SA, Christos P, Berwick M, Coit DG, Katz J, et al. Patterns of detection in patients with cutaneous melanoma. Cancer 2000 Jul 15;89(2):342-347. [doi: 10.1002/1097-0142(20000715)89:2<342::AID-CNCR19>3.0.CO;2-P]

40.   Berwick M, Begg C, Fine J, Roush G, Barnhill R. Screening for Cutaneous Melanoma by Skin Self-Examination. JNCI Journal of the National Cancer Institute 1996 Jan 03;88(1):17-23. [doi: 10.1093/jnci/88.1.17]

41.   Karabatsos G. Comparing the aberrant response detection performance of thirty-six person-fit statistics. Applied Measurement in Education 2003;16(4):277-298.

42.   Eack S, Singer J, Greeno C. Screening for anxiety and depression in community mental health: the Beck Anxiety and Depression Inventories. Community Ment Health J 2008 Dec;44(6):465-474. [doi: 10.1007/s10597-008-9150-y] [Medline: 18516678]

43.   Shear M, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. Am J Psychiatry 2000 Apr;157(4):581-587. [doi: 10.1176/appi.ajp.157.4.581] [Medline: 10739417]

44.   Ramirez BM, Bostic J, Davies D, Rush A, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. Am J Psychiatry 2000 Oct;157(10):1599-1605. [doi: 10.1176/appi.ajp.157.10.1599] [Medline: 11007713]

45.   De Beurs DP, de Vries ALM, de Groot MH, de Keijser J, Kerkhof JFMA. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. J Med Internet Res 2014;16(9):e207 [FREE Full text] [doi: 10.2196/jmir.3511] [Medline: 25213259]

46.   Antal M, Imre A. Computerized adaptive testing: implementation issues. Acta Univ Sapientiae Informatica 2010;2(2):168-183 [FREE Full text]

47.   Mitchell JS, Godoy L, Shabazz K, Horn BI. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]

Abbreviations

BCC: basal cell carcinoma
CAT: computer adaptive testing
DIF: differential item functioning
NAT: nonadaptive testing
PCM: Partial Credit Model
ROC: receiver operating characteristic
RSM: Rating Scale Model
SCC: squamous cell carcinoma
SE: standard error
SEM: standard error of measurement

XSL•FO

RenderX

Original Paper

The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis

Qian Liu[1], MMSc; Weijun Peng[1], MMSc; Fan Zhang[1], MMSc; Rong Hu[1], MPH; Yingxue Li[1], MMSc; Weirong Yan[1], PhD

Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College of Huazhong University of Science &Technology, Wuhan, China

Corresponding Author:
Weirong Yan, PhD
Department of Epidemiology and Biostatistics
School of Public Health
Tongji Medical College of Huazhong University of Science &Technology
Tongji Medical College
No.13 Hangkong Road
Wuhan,
China
Phone: 86 (027)83650713
Fax: 86 (027)83650713
Email: weirong.yan@hust.edu.cn

Abstract

Background: Blended learning, defined as the combination of traditional face-to-face learning and asynchronous or synchronous e-learning, has grown rapidly and is now widely used in education. Concerns about the effectiveness of blended learning have led to an increasing number of studies on this topic. However, there has yet to be a quantitative synthesis evaluating the effectiveness of blended learning on knowledge acquisition in health professions.

Objective: We aimed to assess the effectiveness of blended learning for health professional learners compared with no intervention and with nonblended learning. We also aimed to explore factors that could explain differences in learning effects across study designs, participants, country socioeconomic status, intervention durations, randomization, and quality score for each of these questions.

Methods: We conducted a search of citations in Medline, CINAHL, Science Direct, Ovid Embase, Web of Science, CENTRAL, and ERIC through September 2014. Studies in any language that compared blended learning with no intervention or nonblended learning among health professional learners and assessed knowledge acquisition were included. Two reviewers independently evaluated study quality and abstracted information including characteristics of learners and intervention (study design, exercises, interactivity, peer discussion, and outcome assessment).

Results: We identified 56 eligible articles. Heterogeneity across studies was large ($I^2 \geq 93.3$) in all analyses. For studies comparing knowledge gained from blended learning versus no intervention, the pooled effect size was 1.40 (95% CI 1.04-1.77; $P<.001$; n=20 interventions) with no significant publication bias, and exclusion of any single study did not change the overall result. For studies comparing blended learning with nonblended learning (pure e-learning or pure traditional face-to-face learning), the pooled effect size was 0.81 (95% CI 0.57-1.05; $P<.001$; n=56 interventions), and exclusion of any single study did not change the overall result. Although significant publication bias was found, the trim and fill method showed that the effect size changed to 0.26 (95% CI -0.01 to 0.54) after adjustment. In the subgroup analyses, pre-posttest study design, presence of exercises, and objective outcome assessment yielded larger effect sizes.

Conclusions: Blended learning appears to have a consistent positive effect in comparison with no intervention, and to be more effective than or at least as effective as nonblended instruction for knowledge acquisition in health professions. Due to the large heterogeneity, the conclusion should be treated with caution.

XSL•FO
RenderX

Introduction

Electronic learning (e-learning) has quickly become popular for health education [1-3], especially since the emergence of the Internet has allowed its potential to be realized [4]. E-learning can not only transcend space and time boundaries and improve convenience and effectiveness for individualized and collaborative learning, but also provide reusable and up-to-date information through the use of interactive multimedia [3,5-9]. However, it also suffers from disadvantages such as high costs for preparing multimedia materials, continuous costs for platform maintenance and updating, as well as learners' feelings of isolation in virtual environments [8,10,11]. Traditional learning must be conducted at a specific time and place and is considered vital in building a sense of community [12,13]. Blended learning, defined as the combination of traditional face-to-face learning and asynchronous or synchronous e-learning [14], has been presented as a promising alternative approach for health education because it is characterized as synthesizing the advantages of both traditional learning and e-learning [8,15,16]. Moreover, blended learning has shown rapid growth and is now widely used in education [17,18].

With the introduction of blended learning, increasing research has focused on concerns about its effectiveness. Three original research articles reporting on quantitative evaluations of blended learning were published in the 1990s [19-21], and then many were published after 2000 [16,22-29]. A quantitative synthesis of these studies could inform educators and students about evidence for, and factors influencing, the effectiveness of blended learning.

Rowe et al's systematic review reported that blended learning has the potential to improve clinical competencies among health students [30]. In another systematic review, McCutcheon et al suggested a lack of evaluation of blended learning in undergraduate nursing education [31]. Several reviews have also summarized the evaluation of e-learning in medical education, but none separated blended learning from pure e-learning [32-34]. Furthermore, these systematic reviews were limited to only some areas or branches of health education; there has been no quantitative synthesis to evaluate the effectiveness of blended learning in all professions directly related to human and animal health.

Therefore, our study aimed to identify and quantitatively synthesize all studies evaluating the effectiveness of blended learning for health professional learners who were students, postgraduate trainees, or practitioners in a profession directly related to human or animal health. We conducted two meta-analyses: the first summarized studies comparing blended learning with no intervention, and the second explored blended learning compared with nonblended learning (including pure e-learning and traditional face-to-face learning). We also aimed to explore factors that could explain differences in learning effectiveness across characteristics of participants, interventions, and study designs. Based on previous research, we hypothesized that learning outcomes would be improved through exercises, cognitive interactivity, and peer discussion [35-38]. Exercises contain cases, quizzes, self-assessment test, and other activities requiring learners to apply knowledge acquired from the course [33]. Cognitive interactivity reflects cognitive engagement required for course participation, and multiple practice exercises, essays, and group collaborative projects account for high interactivity [38]. Peer discussion includes instructor-student or peer-peer face-to-face discussion that might arise in a typical lecture, and synchronous or asynchronous online communication such as discussion boards, email, chat, or Internet conferencing [33].

Methods

Reporting Standards

We conducted and reported our study according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [39] (see e-Table 7 in Multimedia Appendix 1) and meta-analyses of observational studies in epidemiology [40].

Eligibility Criteria

Inclusion criteria for studies were based on the PICOS (population, intervention, comparison, outcome, and study design) framework [39]. Studies were included only if they (1) were conducted among health professional learners, (2) used a blended learning intervention in the experimental group, (3) involved a comparison of blended learning with no intervention or nonblended learning, (4) included quantitative outcomes with respect to knowledge assessed with subjective (eg, learner self-report) or objective assessments (eg, multiple-choice question knowledge test) of learners' factual or conceptual understanding of the course, and (5) were randomized controlled trials (RCTs) or nonrandomized studies (NRSs), which are widely used in health profession education [33]. Studies in any language and of any publication type were included. Gray literature was searched in CENTRAL and ERIC.

Studies were excluded if they did not compare blended learning with nonblended learning or no intervention, did not report quantitative outcomes with respect to knowledge, used a single-group posttest-only design, were not conducted with health professional learners, evaluated pure e-learning instead of blended learning, or used the computer only for administrative purposes. Reviews, editorials, or meeting abstracts without original data were also excluded.

Data Sources

To identify relevant studies, we conducted a search of citations in Medline, CINAHL, Science Direct, Ovid Embase, Web of Science, CENTRAL, and ERIC. Key search terms included delivery concepts (eg, blended, hybrid, integrated, computer-aided, computer-assisted; learning, training, education, instruction, teaching, course), participants' characteristics (eg, physician*, medic*, nurs*, pharmac*, dent*, cme, health*), and

study design concepts (eg, compar*, trial*, evaluat*, assess*, effect*, pretest*, pre-test, posttest*, post-test, preintervention, pre-intervention, postintervention, post-intervention). The asterisk (*) was used as a truncation symbol for searching. For instance, evaluat* retrieved entries containing the following words: evaluate, evaluation, or evaluative, etc. E-Table 1 in Multimedia Appendix 1 describes the complete search strategy for each database. The last date of search was September 25, 2014. In addition, all references of included studies were screened for any relevant articles.

Study Selection

Using these criteria, QL and FZ independently screened all titles and abstracts and reviewed the full text of all potentially eligible abstracts. Conflicts between these reviewers were resolved through discussion with other members of the research group until a consensus was obtained.

Data Extraction

QL and FZ developed a form (based on the Cochrane Consumers and Communication Review Group's data extraction template), pilot-tested it on 10 randomly selected included publications, and refined it accordingly. Using the same form, data related to the following issues were extracted independently by QL and FZ: first author's name, year of publication, country where the intervention was conducted, study design, study subjects, sample size, specific health profession of the intervention, comparison intervention, intervention duration, exercises, interactivity, peer discussion, outcome assessment, conflict of interest (whether there was a conflict of interest), and funding from company (whether funding was obtained from a source that had a direct interest in the results). Disagreements were resolved through discussion with another research team member until agreement was reached. If the required data for the meta-analyses were missing from the original report, attempts were made to obtain the information by contacting the corresponding authors by email.

Quality Assessment

Recognizing that many nonrandomized and observational studies would be included, the methodological quality of the studies was evaluated using a modified Newcastle-Ottawa Scale (also called the Newcastle-Ottawa Scale-Education), which is an instrument used to appraise the methodological quality of original medical education research studies, typically in the process of a literature review of a field or topic in medical education [33,41-43]. Each study could receive up to 6 points and was rated in the following five domains:

- Representativeness: the intervention group was "truly" or "somewhat" representative of the average learner in this community (1 point).

- Selection: the comparison group was drawn from the same community as the experimental cohort (1 point).

- Comparability of cohorts (2 points possible): These include nonrandomized two-cohort studies (further classified into "controlled for baseline learning outcome [eg, adjusted for knowledge pretest scores; 1 point]" and "controlled for other baseline characteristics [1 point]") and randomized

studies (further classified into randomized [1 point] and allocation concealed [1 point]).

- Blinding: outcome assessment was blinded (1 point). These include (1) blinded if the assessor cannot be influenced by group assignment; (2) assessments that do not require human judgment (eg, multiple-choice tests or computer-scored performance) are considered to be blinded; (3) one-group studies are not blinded unless scoring does not require judgment or authors describe a plausible method for hiding the timing of assessment; (4) participant-reported outcomes are never blinded.

- Follow-up: subjects lost to follow-up were unlikely to introduce bias; small number lost (75% or greater follow-up) or description provided of those lost (1 point).

In addition, we evaluated the quality of evidence with the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) instrument [44-53]. GRADE identifies five factors that may decrease the quality of evidence of studies, and three factors that may increase it. RCTs start with a high rating and observational studies with a low rating. Ratings are modified downward due to (1) study limitations (risk of bias) [47], (2) inconsistency of results [50], (3) indirectness of evidence [51], (4) imprecision [49], and (5) likely publication bias [48]. Ratings are modified upward due to (1) large magnitude of effect, (2) dose response, and (3) confounders likely to minimize the effect. Evaluating these elements, we determine the quality of evidence as "high" (ie, further research is very unlikely to change our confidence in the estimate of effect), "moderate" (ie, further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate), "low" (ie, further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate), or "very low" (ie, we are very uncertain about the estimate).

Data Synthesis

Analyses were carried out for knowledge outcomes using Stata Version 12.0 and R 3.1.2. The standardized mean difference (SMD; Hedges g effect sizes), converted from means and standard deviations from each study, was used [33,54]. When the mean was available but the standard deviation (SD) was not, we used the mean SD of all other included studies. As the overall scores of included studies were not the same and SMD could eliminate the effects of absolute values, we adjusted the mean and SD so that the average SD could replace the missing value of SD.

The I^2 statistic was used to quantify heterogeneity across studies [55]. When the estimated I^2 was equal to or greater than 50%, this indicated large heterogeneity. As the studies incorporated are functionally different and involve different study designs, participants, interventions, and settings, a random-effects model allowing more heterogeneity was used. Meta-analyses were conducted and forest plots were created. To explore publication bias, funnel plots were created and Begg's tests were performed. To explore potential sources of heterogeneity, we performed multiple meta-regression and subgroup analyses based on factors selected in advance, such as study design, country socioeconomic status, participant type, duration of intervention,

randomization, quality score, exercises, interactivity, peer discussion, outcome assessment, and intervention of the control group. Moreover, we performed sensitivity analyses to test the robustness of findings.

Results

Study Selection

The search strategy identified 4815 citations from the databases, and 642 duplicates were removed. After scanning the titles and abstracts, 225 were found to be potentially eligible. Then, full texts were read for further assessment, and 62 remained. For 12 articles without accessible full texts and 6 without sufficient quantitative data (mean knowledge scores), we tried contacting the authors by email but received no reply. Thus, 56 publications were included, among which one publication compared blended learning with both no intervention and nonblended instruction (Figure 1). No more relevant articles were found by reviewing the references of the included articles. E-Table 2 in Multimedia Appendix 1 includes the references of articles excluded based on full text (n=163) and insufficient quantitative data reported (n=6).

Figure 1. Study selction process.
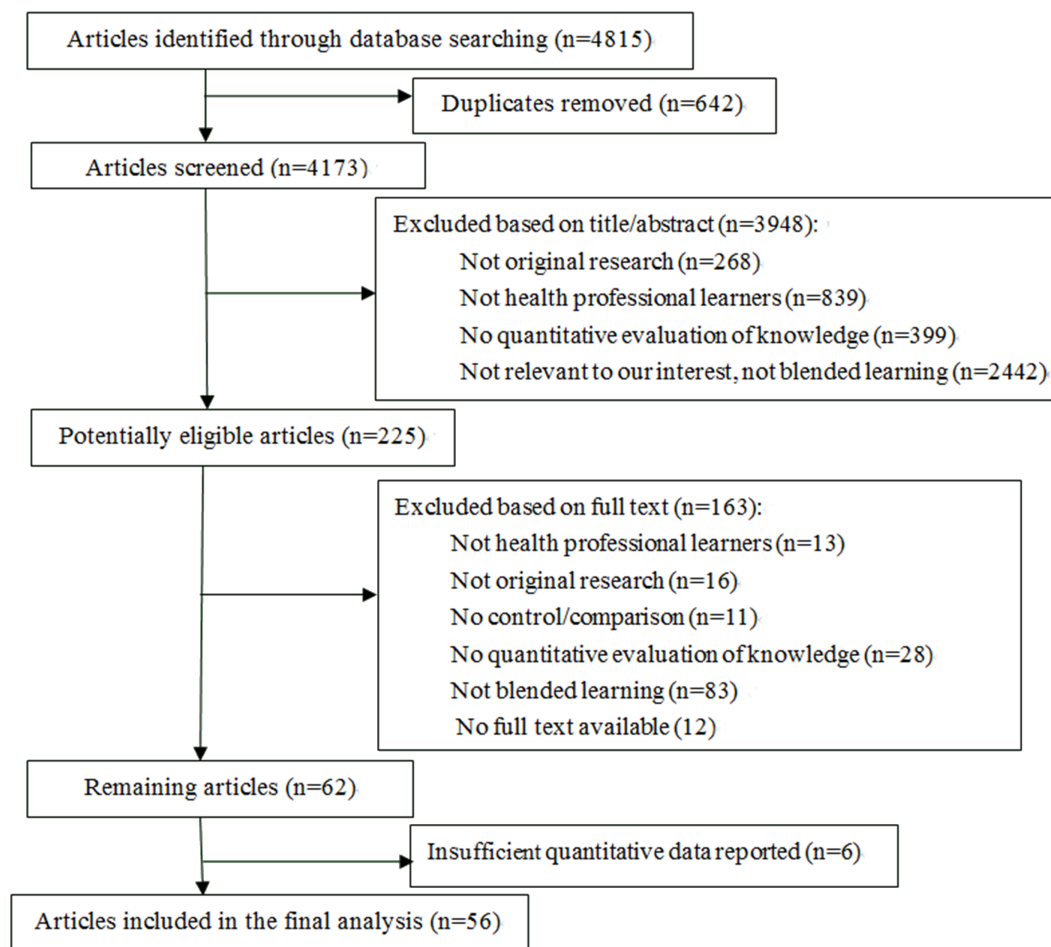
Study Characteristics

In the meta-analysis, we included 13 publications representing 20 interventions published from 2004-2014, which compared blended learning with no intervention and included 2238 health professional participants [22-24,56-65]. The number of participants ranged from 6 [61] to 817 [62], and the duration of the intervention ranged from 24 hours [63] to one semester [58].

We included 44 publications representing 56 interventions comparing blended learning with nonblended learning published from 1991 to 2014 that covered 6110 health profession participants [16,19-21,25,26,28,29,63,66-100]. There was 1 pre-posttest one-group intervention, 27 posttest-only two-group interventions, and 28 pre-posttest two-group interventions. The number of participants ranged from 14 [72] to 609 [84], and the duration ranged from 1 hour [101] to 1 year [77].

Components or features of the study intervention were mostly "Web-based+ face-to-face", "e-learning+ class session", and "Web-based online instruction+ off-line instruction (review of the core contents on the online program, case analysis, small group discussion, and miscellaneous activities)". "Modality or technology" varied, such as "Moodle, on-site workshops", "asynchronous discussion forums, a live audio and text-based online synchronous session (Centra); online modules (Macromedia Breeze)". More than 80% of the interventions were measured using objective assessment, which included multiple choice questions, true or false questions, matching questions, and essays. For most studies, there was no delay between the end of the intervention and the posttest. Table 1

summarizes the key features and e-Table 3 in Multimedia Appendix 1 describes the detailed information.

Table 1. Summary description of included studies.

Study characteristics	No intervention comparison		Nonblended learning comparison	
	Interventions, n (%) (N=20)	Participants, n (N=2238)	Interventions, n (%) (N=56)	Participants, n (N=6110)
Study design				
Pre-posttest 1-group	17 (85.0)	1656	27 (48.2)	97
Posttest 2-group	2 (10.0)	130	28 (50.0)	3468
Pre-posttest 2-group	1 (5.0)	452	1 (1.8)	2545
RCT/NRS				
RCT	2 (10.0)	130	31 (55.4)	2919
NRS	18 (90.0)	2108	25 (44.6)	3191
Country				
Developed	14 (70.0)	1673	44 (78.6)	4489
Developing	6 (30.0)	565	12 (21.4)	1621
Participant				
Medical students	9 (45.0)	887	37 (66.1)	4593
Nursing students	1 (5.0)	69	9 (16.1)	870
Nurses	2 (10.0)	103	5 (8.9)	259
Physicians	6 (30.0)	137	2 (3.6)	256
Public health workers	1 (5.0)	817	1 (1.8)	66
Others	1 (5.0)	225	1 (1.8)	66
Intervention duration				
1 semester	17 (85.0)	2038	43 (76.8)	4578
≥1 semester	3 (15.0)	200	13 (23.2)	1532
Exercises				
Present	15 (75.0)	1273	41 (73.2)	4526
Absent	5 (25.0)	965	15 (26.8)	1584
Interactivity				
High	15 (75.0)	1559	35 (62.5)	4460
Low	5 (25.0)	679	21 (37.5)	1650
Peer discussion				
Present	10 (50.0)	1456	28 (50.0)	3369
Absent	10 (50.0)	782	28 (50.0)	2741
Outcome assessment				
Objective	16 (80.0)	1833	53 (93.6)	5832
Subjective	4 (20.0)	405	3 (6.4)	278
Comparison intervention				
E-learning	NA	NA	5 (8.9)	205
Traditional learning	NA	NA	51 (91.1)	5905
Conflict of interest				
Yes	0	0	2 (3.6)	612
No	20 (100.0)	2238	54 (96.4)	5498
Quality score				
≥4	5 (25.0)	730	47 (83.9)	4965

Study characteristics	No intervention comparison		Nonblended learning comparison	
	Interventions, n (%) (N=20)	Participants, n (N=2238)	Interventions, n (%) (N=56)	Participants, n (N=6110)
4	15 (75.0)	1508	9 (16.1)	1145

Study Quality

All of the intervention groups in the included studies were representative of average learners. Ten percent (2/20) of no-intervention controlled studies and 98% (55/56) of nonblended learning controlled studies selected the control group from the same community as the experimental group. Nearly a third (30%, 6/20) of the no-intervention controlled studies and 46% (26/56) of nonblended learning controlled studies reported blinded outcome assessment. All of the no-intervention controlled studies (100%) and 96% (54/56) of nonblended learning controlled studies reported completeness of follow-up. The mean (SD) quality score was 3.40 (0.82) for no-intervention controlled studies, and 4.45 (0.78) for nonblended learning controlled studies. The results of the quality assessment are shown in e-Table 4 in Multimedia Appendix 1.

Figure 2. Forest plot of blended learning versus no intervention.

Quantitative Data Synthesis

Comparisons With No Intervention

As effect sizes larger than 0.8 were considered to be large [102], the pooled effect size (SMD 1.40; 95% CI 1.04-1.77; Z=7.52, P<.001) suggests a significantly large effect. However, significant heterogeneity was observed among studies (P<.001, I^2=94.8%, 95% CI 93.1-96.0), and individual effect sizes ranged from -0.12 to 4.24. Figure 2 shows detailed results of the meta-analysis. The test of funnel plots (Figure 3) indicated no significant publication bias among studies (Begg's test P=.587). Based on risk of bias and large effect, we graded the quality of evidence as moderate. E-Table 5 in Multimedia Appendix 1 provides the GRADE evidence profile. E-Table 6 in Multimedia Appendix 1 contains the mean, standard difference, and number of participants for both blended learning and no intervention/nonblended learning.

Study ID		SMD (95% CI)	% Weight
Flys, 2012		1.37 (1.17, 1.58)	5.66
Purl, 2010		1.87 (1.62, 2.12)	5.61
Karaksha, 2011		0.95 (0.29, 1.62)	4.84
Buchowski, 2002		4.24 (3.67, 4.81)	5.06
Wallen, 2010		1.11 (0.78, 1.44)	5.50
Weaver, 2014(a)		0.20 (-0.16, 0.56)	5.46
Weaver, 2014(b)		-0.12 (-0.48, 0.24)	5.46
Riesen, 2012		0.34 (-0.02, 0.70)	5.45
Cho,2014		0.95 (0.51, 1.38)	5.32
Cho, 2014		2.14 (1.72, 2.56)	5.35
Pereira, 2008(a)		1.48 (0.63, 2.33)	4.39
Pereira, 2008(b)		3.09 (2.03, 4.16)	3.87
ZKaramizadeh, 2015(a)		1.68 (0.28, 3.08)	3.13
ZKaramizadeh, 2014(b)		0.91 (0.44, 1.39)	5.25
ZKaramizadeh, 2013(c)		2.13 (1.32, 2.94)	4.48
ZKaramizadeh, 2012(d)		2.48 (1.89, 3.07)	5.01
ZKaramizadeh, 2011(e)		1.11 (0.16, 2.07)	4.13
Chandler, 2008		1.72 (1.60, 1.83)	5.73
Baumlin, 2006		0.34 (-0.08, 0.76)	5.38
Cragun, 2005		0.88 (0.26, 1.50)	4.94
Overall (I-squared = 94.8%, p = 0.000)		1.40 (1.04, 1.77)	100.00

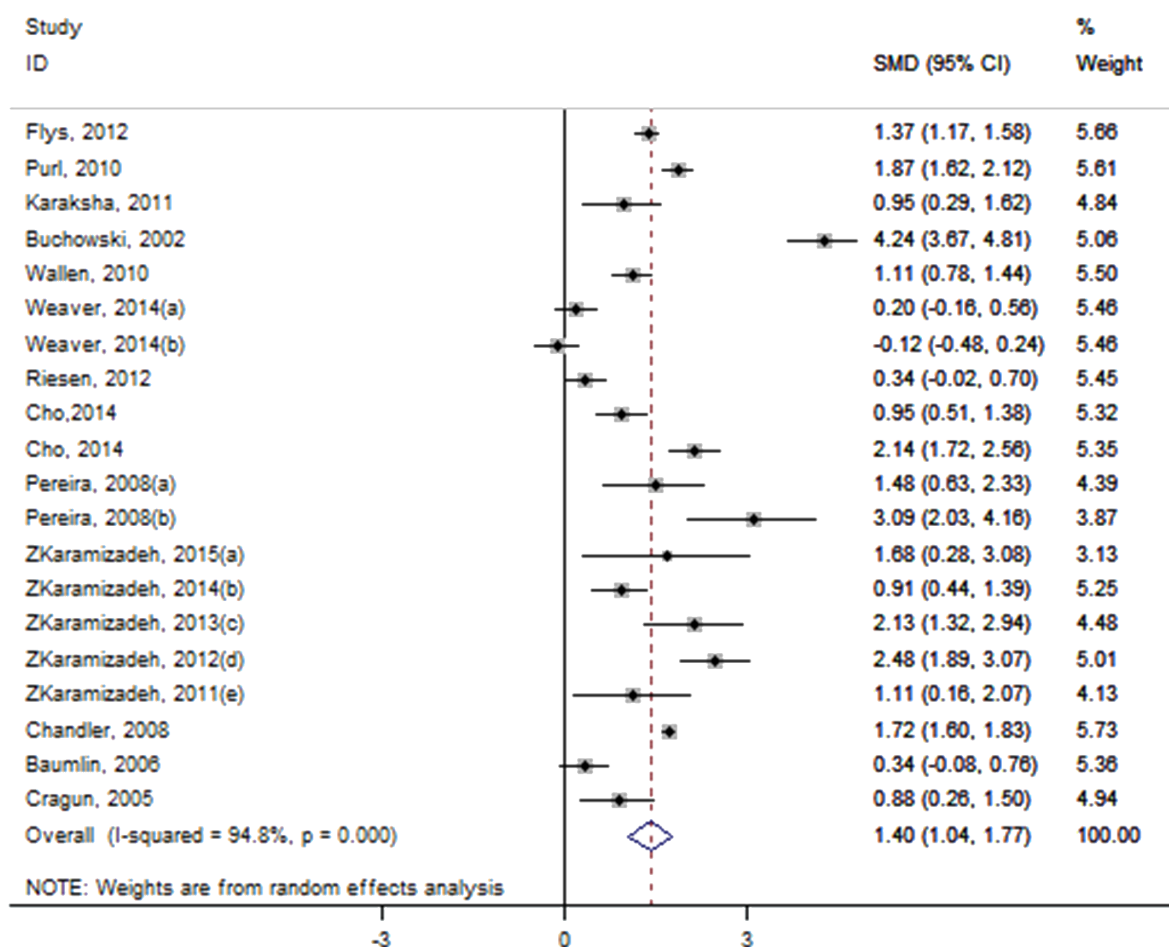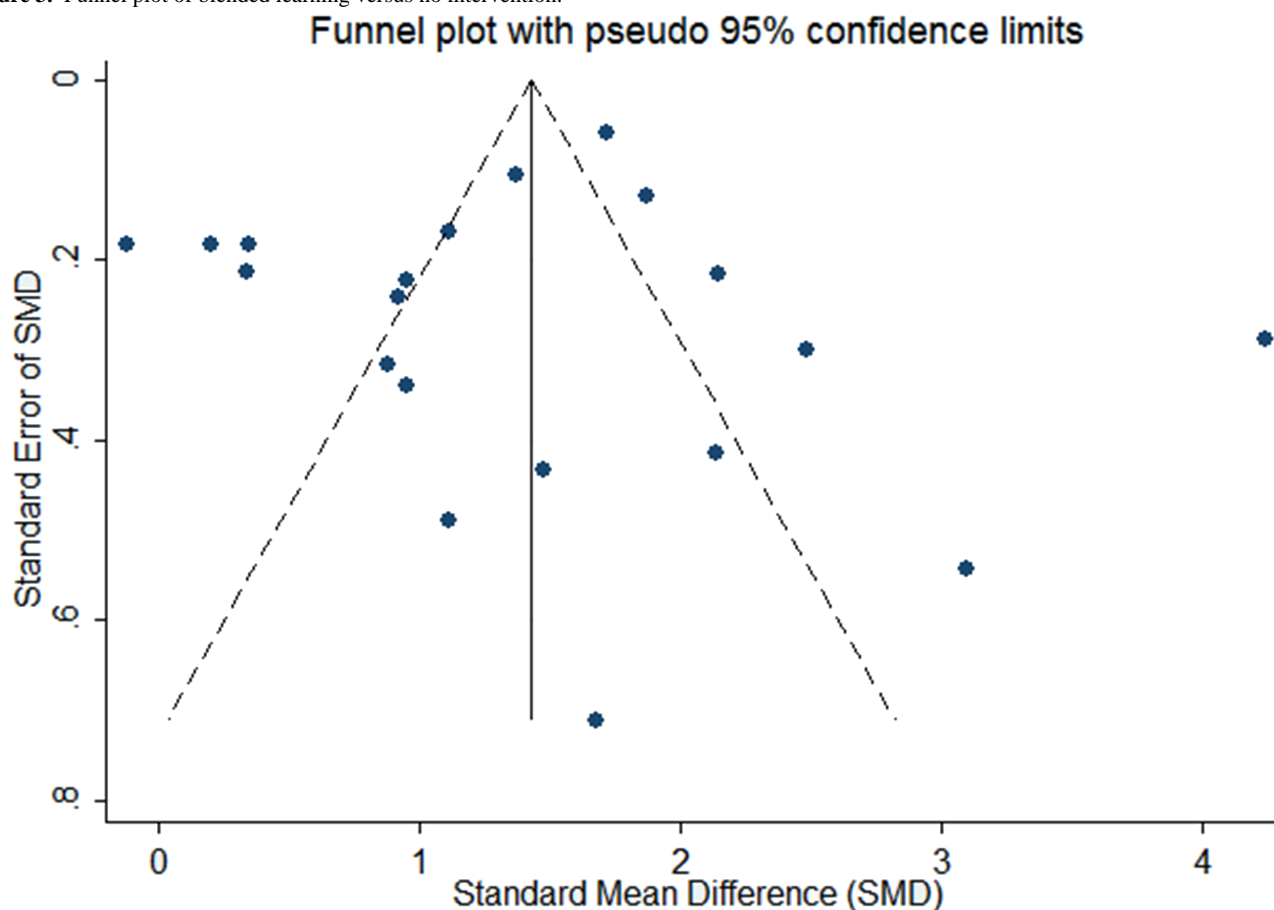NOTE: Weights are from random effects analysis

Figure 3. Funnel plot of blended learning versus no intervention.

Meta-Regression and Subgroup Analysis

We investigated a multiple regression model with each possible source of heterogeneity (I^2_res=85.33%, adjusted R^2=48.89%; I^2_res means residual variation due to heterogeneity) and found that the outcome assessment (P=.03) was a potential source of heterogeneity (Table 2). Studies with objective outcome assessments had larger pooled effect sizes. Furthermore, subgroup analyses were performed to evaluate the sources of heterogeneity. A statistically significant interaction favoring pre-posttest two-groups designs and pre-posttest one-group designs was found (P for interaction<.001), which was consistent with the result of the meta-regression. Statistical differences existed between the groups of participants (P for interaction<.001). Nonrandomized studies had larger effects than randomized ones (P for interaction=.01). The effect size was significantly larger for blended learning with objective assessment than with subjective assessment (P for interaction=.005). However, we did not find support for the hypotheses regarding subgroup interactions across levels of exercises (P for interaction=.92).

Sensitivity Analyses

Exclusion of any single study did not change the overall result, which ranged from 1.24 (95% CI 0.91-1.57) to 1.48 (95% CI 1.14-1.83).

Comparisons With Nonblended Learning

The pooled effect size (SMD 0.81; 95% CI 0.57-1.05; Z=6.59, P<.001) significantly reflected a large effect, and significant heterogeneity was observed among studies (P<.001, I^2=94.6%, 95% CI 93.7-95.5). Figure 4 shows detailed results of the main analysis. The test of asymmetry funnel plot (Figure 5) indicated publication bias among studies (Begg's test P=.01). The publication bias may have been towards larger studies with generally large magnitudes of effects. The trim and fill method indicated that the effect size changed to 0.26 (95% CI -0.01 to 0.54) after adjusting for publication bias, which suggested that blended learning was at least as effective as nonblended learning. Based on risk of bias, publication bias, and large effect, we graded the quality of evidence as low. E-Table 5 in Multimedia Appendix 1 provides the GRADE evidence profile.

Table 2. Subgroup analysis of blended learning versus no intervention.

Subgroup	Interventions, n	Pooled effect sizes (95% CI)	Heterogeneity (I^2), P	Interaction, P[a]	Meta-regression	
					Coef.	P
All interaction	20	1.40 (1.04-1.77)	94.8% (93.1-96.0), P<.001			
Study design						
Posttest 2-groups	2	0.59 (0.00-1.18)	57.0%, P=.13			
Pre-posttest 1 group	17	1.47 (1.05-1.88)	95.0% (93.3-96.3), P<.001	<.001	.27	.81
Pre-posttest 2-groups	1	1.87 (1.62-2.13)	0			
Country						
Developed	14	1.29 (0.83-1.75)	96.0% (94.6-97.1), P<.001	.23	-.22	.90
Developing	6	1.71 (1.20-2.22)	76.5% (47.4-89.5), P=.001			
Participant						
Medical students	9	1.13 (0.32-1.94)	96.8% (95.4-97.8), P<.001			
Nursing students	1	2.14 (1.72-2.56)	0			
Nurses	2	1.05 (0.79-1.91)	0.0%, P=.56	<.001	.05	.82
Physicians	6	1.84 (1.14-2.54)	81.2% (59.7-91.2), P<.001			
Public health workers	1	1.72 (1.60-1.83)	0			
Others	1	1.37 (1.17-1.58)	0			
Intervention duration						
1 semester	17	1.39 (1.10-1.18)	89.2% (84.2-92.6), P<.001	.97	-.33	.69
≥1 semester	3	1.43 (-0.82-3.68)	98.9% (98.1-99.3), P<.001			
Randomization						
Randomized	2	0.59 (.001-1.64)	57.0%, P=.013	.01	.67	.45
Nonrandomized	18	1.49 (1.11-1.87)	94.9% (93.2-96.2), P<.001			
Quality score						
≥4	5	1.89 (1.13-2.66)	96.2% (93.4-97.8), P<.001	.63	-1.05	.29
4	15	1.23 (.77-1.69)	94.3% (92.1-95.9), P<.001			
Exercises						
Present	10	1.28 (0.64-1.90)	95.1% (93.2-96.4), P<.001	.92	-.21	.75
Absent	10	1.53 (1.08-1.99)	89.5% (88.7-96.7), P<.001			
Interactivity						
High	15	1.54 (1.07-2.00)	95.6% (94.0-96.7), P<.001	.20	-1.25	.41
low	5	1.05 (0.44-1.65)	90.9% (81.7-95.5), P<.001			
Peer discussion						
Present	10	1.25 (0.70-1.79)	96.2% (94.2-97.2), P<.001	.11	-.07	.97
Absent	10	1.87 (1.21-2.53)	93.1% (88.6-95.3), P<.001			
Outcome assessment						
Objective	16	1.66 (1.29-2.04)	91.9% (88.4-94.3), P<.001	.005	-2.02	.03
Subjective	4	0.46 (-0.30-1.22)	95.8% (92.1-97.8), P<.001			
Funding from company						
Yes	2	2.29 (-1.53 to 6.11)	99.2%, P<.001	.61	-.93	.37
No	18	1.30 (.97-1.62)	92.7% (88.9-94.7), P<.001			

[a]*P* for interaction means the *P* of heterogeneity between groups.

Figure 4. Forest plot of blended learning versus non-blended learning.
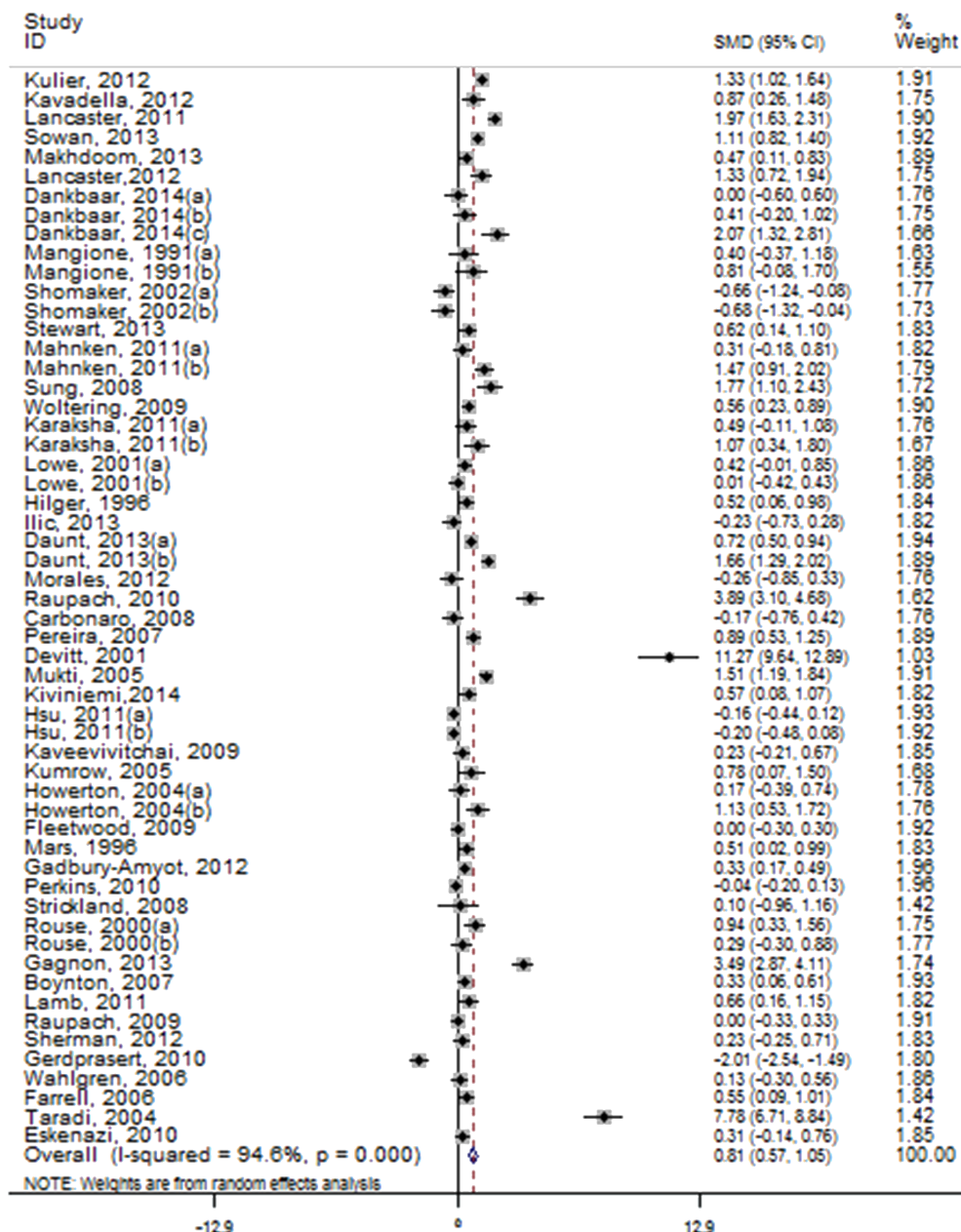
Meta-Regression and Subgroup Analysis

A multiple regression model for each possible source of heterogeneity was conducted (I^2_res=94.59%, adjusted R^2=-26.38%), and no significant source of heterogeneity was found (Table 3). Furthermore, subgroup analyses were performed to evaluate the sources of heterogeneity. We found both pre-posttest two-group studies and pre-posttest one-group studies showed larger effects than posttest-only studies (P for interaction<.001). It was shown that the presence of exercises could yield a larger SMD (P for interaction=.49). Studies with objective assessments yielded a larger effect than studies with subjective assessments (P for interaction=.01). Studies without conflicts of interest yielded a larger effect than those with conflicts of interest (P for interaction<.001). However, high interactivity and presence of peer discussion did not yield larger effect sizes (P for interaction>.85).

Figure 5. Funnel plot of blended learning versus non-blended learning.
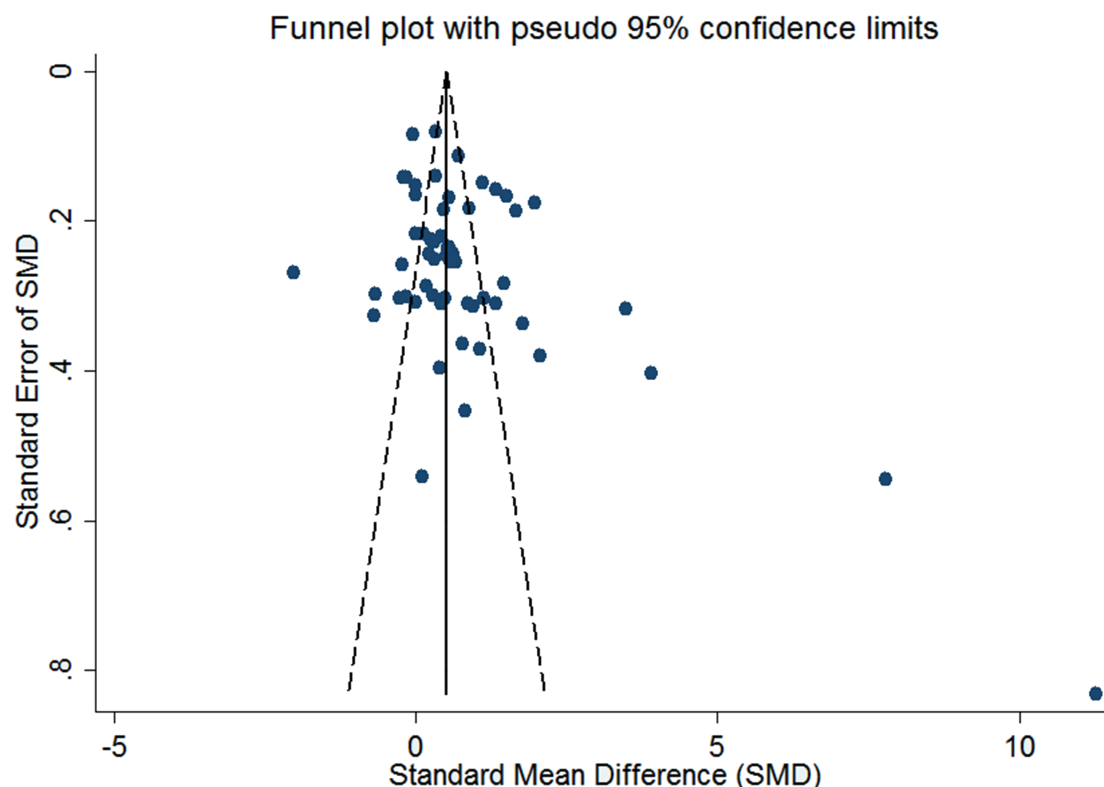
Funnel plot with pseudo 95% confidence limits

Sensitivity Analyses

Exclusion of any single study did not change the overall result, which ranged from 0.70 (95% CI 0.48-0.92) to 0.86 (95% CI 0.63-1.10).

Discussion

Principal Findings

This meta-analysis shows that blended learning has a large consistent positive effect (SMD 1.40, 95% CI 1.04-1.77) on knowledge acquisition compared with no intervention, which suggested that blended learning was very effective and educationally beneficial in health professions. Moreover, we also found that blended learning had a large effect (SMD 0.81, 95% CI 0.57-1.05) in comparison with the nonblended learning group. This means that blended learning may be more effective than nonblended learning, including both traditional face-to-face learning and pure e-learning. Possible explanations could be as follows: (1) compared with traditional learning, blended learning allows students to review electronic materials as often as necessary and at their own pace, which likely enhances learning performance [8,16], and (2) compared with e-learning, blended learning learners are less likely to experience feelings of isolation or reduced interest in the subject matter [8,11,103]. However, publication bias was found in the nonblended learning comparison group, and the trim and fill method showed that the pooled effect size changed to 0.26 (-0.01 to 0.54), which means blended learning is at least as effective as nonblended learning.

To the best of our knowledge, this may be the first meta-analysis to reveal the effectiveness of blended learning for knowledge acquisition in health professions, which includes all those directly related to human and animal health.

However, large heterogeneity was found across studies in both no-intervention and nonblended comparisons, and the subgroup comparisons partially explained these differences. The heterogeneity may be due to variations in study design, outcome assessment, exercises, conflict of interest, randomization, and type of participants. We found that effect sizes were significantly higher for studies using pre-posttest designs than posttest-only designs, which suggested that the former improved learning outcomes relative to the latter. As pretests may inform instructors about the knowledge learners have acquired before the course, which is considered to be one of the most important factors influencing education [104], they allow instructors to determine learning objectives and to prepare course materials accordingly [105]. Therefore, it is necessary for educators to administer pretests to learners to prepare well for courses. We also found that studies with objective assessments yielded a larger effect than those with subjective assessments. In contrast, Cook et al reported no difference between objective and subjective assessments in knowledge scores [33]. This is probably due to differences in personality traits of learners, as people with greater confidence tend to give higher ratings on subjective assessments than people who are less confident [106]. Thus, educators should objectively assess learners instead of using subjective evaluations.

Table 3. Subgroup analysis of blended learning versus nonblended learning.

Subgroup	Interventions, n	Pooled effect sizes (95% CI)	Heterogeneity (I^2), P	Interaction, P	Meta-regression	
					Coef.	P
All interventions	56	0.81 (0.57-1.05)	94.6% (93.7-95.5), P<.001			
Study design						
Posttest 2-groups	27	0.70 (0.32-1.07)	94.0% (92.3-95.3), P<.001	<.001		
Pre-posttest 2-groups	28	.89 (0.58-1.19)	94.5% (93.0-95.6), P<.001		-.001	.99
Pre-posttest 1-group	1	1.97 (1.63-2.32)	0			
Country						
Developed	44	0.80 (0.54-1.01)	93.2% (91.7-94.4), P<.001	.83	.13	.86
Developing	12	0.87 (0.22-1.53)	97.2% (96.2-97.9), P<.001			
Participant						
Medical students	38	0.88 (0.60-1.17)	94.8% (93.6-95.7), P<.001			
Nursing students	9	0.42 (-0.32-1.16)	96.0% (94.0-97.3), P<.001			
Nurses	5	0.87 (0.09-1.65)	87.7% (73.8-94.2), P<.001	.03	-.17	.61
Physicians	2	1.33 (1.05-1.60)	0.0%, P=.996			
Public health workers	1	0.57 (0.08-1.07)	0			
Others	1	0.66 (0.16-1.15)	0			
Intervention duration						
1 semester	43	0.73 (0.45-1.00)	94.5% (93.3-95.5), P<.001	.17	-.29	.68
≥1 semester	13	1.10 (0.63-1.59)	93.9% (91.3-95.8), P<.001			
Randomization						
Randomized	31	0.75 (0.38-1.12)	95.1% (94.0-96.1), P<.001	.63	.29	.69
Nonrandomized	25	0.87 (0.56-1.05)	94.1% (92.3-95.4), P<.001			
Quality score						
≥4	47	0.82 (0.55-1.09)	94.9% (93.9-95.8), P<.001	.99	-.27	.78
4	9	0.83 (0.39-1.26)	90.4% (84.1-94.2), P<.001			
Exercises						
Present	41	0.93 (0.63-1.25)	95.7% (94.9-96.4), P<.001	.49	-.51	.51
Absent	15	0.53 (0.26-0.80)	82.5% (72.2-88.9), P=0.011			
Interactivity						
High	37	0.84 (0.55-1.13)	95.2% (94.2-96.1), P<.001	.85	.48	.60
Low	19	0.78 (0.35-1.23)	93.4% (91.2-95.1), P<.001			
Peer discussion						
Present	28	0.82 (0.46-1.18)	95.9% (94.9-96.7), P<.001	.93	-.43	.96
Absent	28	0.80 (0.48-1.12)	92.7% (90.6-94.4), P<.001			
Outcome assessment						
Objective	53	0.85 (0.61-1.10)	94.8% (93.8-95.6), P<.001	.01	-.91	.47
Subjective	3	0.07 (-0.46 to 0.60)	68.6% (0-90.9), P=.04			
Comparison intervention						
E-learning	5	0.40 (-0.21-1.01)	77.5% (34.8-87.8), P=.23	.17	.69	.52
Traditional learning	51	0.85 (0.60-1.11)	95.0% (94.1-95.8), P<.001			
Conflict of interest						

Subgroup	Interventions, n	Pooled effect sizes (95% CI)	Heterogeneity (I^2), P	Interaction, P	Meta-regression	
					Coef.	P
Yes	2	-0.06 (-0.21 to 0.10)	0.0%	<.001	1.17	.44
No	54	0.85 (0.60-1.10)	94.5% (93.5-95.4), P<.001			

Additionally, effect size was found to be significantly larger for blended courses with exercises versus no exercises, which was consistent with the results of a previous study conducted by Cook et al in 2006, which found that continuity clinics had higher test scores when using a question format compared to a standard format [37]. Thus, it is necessary for educators to include exercises in their teaching, such as cases and self-assessment questions. However, we failed to confirm our hypothesis that presence of peer discussion and high interactivity would yield larger effect sizes. Although we found statistical differences between the RCTs and NRS in the no-intervention comparison, it could probably be due to chance as there were only two RCTs (130 participants) included. Differences between studies with conflicts of interest and those without conflicts of interest in nonblended comparisons could be also due to chance, as only two studies with conflicts of interest (612 participants) were included. The remainder of the high heterogeneity may arise from other characteristics, such as individual learning styles, study intervention, assessment instrument, and ongoing access to learning materials [33,107,108], for which detailed information was not available in the included studies. As Wong et al cited in their review, different modes of course delivery suit different learners in different environments [109].

Our samples consisted of various health professional learners (nurses, medical students, nursing students, physicians, public health workers, and other health professionals) across a wide variety of health care disciplines, such as medicine, nursing, ethics, health policy, pharmacy, radiology, genetics, histology, and emergency preparedness. Moreover, we found medium or large effects for the pooled effect sizes of almost all subgroup analyses exploring variations in study design, participant type, randomization, quality scores, exercises, interactivity, and peer discussion. Thus, our results suggest that health care educators should use blended learning as a teaching component in various disciplines and course settings.

Strengths and Limitations

Our meta-analysis also has several strengths. Evaluations of the effectiveness of blended learning for health professions are timely and very important for both medical educators and learners. We intentionally kept our scope broad in terms of subjects and included all studies with learners from health professions. We searched for relevant studies in manifold research databases up to September 2014. The systematic literature search encompassed multiple databases and had few exclusion criteria. We also conducted all aspects of the review process in duplicate.

However, there are limitations to consider. First, although we searched gray literature in two databases (CENTRAL and ERIC), gray literature indexed by other databases may have been missed, which could be the reason for the observed

publication bias. Second, the quality of meta-analyses is dependent on the quality of data from the included studies. Although the standard deviation of eight interventions was not available due to poor reporting, we used the average standard deviation of other included studies and imputed effect sizes with concomitant potential for error. Third, despite conducting the review and extraction independently and in duplicate, the process was subjective and dependent on the descriptions of the included articles instead of direct evaluation of interventions. Fourth, although the modified Newcastle–Ottawa scale is a useful and reliable tool for appraising methodological quality of medical education research and enhances flexibility for different study designs, it increases the risk of reviewer error or bias due to a certain amount of rater subjectivity. Then, results of subgroup analyses should be interpreted with caution because of the absence of a priori hypotheses in some cases, such as study design, country socioeconomic status, and outcome assessment. Moreover, although the subgroup analyses showed the variability of participant types, socioeconomic status of country, intervention duration, interactivity, peer discussion, and study design of RCT or NRS did not make a difference in the overall results, the large clinical heterogeneity and inconsistent magnitude of effects across studies makes it difficult to generalize the conclusions. In addition, as variability of study interventions, assessment instruments, circumstances and so on, which were not assessed, could be potential sources of heterogeneity, the results of both meta-analyses should be treated with caution. Furthermore, publication bias was found in the meta-analysis with the nonblended comparison. Although we used the trim and fill method for adjustment, the results should be treated with caution.

Implications

Our study has implications for both research on blended learning and education in health professions. Despite the fact that conclusions could be weakened by heterogeneity across studies, the results of our quantitative synthesis demonstrated that blended learning may have a positive effect on knowledge acquisition across a wide range of learners and disciplines directly related to health professions. In summary, blended learning could be promising and worthwhile for further application in health professions. The difference in effects across subgroup analyses indicates that different methods of conducting blended courses may demonstrate differing effectiveness. Therefore, researchers and educators should pay attention to how to implement a blended course effectively. This question could be answered successfully through studies directly comparing different blended instructional methods. Thus, such studies are of critical importance.

Studies comparing blended learning with no intervention suggested that blended learning in health professions might be

invariably effective. However, although observational studies yielded a large effect size, the quality of evidence was lower due to their inherent study design limitations. Additionally, owing to the small number of RCTs, the meta-analysis did not meet the optimal size (imprecision) and therefore, quality of evidence was ranked lower. Thus, despite the consistency of effect and no significant reporting bias, the evidence of the no-intervention comparison was of moderate quality, which means further research is likely to have an impact on our confidence in the estimate of effect and may change the estimate, and RCTs with large samples may modify the estimates. Thus, there is still great value in further research comparing blended learning with no intervention, and RCTs with large samples may modify the estimates. For nonblended comparisons, pooled estimates showed that blended learning is more effective than or at least as effective as pure e-learning and pure traditional learning. However, due to publication bias towards larger studies with generally large magnitudes of effects, the evidence was of low quality, which means further research is very likely to change our estimate. Furthermore, only four studies using e-learning were included. Therefore, the effect of blended learning especially in comparison with e-learning should be evaluated in future research, and studies with small magnitudes of effect should merit publication.

Conclusions

Blended learning appears to have a consistent positive effect in comparison with no intervention and appears to be more effective than or at least as effective as nonblended instruction for knowledge acquisition in health professions. Moreover, pre-posttest study design, presence of exercises, and objective outcome assessment in blended courses could improve health care learners' knowledge acquisition. Due to the large heterogeneity, the conclusion should be treated with caution.

Authors' Contributions

WRY conceptualized and designed the study. QL and FZ performed the review, extraction, and data analysis. QL prepared the first draft of the paper. WRY, WJP, RH, YXL, and FZ contributed to the revision of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

E-tables 1-7.

[PDF File (Adobe PDF File), 228KB - jmir_v18i1e2_app1.pdf]

References

1.   Bediang G, Stoll B, Geissbuhler A, Klohn AM, Stuckelberger A, Nko'o S, et al. Computer literacy and E-learning perception in Cameroon: the case of Yaounde Faculty of Medicine and Biomedical Sciences. BMC Med Educ 2013;13:57 [FREE Full text] [doi: 10.1186/1472-6920-13-57] [Medline: 23601853]

2.   Sun P, Tsai RJ, Finger G, Chen Y, Yeh D. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education 2008 May;50(4):1183-1202. [doi: 10.1016/j.compedu.2006.11.007]

3.   Choules AP. The use of elearning in medical education: a review of the current situation. Postgrad Med J 2007 Apr;83(978):212-216 [FREE Full text] [doi: 10.1136/pgmj.2006.054189] [Medline: 17403945]

4.   Berners-Lee T, Cailliau R, Luotonen A, Nielsen HF, Secret A. The World-Wide Web. Communications of the ACM 1994 Aug;37(8):76-82 [FREE Full text]

5.   Cook DA. Web-based learning: pros, cons and controversies. Clinical Medicine 2007 Jan 01;7(1):37-42. [doi: 10.7861/clinmedicine.7-1-37]

6.   Yu S, Yang K. Attitudes toward Web-based distance learning among public health nurses in Taiwan: a questionnaire survey. Int J Nurs Stud 2006 Aug;43(6):767-774. [doi: 10.1016/j.ijnurstu.2005.09.005] [Medline: 16253261]

7.   Peng Y, Wu X, Atkins S, Zwarentein M, Zhu M, Zhan XX, et al. Internet-based health education in China: a content analysis of websites. BMC Med Educ 2014;14:16 [FREE Full text] [doi: 10.1186/1472-6920-14-16] [Medline: 24467710]

8.   Wu J, Tennyson RD, Hsia T. A study of student satisfaction in a blended e-learning system environment. Computers & Education 2010 Aug;55(1):155-164. [doi: 10.1016/j.compedu.2009.12.012]

9.   Moreira IC, Ventura SR, Ramos I, Rodrigues PP. Development and assessment of an e-learning course on breast imaging for radiographers: a stratified randomized controlled trial. J Med Internet Res 2015;17(1):e3 [FREE Full text] [doi: 10.2196/jmir.3344] [Medline: 25560547]

10. Wu J, Tennyson RD, Hsia T, Liao Y. Analysis of E-learning innovation and core capability using a hypercube model. Computers in Human Behavior 2008 Sep;24(5):1851-1866. [doi: 10.1016/j.chb.2008.02.008]

11. Hara N. Student distress in a Web-based distance education course. Information, Communication & Society 2000 Jan;3(4):557-579. [doi: 10.1080/13691180010002297]

12. Kemp N, Grieve R. Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. Front Psychol 2014;5:1278 [FREE Full text] [doi: 10.3389/fpsyg.2014.01278] [Medline: 25429276]

13. Conole G, de Laat M, Dillon T, Darby J. 'Disruptive technologies', 'pedagogical innovation': What's new? Findings from an in-depth study of students' use and perception of technology. Computers & Education 2008 Feb;50(2):511-524. [doi: 10.1016/j.compedu.2007.09.009]

14. Bonk CJ, Graham CE. The handbook of blended learning: global perspectives, local designs. San Francisco, CA: Pfeiffer; 2006.

15. Thakore H, McMahon T. Virtually there: e-learning in medical education. Clinical Teacher 2006 Dec;3(4):225-228. [doi: 10.1111/j.1743-498X.2006.00114.x]

16. Makhdoom N, Khoshhal KI, Algaidi S, Heissam K, Zolaly MA. 'Blended learning' as an effective teaching and learning strategy in clinical medicine: a comparative cross-sectional university-based study. Journal of Taibah University Medical Sciences 2013 Apr;8(1):12-17. [doi: 10.1016/j.jtumed.2013.01.002]

17. Moore M. Emerging practice and research in blended learning. In: Handbook of distance education. New York: Routledge; 2012.

18. Norberg A, Dziuban CD, Moskal PD. A time‐based blended learning model. On the Horizon 2011 Aug 16;19(3):207-216. [doi: 10.1108/10748121111163913]

19. Mars M, McLean M. Students' perceptions of a multimedia computer-aided instruction resource in histology. S Afr Med J 1996 Sep;86(9):1098-1102. [Medline: 8888778]

20. Rouse DP. The effectiveness of computer-assisted instruction in teaching nursing students about congenital heart disease. Comput Nurs 2000;18(6):282-287. [Medline: 11105402]

21. Mangione S, Nieman LZ, Greenspon LW, Margulies H. A comparison of computer-assisted instruction and small-group teaching of cardiac auscultation to medical students. Med Educ 1991 Sep;25(5):389-395. [Medline: 1758315]

22. Cho K, Shin G. Operational effectiveness of blended e-learning program for nursing research ethics. Nurs Ethics 2013 Nov 19;21(4):484-495. [doi: 10.1177/0969733013505310] [Medline: 24258252]

23. Riesen E, Morley M, Clendinneng D, Ogilvie S, Ann MM. Improving interprofessional competence in undergraduate students using a novel blended learning approach. J Interprof Care 2012 Jul;26(4):312-318. [doi: 10.3109/13561820.2012.660286] [Medline: 22559163]

24. Wallen GR, Cusack G, Parada S, Miller-Davis C, Cartledge T, Yates J. Evaluating a hybrid web-based basic genetics course for health professionals. Nurse Educ Today 2011 Aug;31(6):638-642 [FREE Full text] [doi: 10.1016/j.nedt.2010.11.001] [Medline: 21106279]

25. Sung YH, Kwon IG, Ryu E. Blended learning on medication administration for new nurses: integration of e-learning and face-to-face instruction in the classroom. Nurse Educ Today 2008 Nov;28(8):943-952. [doi: 10.1016/j.nedt.2008.05.007] [Medline: 18599162]

26. Mukti MA, Razali D, Ramli MF, Zaman HB, Ahmad A. Hybrid learning and online collaborative enhance students performance. : IEEE; 2005 Presented at: 5th IEEE International Conference on Advanced Learning Technologies; July 5-8, 2005; Taiwan p. 481-483. [doi: 10.1109/ICALT.2005.166]

27. Seabra D, Srougi M, Baptista R, Nesrallah LJ, Ortiz V, Sigulem D. Computer aided learning versus standard lecture for undergraduate education in urology. J Urol 2004 Mar;171(3):1220-1222. [doi: 10.1097/01.ju.0000114303.17198.37] [Medline: 14767306]

28. Shomaker TS, Ricks DJ, Hale DC. A prospective, randomized controlled study of computer-assisted learning in parasitology. Acad Med 2002 May;77(5):446-449. [Medline: 12010707]

29. Fleetwood J, Vaught W, Feldman D, Gracely E, Kassutto Z, Novack D. MedEthEx Online: a computer-based learning program in medical ethics and communication skills. Teach Learn Med 2000;12(2):96-104. [doi: 10.1207/S15328015TLM1202_7] [Medline: 11228685]

30. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach 2012;34(4):e216-e221. [doi: 10.3109/0142159X.2012.642831] [Medline: 22455712]

31. McCutcheon K, Lohan M, Traynor M, Martin D. A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. J Adv Nurs 2015 Feb;71(2):255-270. [doi: 10.1111/jan.12509] [Medline: 25134985]

32. Jwayyed S, Stiffler KA, Wilber ST, Southern A, Weigand J, Bare R, et al. Technology-assisted education in graduate medical education: a review of the literature. Int J Emerg Med 2011;4:51 [FREE Full text] [doi: 10.1186/1865-1380-4-51] [Medline: 21824405]

33. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]

34. Lahti M, Välimäki M. Is computer assisted learning among nurses or nursing students more effective than traditional learning? - A mini-review. Stud Health Technol Inform 2009;146:842. [Medline: 19593011]

35. Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? JAMA 1999 Sep 1;282(9):867-874. [Medline: 10478694]

36. Mayer RE. The Cambridge handbook of multimedia learning. New York: Cambridge University Press; 2005.

37. Cook DA, Thompson WG, Thomas KG, Thomas MR, Pankratz VS. Impact of self-assessment questions and learning styles in Web-based learning: a randomized, controlled, crossover trial. Acad Med 2006 Mar;81(3):231-238. [Medline: 16501263]

38. Cook DA, McDonald FS. E-learning: is there anything special about the "E"? Perspect Biol Med 2008;51(1):5-21. [doi: 10.1353/pbm.2008.0007] [Medline: 18192762]

39. 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 health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]

40. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [Medline: 10789670]

41. Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: a systematic review. Med Educ 2011 Mar;45(3):227-238. [doi: 10.1111/j.1365-2923.2010.03890.x] [Medline: 21299598]

42. Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. Acad Med 2015 Aug;90(8):1067-1076. [doi: 10.1097/ACM.0000000000000786] [Medline: 26107881]

43. Wells G, Shea B, Petersen J, Welch V, Losos M. Department of Epidemiology and Community Medicine, University of Ottawa, Canada. 2015. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses URL: http://www.medicine.mcgill.ca/rtamblyn/Readings/ The%20Newcastle%20-%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses. pdf [accessed 2015-05-30] [WebCite Cache ID 6YueVvmaa]

44. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]

45. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011 Apr;64(4):395-400. [doi: 10.1016/j.jclinepi.2010.09.012] [Medline: 21194891]

46. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]

47. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 2011 Apr;64(4):407-415. [doi: 10.1016/j.jclinepi.2010.07.017] [Medline: 21247734]

48. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 2011 Dec;64(12):1277-1282. [doi: 10.1016/j.jclinepi.2011.01.011] [Medline: 21802904]

49. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011 Dec;64(12):1283-1293. [doi: 10.1016/j.jclinepi.2011.01.012] [Medline: 21839614]

50. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011 Dec;64(12):1294-1302. [doi: 10.1016/j.jclinepi.2011.03.017] [Medline: 21803546]

51. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 2011 Dec;64(12):1303-1310. [doi: 10.1016/j.jclinepi.2011.04.014] [Medline: 21802903]

52. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011 Dec;64(12):1311-1316. [doi: 10.1016/j.jclinepi.2011.06.004] [Medline: 21802902]

53. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]

54. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods 1996;1(2):170-177. [doi: 10.1037/1082-989X.1.2.170]

XSL•FO

RenderX

55. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

56. Flys T, González R, Sued O, Suarez CJ, Kestler E, Sosa N, et al. A novel educational strategy targeting health care workers in underserved communities in Central America to integrate HIV into primary medical care. PLoS One 2012;7(10):e46426 [FREE Full text] [doi: 10.1371/journal.pone.0046426] [Medline: 23115628]

57. Puri R, Bell C, Evers WD. Dietetics students' ability to choose appropriate communication and counseling methods is improved by teaching behavior-change strategies in computer-assisted instruction. J Am Diet Assoc 2010 Jun;110(6):892-897. [doi: 10.1016/j.jada.2010.03.022] [Medline: 20497778]

58. Buchowski MS, Plaisted C, Fort J, Zeisel SH. Computer-assisted teaching of nutritional anemias and diabetes to first-year medical students. Am J Clin Nutr 2002 Jan;75(1):154-161 [FREE Full text] [Medline: 11756073]

59. Weaver SB, Oji V, Ettienne E, Stolpe S, Maneno M. Hybrid e-learning approach to health policy. Currents in Pharmacy Teaching and Learning 2014 Mar;6(2):313-322. [doi: 10.1016/j.cptl.2013.11.013]

60. Pereira J, Palacios M, Collin T, Wedel R, Galloway L, Murray A, et al. The impact of a hybrid online and classroom-based course on palliative care competencies of family medicine residents. Palliat Med 2008 Dec;22(8):929-937. [doi: 10.1177/0269216308094561] [Medline: 18772211]

61. Karamizadeh Z, Zarifsanayei N, Faghihi AA, Mohammadi H, Habibi M. The study of effectiveness of blended learning approach for medical training courses. Iran Red Crescent Med J 2012 Jan;14(1):41-44 [FREE Full text] [Medline: 22737553]

62. Chandler T, Qureshi K, Gebbie KM, Morse SS. Teaching emergency preparedness to public health workers: use of blended learning in web-based training. Public Health Rep 2008;123(5):676-680 [FREE Full text] [Medline: 18828425]

63. Karaksha A, Grant G, Davey A, Anoopkumar-Dukie S. Development and evaluation of computer-assisted learning (CAL) teaching tools compared to the conventional didactic lecture in pharmacology education. In: Proceedings of EDULEARN11 Conference. 2011 Presented at: EDULEARN11 Conference; July 4-6, 2011; Spain p. 3580-3589.

64. Cragun DL, Couch SC, Prows CA, Warren NS, Christianson CA. A success of a genetics educational intervention for nursing and dietetic students: A model for incorporating genetics into nursing and allied health curricula. J Allied Health 2005;34(2):90-96. [Medline: 16032915]

65. Baumlin KM, Bessette MJ, Lewis C, Richardson LD. EMCyberSchool: an evaluation of computer-assisted instruction on the Internet. Acad Emerg Med 2000 Aug;7(8):959-962 [FREE Full text] [Medline: 10958144]

66. Mahnken AH, Baumann M, Meister M, Schmitt V, Fischer MR. Blended learning in radiology: is self-determined learning really more effective? Eur J Radiol 2011 Jun;78(3):384-387. [doi: 10.1016/j.ejrad.2010.12.059] [Medline: 21288674]

67. Raupach T, Münscher C, Pukrop T, Anders S, Harendza S. Significant increase in factual knowledge with web-assisted problem-based learning as part of an undergraduate cardio-respiratory curriculum. Adv Health Sci Educ Theory Pract 2010 Aug;15(3):349-356 [FREE Full text] [doi: 10.1007/s10459-009-9201-3] [Medline: 19774475]

68. Devitt P, Smith JR, Palmer E. Improved student learning in ophthalmology with computer-aided instruction. Eye (Lond) 2001 Oct;15(Pt 5):635-639. [doi: 10.1038/eye.2001.199] [Medline: 11702976]

69. Hsu L, Hsieh S. Effects of a blended learning module on self-reported learning performances in baccalaureate nursing students. J Adv Nurs 2011 Nov;67(11):2435-2444. [doi: 10.1111/j.1365-2648.2011.05684.x] [Medline: 21592188]

70. Kaveevivitchai C, Chuengkriankrai B, Luecha Y, Thanooruk R, Panijpan B, Ruenwongsa P. Enhancing nursing students' skills in vital signs assessment by using multimedia computer-assisted learning with integrated content of anatomy and physiology. Nurse Educ Today 2009 Jan;29(1):65-72. [doi: 10.1016/j.nedt.2008.06.010] [Medline: 18703258]

71. Howerton WB, Enrique PRT, Ludlow JB, Tyndall DA. Interactive computer-assisted instruction vs. lecture format in dental education. J Dent Hyg 2004;78(4):10. [Medline: 16197750]

72. Strickland S. The effectiveness of blended learning environments for the delivery of respiratory care education. J Allied Health 2009;38(1):E11-E16. [Medline: 19753407]

73. Llambí L, Esteves E, Martinez E, Forster T, García S, Miranda N, et al. Teaching tobacco cessation skills to Uruguayan physicians using information and communication technologies. J Contin Educ Health Prof 2011;31(1):43-48. [doi: 10.1002/chp.20100] [Medline: 21425359]

74. Perkins GD, Fullerton JN, Davis-Gomez N, Davies RP, Baldock C, Stevens H, et al. The effect of pre-course e-learning prior to advanced life support training: a randomised controlled trial. Resuscitation 2010 Jul;81(7):877-881. [doi: 10.1016/j.resuscitation.2010.03.019] [Medline: 20409627]

75. Lancaster JW, McQueeney ML, Van Amburgh JA. Online lecture delivery paired with in class problem-based learning … Does it enhance student learning? Currents in Pharmacy Teaching and Learning 2011 Jan;3(1):23-29. [doi: 10.1016/j.cptl.2010.10.008]

76. Sowan AK, Jenkins LS. Use of the seven principles of effective teaching to design and deliver an interactive hybrid nursing research course. Nurs Educ Perspect 2013;34(5):315-322. [Medline: 24245382]

77. Lancaster JW, Wong A, Roberts SJ. 'Tech' versus 'talk': a comparison study of two different lecture styles within a Master of Science nurse practitioner course. Nurse Educ Today 2012 Jul;32(5):e14-e18. [doi: 10.1016/j.nedt.2011.09.018] [Medline: 22071277]

78. Dankbaar MW, Storm DJ, Teeuwen IC, Schuit SE. A blended design in acute care training: similar learning results, less training costs compared with a traditional format. Perspect Med Educ 2014 Sep;3(4):289-299 [FREE Full text] [doi: 10.1007/s40037-014-0109-0] [Medline: 24477971]

79. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. Arch Dis Child Fetal Neonatal Ed 2013 Mar;98(2):F141-F144. [doi: 10.1136/archdischild-2011-301252] [Medline: 22684155]

80. Lowe CI, Wright JL, Bearn DR. Computer-aided Learning (CAL): an effective way to teach the Index of Orthodontic Treatment Need (IOTN)? J Orthod 2001 Dec;28(4):307-311. [doi: 10.1093/ortho/28.4.307] [Medline: 11709597]

81. Arroyo-Morales M, Cantarero-Villanueva I, Fernández-Lao C, Guirao-Piñeyro M, Castro-Martín E, Díaz-Rodríguez L. A blended learning approach to palpation and ultrasound imaging skills through supplementation of traditional classroom teaching with an e-learning package. Man Ther 2012 Oct;17(5):474-478. [doi: 10.1016/j.math.2012.04.002] [Medline: 22579034]

82. Kiviniemi MT. Effects of a blended learning approach on student outcomes in a graduate-level public health course. BMC Med Educ 2014;14:47 [FREE Full text] [doi: 10.1186/1472-6920-14-47] [Medline: 24612923]

83. Kumrow DE. Evidence-based strategies of graduate students to achieve success in a hybrid Web-based course. J Nurs Educ 2007 Mar;46(3):140-145. [Medline: 17396555]

84. Gadbury-Amyot CC, Singh AH, Overman PR. Teaching with technology: learning outcomes for a combined dental and dental hygiene online hybrid oral histology course. J Dent Educ 2013 Jun;77(6):732-743 [FREE Full text] [Medline: 23740910]

85. Gagnon M, Gagnon J, Desmartis M, Njoya M. The impact of blended teaching on knowledge, satisfaction, and self-directed learning in nursing undergraduates: a randomized, controlled trial. Nurs Educ Perspect 2013;34(6):377-382. [Medline: 24475598]

86. Boynton JR, Green TG, Johnson LA, Nainar SMH, Straffon LH. The virtual child: evaluation of an internet-based pediatric behavior management simulation. J Dent Educ 2007 Sep;71(9):1187-1193 [FREE Full text] [Medline: 17761625]

87. Kulier R, Gülmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training: a randomized trial. JAMA 2012 Dec 5;308(21):2218-2225. [doi: 10.1001/jama.2012.33640] [Medline: 23212499]

88. Kavadella A, Tsiklakis K, Vougiouklakis G, Lionarakis A. Evaluation of a blended learning course for teaching oral radiology to undergraduate dental students. Eur J Dent Educ 2012 Feb;16(1):e88-e95. [doi: 10.1111/j.1600-0579.2011.00680.x] [Medline: 22251359]

89. Woltering V, Herrler A, Spitzer K, Spreckelsen C. Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation. Adv Health Sci Educ Theory Pract 2009 Dec;14(5):725-738. [doi: 10.1007/s10459-009-9154-6] [Medline: 19184497]

90. Ilic D, Hart W, Fiddes P, Misso M, Villanueva E. Adopting a blended learning approach to teaching evidence based medicine: a mixed methods study. BMC Med Educ 2013;13:169 [FREE Full text] [doi: 10.1186/1472-6920-13-169] [Medline: 24341502]

91. Carbonaro M, King S, Taylor E, Satzinger F, Snart F, Drummond J. Integration of e-learning technologies in an interprofessional health science course. Med Teach 2008 Feb;30(1):25-33. [doi: 10.1080/01421590701753450] [Medline: 18278648]

92. Pereira JA, Pleguezuelos E, Merí A, Molina-Ros A, Molina-Tomás MC, Masdeu C. Effectiveness of using blended learning strategies for teaching and learning human anatomy. Med Educ 2007 Feb;41(2):189-195. [doi: 10.1111/j.1365-2929.2006.02672.x] [Medline: 17269953]

93. Raupach T, Muenscher C, Anders S, Steinbach R, Pukrop T, Hege I, et al. Web-based collaborative training of clinical reasoning: a randomized trial. Med Teach 2009 Sep;31(9):e431-e437. [Medline: 19811180]

94. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning: outcomes for staff development. J Nurses Staff Dev 2012 Jul;28(4):186-190. [doi: 10.1097/NND.0b013e31825dfb71] [Medline: 22821022]

95. Gerdprasert S, Pruksacheva T, Panijpan B, Ruenwongsa P. Development of a web-based learning medium on mechanism of labour for nursing students. Nurse Educ Today 2010 Jul;30(5):464-469. [doi: 10.1016/j.nedt.2009.10.007] [Medline: 19931950]

96. Wahlgren C, Edelbring S, Fors U, Hindbeck H, Ståhle M. Evaluation of an interactive case simulation system in dermatology and venereology for medical students. BMC Med Educ 2006;6:40 [FREE Full text] [doi: 10.1186/1472-6920-6-40] [Medline: 16907972]

97. Farrell MJ, Rose L. Use of mobile handheld computers in clinical nursing education. J Nurs Educ 2008 Jan;47(1):13-19. [Medline: 18232610]

98. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ 2005 Mar;29(1):35-39 [FREE Full text] [doi: 10.1152/advan.00026.2004] [Medline: 15718381]

99. de Sousa EE, de Arruda MM, Ferreira MJ. Oral health promotion through an online training program for medical students. J Dent Educ 2011 May;75(5):672-678 [FREE Full text] [Medline: 21546601]

100. Hilger AE, Hamrick HJ, Denny FW. Computer instruction in learning concepts of streptococcal pharyngitis. Arch Pediatr Adolesc Med 1996 Jun;150(6):629-631. [Medline: 8646314]

101. Teeraporn C. Computer-aided learning for medical chart review instructions. Afr J Pharm Pharmacol 2012 Jul 22;6(27):2061-2067. [doi: 10.5897/AJPP12.269]

102. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.

103. Maki RH, Maki WS, Patterson M, Whittaker PD. Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. Behav Res Methods Instrum Comput 2000 May;32(2):230-239. [Medline: 10875167]

104. Cruickshank DR, Jenkins DB, Metcalf KK. The act of teaching. Boston: McGraw-Hill; 2006.

105. Hartley J, Davies IK. Preinstructional Strategies: The Role of Pretests, Behavioral Objectives, Overviews and Advance Organizers. Review of Educational Research 1976 Jan 01;46(2):239-265. [doi: 10.3102/00346543046002239]

106. Wilson JD. Subjective assessment. In: Evaluation of human work: a practical ergonomics methodology. Washington: Taylor & Francis; 1995.

107. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med 2010 May;85(5):909-922. [doi: 10.1097/ACM.0b013e3181d6c319] [Medline: 20520049]

108. Manochehr NN. Computers in Higher Education Economics Review. 2006. The influence of learning styles on learners in e-learning environments: An empirical study URL: http://www.webcitation.org/6Yug8S5O3 [accessed 2015-05-30]

109. Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. BMC Med Educ 2010;10:12 [FREE Full text] [doi: 10.1186/1472-6920-10-12] [Medline: 20122253]

Abbreviations

GRADE: Grades of Recommendation, Assessment, Development, and Evaluation
PICOS: population, intervention, comparison, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SD: standard deviation
SMD: standardized mean difference

XSL•FO
RenderX

Original Paper

Health Advice from Internet Discussion Forums: How Bad Is Dangerous?

Jennifer Cole[1], MA; Chris Watkins[2], PhD; Dorothea Kleine[3], PhD

[1]H2B2, Department of Computer Science, Royal Holloway, University of London, Egham, Surrey, United Kingdom

[2]Department of Computer Science, Royal Holloway University of London, London, United Kingdom

[3]ICT4D Centre, Department of Geography, Royal Holloway, University of London, London, United Kingdom

Corresponding Author:
Jennifer Cole, MA
H2B2
Department of Computer Science
Royal Holloway, University of London
112 McCrea Building
Egham, Surrey, TW20 0EX
United Kingdom
Phone: 44 7885 655695
Fax: 44 1784 443421
Email: Jennifer.Cole.2013@live.rhul.ac.uk

Abstract

Background: Concerns over online health information–seeking behavior point to the potential harm incorrect, incomplete, or biased information may cause. However, systematic reviews of health information have found few examples of documented harm that can be directly attributed to poor quality information found online.

Objective: The aim of this study was to improve our understanding of the quality and quality characteristics of information found in online discussion forum websites so that their likely value as a peer-to-peer health information–sharing platform could be assessed.

Methods: A total of 25 health discussion threads were selected across 3 websites (Reddit, Mumsnet, and Patient) covering 3 health conditions (human immunodeficiency virus [HIV], diabetes, and chickenpox). Assessors were asked to rate information found in the discussion threads according to 5 criteria: accuracy, completeness, how sensible the replies were, how they thought the questioner would act, and how useful they thought the questioner would find the replies.

Results: In all, 78 fully completed assessments were returned by 17 individuals (8 were qualified medical doctors, 9 were not). When the ratings awarded in the assessments were analyzed, 25 of the assessments placed the discussion threads in the highest possible score band rating them between 5 and 10 overall, 38 rated them between 11 and 15, 12 rated them between 16 and 20, and 3 placed the discussion thread they assessed in the lowest rating band (21-25). This suggests that health threads on Internet discussion forum websites are more likely than not (by a factor of 4:1) to contain information of high or reasonably high quality. Extremely poor information is rare; the lowest available assessment rating was awarded only 11 times out of a possible 353, whereas the highest was awarded 54 times. Only 3 of 78 fully completed assessments rated a discussion thread in the lowest possible overall band of 21 to 25, whereas 25 of 78 rated it in the highest of 5 to 10. Quality assessments differed depending on the health condition (chickenpox appeared 17 times in the 20 lowest-rated threads, HIV twice, and diabetes once). Although assessors tended to agree on which discussion threads contained good quality information, what constituted poor quality information appeared to be more subjective.

Conclusions: Most of the information assessed in this study was considered by qualified medical doctors and nonmedically qualified respondents to be of reasonably good quality. Although a small amount of information was assessed as poor, not all respondents agreed that the original questioner would have been led to act inappropriately based on the information presented. This suggests that discussion forum websites may be a useful platform through which people can ask health-related questions and receive answers of acceptable quality.

Introduction

Background

Over the past 2 decades in England and Wales, consultation rates within general practitioners' (GP) surgeries have increased from approximately 220 million in 1995 to 300 million in 2008 and are estimated currently at approximately 340 million [1]. Over the last decade, the number of attendances at accident and emergency (A&E) units in the National Health Service has increased more than 30%, from 14 million per year prior to 2003/2004 to 21.7 million in 2013/2014, and numbers are continuing to grow [2]. Pressure on GP surgeries may be one of the reasons for the increasing pressure on hospital A&E departments: 22% of patients report that it is not easy to get through to their GP's surgery on the telephone and 9.8% of people who are unable to get a convenient GP appointment go to an A&E walk-in center instead [1].

Since 2008, online health-seeking information in the United Kingdom has increased dramatically, from 18% of UK adults saying they looked for health information online in 2008 to 43% in 2013, with an increase of 59% among the 25 to 29 years age group [3]. Health information seeking represented one of the fastest growing areas of Internet use measured by the UK government during the period from 2008 to 2013 and in 2014; 8% of people aged 16 to 35 years and 15% of those aged 55 to 64 years made a GP appointment using the Internet. The United Kingdom is the second highest country globally for Internet health searches; in a recent survey, "Google my symptoms" was a more common first action than "book a doctor's appointment" or "visit a pharmacy for advice" [4]. In 2014, the number of health searches carried out in the United Kingdom increased by 19% [4].

Quality Considerations for Online Health-Seeking Behavior

Supporting individuals to shift at least some of their health-seeking behavior from a face-to-face consultation with a medically qualified practitioner to seeking information online, both before and following diagnosis, provides opportunities to relieve the pressure on GP surgeries and A&E departments. However, it is also dependent on the information found online being of sufficiently high quality that following it does not pose a health risk. Prior studies of health information online have shown that it is of variable quality [5-10]. Although much concern has been expressed over this [11-15], few examples of actual rather than potential harm have been documented [16,17].

Internet users often seek disease-specific information [18,19], including information that will enable them to diagnose a particular health problem [20]. Because trusted brands play an important role in health-seeking behavior [21-23], one way to make health-seeking behavior more comfortable for the Internet user may be to encourage them to turn to known and trusted websites when seeking health information, leveraging trusted brands to help them feel confident about the information they find there. If the brand is not health-specific, but is a source of information on a range of topics that the patient already trusts, they may be more likely to turn to it for information when they engage in online health-seeking behavior for the first time. Respondents to a short study on health information-seeking behavior during the 2014-2015 Ebola crisis in West Africa largely did not begin to use new modes of communication to seek out health information. Instead, they searched for health information through platforms and media they were already familiar with, turning first to trusted health and information brands, such as the World Health Organization, the BBC, and government ministries of health in addition to knowledgeable friends [24].

Characteristics to Support Health-Seeking Behavior

Online discussion forums have a number of characteristics that could benefit online health information seekers. Discussion is known to enable better learning and absorption of knowledge [25,26] and this has been identified as a benefit of discussion forums in general [27] and of online discussion forums specifically [28]. The emergence of Web 2.0 has provided new opportunities to gain and share knowledge about health issues [29-31] and discussion forums display positive attributes relating to all 4 website characteristics (source, medium, message, and receiver) that have been identified as important to engendering trust during online health-seeking behavior [32]. In particular, discussion forums can act as both the medium for and source of health information. Because both doctors and friends can be accessed through online discussion forums, the Internet should not be seen as a competing category to face-to-face interaction with such sources, but rather an enabler of it.

A weakness of the existing literature is the tendency to approach the Internet as if it is a homogenous environment where every website can be trusted or mistrusted equally until trust is added on by accreditation seals or source authority. This does not consider whether some characteristics, such as voted discussion forums that offer the ability to counter a previous post with more accurate information or to fill-in the information missing from a previous incomplete answer, make discussion forums inherently more conducive to the transfer of good quality information than other types of websites. All spaces exert influences on the choices that people make in those spaces. The more designers, owners, operators, and users of online discussion forums are aware of what these influences are likely to be, the more able they will be to consider how they can influence users' choices [33,34].

The aim of this study is to provide an assessment of the quality and quality characteristics of information found in online discussion forums so that doctors, patients, and health care policymakers can better understand the online discussion forum environment and the information found there.

Methods

Selection Criteria

Our study involved UK-qualified medical doctors and UK (London)-based nonmedically qualified individuals assessing the information found in 3 online discussion forums (Reddit [35], Mumsnet [36], and Patient [37]) relating to 3 health conditions. We selected 3 health conditions that affect a high number of individuals in the United Kingdom: diabetes, chickenpox, and human immunodeficiency virus (HIV). According to the most recent figures from Public Health England, an estimated 107,800 individuals in the United Kingdom were living with HIV in 2013 [38]. An estimated 3.2 million (7%) of the UK population is living with diabetes [39], of whom 10% have type 1 diabetes and the remaining 90% have type 2. An estimated 90% of all Britons will have had chickenpox by the age of 15 [40], although no exact figures on infection exist for the United Kingdom because not all cases receive clinical attention. It is important to note that although US health policy positions chickenpox as a dangerous disease for which childhood vaccination is recommended [41], this is not the case in the United Kingdom, where it is positioned as a mild childhood disease for which vaccination is only necessary for high-risk groups. Professional medical consultation is considered necessary only in cases of complications listed on the website of the UK National Health Service [42].

There is evidence that a high volume of health-seeking information occurs in relation to all 3 conditions. Diabetes and HIV both feature in the top 10 most searched for diseases on Google (diabetes at number 2 with more than 9 million monthly global searches in 2013, HIV at number 4 with more than 6 million, and acquired immune deficiency syndrome [AIDS] at number 6 with 5 million) [43]. Although chickenpox appears lower on the list (at 43 with more than half a million global monthly searches), it is one of only a handful of communicable diseases found there and is the most significant childhood disease in the United Kingdom. Vaccination is uncommon (ie, the majority of the population are likely to catch it), but many parents do seek professional advice when their children develop symptoms. Therefore, it was likely that a considerable volume of health information would exist online for these 3 conditions and that forums where health is discussed were likely to have discussion threads related to them.

Selection of Discussion Forum Websites and Discussion Forum Threads

Three online discussion forum websites were selected based on their popularity and common usage by the UK population (rather than among specialist interest groups or social media superusers). These included 2 general discussion websites (Reddit and Mumsnet) and a health-specific site (Patient). We investigated each of the 3 websites to see if their message forums had existing discussion threads related to these conditions and found that all 3 health conditions were discussed on all 3 forums.

We selected specific discussion threads for the survey subjectively by undertaking a basic search inside each selected website on the chosen health conditions between February 15

and 17, 2015, and reading through the returned results to find questions for which we felt the original poster could (and probably should) have sought advice from a qualified medical practitioner. Discussion threads were rejected if the question did not require a medical or scientific reply (eg, a diabetic asking others whether they thought disclosing his diabetes on job applications would be disadvantageous) and if the question had received less than 2 replies. We selected 25 suitable questions (Reddit: n=9; Mumsnet: n=8; Patient: n=8; diabetes: n=8; HIV: n=9; chickenpox: n=8) according to the order they appeared in the search results, favoring discussions in which the question had been posted within the previous 12 months. If no suitable questions matching these criteria appeared within the first 50 search results, questions posted earlier or from beyond the first 50 search results were selected. Eleven of 25 questions selected appeared to be prediagnosis and were asking if symptoms they or a friend/family member were exhibiting might be indicative of a certain condition (eg, diabetes), 9 of 25 appeared to be postdiagnosis asking for advice on how to act in light of a condition (eg, whether certain exercise routines were suitable for diabetics), and 4 were asking for general advice on topics such as vaccination. Each question and the discussion thread that followed it was then assessed by more than one assessor. In total, 79 assessments were returned (mean 3.2, range 2-7 for the 25 questions).

Selection of Study Participants

We aimed to have the information in the forums assessed for quality by UK-qualified medical doctors (1 GP, 2 hospital infection specialists, 1 hospital-based diabetes consultant, 4 who did not give exact details) and also by London-based individuals who were not medically qualified, but who had experience with the health issue being discussed as a patient or as a carer of a patient.

The majority of the doctors were recruited through Ashford and St Peter's Hospital, which has links with the Health, Human Body and Behaviour (H2B2) program at Royal Holloway, University of London. Two other medical doctors, known personally by the authors, were also invited to participate.

The nonmedically qualified participants for diabetes were recruited by contacting the chairs of 2 (offline) support groups for diabetics. Contact addresses for support groups are given on the website of the diabetes support charity Diabetes UK, which enabled group coordinators to be contacted personally and asked to take part. For chickenpox, parents of children in the common age group for contracting chickenpox (age 2-10 years) were recruited through the Parents and Friends Associations of 2 local West London schools (Lovelace Primary School in Chessington and Putney Girls High School). The Terrence Higgins Trust, a charity that supports people living with AIDS and HIV, was also approached and asked to contact people living with HIV who would be willing to take part, but they did not reply. As such, no HIV-positive patients participated and the questions relating to HIV were answered by doctors only. Participants were self-selecting and, therefore, may be subject to selection bias. Demographic data collected on the participants was minimal: it recorded whether or not they were medically qualified and confirmed they were adults older than

XSL•FO

RenderX

age 18 years before taking part, but no other particulars were recorded because these were not deemed necessary for this part of the study. Participants were given the option of taking part anonymously; 12 chose to disclose no other information than their level of qualification. Five of the doctors, but only one of the nonmedically qualified participants, provided a contact email address.

Participants were sent, by email, a list of paired URL links for each discussion thread they were asked to assess. One linked to the actual online forum discussion thread, which they saw in situ with no modification made to it for the sake of the study, and the other to an online assessment form. Each discussion thread was assessed against the same criteria. The discussion threads were assessed according to 5 criteria and the participant responded by rating the information from highest quality to lowest quality (range 1-5) on:

1. The medical/scientific accuracy of the information found there [5,7,9];
2. The medical/scientific completeness of the information [9,10,44];
3. How sensible they considered the answers provided to be;
4. Whether they thought someone reading the website would act appropriately based on the information provided; and
5. How useful they felt the answers given would be to the original poster.

An additional question was asked to check that the respondents found the discussions easy to follow; only 4 assessments recorded any level of difficulty in following the discussions.

The responses assessed perceived factual quality of the answer (accuracy and completeness), gave a subjective assessment on that information (how sensible was it?), and subjective assessments of how the reader might respond (would they act appropriately and would they find the information useful?). We included this differentiation in the questions because, although many previous studies have criticized online health information for being of poor or variable quality [11,14,15], far less have found actual evidence of poor information leading to inappropriate or dangerous health decisions being made [16,17,45,46]. Because even fewer studies focused on how likely it is that Internet discussion forum readers will take action based on the information they found there, exploring perceptions around this is of particular interest.

In each assessment, the discussion threads could be assigned 1 of 5 rating values, for which the highest (1) related to the best quality information and the lowest (5) to information that was considered to be inaccurate or ill advised. The criteria for marking were consistent across each health topic and website, and provided a potential overall score of between 5 (5*1, top rating for each criteria) and 25 (5*5, lowest rating for each criteria) to each discussion forum thread.

Participants were invited to participate between May 12 and June 4, 2015, and were given 2 to 3 weeks to reply. The final survey assessments were accepted on June 13, 2015. Participants were sent a mean 8 (range 7 to 25) discussion threads to assess (each of which required assessments of the 5 separate criteria) based on their particular area of experience or medical expertise, with only one participant—a recently retired GP—offered all surveys to complete. The assessments were completed, and results collected, using the free online survey software SmartSurvey. A generic version of the assessment questionnaire is available in Multimedia Appendix 1.

Results

Survey Data Returned

A total of 79 assessments of discussion threads were returned at least partially completed (as of June 13, 2015). For 78 assessments, all 5 criteria were assessed and rated, but on 1 assessment, 2 of the criteria were skipped. Seventeen separate individuals took part, 8 of whom identified themselves as medically qualified. The qualified medical doctors completed 58 of 79 (73%) returned surveys and 21 of 79 (27%) surveys were completed by the nonmedically qualified respondents.

Tables 1-3 show the data from all survey responses by health condition and website. Multimedia Appendix 2 includes a visual representation of the quality scores and a reference to the actual question as it appeared on the Internet discussion forum website (as the assessors saw it when they made their assessment). Figure 1 shows a visual comparison of ratings in each category across all websites and health conditions.

Table 1. Quality score data from all survey responses for diabetes-related questions.

Website, question, and respondent	Quality score (1=high, 5=low)					Total
	Accurate	Complete	Sensible	Appropriate	Useful	
Reddit						
Q1. First party since being diagnosed, need advice?						
Doctor 1	2	1	1	1	1	6
Doctor 2	2	2	1	1	1	7
Public 9	2	2	1	2	2	9
Q2. Advice for exercise and midnight lows?						
Doctor 1	2	3	3	2	2	12
Doctor 2	2	2	1	1	1	7
Mumsnet						
Q3. Are anger outbursts normal with diabetes?						
Doctor 1	2	3	2	2	2	11
Doctor 2	3	3	2	2	2	12
Public 1	2	2	2	2	2	10
Public 9	1	4	3	3	2	13
Q4. Signs of diabetes or paranoid Mummy?						
Doctor 1	2	2	2	3	3	12
Doctor 2	2	2	1	1	1	7
Public 1	2	3	2	2	2	11
Patient						
Q5. Longer to get over a cold with diabetes?						
Doctor 1	5	4	4	3	3	19
Doctor 3	3	3	2	2	2	12
Public 2	3	1	3	3	4	14
Public 3	4	3	3	2	2	14
Q6. Can this be diabetes?						
Doctor 1	2	3	3	3	3	14
Public 2	3	1	3	2	4	13
Q7. Do I have Type 1 diabetes?						
Doctor 1	2	2	1	1	1	7
Doctor 2	2	2	1	1	1	7
Public 1	2	2	2	2	2	10
Q8. Diabetes: advice please?						
Doctor 1	2	3	2	3	2	12
Public 2	3	1	__[a]	__[a]	4	—

[a] Scores are missing because respondent did not answer for these criteria.

Table 2. Quality score data from all survey responses for HIV-related questions.

Website, question, and respondent	Quality score (1=high, 5=low)					Total
	Accurate	Complete	Sensible	Appropriate	Useful	
Reddit						
Q9. I found out I had HIV; not clear about the stage						
Doctor 1	2	3	2	2	2	11
Doctor 4	2	2	2	2	2	10
Q10. HIV and depression						
Doctor 1	3	5	4	4	5	21
Doctor 4	3	4	3	3	2	15
Q11. Question about HIV and personal fitness						
Doctor 1	2	2	2	1	1	8
Doctor 4	2	1	1	1	1	6
Q12. FAQ: Worried? Risk, testing, and anxiety						
Doctor 1	1	2	1	1	2	7
Doctor 4	1	2	1	2	1	7
Mumsnet						
Q13. Babysitter has just announced he's HIV positive						
Doctor 1	3	2	3	3	4	15
Doctor 4	3	3	2	2	2	12
Q14. Question about HIV (and partner)						
Doctor 1	3	2	3	3	3	14
Doctor 4	2	3	2	2	2	11
Q15. Children and HIV						
Doctor 1	3	2	3	2	3	13
Doctor 4	2	4	2	3	3	14
Patient						
Q16. HIV question						
Doctor 1	2	2	2	2	2	10
Doctor 4	2	3	1	2	2	10
Q17. HIV infection: intestinal yeast after 4 months?						
Doctor 1	3	3	3	3	4	16
Doctor 4	4	3	3	2	2	14

Table 3. Quality score data from all survey responses for chicken pox–related questions.

Website, question, and respondent	Quality score (1=high, 5=low)					Total
	Accurate	Complete	Sensible	Appropriate	Useful	
Reddit						
Q18. Is this chickenpox? Help!!						
Doctor 1	5	4	4	5	5	23
Doctor 4	3	3	3	2	2	13
Doctor 5	4	4	3	3	4	18
Doctor 6	5	4	4	2	3	18
Public 4	3	4	3	2	3	15
Public 5	4	2	2	2	2	12
Public 6	3	2	2	1	2	10
Q19. Did you give your child the chickenpox vaccine?						
Doctor 1	2	2	2	1	2	9
Doctor 4	2	2	1	2	1	8
Doctor 5	2	2	1	1	1	7
Doctor 6	3	3	4	2	3	15
Public 4	3	3	3	3	3	15
Public 5	2	2	1	1	1	7
Public 6	3	4	3	3	3	16
Q20. Chickenpox: why more dangerous to adults?						
Doctor 1	3	3	3	3	4	16
Doctor 5	3	4	3	3	4	17
Doctor 7	3	3	3	3	3	15
Q21. Chickenpox: is 5 months too young to expose?						
Doctor 1	3	3	3	3	3	15
Doctor 4	2	2	2	2	2	10
Doctor 5	5	4	4	5	5	23
Public 4	3	3	3	3	3	15
Public 7	3	3	2	2	2	12
Public 8	4	4	4	3	3	18
Q22. Has your child had the chickenpox vaccine?						
Doctor 1	3	4	3	3	3	16
Doctor 4	2	2	2	1	2	9
Doctor 5	3	3	3	3	3	15
Public 6	3	4	3	3	3	16
Public 7	3	2	4	3	3	15
Q23. Is it normal to be so very ill with chickenpox?						
Doctor 1	4	4	3	2	3	16
Doctor 4	2	2	3	1	2	10
Doctor 8	2	2	1	1	2	8
Public 4	3	3	1	3	3	13
Q24. Should toddler get the chickenpox vaccine?						
Doctor 1	2	2	2	3	3	12

Website, question, and respondent	Quality score (1=high, 5=low)					Total
	Accurate	Complete	Sensible	Appropriate	Useful	
Doctor 4	2	3	2	2	2	11
Doctor 5	1	2	1	3	4	11
Q25. Strange symptom with chickenpox						
Doctor 1	4	4	4	2	4	18
Doctor 4	3	3	3	2	2	13
Doctor 5	3	4	2	3	2	14

On average (excluding the retired GP), medically qualified respondents completed 5 surveys each and nonmedically qualified respondents completed 2 to 3 assessments each. When asked, the reason why respondents did not complete all assessments they were offered was lack of time. In total, this provided 393 criteria ratings across all 79 assessments (see Table 4).

Table 4. Number of individual criteria assessed (5 assessments per discussion forum thread) out of 393.

| Condition	Discussion forum, n			Total, n
	Reddit	Mumsnet	Patient	
Diabetes	25	35	53[a]	113
HIV	40	30	20	90
Chickenpox	85	75	30	190
Total	150	140	103	393

[a] Two questions on a survey were skipped.

Of the 353 assessments made overall, assessors rated the majority as a score of 2 (n=149) or 3 (n=137) on the scale of 1 to 5. Some information was rated the highest score of 1 (n=54) and a smaller proportion was rated the lowest scores (n=42 for a score of 4; n=11 for a score of 5). The lowest possible rating was given only 11 times out of 393 (2.8%) across the entire survey in comparison to 54 instances (13.5%) in which the highest possible rating was given. No discussion thread was given the lowest possible rating across all 5 categories (which would have given it an overall score of 25 suggesting that the information was inaccurate, incomplete, and likely to lead to harm) or the highest possible rating (which would have given it an overall score of 5 suggesting that the information was entirely accurate, complete, and would lead to appropriate action being taken).

When ratings were considered across all 5 criteria on which discussion forums were assessed and grouped into 4 value groups (5-10: threads were predominantly given 1 of 2 highest ratings for all criteria; 11-15: threads were largely given the high-middle ratings; 16-20: threads were often rated in the lower categories; 20-25: threads were assessed poorly on all 4 criteria) across the 79 surveys completed, 25 scored between 5 and 10, 38 scored between 11 and 15, 12 scored between 16 and 20, and 3 scored between 21 and 25. The one assessment that was only partially completed and could not be awarded an overall score was excluded from this stage.

High ratings were awarded more often than low ratings by a factor of 4:1. Ratings of 1 or 2 were awarded 203 of 393 times (51.7%), whereas ratings of 4 or 5 were awarded 53 of 393 times (13.5%). Overall ratings of 5 to 10 or 11 to 15 were awarded 63 of 78 times (81%), whereas ratings of 16 to 20 or 21 to 25 were awarded 15 of 78 times (19%).

Figure 1. Comparison of ratings (scores) awarded to each category across all websites and health conditions.

Times each score was awarded against criteria
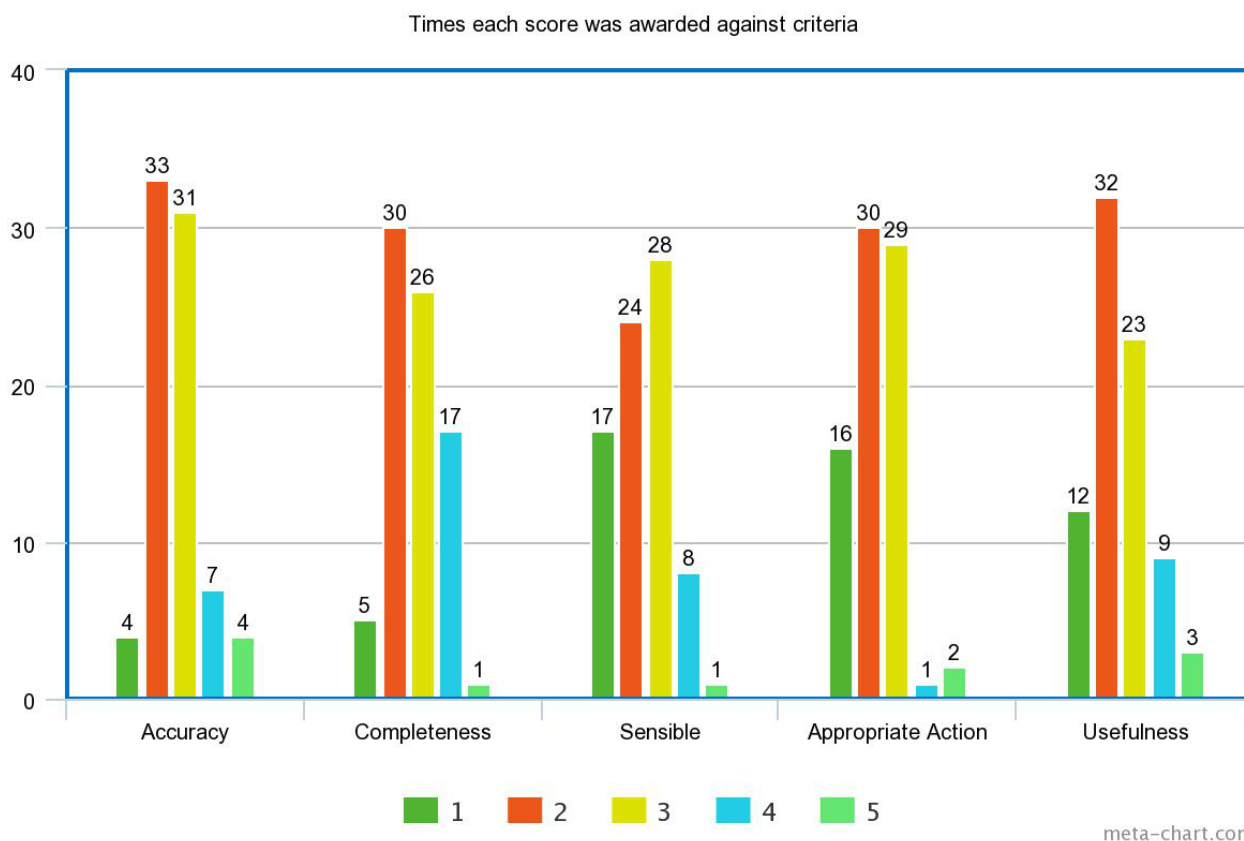
How Is "Bad" Information "Bad"?

Previous studies tended to assess information against a single criterion for quality (ie, completeness of information or accuracy of information) across all websites available regardless of their characteristics. This study enabled a comparison across a number of criteria and considered how website characteristics specific to discussion forums may influence this.

Although 18 of 79 surveys (23%) were considered to be incomplete (receiving scores of 4 or 5), covering "very little" or "none" of the information the assessor would expect to see, a smaller number than this (11/79, 15%) considered the information given to be "somewhat" or "very" medically inaccurate (receiving scores of 4 or 5 for those criteria). The findings are consistent with the existing literature in that information marked on quality is generally more related to the incompleteness of the information rather than actual inaccuracy [5,10,45]. An even smaller number (3/79, 4%) thought that poor information may lead to someone acting in a way that may put their health at risk. The results suggest that even if information is considered to be inaccurate or incomplete, this may not necessarily result in poor advice being given.

Because all message threads were rated by at least 2 assessors, and some were rated by 5 to 7 different assessors, this offered an opportunity to note whether information assessed as poor was marked consistently (or inconsistently) by all assessors, which would suggest an element of subjectivity in the assessment. In general, assessors scored consistently. For example, Q1 was awarded overall scores of 6, 7, and 9 by its 3 assessors; Q3 scored 10, 11, 12, and 13; and Q20 scored 15, 16,

and 17. Some threads scored consistently high by all assessors (eg, Q1, Q7, Q12) and some scored consistently average (Q3, Q20, Q25). No discussion thread was consistently awarded the lowest band range by all its assessors.

Assessors who marked more than 3 discussion threads awarded marks reasonably evenly across the possible rating bands: Doctor 1, the only respondent who assessed all 25 threads, gave ratings ranging from 6 to 24; Doctor 4 gave a range from 6 to 15; and Doctor 6 gave a range from 7 to 23. The nonmedically qualified respondents tended to be more cautious awarding ratings marks; they awarded across a range of 7 to 18. Only 1 of the 10 lowest-rating assessments was returned by a nonmedically qualified respondent (Q21/Public 8). This could be seen as nonmedically qualified respondents being less able to recognize poor quality information or less confident about highlighting it as poor quality. However, the nonmedically qualified respondents tended to agree with the medically qualified respondents when assessing information as being inaccurate and/or incomplete, but they gave different responses on how likely someone would be to act inappropriately on the information provided (eg, see Q21/Doctor 5 and Public 8).

In the 11 cases in which the lowest possible rating was returned for a discussion thread, other assessors (including doctors) of the same thread rated it more favorably. If a thread was rated in the highest score group (5-10) by one respondent, the lowest rating it received from any of the other assessors was 16, with the largest range 7 to 16 (Q19), whereas the 3 threads that rated in the lowest band (21-25) received a much broader range of scores (Q18: 10-24; Q21: 10-23). When information was considered to be of middling quality overall, it was more likely

that some assessors would consider it to be very poor. However, in most cases, low ratings were outliers in a broad range; for instance, the low rating of 23 awarded to Q21 by assessor Doctor 5 and the rating of 24 awarded to Q18 by assessor Doctor 1 were outliers to a range of 10 to 18 in both cases from the same thread's other assessors.

Assessment of Quality by Message Forum

There was some variation in quality between the message forums, with Reddit containing the highest quality information (rating of 1: n=33) more often than either Mumsnet (n=9) or Patient (n=12), but also being more likely to contain the lowest quality information (rating of 5: Reddit=7; Mumsnet=3; Patient=1) than the other 2 websites.

Assessment of Quality by Health Condition

Message threads related to chickenpox were less likely to be awarded high ratings (score of 1 or 2) than discussion threads related to either HIV or diabetes. The middle rating (score of 3) was most often awarded to discussions on chickenpox, whereas both HIV and diabetes were mostly likely to be rated 2. Eight of the total 11 lowest ratings were awarded against chickenpox discussion threads and 8 of the 10 lowest-rated threads overall were chickenpox threads.

Figure 2 displays all scores of the individual health topics to show the variation in results across the different health conditions.

Figure 2. Overall ratings (scores) awarded by health condition.

Times each score was awarded against health condition
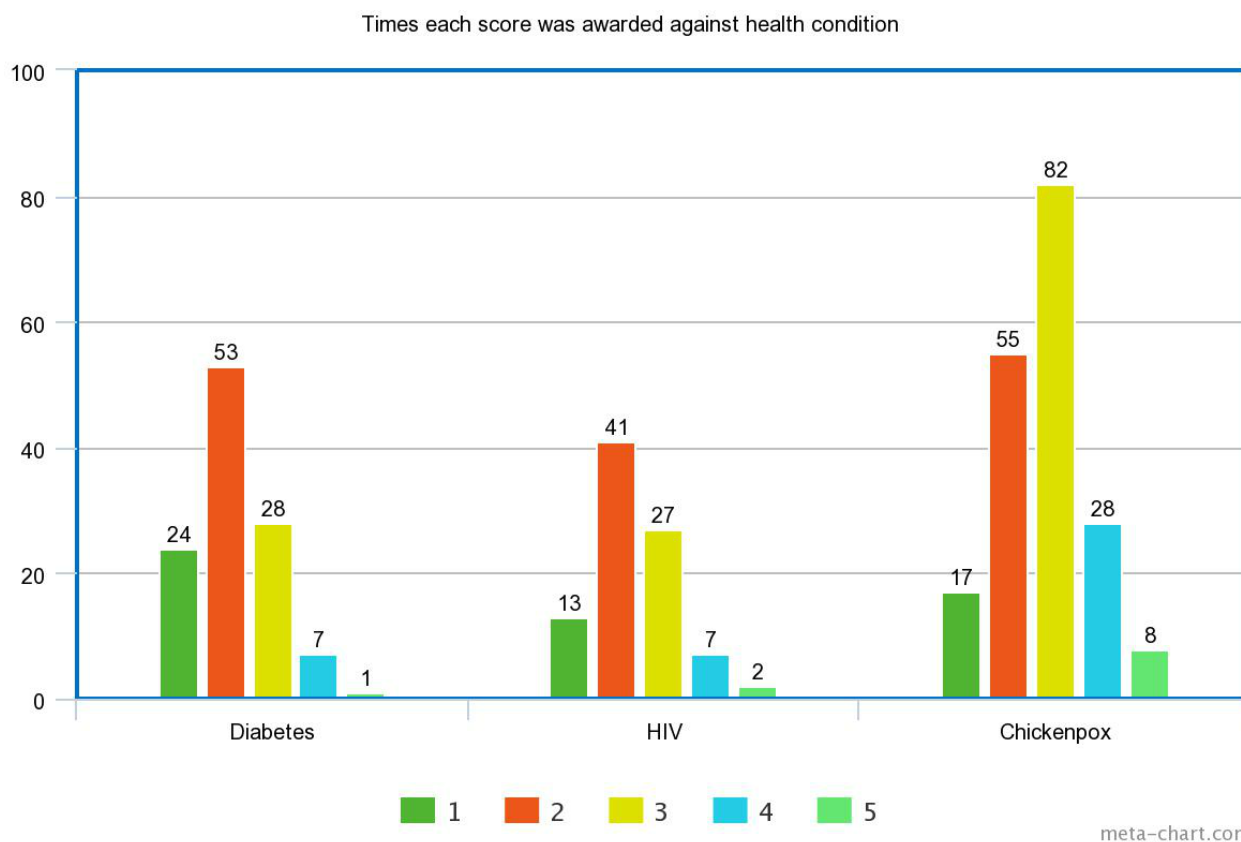
Discussion

Principal Results

The results of this survey suggest that, in general, the health information found in discussion forums is of reasonably good quality and only rarely does it contain information that is very inaccurate (4/79) and which some reviewers (3/79) feel may lead someone to act in a way that may put their health at risk.

These results are broadly consistent with those found elsewhere in the existing literature on the completeness and accuracy of health information found online. Previous studies on a diverse range of health conditions, including cancer, managing fever in children, and childbirth, have consistently suggested that approximately 60% to 70% of information is generally of good quality [5,10,15] with only approximately 5% to 7% considered

genuinely inaccurate [5,15,45]. The results also suggest that people may be more able to make sensible decisions when faced with poor quality information than doctors give them credit for. This warrants further study.

Rating the discussion threads on different criteria enables us to look more closely than previous studies at how, where, and why poor ratings are awarded. It is interesting to note, for example, that the controversial discussions around vaccination and herbal/natural remedies in the chickenpox discussions led to 36 separate low ratings of 4 or 5 being awarded. These were more often awarded against the inaccuracy of the information (n=11) or incompleteness (n=18) than the information being likely to lead the poster to make a somewhat inappropriate or very ill-advised decision (n=3).

It was not within the scope of the study to compare the questions that appeared to be asked prediagnosis with those that appeared to have been asked postdiagnosis. It was also not within the scope to compare those questions that were asked about the more serious conditions (ie, HIV and diabetes) with those that were asked about the milder condition (ie, chickenpox). We appreciate that these may be important factors in influencing the replies given and they warrant further research.

Q18. Reddit/Chickenpox

The discussion thread that was rated most poorly was Q18 ("Is this Chickenpox? Help!!") for chickenpox on Reddit, on which a parent had posted a photo of spots their child had developed and asked, "Is this chickenpox?" Two respondents considered the information given to be very medically/scientifically inaccurate, one of whom also considered the information to be very ill advised, likely to lead the questioner to make a very ill-advised decision, and to act in a way that may put their health in danger.

In total, 8 respondents completed this questionnaire. Although more than half (4/7) considered the information given to be "somewhat" or "very" scientifically inaccurate and to cover "very little" of the medical information they would expect to see, 6 of 7 respondents did not think this would actually lead to harmful behavior. It is also worth noting that some posters did encourage the original questioner to go to the doctor, who later posted an update to say that they had taken this course of action. This is particularly interesting because it provides proof that although the information was assessed by some experts to be poor, it did not lead to dangerous behavior and the original poster was capable of sorting the sensible advice from the mix of replies given.

Q21. Mumsnet/Chickenpox

On Mumsnet, Q21 ("Chicken pox–is 5 months too young to expose?") for chickenpox returned 3 lowest possible ratings against the inaccuracy of the information and in the 2 categories relating to how the poster might act. The survey was completed by 6 respondents in total and the low scores were awarded by only 1 of the 6; the other 5 rated the information more favorably. The discussion related to a parent's question about the safety of exposing their 5-month-old child to someone who was infected with chickenpox in the hope of getting the disease "out of the way." The discussion contained a range of views, from some parents who thought there would be little harm in it (largely due to experience of their own children having had the

disease at a similar age with no problems arising) to those who considered it dangerous. Several replies actively discouraged the parent from exposing an infant so young. None of the discussions displayed antivaccination viewpoints and none actively encouraged the mother to go ahead. At the end of the discussions, the original poster summarized her understanding of the discussions and stated that, after reading the advice, she thought that the ideal age to catch chickenpox "is 2 to 6 years," suggesting that she had been convinced that deliberately exposing a 5-month-old child would not be a good idea. Therefore, it is difficult to understand why one assessor felt that the poster would have made an ill-advised decision that would have put [her child's] health at risk, rather than taking the view of another assessor who made the qualitative comment: "I think she came to the right conclusion based on the information given."

Q10. Reddit/HIV

Two lowest ratings were given in response to Q10 ("I am somewhat prone to depression, but even more so now that I am HIV+. How do you guys deal with it?") for HIV on Reddit, which asked about links between HIV diagnosis and depression. One respondent ranked this discussion in the lowest categories for completeness of information and how the poster might act based on information provided. A qualitative response given in the comments box explained that the low ratings had been given because none of the replies encouraged the poster to seek professional help, which the respondent (a GP) believed they needed. Therefore, it was not so much that poor information was given, but that the appropriate good information was not. Another respondent (a hospital consultant) marked the discussion more favorably.

Q5. Patient/Diabetes

The final low rating was recorded against a discussion on whether diabetes affects a person's ability to recover from a cold ("Does it take longer to get over a cold if you have type 2 diabetes?"). One of 4 assessors felt that the information given was "very medically/scientifically inaccurate," but in this case, they did not feel that the information would lead the poster to make an ill-advised decision or to act in a way that may put their health at risk.

Because 4 of 79 surveys were responsible for all 11 instances of low ratings and just one of those (Q18. Reddit/chickenpox) was responsible for 5 of 11 low ratings (see Table 5), this warrants further research.

Table 5. Discussion forums returning the lowest possible ratings.

Header	Inaccurate	Incomplete	Ill advised	Make bad decision	Dangerous to health
Q5. Patient/diabetes	1				
Q10. Reddit/HIV		1			1
Q18. Reddit/chickenpox	2		1	1	1
Q21. Mumsnet/chickenpox	1			1	1

There was no discussion thread that was consistently rated in the lowest (or even lowest plus second lowest) categories by

all its respondents, suggesting that what constitutes poor information is as much a subjective judgment on the part of the

reviewer as an absolute. Respondents disagreed more on how people who read the information may act based on it than on the accuracy or completeness of the information. Previous studies have suggested that there is an element of subjectivity attached to assessments of quality [45] and although the results of this study uphold this, further exploration is warranted of how health information in online discussion forums is received and acted on.

Limitations of the Study

A number of limitations have to be taken into account when considering the results presented here. Firstly, the sample size was very small, consisting of only 17 individuals from a limited demographic (UK adults in West London). This cannot be considered to be representative; a much larger sample would need to be surveyed to ensure results could be applied more generally.

Secondly, recruiting respondents to the study was difficult, especially recruiting nonmedically qualified respondents. The diabetes support groups contacted were nervous about involving their members in a study that may direct them to incorrect and potentially harmful information. The HIV support charities contacted did not respond. Although it was easier to recruit parents to assess the chickenpox discussion, the numbers recruited were still fewer than hoped for. A larger future study would need to consider more efficient ways of recruiting higher numbers of participants.

Comparison With Prior Work

These results support other studies of online health information that found although online health information is of variable quality, the majority of it is of reasonably high quality with only a very small proportion considered to be factually incorrect (4/79) or potentially harmful to health (3/79 assessors thought the poor information given may lead someone to "act in a way that may put their health at risk"). Only 3 assessors awarded a discussion thread the lowest rating band overall, whereas 25 assessments rated a message thread in the highest band. This is broadly consistent with previous assessments about the quality of online health information in general [5,10,15,16,45].

Conclusions

The results of this study suggest that discussion forums are capable of producing health information of reasonably high quality. Of the 79 threads, 68 were assessed to contain at least some medically/scientifically accurate information and 61 of 79 were considered to contain at least some of the medical information that would be expected.

On only 3 occasions did an assessor think someone might make a "somewhat" or "very" ill-advised decision based on the information provided and there were only 3 occasions in which assessors felt the questioner may be led to act in a way that could put their health at risk. In each case, only one of the assessors felt this way when others did not; in the case of 2 of the 3 lowest ratings, comments made in the discussion forum by the original poster could be interpreted as meaning that they were not going to take a potentially harmful course of action. This challenges the assumption that the presence of poor information is automatically harmful.

The forums that contained the most inaccurate or controversial information also contained counterbalancing comments that appear able to dilute the potentially harmful consequences of the poor quality information. Comments made by the original poster and the majority of the respondents suggest that the better quality information was the more influential. This, in particular, warrants further study.

Online discussion forums do seem to be able to provide an opportunity for online health information seekers to access health information of acceptable quality. The findings suggest that there is merit in further exploring the possibilities of online discussion forums for providing peer-to-peer health information. In particular, there is a need to develop a better understanding of whether, and how, the small amount of incorrect or ill-advised information provided in a minority of answers is likely to result in adverse health outcomes or whether the discussion forum characteristics enable such messages to be counteracted and diluted. Most previous studies have tended to hone in on the small minority of poor quality examples and overemphasize the potentially detrimental impact they may have. This is despite the small number of studies that have found evidence of actual harm caused by poor health information found on the Internet. Thus, further analysis of the relationship between poor information and patient interpretation/action is crucial.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary File 1: Example Assessment Questionnaire.

XSL•FO

RenderX

[PDF File (Adobe PDF File), 40KB - jmir_v18i1e4_app1.pdf]

Multimedia Appendix 2

Data from all survey results including links to the actual question as it appeared on the discussion forum website (as the assessors saw it when they made their assessment).

[PDF File (Adobe PDF File), 101KB - jmir_v18i1e4_app2.pdf]

References

1. NHS England Analytical Service. Improving General Practice–A Call to Action: Evidence Pack. 2013 Aug. URL: http://www.england.nhs.uk/wp-content/uploads/2013/09/igp-cta-evid.pdf [accessed 2015-08-18] [WebCite Cache ID 6arzo8ZfI]

2. The King's Fund. 2015 Oct 08. What's going on in A&E? The key questions answered URL: http://www.kingsfund.org.uk/projects/urgent-emergency-care/urgent-and-emergency-care-mythbusters [accessed 2015-08-18] [WebCite Cache ID 6as0ecNqj]

3. Office for National Statistics. 2013 Aug 08. Internet access-households and individuals, 2013 URL: http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2013/stb-ia-2013.html [accessed 2015-08-18] [WebCite Cache ID 6as1KJvqP]

4. Last Night The Internet Saved My Life: Digital Health Report 2015. Stourbridge, UK: Pushdoctor; 2015 Aug. URL: https://www.pushdoctor.co.uk/Resources/PushDoctor-Health-report.pdf [accessed 2015-08-18] [WebCite Cache ID 6as0FAW8K]

5. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. BMJ 1997 Jun 28;314(7098):1875-1879 [FREE Full text] [Medline: 9224132]

6. Suarez-Almazor ME, Kendall CJ, Dorgan M. Surfing the Net--information on the World Wide Web for persons with arthritis: patient empowerment or patient deceit? J Rheumatol 2001 Jan;28(1):185-191. [Medline: 11196523]

7. Air M, Roman SA, Yeo H, Maser C, Trapasso T, Kinder B, et al. Outdated and incomplete: a review of thyroid cancer on the World Wide Web. Thyroid 2007 Mar;17(3):259-265. [doi: 10.1089/thy.2006.0300] [Medline: 17381360]

8. Lupiáñez-Villanueva F. Health and the internet: beyond the quality of information. Rev Esp Cardiol 2011 Oct;64(10):849-850 [FREE Full text] [doi: 10.1016/j.recesp.2011.06.014] [Medline: 21839571]

9. Lawrentschuk N, Sasges D, Tasevski R, Abouassaly R, Scott AM, Davis ID. Oncology health information quality on the Internet: a multilingual evaluation. Ann Surg Oncol 2012 Mar;19(3):706-713. [doi: 10.1245/s10434-011-2137-x] [Medline: 22146882]

10. Whitelaw N, Bhattacharya S, McLernon D, Black M. Internet information on birth options after caesarean compared to the RCOG patient information leaflet; a web survey. BMC Pregnancy Childbirth 2014;14:361 [FREE Full text] [doi: 10.1186/1471-2393-14-361] [Medline: 25306551]

11. Ahmad F, Hudak P, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with Internet-based health information? J Med Internet Res 2006;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]

12. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]

13. Sillence E, Hardy C, Briggs P. Why don't we trust health websites that help us help each other?: An analysis of online peer-to-peer healthcare. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris p. 396-404 URL: http://dl.acm.org/citation.cfm?id=2464488 [doi: 10.1145/2464464.2464488]

14. Grant L, Hausman BL, Cashion M, Lucchesi N, Patel K, Roberts J. Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites. J Med Internet Res 2015;17(5):e133 [FREE Full text] [doi: 10.2196/jmir.4153] [Medline: 26024907]

15. Biermann JS, Golladay GJ, Greenfield ML, Baker LH. Evaluation of cancer information on the Internet. Cancer 1999 Aug 1;86(3):381-390. [Medline: 10430244]

16. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]

17. Crocco AG, Villasis-Keever M, Jadad AR. Analysis of cases of harm associated with use of health information on the internet. JAMA 2002 Jun 5;287(21):2869-2871. [Medline: 12038937]

18. Dickerson S, Reinhart AM, Feeley TH, Bidani R, Rich E, Garg VK, et al. Patient Internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 2004;11(6):499-504 [FREE Full text] [doi: 10.1197/jamia.M1460] [Medline: 15298993]

19. Schwartz KL, Roe T, Northrup J, Meza J, Seifeldin R, Neale AV. Family medicine patients' use of the Internet for health information: a MetroNet study. J Am Board Fam Med 2006;19(1):39-45 [FREE Full text] [Medline: 16492004]

20. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]

XSL•FO
RenderX

21. Luo WM, Najdawi M. Trust-building measures. Commun ACM 2004 Jan 01;47(1):108-113. [doi: 10.1145/962081.962089]
22. Ye Y. Correlates of consumer trust in online health information: findings from the health information national trends survey. J Health Commun 2011 Jan;16(1):34-49. [doi: 10.1080/10810730.2010.529491] [Medline: 21086209]
23. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2014 Dec 16:e1. [doi: 10.1177/1460458214559432] [Medline: 25518944]
24. Cole J, Watkins C. International employees' concerns during serious disease outbreaks and the potential impact on business continuity: Lessons identified from the 2014-15 West African Ebola outbreak. J Bus Contin Emer Plan 2015;9(2):149-162. [Medline: 26642172]
25. Laurillard D. Rethinking University Teaching: A Framework for the Effective Use of Educational Technology. London: Routledge/Falmer; 2002.
26. Laurillard D. A conversational framework for individual learning applied to the 'learning organisation' and the 'learning society'. Syst Res Behav Sci 1999;16(2):113-122 [FREE Full text] [doi: 10.1002/(SICI)1099-1743(199903/04)16:23.0.CO;2-C]
27. Rowntree D. Teaching and learning online: a correspondence education for the 21st century? Brit J Educ Technol 1995;26(3):205-215. [doi: 10.1111/j.1467-8535.1995.tb00342.x]
28. Thomas MJ. Learning within incoherent structures: the space of online discussion forums. J Comput Assist Lear 2002;18:351-366. [doi: 10.1046/j.0266-4909.2002.03800.x]
29. Sundar S, Rice R, Kim H, Sciamanna C. Online health information: conceptual challenges and theoretical opportunities. In: Thomas T, Parrott R, Nusbaum J, editors. Handbook of Health Communication, 2nd edition. London: Routledge; 2011:108-112.
30. Hajli M, Bugshan H, Lin X, Featherman M. From e-learning to social learning–a health care study. Euro J Train Dev 2013 Nov 04;37(9):851-863. [doi: 10.1108/EJTD-10-2012-0062]
31. Groselj D. A webometric analysis of online health information: sponsorship, platform type and link structures. Online Inform Rev 2014 Feb 25;38(2):209-231. [doi: 10.1108/OIR-01-2013-0011]
32. Wathen CN, Burkell J. Believe it or not: factors influencing credibility on the Web. J Am Soc Inf Sci 2002;53(2):134-144. [doi: 10.1002/asi.10016]
33. Sunstein TA, Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness. New York: Penguin; 2009.
34. Munson SA, Cavusoglu H, Frisch L, Fels S. Sociotechnical challenges and progress in using social media for health. J Med Internet Res 2013;15(10):e226 [FREE Full text] [doi: 10.2196/jmir.2792] [Medline: 24148206]
35. Reddit. 2015. URL: https://www.reddit.com/ [accessed 2015-12-18] [WebCite Cache ID 6dryPs7g3]
36. Mumsnet. URL: http://www.mumsnet.com/ [accessed 2015-12-28] [WebCite Cache ID 6e7XvqM9X]
37. Patient. URL: http://patient.info/ [accessed 2015-12-28] [WebCite Cache ID 6e7Y0CqlL]
38. HIV in the United Kingdom: 2014 Report. London: Public Health England; 2014 Nov. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/401662/2014_PHE_HIV_annual_report_draft_Final_07-01-2015.pdf [accessed 2015-08-18] [WebCite Cache ID 6as2XGatN]
39. State of the Nation: Challenges for 2015 and Beyond. London: Diabetes UK; 2015. URL: https://www.diabetes.org.uk/Documents/About%20Us/What%20we%20say/State%20of%20the%20nation%202014.pdf [accessed 2015-08-18] [WebCite Cache ID 6as2DvLUe]
40. Public Health Wales. 2015 Jan 10. Chickenpox URL: http://www.wales.nhs.uk/sitesplus/888/page/43921 [accessed 2015-08-18] [WebCite Cache ID 6as2iXTfq]
41. Centers for Disease Control and Prevention. 2011 Nov 16. Chickenpox (varicella) URL: http://www.cdc.gov/chickenpox/about/overview.html [accessed 2015-10-15] [WebCite Cache ID 6cIJcDcwE]
42. NHS Choices. 2014 Jul 28. Chickenpox URL: http://www.nhs.uk/conditions/chickenpox/Pages/Introduction.aspx [accessed 2015-10-15] [WebCite Cache ID 6cIJle48B]
43. PharmaForward LLC. Most Searched for Diseases on Google. 2013 May. URL: http://www.pharmaforward.com/wp-content/uploads/2013/05/top50conditions_may2013.pdf [accessed 2015-08-18] [WebCite Cache ID 6as2trxdr]
44. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [Medline: 12020305]
45. 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]
46. Bansil P, Keenan NL, Zlot AI, Gilliland JC. Health-related information on the Web: results from the HealthStyles Survey, 2002-2003. Prev Chronic Dis 2006 Apr;3(2):A36 [FREE Full text] [Medline: 16539777]

Abbreviations

A&E: accident and emergency
AIDS: acquired immune deficiency syndrome

GP: general practitioner

HIV: human immunodeficiency virus

XSL•FO

RenderX

Original Paper

Website Sharing in Online Health Communities: A Descriptive Analysis

Chinmoy Nath[1], PhD; Jina Huh[2], PhD; Abhishek Kalyan Adupa[1]; Siddhartha R Jonnalagadda[1], PhD

[1]Division of Health and Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States
[2]Department of Biomedical Informatics, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Siddhartha R Jonnalagadda, PhD
Division of Health and Biomedical Informatics
Department of Preventive Medicine
Northwestern University Feinberg School of Medicine
750 N Lakeshore Drive, Floor 11
Chicago, IL, 60611
United States
Phone: 1 312 503 2826
Fax: 1 312 503 2833
Email: sid@northwestern.edu

Abstract

Background: An increasing number of people visit online health communities to seek health information. In these communities, people share experiences and information with others, often complemented with links to different websites. Understanding how people share websites can help us understand patients' needs in online health communities and improve how peer patients share health information online.

Objective: Our goal was to understand (1) what kinds of websites are shared, (2) information quality of the shared websites, (3) who shares websites, (4) community differences in website-sharing behavior, and (5) the contexts in which patients share websites. We aimed to find practical applications and implications of website-sharing practices in online health communities.

Methods: We used regular expressions to extract URLs from 10 WebMD online health communities. We then categorized the URLs based on their top-level domains. We counted the number of trust codes (eg, accredited agencies' formal evaluation and PubMed authors' institutions) for each website to assess information quality. We used descriptive statistics to determine website-sharing activities. To understand the context of the URL being discussed, we conducted a simple random selection of 5 threads that contained at least one post with URLs from each community. Gathering all other posts in these threads resulted in 387 posts for open coding analysis with the goal of understanding motivations and situations in which website sharing occurred.

Results: We extracted a total of 25,448 websites. The majority of the shared websites were .com (59.16%, 15,056/25,448) and WebMD internal (23.2%, 5905/25,448) websites; the least shared websites were social media websites (0.15%, 39/25,448). High-posting community members and moderators posted more websites with trust codes than low-posting community members did. The heart disease community had the highest percentage of websites containing trust codes compared to other communities. Members used websites to disseminate information, supportive evidence, resources for social support, and other ways to communicate.

Conclusions: Online health communities can be used as important health care information resources for patients and caregivers. Our findings inform patients' health information–sharing activities. This information assists health care providers, informaticians, and online health information entrepreneurs and developers in helping patients and caregivers make informed choices.

(J Med Internet Res 2016;18(1):e11) doi:10.2196/jmir.5237

KEYWORDS

Introduction

Increased access to online health information can empower patients to manage health better. A survey of US cancer patients showed that 92% of patients believed the Internet empowered them to make better health decisions and helped them communicate with their physicians [1]. Patients increasingly participate in online health communities and seek online health information; currently, more than 70,000 websites provide health information [2]. By May 2005, Yahoo! Groups [3] had listed more than 68,000 online support groups in their Health and Wellness section. Online health communities have been identified as one of the primary methods of online health information seeking for both consumers and members of their social networks [4-6]. Patients share their experiences and exchange emotional support and information through online health communities in the context of varying illnesses (eg, heart disease [7], rare diseases [8]). Patients also share resources for health information, including websites. Despite all the positive aspects of using online health information, it can be overwhelming, conflicting, and confusing for patients to find relevant, validated information [9]. Providing information to patients about the relevance and the validity of the websites posted in online health communities can assist in meeting the health information needs of patients while seeking online health information.

Members of online health communities, in addition to peer patients' psychosocial support, increasingly share health information resources, such as links to websites. Gustafson et al [10] showed that informational support in online health communities has the potential to affect health care consumers' decision making. Nambisan [11] studied the impact of empathy perceived by patients in an online health community based on their information-seeking effectiveness and social support. In online health communities, patients not only learn from peer patients, but also from online community moderators. Huh et al [12] showed that patients gained informational support from the community moderators. Further, other researchers have explored the assessment strategies for Internet information quality and readability [13], automated detection of conformity with the HONcode [14], computer-aided analysis of online social support [15], use of text mining and visualization for understanding smoking behavior [16], and analysis of top-level domain assignments [17].

These studies point to the importance of studying health information–sharing practices in online health communities. However, we lack knowledge around what kinds of information resources are being shared. One shared information resource in online health communities that we can easily capture is websites shared in the form of weblinks. We do not know what kinds of websites are being shared as an information resource and the context around how those resources are being shared.

Analyzing websites shared in online health communities should include the quality and purposes of these websites, who posts these websites, and whether there are any community differences. Investigating these issues around websites shared in online health communities will provide implications for developing how patients can appropriately navigate the online environment to locate relevant, high-quality health information.

Our research questions were:

1. Website categories: what kinds of websites are being shared in online health communities?
2. Information quality: what is the information quality of the websites being shared in online health communities?
3. Poster information: who are posting to those websites?
4. Community differences: how do communities post websites differently?
5. Context of website sharing: what are the contexts in which websites are being shared?

Methods

Data Collection

To answer our research questions, we chose the WebMD online health communities to investigate website-sharing practices. We chose WebMD because the community posts are publicly available and it is one of the most active online health communities online. We chose 10 WebMD online health communities on addiction, attention-deficit/hyperactivity disorder (ADHD), breast cancer, diabetes, weight loss, fit kids, heart disease, multiple sclerosis (MS), pain management, and sexual health. Our inclusion criteria for selecting these communities included being ranked within the top 15 communities in terms of total posting activity and having at least one health professional moderator and one staff moderator.

WebMD [18] is one of the few online health communities that offers both health professional and staff moderators. Health professional moderators at WebMD have clinical backgrounds in medicine, nursing, or nutrition. Staff moderators do not have clinical backgrounds, but facilitate and monitor conversations. We considered having enough moderator participation as criteria for choosing the community because we wanted to look at potential poster group differences in sharing websites.

We downloaded all posts from the 10 WebMD communities, which included 288,349 posts from June 2007 to February 2014. We received a letter from Northwestern University's Institutional Review Board (IRB) that this study is not regulated by the IRB because our study is equivalent to the observation of public behavior.

URL Extraction

To extract websites shared in WebMD online health communities, we extracted the URLs from each post using a regular expression pattern shown in the following:

"https?://[-.w.]*(:d+)?

([w/_-.=?&%+@^~!#$]*)?

[^www]|www.(:d+)?

([w/_-.=?

&%+@^~!#$]*)?[^www]."

Regular expressions are formal representations of text character patterns that represent a sequence of characters appearing in a

text document with functionalities such as set operations (eg, union, intersection, negation), boundary matches, quantifiers (eg, at least once, exactly *n* times), and logical operators. We selected all posts containing at least one URL with this pattern. Many community members mentioned website names (eg, YouTube [19], Facebook [20]), but not the URLs linking to the website. We excluded such mentions of website names not following the conventional URL pattern as shown. Figure 1 shows the process of extracting and analyzing URLs in the dataset.

Figure 1. Process description from data collection to URL extraction and analysis.
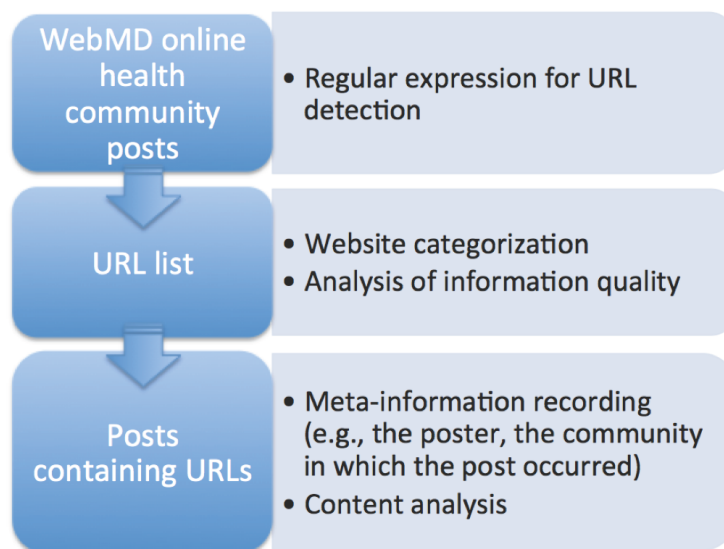
Website Categorization

We then developed a mechanism to classify the websites identified from the URLs based on their top-level domain (TLD) names [21]. URLs and other resources connected to the Internet (servers, computer) are hierarchically separated by the dot (".") symbol. For example, the hierarchy for "en.wikipedia.org" is "org → Wikipedia → en." TLDs are the top-most level in the hierarchy (org for "en.wikipedia.org"). Sometimes the TLDs are country code TLDs (ccTLD) (eg, "health.wa.gov.au"). In such cases, the ccTLD is ignored and the next domain name is considered as the TLD. We categorized a website as a ".gov" website if the TLD was .gov, an ".edu" website if the TLD was .edu, and a ".org" website if the TLD was .org. We also classified URL lists based on whether they could be considered as social media. It is difficult to identify social media websites from the domain names, let alone defining what a social media website is. For instance, Facebook is a representative website for social media. However, other generic websites, such as NYTimes.com, can also include social media features where the readers can interact online. To operationalize categorizing websites into social media, we selected the top 15 social media websites from eBizMBA [22], which included Facebook, Twitter, LinkedIn, Pinterest, Google Plus+, Tumblr, Instagram, VK, Flickr, Vine, Meetup, Tagged, Ask.fm, MeetMe, and ClassMates. For the URLs that contained these websites in their domain names (eg, www.facebook.com/pages/[...]), we categorized the URLs into "social media" websites.

Analysis of Information Quality

To assess the quality of information shared in each website, we used the total number of trust codes assigned to the website. *Trust codes* refers to official validations the website has fulfilled in terms of health information quality requirements. The accredited agencies we used in our analysis conducting such validations included Health On the Net Foundation [23], True Ultimate Standards Everywhere Inc [24], Utilization Review Accreditation Commission [25], GuideStar USA Inc [26], National Committee for Quality Assurance [27], and National Health Council [28]. Accordingly, a website can have multiple validations through the form of trust codes that appears on their website, given by these agencies as evidence that they have fulfilled the requirements as a safe health information-sharing website.

Because of the overwhelming number of URLs extracted, we developed a systematic way to efficiently examine the information quality. If a website was mentioned 3 times or more from at least one WebMD community, we hand coded for assignment of trust codes. Then, we collected a list of commonly occurring keywords from the URLs of the websites identified to contain trust codes. Examples included "med," "help," "doc," "Rx," and "MD." To assess the validity of the websites mentioned less than 3 times from one of the communities, we selected only those websites with URLs containing the previously listed keywords. Two authors (CN and AKA) hand coded trust codes for the websites that were mentioned at least 3 times or whose URLs contained these keywords. The interannotator agreement based on the kappa for assigning trust

codes to these 1229 URLs was .948 (95% CI .932-.964), which is considered very good agreement [29]. We recorded the total number of trust codes for each website collected.

Metainformation Recording: The Posters and the Community

To understand poster characteristics of website sharing, we aggregated community members into 3 groups: patient members, staff moderators, and health professional moderators. We used the list of staff moderators and health professional moderators' usernames available on the WebMD website to identify these 3 poster groups. We then ranked all patient members based on their total posting frequency. We then subgrouped patient members as the following: (1) high-posting members (posters in the upper quartile of the list), (2) medium-posting members (posters in the interquartile of the list), and (3) low-posting members (posters in the lower quartile of the list). We also retained the information on which community the post came from (eg, diabetes vs heart disease).

Qualitative Content Analysis

To qualitatively understand when and how community members shared URLs, we conducted a simple random selection of 5 threads among the conversation threads that included at least one URL in either the thread-initiating post or the replies from each of the 10 communities, resulting in a total of 50 threads.

The number of replies to these threads varied between 2 and 15, resulting in a total of 386 posts for the qualitative analysis. We analyzed the post content using open coding analysis [30] for identifying emerging themes for sharing the URLs.

Results

We extracted 25,448 URLs from 8714 unique posts out of the total 288,349 posts (3.02%) in all 10 communities (frequently shared websites shown in Multimedia Appendix 1). On average, posts in a community contained 1.99 (SD 1.14) URLs. Of all retrieved URLs, 94.83% (24,132/25,448) were posted in the replies.

Subsequently, we describe the categories of shared websites, the information quality of the shared websites, and findings around poster group and community differences in website-sharing behavior. We end with overall frequently shared websites and the context in which these URLs were shared.

Results on Website Categories

Our categorization criteria using TLDs resulted in 6 categories (in the order of appearance from high to low): .com websites, WebMD websites, .org websites, .gov websites, .edu websites, and social media websites. Figure 2 shows the website categories shared from our data and the content for each website category.

Figure 2. Trust code points across the website categories. The x-axis indicates the website categories and the y-axis indicates the percentage of websites with trust code points.
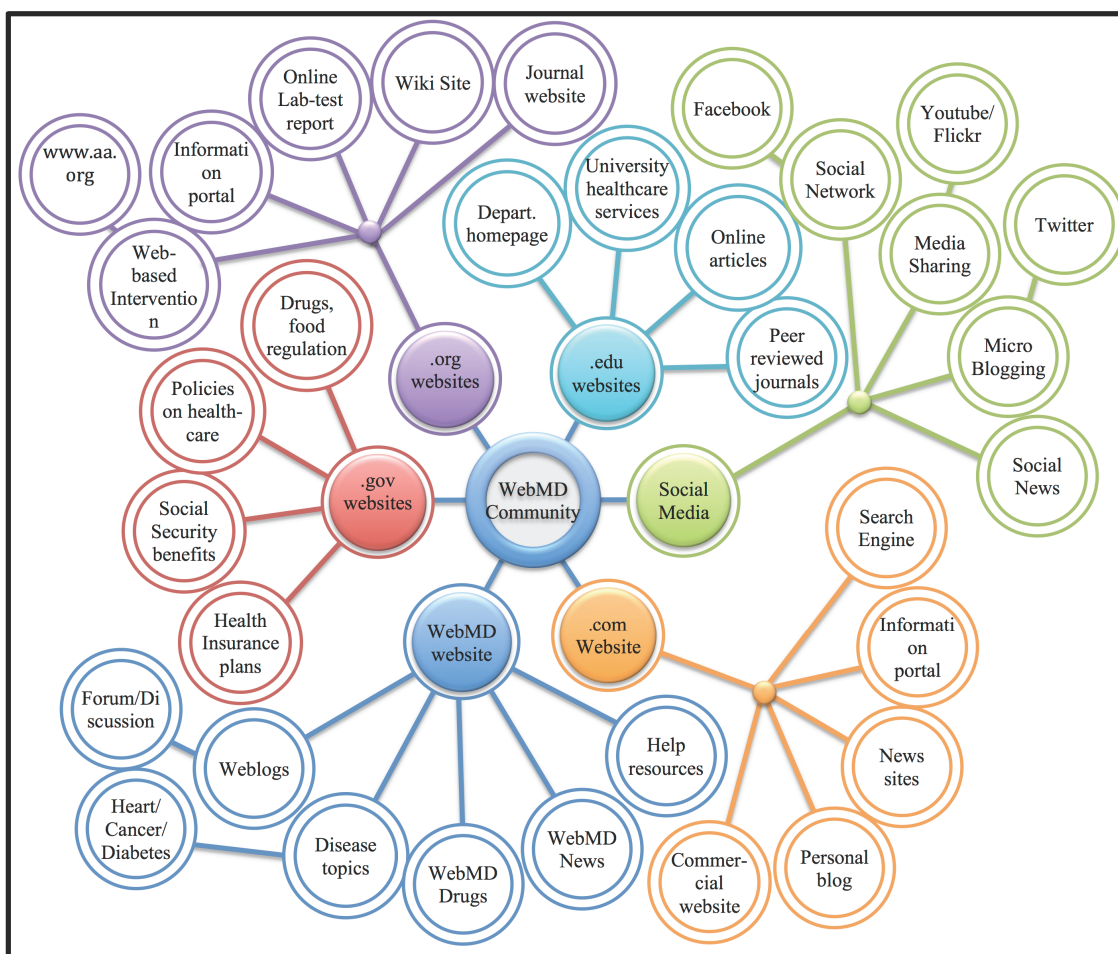
.com Websites

Out of the total of 25,448 URLs extracted, 15,056 URLs (59.16%) belonged to the .com websites. Websites classified in this category included search engines, information portal websites on drugs or medical tests, personal blogs, and commercial websites (eg, chemosavvy [31], Drugs.com [32], LIVESTRONG.COM [33], and Michelle's Road to Recovery [34]).

WebMD Websites

In all, 5905 of 25,448 URLs (23.20%) belonged to the WebMD websites. Because our data came from the WebMD online health communities, those community members often shared resources they found from WebMD. To address this bias, we separated the URLs from the WebMD website as its own category, rather than including it as part of the .com websites. The identified WebMD websites included information on information on various diseases, drug information, news on health, or resources for crisis assistance.

.org Websites

Another 3369 of 25,448 URLs (13.24%) belonged to the .org websites. The URLs in this category included nonprofit organizations representing community members' disease foci (eg, CHADD [35] for ADHD, Breastcancer.org [36] for breast cancer) and for-profit organizations related to the disease foci (eg, Joslin Diabetes Center [37] for diabetes). Other .org websites included wikis (eg, WikiEducator [38]), websites designed to help users understand laboratory test results (eg, Lab Tests Online [39]), journal websites (eg, American Medical Association [40], The American Journal of Clinical Nutrition [41]), and Web-based intervention websites.

.gov Websites

Of the 25,448 URLs, 930 (3.65%) belonged to government websites (eg, Centers for Disease Control and Prevention [CDC] [42], US Food and Drug Administration [FDA] [43]). These websites included information about government policies on health insurance plans, social security benefits, information on drugs, and food and health care.

.edu Websites

Another 149 of 25,448 URLs (0.59%) belonged to educational websites (eg, Perelman School of Medicine, University of Pennsylvania [44], University of South Florida [45]). Educational websites contained a university department's website (introduction to the department), news related to innovative therapeutic research, online educational resources, and journal articles published by university faculty.

Social Media Websites

Finally, 39 of 25,448 URLs (0.15%) belonged to social media websites, including social networking sites such as Facebook, media-sharing apps such as YouTube [19] and Flickr [46], and microblogging sites such as Twitter [47].

Information Quality of the Websites
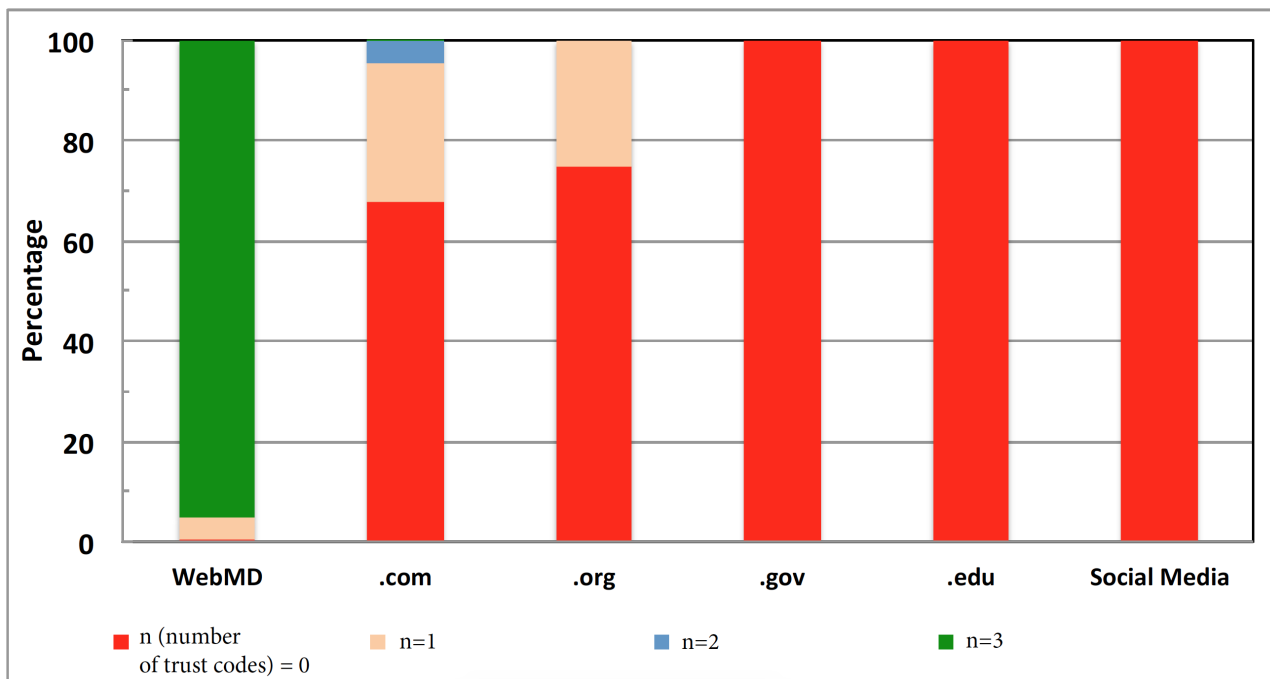
Overall Information Quality

We found at least one trust code in 4875 URLs: 32.38% (1901/5872) of all URLs shared under the .com websites, 99.38% (5836/5872) of all URLs for the WebMD websites, and 25.20% (849/5872) of URLs for the .org websites (Figure 3). In terms of the average number of trust codes per URL, the .com websites had 0.36 trust codes per post (SD 0.57; n=15,056), 2.89 (SD 0.45; n=5872) for the WebMD website, and 0.25 (SD 0.43; n=3369) for the .org website.

For the rest of the website categories, all 930 URLs (100%) of the .gov websites, all 149 URLs (100%) of the .edu websites, and all 39 URLs (100%) of the social media websites did not contain any trust codes. A potential reason for this result is that social media websites contain information that can be posted without validation of their truthfulness.

To examine the quality of .edu websites with regards to their institutions' expertise and existing work in disseminating health-related research, we investigated the number of systematic reviews published and indexed in PubMed. We focused on systematic reviews because they represent institutions with authors that synthesize evidence as opposed to focusing on primary literature. Our algorithm first retrieved the abstracts of all the 266,296 systematic reviews (as of November 11, 2015) using the clinical queries filter in PubMed [48]. Each abstract has an affiliation sentence that is often accompanied by the email address of the corresponding author(s). We used a simple regular expression (@?([^. <]+.)*[^. <]+) to extract the TLDs of the authors' institution. We separated the .edu TLDs from this list and ranked the list as shown in Multimedia Appendix 2. We found that only 25 of 149 (16.8%) .edu URLs in our dataset were from educational institutions that did not have at least one systematic review.

We also separated the .gov TLDs from the list of TLDs extracted from systematic review affiliation sentences and ranked the list as shown in Multimedia Appendix 3. However, very few .gov institutions publish research and systematic reviews (eg, CDC, National Institutes of Health [NIH], Department of Veterans Affairs, Department of Health and Human Services, Agency for Healthcare Research and Quality). Many .gov websites (eg, cancer.gov and whitehouse.gov) that might contain reliable information are not in our list.

Figure 3. Trust code points across the website categories. The x-axis indicates the website categories and the y-axis indicates the percentages of websites with trust code points.
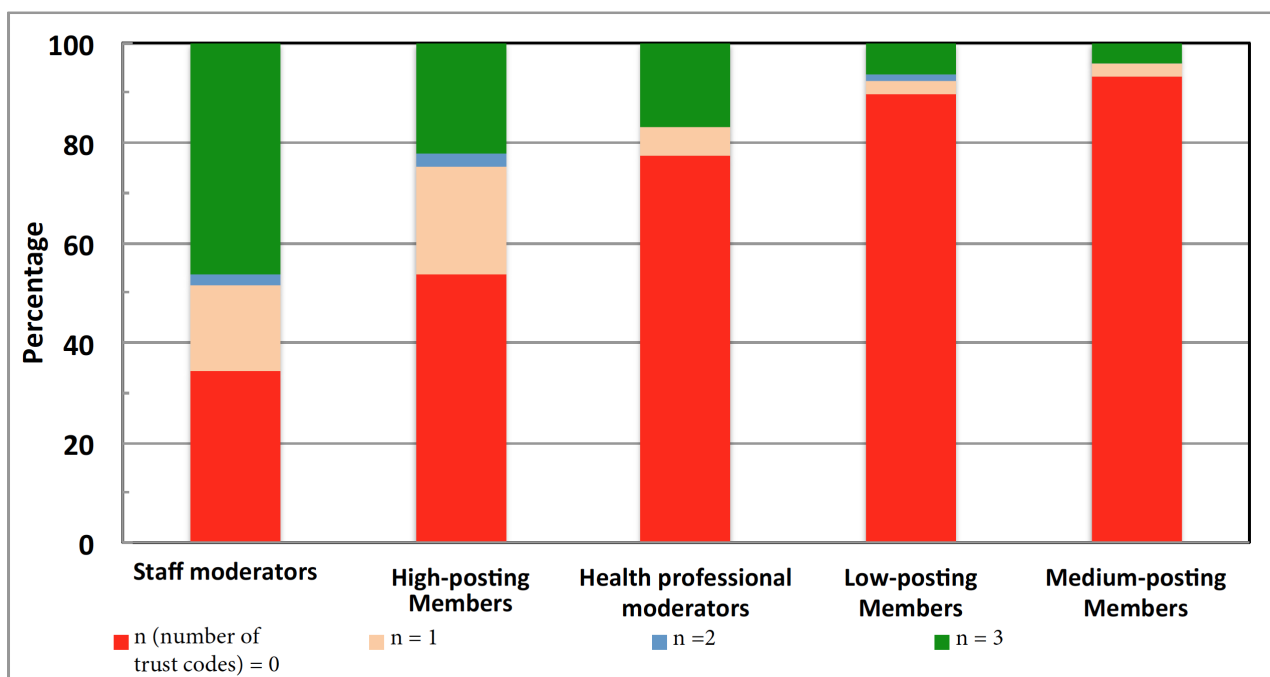

Poster Groups and Information Quality

There were 9671 high-posting members, 19,362 medium-posting members, and 9671 low-posting members. There were 88 staff moderators and 31 health professional moderators.

We found at least one trust code in 46.11% (4459/9671) of all URLs posted by the high-posting members, 6.67% (1911/19,362) for the medium-posting members, and 10.11% (978/9671) for the low-posting members. For the moderators,

we found 66% (58/88) of all URLs posted by staff moderators and 23% (7/31) for the health professional moderators contained at least one trust code (Figure 4). The average trust code numbers per URLs shared followed the same order: the staff moderators ranked the highest (n=474 trust codes; mean 1.61, SD 1.36), followed by high-posting members (n=24,252; mean 0.92, SD 1.20), health professional moderators (n=180; mean 0.56, SD 1.12), low-posting members (n=188; mean 0.24, SD 0.77), and medium-posting members (n=360; mean 0.15, SD 0.60).

Figure 4. Trust code points among various member groups. The x-axis indicates the member groups and the y-axis indicates the percentage of website with trust code points.

Community Differences and Information Quality

Table 1 shows the total number of posts, the number of posts containing URLs, and the mean number of URLs per post for each community. In the heart disease community, 3107 of the total 14,033 posts (22.14%) contained at least one URL. In other WebMD communities, less than 6% (mean 2.71%, SD 1.47%) of total posts contained URLs. On average, the heart disease community shared more than one URL per post, whereas other WebMD communities shared fewer than one URL per post. The heart disease community's total number of posts were fewer than many of the other communities. However, the total number of URLs shared in the heart disease community alone (n=16,146) was more than all URLs shared combined in other communities (n=9315).

Table 1. Total number of posts, posts containing URLs, URLs, and mean URLs per post for each WebMD community.

Community	Total posts	Posts containing URLs, n (%)	Total URLs	URLs per post
Heart	14,033	3107 (22.14)	16,146	1.15
Diabetes	71,168	2079 (2.92)	3586	0.05
Weight loss	58,344	956 (1.64)	1474	0.03
Breast cancer	26,653	729 (2.74)	1376	0.05
Sexual health	68,113	677 (0.99)	849	0.01
MS	28,267	527 (1.86)	848	0.03
ADHD	9637	363 (3.77)	697	0.07
Pain	8108	203 (2.50)	373	0.05
Addiction	3806	61 (1.60)	95	0.02
Fit kids	220	12 (5.5)	17	0.08

For information quality, the heart disease community had the highest percentage of URLs containing at least one trust code (61.60%, 9947/16146). The next in line was the fit kids community (41%, 7/17), followed by the weight loss community (31.47%, 464/1474), the pain community (28.4%, 106/373), the addiction community (27%, 26/95), the sexual health community (20.8%, 177/849), the diabetes community (18.07%, 648/3586), the breast cancer community (11.7%, 66/566), the MS community (10.5%, 89/848), and the ADHD community (7.2%, 50/697) (Figure 5). The mean trust codes per post was highest for the heart community (n=16,146 trust codes; mean 1.23, SD 1.24). The next in line was the weight lost community (n=1474; mean 0.71, SD 1.18), followed by the fit kids community (n=17; mean 0.65, SD 0.97), the pain community (n=373; mean 0.56, SD 1.04), the addiction community (n=95; mean 0.49, SD 0.95), the sexual health community (n=849; mean 0.43, SD 0.95), the diabetes community (n=3586; mean 0.38, SD 0.90), the breast cancer community (n=566; mean 0.25, SD 0.78), the MS community (n=848; mean 0.19, SD 0.65), and the ADHD community (n=697; mean 0.18, SD 0.69).

Figure 5. Trust code points in various online health communities. The x-axis indicates the online health communities and the y-axis indicates the percentage of websites with trust codes points.
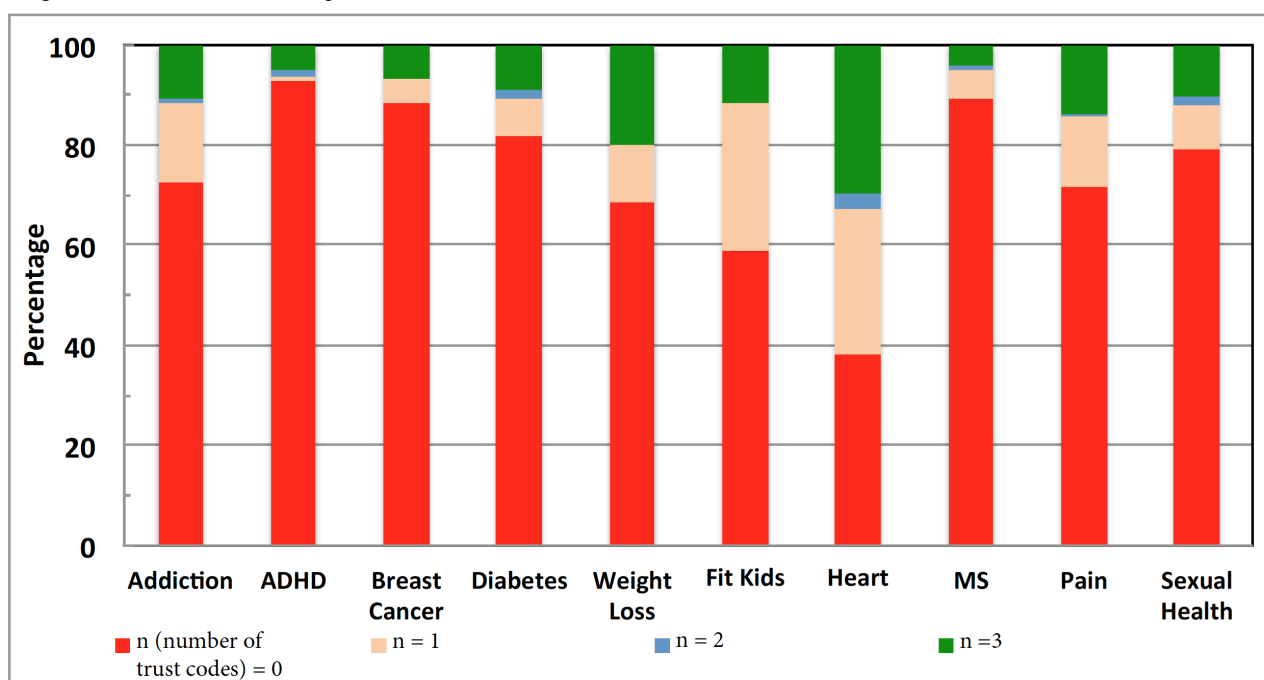
Content Analysis

Within our dataset for content analysis, some threads had more than one post containing URLs. There were 64 posts that contained URLs (mean 1.28, SD 2.06 posts containing URLs per thread). Members used URLs in a number of ways: to disseminate information (41/64, 64%), to find alternative means of communication (5/64, 8%), to extend social support (6/64, 9%), and to use as a resource for supporting evidence (12/64, 18%). Of the 64 posts containing URLs, 13 were posted by the WebMD moderators, who mostly did so to disseminate information (7/13, 54%) and help members extend social support (4/13, 31%). WebMD moderators also posted URLs as a resource for supporting evidence (2/13, 15%). Only 3 of 64 (7%) URLs containing posts were thread-initiating posts, showing that members in our content analysis sample shared URLs to respond to others' questions, rather than voluntarily disseminating information.

"Disseminate information" was the most frequently occurring theme. This theme included solutions and answers to questions, news, new study findings, product information, and education materials for new members. For instance, a member in the WebMD pain management community initiated a thread regarding an article against the petition to the FDA to modify the label on opioid medications. The petition called for constraining opioid use for non-cancer patients, which would be of concern to those dealing with pain. The members developed the thread into a heated debate about the effectiveness of such petitions.

Another example of disseminating information occurred when members asked questions that showed their early status in their illness and other members shared resources that new patients needed to know. In the WebMD breast cancer community, a member posted that she was just told to either have a biopsy or have a lump removed. To this post, another member responded with URLs regarding information about insurance as well as websites (eg, breastcancer.org) and phone numbers the member could call to get additional help on being newly diagnosed with breast cancer.

The members also posted URLs that allowed alternative means of communication, such as a separate forum or thread to share experiences. In the WebMD fit kids community, a member posted a WebMD forum link through which other members could share their ideas on, for instance, cooking with kids to help them make healthy eating choices. A member started a thread with the following message: "The topic is cooking with kids! How do you have fun with your kids in the kitchen and help them learn about healthy choices? Share your favorite ideas and recipes here! [URL]"

To those posters who were seeking social support, members responded with URLs through which the members could seek further emotional and informational support. For instance, in the WebMD sexual health community, a member posted her concerns that her fiancé was positive for the human immunodeficiency virus (HIV) and, although she loves him, she is concerned about their future sexual health. To this post, another member provided a link to an HIV community, stating that she could receive better support in that community.

Lastly, members used URLs to support, add, or clarify their arguments stated in their posts. Standards, guidelines, and definitions (eg, food portions for kids) were shared from authoritative, government websites. Members also used research articles listed on PubMed as evidence when they talked about the efficacy of certain treatments. Among the repliers, sometimes URLs were used to debate opposite opinions. For instance, in the ADHD community on WebMD, members debated how medications were harmful or benign based on research study findings linked with URLs.

In summary, members used URLs in many ways that triggered conversations, enriched their discussions, supported arguments, and added validity to various types of information shared.

Frequently Shared Websites

Of the top 50 frequently shared websites across all communities, 86% (43/50) contained health information (eg, www.hrspatients.org, www.healingwell.com); 50% (21/43) of these health information-providing websites were either .gov websites or websites certified with trust codes (eg, NIH [49]). Of the top 50 ranked websites, 54% (27/50) belonged to the .com website category, 4% (2/50) were blog websites (eg, The Life of Teddybear's Owner [50]), 2% (1/50) were social media websites (eg, YouTube [19]), and another 2% (1/50) were .org websites (eg, Northwestern Medicine [51]). All these websites are included in Multimedia Appendix 1.

Table 2 shows the top 10 frequently shared websites. These websites included 6 .com websites, 3 .org websites containing at least one trust code, and one .gov website. The top 10 websites belonging to the .com websites included information on health and health care, drugs and pharmacy information, community support, and repositories of personal medical information and healthy lifestyle information, including food, nutrition, and physical exercise. The .org websites belonged to a health care institution and disease-specific nonprofit organizations. The .gov website included in this list was a leading public health institute conducting research and providing information for control and prevention of diseases.

Table 2. The top 10 most-shared websites.

Websites	Occurrences, n
WebMD [18]	1739
Mayo Clinic [52]	544
HeartSite.com [53]	446
HealingWell.com [54]	290
myOptumHealth [55]	216
Heart Rhythm Society [56]	112
American Heart Association [57]	104
FITDAY [58]	94
ehealthMD [59]	89
US National Library of Medicine [60]	88

Discussion

In this paper, we present the website-sharing practices of online health community users. Using objective measures, such as post frequency, TLD, and trust code assignment, we learned about the kinds of websites shared, the information quality of these websites, the posters of these websites, and the community differences. We also show the context in which these websites were shared. Subsequently, we discuss the implications and practical applications of our findings.

Website Categories: .com Websites and Internal Resource Use

The majority of websites shared were .com websites. The community members rarely shared links to the popular social media websites. The .com websites contained a variety of content areas spanning from news, access to portal, and to personal blogs. Approximately one-third of these .com websites had at least one trust code assignment, meaning that at least one-third of these .com websites belonged to validated health information–sharing websites. Considering that the majority of websites shared were .com websites, more sophisticated methods to detect the content of the websites will help us understand the kinds of information community members attempt to share. For instance, the TLD can be further analyzed to understand whether it contains health-related keywords. The content on the main page can be scraped and automatically analyzed to generate topic distributions of the websites shared in online health communities. This information can then inform community members as well as researchers and practitioners whose goal is to develop better systems that can help patients gain high-quality information.

WebMD websites were ranked as the second most-shared website. This finding shows that the community members increasingly used the resources housed in their parent website. This finding shows the importance in choosing the parent website environment for establishing online health communities. The information quality of the websites shared in online health communities can be influenced by the quality of the parent website.

Information Quality: Rethinking Information Quality Detection

For this study, we focused on objective, efficient methods to understand website-sharing practices. The scope of our technique involved using (1) the posting frequency to understand the overall prevalence of various website-sharing practices, (2) the TLDs, which is extremely limited information, to categorize websites, and (3) the assignment of trust codes to assess the information quality. Our approach was helpful in gaining objective and efficient assessment over information quality. However, we faced a few difficulties in detecting the quality of .gov, social media, and .edu website categories. In the case of WebMD websites and other health information portals, their primary goal was to deliver health information to patients. Such health websites inevitably need to add trust codes to reassure that the visitors understand the quality of the website. The .gov and .edu websites are not found to have accreditations in general, but they might be trusted when associated with institutions with solid reputations [61]. None of the social media websites contained trust codes because they do not have an obligation to validate their health information quality; their primary focus is not necessarily sharing health information.

To further develop automated information quality detection, we need to rethink what is high-quality information.

Our measurement of information quality of the shared websites does not address potential unanticipated benefits that websites without trust codes can provide patients. For instance, Nambisan [11] showed that the key gratification for patients from online health communities is perceived empathy. Perceived empathy has the potential to directly affect the success of the treatment and it could supplement a caregiver's provided empathy, which is expensive and time consuming. Choi [62] reported that people increasingly share information about health care institutions through videos. She reported that traffic from YouTube to hospital sites increased 119% over a year in the 2012 Google/Compete Hospital Study; 30% of patients who watched a video made an appointment with that hospital.

Future research should investigate information quality methods for each website category and contexts in which websites are shared. Our qualitative analysis of the website-sharing context

XSL•FO

RenderX

indicated how information quality only matters half of the time people share websites—only when they want to disseminate information and use websites as supportive evidence. We need systems that would identify these varying needs before making uniform decisions about information quality of the shared resources.

In some threads, exchanges of appreciation and greetings took place after members shared websites. In such situations, the website itself acted as a catalyst for social networking among members. Members also shared websites linked with social media as a platform for sharing health information. Members posted Facebook webpages restricted for chemotherapy patients sharing various experiences and information, social medial profiles maintained by independent organizations to assist decision making in medical care, and websites maintained by research groups to assist others in exploring advanced health care topics. Social media websites can play an important role in disseminating what we traditionally consider "validated information" along with empowering anecdotes.

Poster Differences: Activity and Role of the Poster and Information Quality

Depending on how active the poster is, the quality and quantity of websites can differ. High frequency posters and moderators shared higher number of websites assigned with trust codes compared to the medium- to low-posting posters. It could be that the high-posting users and moderators share more health information-related websites than the lower posting users. Another explanation is that high-posting users and moderators take on the information dissemination role, which forces them into sharing validated health information websites. Fox et al [63,64] showed that the more experienced an Internet user is, the more likely they will search for health information online. Oh [65] showed that altruism is the most influential motivation and personal gain is the least motivating factor for responders to health questions in online health communities. Accordingly, because of the altruistic motivation, high-posting community members might be motivated to make sure they share Web resources with high-quality information. When developing information quality assessment tools or guidelines for online health communities, our findings inform it is important to take into account the posting frequency of posting members.

Community Differences: Disease Differences and Information Needs

We learned that the heart community shared the highest number of websites and the most websites with trust codes. Heart disease, with its potential to generate urgency in treatment, might push the community members to share websites that focus on validated health information. On the other hand, the ADHD community shared the least proportion websites with trust codes. Patients and caregivers with ADHD often face disagreement with their providers regarding diagnosis and treatment [66]. Thus, patients and caregivers of ADHD patients might share more controversial information sources. The addiction community shared several websites providing online intervention programs. In the weight loss community, diet and nutrition intervention programs were shared. Similarly, in the sexual health community, websites with intervention for

sex-addicted patients were shared. Christakis and Fowler [67] showed that smoking and alcohol cessation programs and weight loss interventions that provide peer support (ie, that modify the person's social network) are more successful than those that do not. Depending on the disease and the patients' relationship and existing challenges around health care could be reflected in their website-sharing practices. Again, the definition of what is high information quality in online health communities is highly situated.

Addressing Situated Quality: Practical Applications for the Stakeholders

Our findings inform a number of stakeholders, including health care practitioners, patients and caregivers, researchers, and online health information system entrepreneurs and developers. We discuss how situated quality should be addressed in health information sharing in online health communities.

The health care practitioners can learn from our frequently shared websites and descriptive results about what kinds of information patients navigate through. Based on our findings, health care practitioners can either redirect or encourage their patients about the websites they should be cautious of or further investigate. Patients and caregivers can use our findings to guide their future use of online health communities and think about what provisions should be made when using online health communities.

Researchers should further examine ways to improve information quality detection and understanding situated quality, the information quality that is a suitable guideline depending on the disease context and the motivation for sharing information. The online health community entrepreneurs and developers should think about the following when helping to improve information sharing practices in online health communities:

1. Develop real-time assessment of the categories and information quality of shared websites using our techniques: this information can be used for moderators in improving quality of posts.
2. Develop ways to further categorize .com websites in a meaningful manner.
3. Develop a situated information quality assessment tool based on poster characteristics, TLDs, trust codes, and context of posting (eg, thread initiator post vs reply).
4. Aggregate and summarize all websites for all community members to use.
5. In sharing summarized list of websites, reflect the situated context of the posts in which the websites came from.

One of the limitations of this study is that we were unable to collect demographic information on the patients because of WebMD privacy settings. Such patient profiles can further add the situated needs in why patients share websites. Also, many members posted the website's name alone without mentioning URLs. Our algorithm ignored websites that did not follow the regular expression pattern we designed. Because of our semiautomated search for trust codes on websites, it is possible we missed that some of the websites included trust codes. More

sophisticated information quality assessment methods can be developed using our findings.

Conclusions

Online health communities have emerged as one of the core places that patients visit to gain health care information resources and social support. We observed that sharing websites played a vital role in building networks among members of online health communities. We analyzed different contexts under which website sharing takes place and how different Web resources serve members' informational and emotional needs. We summarized the most frequent Web resources disseminated over 10 online health communities. Health care practitioners, content developers, and informaticians can use our findings to further understand how patients share websites online. Our findings might help these stakeholders to design systems that can help patients and caregivers make more informed choices.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of top 50 websites shared in ten different online health communities.

[XLSX File (Microsoft Excel File), 69KB - jmir_v18i1e11_app1.xlsx]

Multimedia Appendix 2

Educational websites sorted by frequency of systematic review email addresses.

[XLSX File (Microsoft Excel File), 69KB - jmir_v18i1e11_app2.xlsx]

Multimedia Appendix 3

Government websites sorted by frequency of systematic review email addresses.

[XLSX File (Microsoft Excel File), 45KB - jmir_v18i1e11_app3.xlsx]

References

1. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
2. Grandinetti D. Doctors and the Web. Help your patients surf the Net safely. Med Econ 2000 Mar 6;77(5):186-188. [Medline: 10848409]
3. Yahoo! Groups. URL: https://groups.yahoo.com/neo [accessed 2015-10-08] [WebCite Cache ID 6c7Q9iwiS]
4. Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]
5. 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]
6. Cotten S. Implications of Internet technology for medical sociology in the new millennium. Sociol Spectrum 2001 Jul;21(3):319-340. [doi: 10.1080/027321701300202019]
7. Bonniface L, Green L. Finding a new kind of knowledge on the HeartNET website. Health Info Libr J 2007 Dec;24 Suppl 1:67-76 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00742.x] [Medline: 18005296]
8. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
9. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [Medline: 15224975]
10. Gustafson DH, Hawkins R, Pingree S, McTavish F, Arora NK, Mendenhall J, et al. Effect of computer support on younger women with breast cancer. J Gen Intern Med 2001 Jul;16(7):435-445 [FREE Full text] [Medline: 11520380]
11. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]

XSL•FO
RenderX

12. Huh J, McDonald D, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]

13. Hanna K, Brennan D, Sambrook P, Armfield J. Third molars on the Internet: a guide for assessing information quality and readability. Interact J Med Res 2015;4(4):e19 [FREE Full text] [doi: 10.2196/ijmr.4712] [Medline: 26443470]

14. Boyer C, Dolamic L. Automated detection of HONcode website conformity compared to manual detection: an evaluation. J Med Internet Res 2015;17(6):e135 [FREE Full text] [doi: 10.2196/jmir.3831] [Medline: 26036669]

15. Wang Y, Kraut R, Levine J. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. J Med Internet Res 2015;17(4):e99 [FREE Full text] [doi: 10.2196/jmir.3558] [Medline: 25896033]

16. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]

17. Eysenbach G. The new health-related top-level domains are coming: will cureforcancer.health go to the highest bidder? J Med Internet Res 2014;16(3):e73 [FREE Full text] [doi: 10.2196/jmir.3358] [Medline: 24598704]

18. WebMD. URL: http://www.webmd.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7SQJb9r]

19. YouTube. URL: http://https://www.youtube.com [accessed 2016-01-06] [WebCite Cache ID 6eLemnWHC]

20. Facebook. URL: http://https://www.facebook.com [accessed 2016-01-06] [WebCite Cache ID 6eLepGhE4]

21. Postel J. RFC 1591: Domain Name System Structure and Delegation. Marina del Rey, CA: USC/Information Sciences Institute; 1994.

22. eBizMBA. 2015 Aug. The eBusiness guide top 15 most popular social networking sites URL: http://www.ebizmba.com/articles/social-networking-websites [accessed 2015-08-26] [WebCite Cache ID 6b4jzr66Z]

23. Health on the Net Foundation. Operational definition of the HONcode principles URL: http://www.hon.ch/HONcode/Webmasters/Guidelines/guidelines.html [accessed 2015-08-26] [WebCite Cache ID 6b4kPWngr]

24. True Ultimate Standards Everywhere, Inc. Privacy assessments & certifications URL: https://www.truste.com/business-products/dpm-services/ [accessed 2015-08-26] [WebCite Cache ID 6b4iCrDK2]

25. Utilization Review Accreditation Commission. Accreditation & management URL: https://www.urac.org/accreditation-and-measurement/measurement/ [accessed 2015-08-26] [WebCite Cache ID 6b4haEw4b]

26. GuideStar. Business solutions URL: http://www.guidestar.org/rxg/products/business-solutions/index.aspx [accessed 2016-01-06] [WebCite Cache ID 6eLfGmc8h]

27. National Committee for Quality Assurance. Accredition Programs URL: http://www.ncqa.org/Programs/Accreditation.aspx [accessed 2015-08-26] [WebCite Cache ID 6b4il5nsu]

28. National Health Council. Standards of Excellence Certification Program. URL: http://www.nationalhealthcouncil.org/resources/standards-excellence-certification-program [accessed 2015-08-26] [WebCite Cache ID 6b4j9XjqP]

29. Fleiss J, Levin B, Paik M. Statistical Methods for Rates and Proportions. Hoboken, NJ: J Wiley-Interscience; 2003.

30. Corbin J, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21. [doi: 10.1007/BF00988593]

31. ChemoSavvy. URL: http://www.chemosavvy.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7R0LPIX]

32. Drugs.com. URL: http://www.drugs.com [accessed 2015-10-08] [WebCite Cache ID 6c7TnJz86]

33. LIVESTRONG.COM. URL: http://www.livestrong.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7Tf1y78]

34. Michelle's Road to Recovery. URL: http://www.mch-breastcancer.blogspot.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7TvcqaX]

35. CHADD: The National Resource on ADHD. URL: http://www.chadd.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7UAdMDz]

36. Breastcancer.org. URL: http://www.breastcancer.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7UKrlU8]

37. Joslin Diabetes Center. URL: http://www.joslin.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7UbpF5G]

38. WikiEducator. URL: http://wikieducator.org/Main_Page [accessed 2015-10-08] [WebCite Cache ID 6c7Ui2Glz]

39. Lab Tests Online. URL: https://labtestsonline.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7UvSdr1]

40. American Medical Association. URL: http://www.ama-assn.org/ama [accessed 2015-10-08] [WebCite Cache ID 6c7V4bQwX]

41. The American Journal of Clinical Nutrition. URL: http://ajcn.nutrition.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7VAGvDq]

42. Centers for Disease Control and Prevention. URL: http://www.cdc.gov [accessed 2015-10-08] [WebCite Cache ID 6c7VNurQ1]

43. US Food and Drug Administration. URL: http://www.fda.gov [accessed 2015-10-08] [WebCite Cache ID 6c7VXnrES]

44. Perelman School of Medicine University of Pennsylvania, Department of Psychiatry, Penn Behavioral Health. URL: http://www.med.upenn.edu/add/adhd.html [accessed 2015-10-08] [WebCite Cache ID 6c7VkAYFL]

45. TRIGR (Trial to Reduce IDDM in the Genetically at Risk). URL: http://www.trigr.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7Vx1Dln]

46. Flickr. URL: http://http://www.Flickr.com [accessed 2016-01-06] [WebCite Cache ID 6eLgFUBSF]

47. Twitter. URL: https://twitter.com/?lang=en [accessed 2015-10-08] [WebCite Cache ID 6c7WWRO5D]

XSL·FO
RenderX

48.   Haynes RB, Wilczynski N. Finding the gold in MEDLINE: clinical queries. ACP J Club 2005;142(1):A8-A9. [Medline: 15656539]

49.   National Institutes of Health. URL: http://www.nih.gov [accessed 2015-10-08] [WebCite Cache ID 6c7WkoXrV]

50.   The Life of Teddybear's Owner. URL: http://www.mybearyspecial.blogspot.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7XiF9tg]

51.   Cadence Health. URL: https://www.cadencehealth.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7UTJYhO]

52.   Mayo Clinic. URL: http://www.mayoclinic.org/ [accessed 2015-10-08] [WebCite Cache ID 6c7XtXXIQ]

53.   Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform 2006 Dec;39(6):589-599 [FREE Full text] [doi: 10.1016/j.jbi.2005.11.004] [Medline: 16359928]

54.   HealingWell.com. URL: http://www.healingwell.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7YCaG8x]

55.   myOptumHealth. URL: http:5client.myoptumhealth.com/myoptumhealth [accessed 2016-01-06] [WebCite Cache ID 6eLgadeYM]

56.   Heart Rhythm Society. Patient resources URL: http://www.hrsonline.org/Patient-Resources [accessed 2015-10-08] [WebCite Cache ID 6c7Yc3sCl]

57.   American Heart Association. URL: http://www.heart.org/HEARTORG/ [accessed 2015-10-08] [WebCite Cache ID 6c7YpKoSb]

58.   FitDay Free Diet and Weight Loss Journal. URL: http://www.fitday.com/ [accessed 2015-10-08] [WebCite Cache ID 6c7YwjOLR]

59.   Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]

60.   US National Library of Medicine. URL: https://www.nlm.nih.gov/ [accessed 2015-10-08] [WebCite Cache ID 6c7ZFhtpn]

61.   National Institute on Aging. 2014 Dec. Online health information: can you trust it? URL: https://www.nia.nih.gov/health/publication/online-health-information [accessed 2015-11-24] [WebCite Cache ID 6dHk3NRA0]

62.   Choi M. Social media in clinical practice. Healthcare Inform Res 2015;21(2):138. [doi: 10.4258/hir.2015.21.2.138]

63.   Fox S, Fallows D. Pew Research Center. 2003 Jul 16. Internet health resources URL: http://www.pewinternet.org/2003/07/16/internet-health-resources/ [accessed 2016-01-06] [WebCite Cache ID 6eLhQ1p2G]

64.   Fox S, Rainie L. Pew Research Center. 2000 Nov 26. The online health care revolution URL: http://www.pewinternet.org/2000/11/26/the-online-health-care-revolution/ [accessed 2016-01-06] [WebCite Cache ID 6eLhfhLJ7]

65.   Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci 2011 Nov 01;63(3):543-557. [doi: 10.1002/asi.21676]

66.   Leslie LK, Plemmons D, Monn AR, Palinkas LA. Investigating ADHD treatment trajectories: listening to families' stories about medication use. J Dev Behav Pediatr 2007 Jun;28(3):179-188. [doi: 10.1097/DBP.0b013e3180324d9a] [Medline: 17565284]

67.   Christakis NA, Fowler JH. Social network visualization in epidemiology. Nor Epidemiol 2009;19(1):5-16 [FREE Full text] [Medline: 22544996]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ccTLD: country code top-level domain
CDC: Centers for Disease Control and Prevention
FDA: US Food and Drug Administration
HIV: human immunodeficiency virus
IRB: Institutional Review Board
MS: multiple sclerosis
NIH: National Institutes of Health
TLD: top-level domain

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Natural Language Search Interfaces: Health Data Needs Single-Field Variable Search

Caroline Jay[1*], PhD; Simon Harper[1*], PhD; Ian Dunlop[1*]; Sam Smith[2], PhD; Shoaib Sufi[1]; Carole Goble[1]; Iain Buchan[2], MD, FFPH, FACMI

[1]Information Management Group, School of Computer Science, University of Manchester, Manchester, United Kingdom

[2]Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Simon Harper, PhD
Information Management Group
School of Computer Science
University of Manchester
Kilburn Building
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1612750599
Fax: 44 1612756213
Email: simon.harper@manchester.ac.uk

Abstract

Background: Data discovery, particularly the discovery of key variables and their inter-relationships, is key to secondary data analysis, and in-turn, the evolving field of data science. Interface designers have presumed that their users are domain experts, and so they have provided complex interfaces to support these "experts." Such interfaces hark back to a time when searches needed to be accurate first time as there was a high computational cost associated with each search. Our work is part of a governmental research initiative between the medical and social research funding bodies to improve the use of social data in medical research.

Objective: The cross-disciplinary nature of data science can make no assumptions regarding the domain expertise of a particular scientist, whose interests may intersect multiple domains. Here we consider the common requirement for scientists to seek archived data for secondary analysis. This has more in common with search needs of the "Google generation" than with their single-domain, single-tool forebears. Our study compares a Google-like interface with traditional ways of searching for noncomplex health data in a data archive.

Methods: Two user interfaces are evaluated for the same set of tasks in extracting data from surveys stored in the UK Data Archive (UKDA). One interface, Web search, is "Google-like," enabling users to browse, search for, and view metadata about study variables, whereas the other, traditional search, has standard multioption user interface.

Results: Using a comprehensive set of tasks with 20 volunteers, we found that the Web search interface met data discovery needs and expectations better than the traditional search. A task × interface repeated measures analysis showed a main effect indicating that answers found through the Web search interface were more likely to be correct ($F_{1,19}$=37.3, P<.001), with a main effect of task ($F_{3,57}$=6.3, P<.001). Further, participants completed the task significantly faster using the Web search interface ($F_{1,19}$=18.0, P<.001). There was also a main effect of task ($F_{2,38}$=4.1, P=.025, Greenhouse-Geisser correction applied). Overall, participants were asked to rate learnability, ease of use, and satisfaction. Paired mean comparisons showed that the Web search interface received significantly higher ratings than the traditional search interface for learnability (P=.002, 95% CI [0.6-2.4]), ease of use (P<.001, 95% CI [1.2-3.2]), and satisfaction (P<.001, 95% CI [1.8-3.5]). The results show superior cross-domain usability of Web search, which is consistent with its general familiarity and with enabling queries to be refined as the search proceeds, which treats serendipity as part of the refinement.

Conclusions: The results provide clear evidence that data science should adopt single-field natural language search interfaces for variable search supporting in particular: query reformulation; data browsing; faceted search; surrogates; relevance feedback; summarization, analytics, and visual presentation.

Introduction

Data science spans many domains of application. For the health care and health sciences domains, it has the potential to bring researchers into "just-in-time collaboration" over shared data and data behaviors. The "big data" or "broad data" of data science lends itself naturally to secondary data analysis (using existing data to answer new research questions), which is traditionally associated with research data archives. Indeed, this data reuse is critical for both application and data mashups and acknowledges the cross-disciplinarity of the gathered data and their importance in combinatorial use. Using archived data such as annual health surveys, data discovery, particularly the discovery of key variables and their inter-relationships, is important for analyzing data and interpreting results properly. In addition, data science makes no assumptions concerning the domain expertise of a particular scientist, whose interests may intersect many domains and thereby enrich the research.

Secondary data analysis has a number of key functions [1] in relation to data science: it allows researchers to link datasets to answer questions that the files could not address adequately in isolation [2]; it creates opportunities to explore associations between factors that were not anticipated at the time of data collection [3]; and it has value from an ethical perspective by increasing the potential benefits to society arising from public investment in the collection of the original data [1]. Although secondary data analysis is essential to many areas of science and policy research, it is often impeded by difficulties in data discovery; besides, finding the most appropriate data to use for analysis can be problematic. Typically, the researcher needs to find a handful of appropriate variables among collections of thousands, often spread across multiple datasets such as successive years of a repeated survey. Current data archive information systems do not optimally support this search process; indeed, they make a presumption that their users are experts within the domain, and therefore, provide complex advanced interfaces to support these "experts." These interfaces hark back to a time when searches needed to be accurate first time as there was a high computational cost associated with each search. In this case, data scientists share more in common with the "Google generation" than with their single-domain, single-tool forebears.

Anecdotally, although following a Web search interface design (best expressed by "Google," "Bing," "Ask," etc.) would seem like the best practice, there is little empirical evidence to support such a claim. While the need to improve access to data for research purposes is recognized [4], no studies to date have directly examined how the user interface of tools providing access to archives impacts on the researcher's ability to discover and extract relevant data. Here, we report the results of a study conducted in collaboration with the UK Data Archive (UKDA) [5], the largest collection of digital research data in the social sciences and humanities in the UK. At the time of the study, access to data stored in the UKDA—including government and other large-scale surveys —was formally provided by the Economic and Social Data Service (ESDS [6]), a data archiving and dissemination service supporting the secondary use of data in both research and teaching. ESDS provided access to a wealth of data and had more than 250 research institutions registered to use its services.

Searching and accessing data from the UKDA has not been easy for two main reasons [7]. First, researchers had to work out which of the more than 5000 datasets stored in the UKDA could most appropriately be used to answer their research question. Getting a sufficient overview of what was available in each set was difficult, and researchers often picked certain datasets simply because they were familiar with them [8]. Identifying the appropriate variables within a dataset was a second problem. Surveys typically contain hundreds, if not thousands of variables (the Health Survey for England 2007, for example, contains more than 2000), and variable labels may not obviously reflect their content. To accurately identify variables of interest, the researcher must read the original questionnaire alongside supporting documentation, a process that can take days or weeks of work, and which may ultimately be fruitless: until the researcher has completed the process they do not necessarily know whether the dataset can answer their research question. Although both fully understanding a dataset and reading its documentation are important to the research process, it would save researchers a great deal of time if they could limit this in-depth exploration to datasets that were likely to be useful to them. Understanding other aspects of data use, such as how derived variables have been constructed, or how data from a number of years can be compared, is also problematic.

Current systems can be thought of as divided into two categories: (1) those that use a traditional advanced search interface [9], which expects accurate queries, patient users, and moderated and homogeneous data; and (2) those that use a Web search interface, which expects vague queries, impatient users, and an enormous and rapidly expanding collection of unmoderated and heterogeneous data [10]. We suggest that variable search for secondary analysis has more in common with the hostile search environment of the modern Web than it does with traditional search.

In this study, we compare two interfaces: one based on a "Google-like" Web search interface that enables users to browse, search for, and view metadata for individual factors and variables; the other a traditional "advanced" search user interface

(which presumes the user knows what they are looking for). Although more data archives do now have this kind of interface, our study is important because there is very little empirical work in this area.

Our hypothesis is that variables will be easier to find in research data archives via a single-field natural language search interface, conforming to Marchionini's Human-Computer Information Retrieval (HCIR) framework [11] and in particular supporting query reformulation; data browsing; faceted search; surrogates; relevance feedback; summarization, analytics, and visual presentation.

Background

There are numerous websites that provide access to the results of large-scale surveys (eg, the Office for National Statistics [12] in the UK, Eurostat [13] in the European Union, and the Bureau of Labor Statistics [14] and Inter-University Consortium for Political and Social Research (IPCSR) [15] in the United States). Until recently, the majority of survey repositories primarily used traditional search for the discovery of entire datasets, although the inclusion of Web search interfaces for variable data discovery is becoming more common. Both the IPCSR website and the Rand Survey Metadata Repository [16] provide access to a number of quantitative surveys conducted around the world and offer a facility for searching datasets at the level of variables. As detailed in the "Study Impact" section, following this study the UKDA now also supports variable data discovery using a Web search interface.

Traditional Search Interfaces

Traditional "advanced" search, and the interfaces that facilitate it, is based on a number of long-held premises. The most noteworthy in this context are the presumptions that the interface can expect accurate queries, that users are patient, and that the data will be moderated and homogeneous [9]. In some cases, especially within the scientific research domain, these assumptions hold true. In other cases, however, they do not reflect reality. This seems especially to be the case with regard to searches of variable datasets that seem to have more in common with the heterogeneity of the open Web. Increasingly, traditional search interfaces focused on delivering well-curated datasets (often already known to the user) are now looking for novel ways to fill the user expectation gap [17]. These systems are increasingly recognizing that providing access to relevant information adapted to the needs and the context of the user is a real challenge [18] and that contextual results are becoming more important. Furthermore, evidence suggests that the traditional search model predicated on users searching for particular information, the so-called information need, may not be as important as navigational searches [19]. Indeed, understanding the underlying goals of user searches is becoming increasingly important; for example, the previously unexplored "resource-seeking" [20] goal may account for a larger fraction of Web searches than previously thought.

Traditional search expects the user to have well-defined boundaries for the information they seek, along with a good knowledge of the terms and meta-data that may be used to describe that information. This is increasingly not the case,

especially in the context of variable data discovery and user-centered approaches [21] so common in the broad domain of data science.

Web Search Interfaces

Web search, and its offshoot of HCIR, recognize the deficiencies in the traditional search model, and thus expects vague queries, impatient users, and an enormous and rapidly expanding collection of unmoderated and heterogeneous data [10]. Indeed, the model of traditional search is changing, with the widespread use of Web search engines, employment of simple queries, and decreased viewing of results pages—changes that have resulted from algorithmic enhancements by Web search engine companies [22]. Large providers, such as Google, run around 10,000 experiments each year in an attempt to refine both the search engine and the search interface and interactivity [23]. We could conclude that the high level of experimentation makes Web search engines de facto best practice for all other search instruments with traditional interfaces not being able to match Google's ability to adapt and refine their algorithms and interactions. This is a trend we can see in search result clustering [24] for instance.

It is therefore not surprising that about 85% of Internet users surveyed claim to use search engines and search services to find specific information [25]. These users have expectations that bleed from Web search into all other areas that require search. To a naïve user, all search activity is the same [18]. In this case, we suggest that variable search for secondary analysis has more in common with the extremely hostile search environment of the modern Web than it does with traditional search.

Faceted Search and the Google Generation

The move from traditional search to Web search may be a result of changes in user attitudes and needs. The "Google generation" appears to behave very differently to older generations [26]. They are less confident about their searching prowess, demonstrated by the fact that they viewed fewer pages, visited fewer domains, and undertook fewer searches than older users [27]. In addition, tellingly, their search statements were much more the product of cut and paste. These characteristics—of relying less on working memory and demonstrating lower competence at multitasking—has knock-on implications for researching in an online environment [26,28]. To overcome some of these limitations, we have seen a rise in faceted search, which combines query and browse strategies and interactively provides an iterative way to refine search results [29]. Faceted search allows users to start very generally and then iteratively refine their searches by allowing them to apply multiple filters selectively. These filters can be based on taxonomies [30], simple classifications systems [29], or other spatial locations [31]—in some cases, they are generated from search results sharing some common overlap [30]. This faceted approach dovetails into the evolving behaviors of the Google generation, and assists in complex decision making [32].

Beyond Web Search

For reasons ranging from an obligation to curiosity, Web search is now moving beyond the individual and into the social domain [33]. Users have a strong inclination to seek information from

others during the search process. Indeed, search systems using statistical analytics over traces left behind by others can help support the search experience [34]. Furthermore, result clustering based on social networks in a crowdsourcing role [35] and grouped clusters displaying multiple tabbed search results [36] are also being increasingly used. These advances suggest a social component to dataset and variable retrieval will, in the future, be expected.

Context of This Study

Access to the UKDA via ESDS was set up primarily to facilitate the discovery and download of entire datasets, and as such shares much with traditional advanced search interfaces. The system provided several ways in which users could access individual variable descriptions, including a dedicated variable search facility, but anecdotal evidence indicated that these were difficult to use and not an adequate substitute for reading the complete survey documentation. Recognizing these issues, the UKDA decided to work with the University of Manchester (UK) to develop a Web search interface [37] as part of the Economic and Social Research Council-funded Obesity eLab project [38].

This interface was designed to simplify the process of accessing survey data, by enabling people to look for variables of interest through a familiar, and potentially more suitable, interface. Researchers typed a query into a single search box and then browsed relevant results. Variables were displayed in a tabular format, with the description shown prominently, allowing users to see at a glance whether the variable was relevant to their research question.

Although there is a large body of research examining search behavior in information retrieval [39,40], there is little that directly examines, from a user's perspective, how best to retrieve variable data from archived surveys. Our solution, called "MethodBox," initially emerged from the need to understand HCIR as it related to variable data discovery. A requirements analysis was conducted to understand the difficulties users experienced with the existing traditional search interface to UKDA, and to pinpoint new features that would help users to identify variables and datasets that could be used to answer research questions. The MethodBox Web search user interface

was then designed to make the search process as straightforward as possible for novice users, reflecting the fact that most of their information retrieval experience will have come from the Web [26,41].

Current Traditional Search Interface

At the time of the study, access to data stored in the UKDA, including government and other large-scale surveys, was formally provided through a website hosted by the ESDS, which provided numerous facilities for searching the UKDA catalogues.

On the home page (Figure 1), the simple search allowed users to search all fields in a record for keywords or phrases.

The resulting surveys were listed on the *catalogue search* page (Figure 2), and searches could then be refined using the *catalogue search* form. To access the variables in the survey, the user clicked through to the "Survey Description/Documentation," and then followed the "Variable List" link at the top of the page, which provided a list of all the variables in the dataset (Figure 3). Variable details were provided on a separate page when the user selected the name and clicked "show variable." The variable search (Figure 4) contained a single search box, and returned surveys that contain variables matching the keywords in a list underneath. Users could click through to the survey description and view the variable list as before, or click the link on the left of the result to go straight to the list of variables. The Nesstar tool allowed users to search and browse surveys in a tree view (Figure 5), and the ESDS government variable search returned a list of variables that matched search terms just from the government surveys.

In addition to the search facilities, there were numerous routes through which users could browse the available surveys, such as the "browse by subject" and "major studies" pages. Lists of variables could then be accessed from the study description pages. The ESDS website, like many sites, was frequently edited and upgraded; the study was conducted between September 27, 2011, and November 3, 2011, a period during which there were no major changes to the functionality offered by the site.

Figure 1. Simple Search -- Can be seen on the right-hand side.
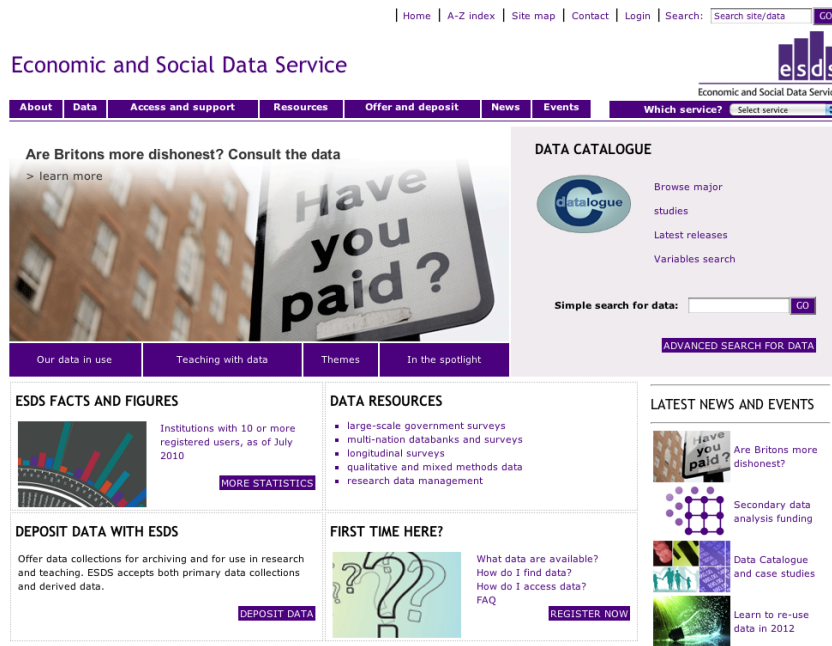

Figure 2. Orientation -- The search form is at the top of the page and the results are returned underneath. To view the variables, the user must click the 'Study Description/Documentation', then use the 'Variable List' link at the top of the page.
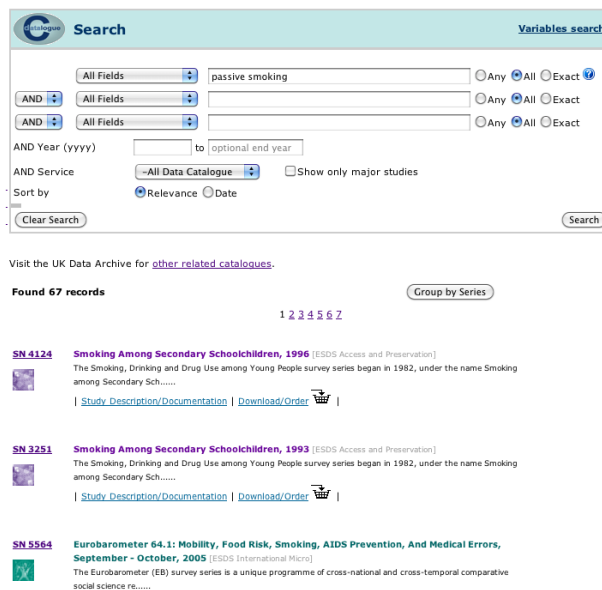
Figure 3. BSAS 2009 -- The variable list in the British Social Attitudes Survey 2009. To view a variable, the user selects one from the list box and clicks 'show variable'.
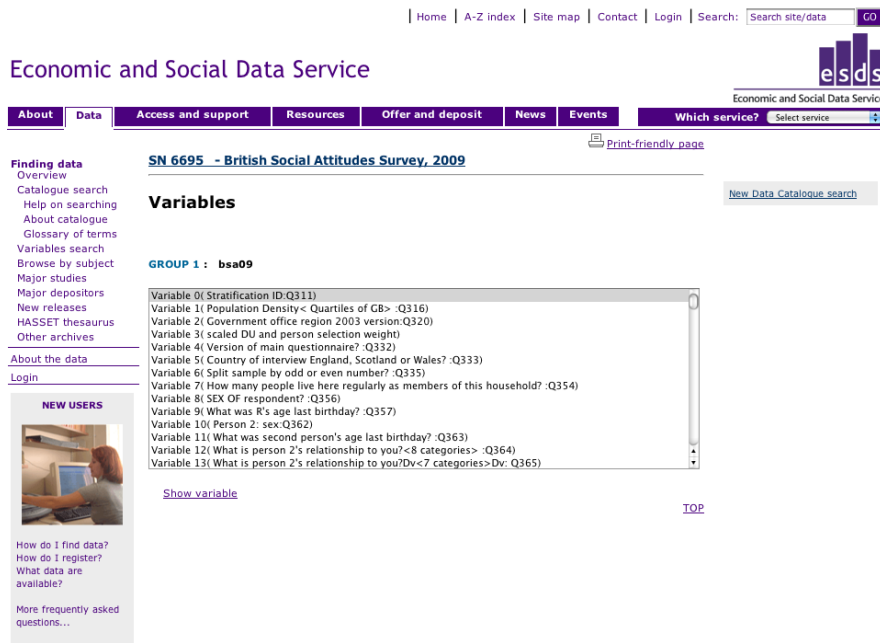

Figure 4. Variable Orientation -- The search box is at the top of the page, and the results are returned underneath. To view the list of variables, the user clicks the 'Variables in...' link on the left hand side of each result, which provides a list of all the variables in the dataset.
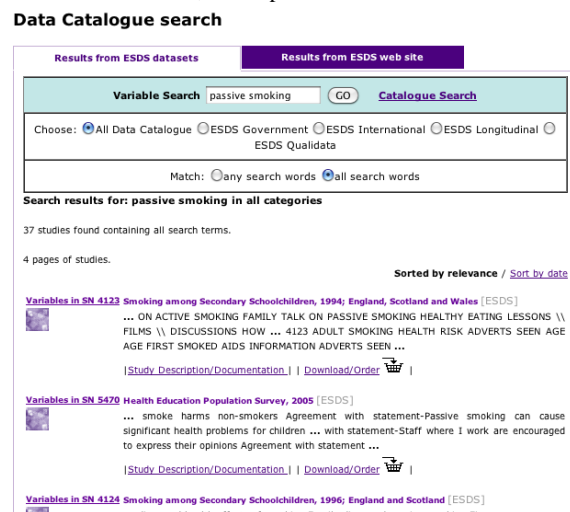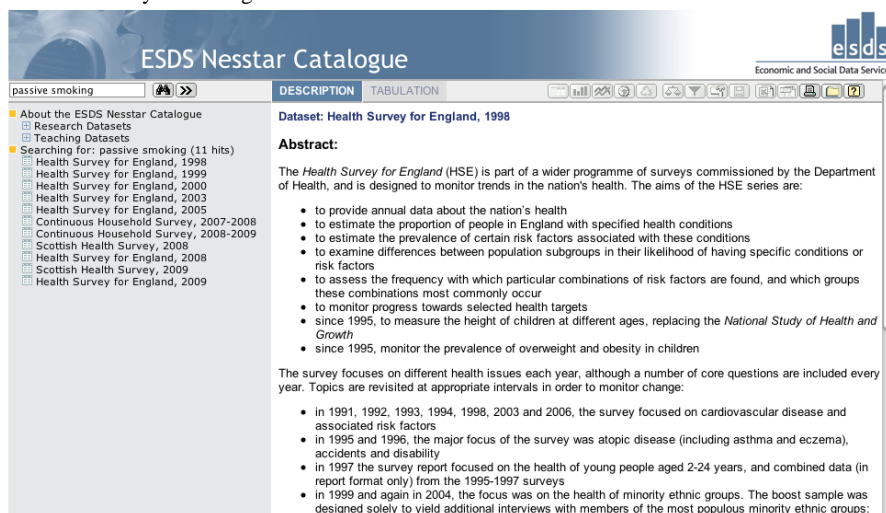

Figure 5. The Nesstar interface -- Surveys matching the search terms are listed in the menu on the left hand side.

Comparable Web Search Interface

The Web search interface (MethodBox) was designed to simplify the process of accessing survey data by enabling people to look for variables of interest through a straightforward "Web search" interface embedded in a scientific social network (Figure 6). Researchers typed a query in a single search box and then browsed relevant results. Variables were displayed in a table format, which could be reordered according to a number of categories. The variable description was displayed prominently, allowing users to see at a glance whether the variable was relevant to their research question. Variables of interest could be selected and then downloaded to the user's desktop.

A clear priority identified in the requirements analysis was a fast and straightforward means of identifying variables that are relevant to a particular research question. To achieve this, MethodBox assimilates all the required information about a variable, including its name, values and metadata, using the Data Documentation Initiative (DDI) [42] XML files available through the ESDS Nesstar service, and through processing the dataset documentation with the Utopia PDF parser [43]. This process allowed MethodBox to treat variables as first-class citizens in their own right. Users could also upload their own data files and add metadata in the DDI XML format. Assets inside MethodBox were indexed using Apache Solr, allowing users to search variable names and metadata quickly and easily, as well as the surveys, data analysis scripts, data extracts (subsets of variables created by other users), publications, and user profiles also held by MethodBox.

The MethodBox user interface was designed to correspond to the common mental model of an online Web search interface: a box for entering terms, a button to run the search, and a list of results [27]. The home page consisted of a single, "Google-style" search box, with checkboxes underneath to allow users to specify what they wanted to search (see Figure 7). All categories (surveys, variables, methods, data extracts, and publications) were selected by default. Matching results were returned in a table format. Results were initially ordered according to relevance, but could be sorted, for example, by year or survey, by clicking the table headers. If there were matching results in more than one category, these were displayed in separate tabs (Figure 8). Variable details could be accessed by clicking the arrow to the left of the result, which provided them in a dropdown box, or by clicking the variable title, which showed them on a separate page. Users could also select and search a subset of surveys (Figure 8) or navigate to a complete list of the variables from a link on the survey description page. If users were logged in, they could add any number of variables to their "shopping" cart, before downloading this subset of data to their desktop as a "data extract."

Figure 6. MethodBox Home -- Users type a query in the central search box, and can modify what is searched (surveys, variables etc.) using the tick boxes underneath.

Figure 7. Survey Results -- Categories that contain results are displayed in separate tabs. Results are displayed in a table and ordered according to relevance, but can be sorted by clicking the table headers.
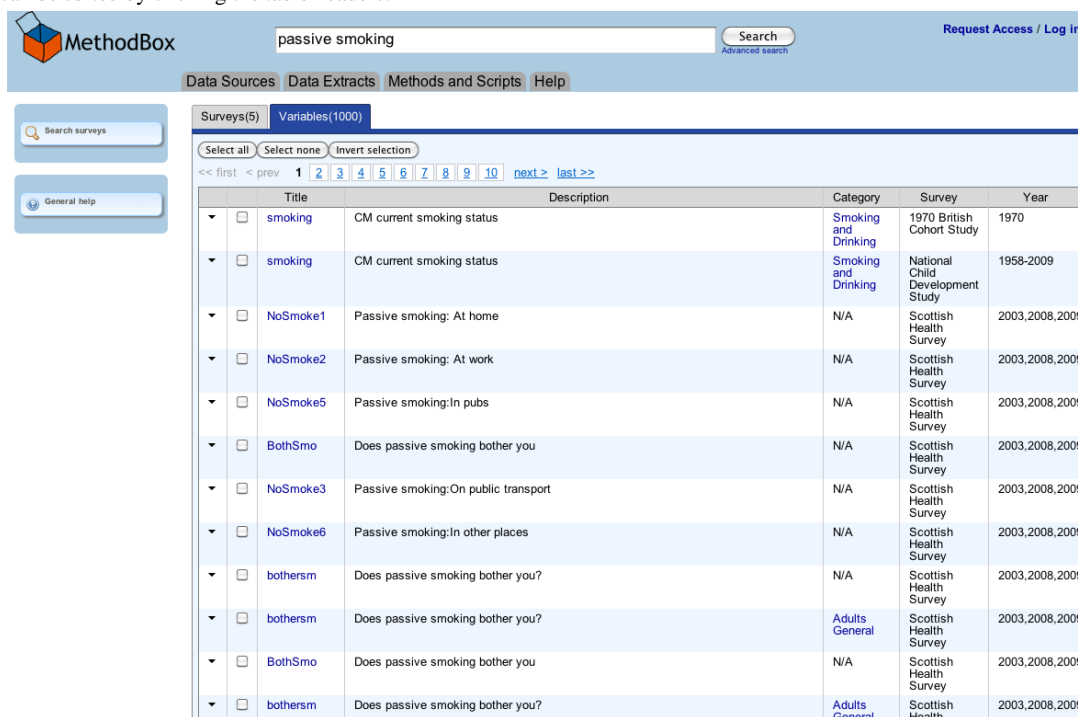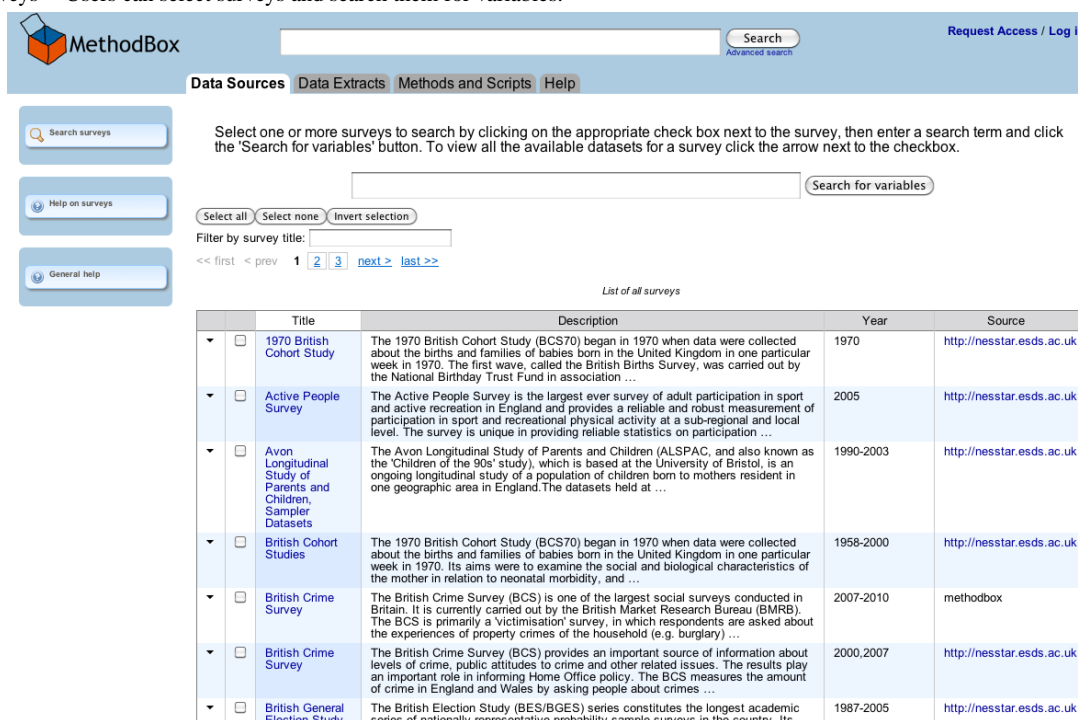
Figure 8. Surveys -- Users can select surveys and search them for variables.

Methods

The aim of the evaluation was to understand whether the Web search interface provided more effective, efficient, and satisfactory access to variable data stored in the UKDA than the traditional search interface.

Hypothesis

The purpose of the study was to evaluate the Web search approach, and as such the broad hypothesis was that users would find the process of discovering variable data to be easier using the Web search interface than the traditional search interface. In particular, we hypothesized that the Web search interface would be

- More effective—users would find variables that more accurately matched their research questions, and would have more confidence in the results.
- More efficient—users would be able to find relevant variables more quickly.

- More satisfactory—users would rate the interface as more learnable, easier to use, and generally more satisfactory.

As the study provided an empirical comparison of various approaches to finding variable data, however (ie, neither website tied users to following a single "route" to data discovery), it was expected that the qualitative results in particular would help to identify the features and functionality that participants either liked or disliked and the reasons why, thus contributing to future user interface development.

Tasks

Search tasks were developed in collaboration with the ESDS government team (ESDSG) [44] based at the University of Manchester, and were designed to reflect the kind of research questions that people may seek to answer using the survey data stored in UKDA. ESDSG designed the format for the tasks, and the details of each were decided in a discussion involving both the MethodBox and ESDSG teams. This was part of a wider initiative between medical and social research funding bodies to improve the use of social data in medical research. Participants were asked to find a variable that could be used to answer the following questions:

1. What proportion of people in Scotland believe Jesus was the son of God? *(hereafter referred to as the Belief task)*
2. What proportion of people in Wales speak Welsh fluently?

   [the Welsh task]

1. What proportion of people in Northern Ireland have a bus link to local shops and services?

   [the Transport task]

1. What proportion of the British population have private health insurance?

   (the Health care task)

When they had found a variable that they felt gave a satisfactory answer, they were asked to say so. They were free to stop at any point if they did not think it was possible to find a variable that would answer the question.

Evaluation

As there are numerous ways of accessing variable data using both interfaces, participants started every task on the home page of the site and were free to navigate around and use resources as they wished. Participants were given a number of focused questions, and asked to find data with which to answer them. Searching through surveys to gain a complete picture of all the data available to answer a question would be very time consuming, potentially taking days or weeks [7]. As a proxy measure participants were therefore asked to locate a single variable that provided as complete an answer as possible. All the tasks were completed using both the interfaces, providing a direct comparison between the two.

Experimental Design

Secondary data analysis involves researchers who are not the originators of the data. Such data are conventionally stored in archives so that a wide variety of researchers can access and reuse them. Researchers usually approach the archive with a specific question and relevant variables in mind. In the health and social sciences domain, large, complex population-based surveys are heavily reused in this way. Thus, the speed and quality of the data presentation and the facilitation of variable discovery and high task performance are critical.

To discover which type of interface best supported users undertaking secondary data analysis a repeated measures design was used. Participants searched for data to answer the same 4 questions using both interfaces. Participants were asked to approach each search afresh: that is, to look for any data with which to answer the question and choose what they thought was most useful, rather than to search only for the name of a specific variable or dataset that they knew would provide the answer. Participants completed all the tasks using one of the interfaces first, then had a short break while they answered questions about the experience they had just had, before completing the tasks in the same order using the other interface. The order of the tasks was varied according to a Latin square. The design was counter-balanced, so for every participant who completed the tasks in a given order using the traditional search interface first, another completed the tasks in the same order but using the web search interface first.

Participants

A total of 5 male and 15 female participants between the ages of 18 and 35 took part in the evaluation; 20 users are seen as a satisfactory sample size for understanding human interaction with a software system in this domain [45]. All participants were working or studying in the areas of social science or health science, and had some experience of secondary data analysis. A total of 11 participants had 1-year experience or less, 5 had 2-3-year experience, and 4 had 4 or more years' experience. Participant's previous experience with the particular tools assessed in the evaluation was very limited. Among the study participants, 1 had used both MethodBox and ESDS before a few times, 1 had used MethodBox once, and 5 had occasionally used ESDS. It should be noted that the participants who had previously used MethodBox would have encountered an earlier version with a different user interface. Other online resources participants used to look for data included the Office for National Statistics or Casweb (4 participants), Survey Question Bank (2 participants), medical databases (4 participants), European Data Centre for Work and Welfare (1 participant), and EuroStat (1 participant). A total of 5 participants worked mainly with data they had collected themselves or which came from colleagues or supervisors. Finally, our participants were between 18 and 35 years of age, as we wished to focus on digital natives and thereby make our evidence more portable to future searchers/users. However, we suggest that this focus did not adversely skew our study. While our participants were all within the 18-35-year age group, a prior work [46] showed that as familiarity increases task performance over the age ranges 20-59 harmonizes. In our context, it is unlikely that our user population would include workers much over 65 years of age. Further, even for groups over 60 years of age, no significant age-related differences in tag-based search interfaces (such as our Google-like faceted browsing) have been found [47], although differences have been found in hierarchy-based search (such as our traditional system) [47].

Metrics and Data Collection

Experimental sessions were audio and video recorded and participants' eye movements and mouse movements/keystrokes were tracked using Tobii Studio Professional software. Task completion times and correctness scores were calculated, and participants' behavior and comments during the sessions were documented and analyzed. Eye-tracking data were used to provide insight into situations that could not be understood using the other measures alone; for example, to determine whether a participant was ignoring a matching variable that had appeared, or had not seen it.

After each task, participants were asked to rate, on a scale of 1-7 (with 1 being "not at all" and 7 being "very") how confident they were that they had found a satisfactory answer, and how easy they found it to obtain their answer. After they had completed all the 4 tasks using the single interface, they were asked to rate, on a scale of 1-7 (with 1 being "not at all" and 7 being "very"), how easy they found it to learn how to use the interface, how easy it was to find data using the interface, and their overall satisfaction with the interface. They were also asked to state what they liked and did not like about the interface.

After they had completed the tasks, participants were asked to state which interface they preferred using for finding variable data, and to provide a reason for this.

Set Up and Procedure

The Web search interface provided an alternative view on the data stored in the UKDA, but did not provide access to the same amount of data as the traditional search interface (eg, census data were not available through the web search interface). The questions were designed, so relevant answers could be found in the UK government surveys that can be accessed through both the web search interface and the traditional search interface. Both sites were checked to confirm that at least one variable (the same in each case) containing all the information required to answer each question could be found. Because both sites were live and independently updated, it is possible that they contained other, potentially different matches.

Participants completed a consent form and an entry questionnaire about their previous experience of finding quantitative survey data for secondary analysis. They were then asked to consult the appropriate help documentation for the first interface. For the traditional search interface, this involved reading the online "Help on Searching the Data Catalogue" [44] document (participants were directed in particular to the section on searching for variables) and in the case of the web search interface, reading the "About" [37] page and watching the help video. Participants completed the 4 search tasks using the first interface, providing confidence and ease-of-use ratings after each task, and learnability, overall ease of use, and satisfaction ratings when they had completed all the tasks. They then repeated this process using the second interface. Finally, they stated their preference for either the Web search or traditional search interface, and provided a reason for this.

Statistical Analysis

The usability metrics were analyzed using a task × interface repeated measures generalized linear model (GLM) procedure and the Greenhouse-Geisser correction was made when looking at task effects. For the overall (all tasks) scores of each aspect of usability, the differences between interfaces were summarized with a confidence interval for the mean difference and a paired Student t test P value. The distributions of some metrics were a little asymmetrical, and therefore, sensitivity analyses were performed using alternative permutative nonparametric methods, which gave almost identical results. We present the main effects of the parametric analyses with a 95% confidence interval unless otherwise stated. We use a 5% statistical significance level and 0.1% high significance level. Post hoc pairwise comparisons were adjusted for multiple testing. Calculations were performed in StatsDirect 3.0 and SPSS 19.

Results

Observations

When using the Web search interface participants started each task on the home page containing the main search box. Just over half of the participants used the checkboxes underneath the search box at least once to restrict the search to variables (6 participants) or variables and surveys (7 participants).

Participants looked for variables by entering terms into either the main search box (in the center of the home page and at the top of the page throughout the rest of the site) or the variable search box on the survey tab. A total of 13 participants chose at least one variable after only a single search; the rest of the time participants performed two or more searches before they found an answer they were happy with. A total of 5 participants chose to search within particular surveys at least once; 3 participants reordered the results table at least once and 7 clicked the "show all variables" link for a particular survey, although only 3 ended up choosing a variable from this list, with the rest returning to the search facility. Half of the participants looked only at the first page of results before either choosing a variable or searching again; the other half looked beyond the first page for 1 task (4 participants), 2 tasks (5 participants), or 3 tasks (1 participant). Observation of participants' eye movements showed that they made a decision about whether or not to view a variable's details primarily by glancing at the "description" column of the results table.

In a number of instances, participants found a variable that answered the question quickly, but did not choose it straightaway. A total of 7 participants hesitated to choose a variable because there were several in the results that would answer the question. In 5 other cases, the eye-tracking data showed that participants saw a correct answer in the first set of search results, but spent some time looking around the page or other parts of the site before choosing the variable as their answer.

A total of 3 participants failed to find (in their opinion) a satisfactory variable in 1 task using the web search interface, and 1 failed in 2 tasks.

When using the traditional search interface, 4 participants failed to complete 1 task, 2 failed to complete 2 tasks, 1 failed to complete 3 tasks, and 4 did not complete any.

There was a much greater variety in the way that participants used the traditional search interface. A total of 4 participants used only the variable search and 2 used only the catalogue search; the remaining participants used a combination of the simple search on the home page (11 participants), the variable search (14 participants), and the catalogue search (13 participants). As much as 11 participants used more than 1 search facility within the same task and 14 used different search facilities across different tasks; 7 participants tried Nesstar, but only 1 participant found a satisfactory variable using this tool. A total of 10 participants chose to use the browsing facilities (such as the "browse by subject" page) to access study descriptions, in addition to searching. None of the participants accessed the government variable search, possibly because there were no prominent links to it from the home page or help documentation.

A total of 7 participants consulted the help documentation when using the traditional search interface, compared with 2 when using the Web search interface; 6 participants used the browser's "Ctrl + F" command at some point to locate text within a page with the traditional search interface, whereas only 2 participants used this approach with the Web search interface.

A total of 13 participants chose to look beyond the first page of the results following a search. Because variables had to be located within a list of all the variables in the dataset, it was typical for participants to spend a long time scrolling before they reached the answer.

Performance

The performance measures were the correctness of the results and the time taken to complete the task.

Because the traditional search interface provides access to a greater volume of data (and the Web search interface a subset of these data), it is possible that it may contain more relevant variables, increasing the chance that participants may find a correct answer. However, it is also possible that this may have a negative impact on task completion times, as the larger data collection may take longer to search.

Table 1. Mean (SD) correctness scores for each task.

Task	Web search interface	Traditional search interface
Belief	2.40 (0.88)	1.05 (0.94)
Welsh	2.45 (0.76)	1.65 (1.09)
Transport	1.95 (1.00)	0.55 (0.94)
Health care	2.70 (0.92)	0.85 (1.23)
Overall	2.38 (0.89)	1.03 (1.05)

Correctness was scored out of 3. If participants found any variable containing all the required information, a score of 3 was given; finding a variable that contained most of the information received a score of 2; finding a variable that contained some of the information received a score of 1; failing to find a relevant variable received a score of 0. Participants were not asked to consider year as part of the search criteria. An investigator from each of the MethodBox and ESDSG teams rated correctness, and reached a consensus about the appropriate value where there was disagreement.

Table 2. Mean (SD) completion times in seconds for each task.

Task	Web search interface	Traditional search interface
Belief	159.1 (110.5)	243.5 (159.1)
Welsh	143.9 (80.6)	202.8 (148.9)
Transport	208.0 (161.5)	309.8 (153.7)
Health care	163.0 (100.1)	313.8 (135.9)
Overall	168.5 (113.2)	267.5 (149.4)

Table 1 shows the mean correctness values for each task. A task × interface repeated measures GLM procedure shows a main effect of interface, indicating that answers found through the Web search interface were more likely to be correct ($F_{1,19}=37.3$, $P<.001$) and a main effect of task ($F_{3,57}=6.3$, $P<.001$), with post hoc pairwise comparisons showing that participants obtained significantly lower scores in the transport task than any of the others. There was also a task × interface interaction effect ($F_{3,57}=3.3$, $P=.028$), which reflects the fact that while correctness scores were lowest for both interfaces in the transport task,

scores for the health care task were the second lowest using the traditional search interface, but highest using the search engine interface.

Observations of the search process show that while participants encountered, on aggregate, more than 5 variables that would provide a correct answer in the Welsh task, and more than 20 variables that would answer the health care question, in the case of the belief and transport tasks, all participants who achieved a score of 3 chose the same, single variable, which was the only correct answer to appear during any search. The correctness

XSL•FO
RenderX

results for the Web search interface, which showed participants achieved the highest scores for the health care task, followed by the Welsh, belief, and finally transport tasks, broadly reflect this fact. When using the traditional search interface, however, participants obtained the second lowest score for the health care insurance task, and therefore, the correctness scores do not appear to vary simply as a function of the number of available answers. In fact, the more important factor appears to be the position of the answer in the variable list; by contrast, the answers chosen in the Welsh task were right at the top, the variables relating to health care were much further down, and many participants simply gave up on the dataset before they got to them.

Table 2 shows the mean completion times for each task. A task × interface repeated measures GLM procedure shows that

participants completed the task significantly faster using the Web search interface ($F_{1,19}$=18.0, P<.001). There was also a main effect of task ($F_{2,38}$=4.1, P=.025). Post hoc pairwise comparisons indicate that this was due to the transport and health care tasks taking significantly longer time to complete than the Welsh task.

A task order × interface repeated measures GLM procedure was conducted to check for task order effects. There was a main effect of interface, showing that people completed the tasks significantly faster using the Web search interface ($F_{1,16}$=8.6, P=.01), but order did not have a significant effect at the 5% level ($F_{3,48}$=2.2, P=.1), and there was no interaction effect ($F_{3,48}$=0.6, P=.6), indicating that there was no significant difference in the rate at which participants learned to use the interfaces.

Table 3. Mean (SD) ratings for overall interface learnability, ease of use, and satisfaction.

Variable	Web search interface	Traditional search interface
Learnability	5.55 (0.94)	4.05 (1.23)
Ease of use	5.88 (0.76)	3.70 (1.80)
Satisfaction	5.78 (0.87)	3.15 (1.66)

Overall Ratings

After participants had completed all the tasks using an interface, they were asked to rate on a scale of 1-7 its overall learnability, its overall ease of use, and their overall satisfaction with it (Table 3). Paired comparisons showed that the Web search interface received significantly higher ratings than the traditional search interface for overall learnability (P=.002, 95% CI [0.6-2.4]), ease of use (P<.001, 95% CI [1.2-3.2]), and satisfaction (P<.001, 95% CI [1.8-3.5]). It is interesting to note

that whereas there is only a 1.5-point difference between the traditional search interface and the Web search interface for learnability, for ease of use this rises to 2.2 points, and for overall satisfaction it rises to 2.6 points.

Confidence and Ease-of-Use Ratings for each Task

After completing each task, participants rated on a scale of 1-7 how confident they were that the variable they had found answered the question, and how easy it was to find the answer.

Table 4. Mean (SD) confidence ratings for each task.

Task	Web search interface	Traditional search interface
Belief	5.65 (1.90)	2.73 (2.47)
Welsh	4.78 (2.00)	4.30 (2.62)
Bus	4.85 (1.87)	1.78 (2.26)
Health care	4.63 (1.99)	2.25 (2.00)
Overall	4.98 (2.77)	1.94 (2.34)

Table 4 shows the mean confidence ratings for each task. A task × interface repeated measures GLM procedure indicates that participants were significantly more confident about their answers when using the Web search interface ($F_{1,19}$=18.8, P<.001). Post hoc pairwise comparisons show that participants were significantly more confident about their answers in the Welsh task than in the transport or health care tasks, and significantly less confident about their answers in the health care task than in the belief task ($F_{2,38}$=4.7, P=.015). A task × interface interaction effect ($F_{3,57}$=4.4, P<.01) indicates that the

confidence rating varied according to the interface: in the Welsh task, participants had a similar level of confidence in their answer, but for all other tasks it was much higher when using the Web search interface.

Table 5 shows the mean ease-of-use ratings for each task. A task × interface repeated measures GLM procedure shows that participants found the Web search interface significantly easier to use ($F_{1,19}$=14.0, P<.001). There was no significant effect of task ($F_{3,57}$=2.2, P=.1).

Table 5. Ratings: mean (SD) ease of use ratings for each task.

Task	Web search interface	Traditional search interface
Belief	5.08 (1.78)	2.95 (2.39)
Welsh	4.85 (1.87)	3.88 (2.42)
Bus	5.03 (1.92)	1.90 (2.31)
Health care	4.70 (2.11)	2.68 (2.19)
Overall	4.92 (1.92)	2.85 (2.33)

Qualitative Feedback

Participants were asked for qualitative feedback at two points: after they had completed all the tasks with an interface, they were asked to say what they liked and disliked about it; and at the end of the study, they were asked which interface they preferred and why. In addition, participants made occasional remarks about the interfaces while they were completing the tasks; these comments are also included in the analysis that follows.

Completed Tasks and Remarks

A total of 18 participants said they preferred using the Web search interface to search for and access variables; 9 participants stated this was because it was more user-friendly or easier to use. According to a participant

It's easier to find variables and the information is clearer. In [the Traditional Search interface], the information is in another file or in another link. [In the Web Search interface] it's just there so I can see it easily. [Participant Number 18, Female].

A total of 7 participants mentioned that the search process was quicker when using the Web search interface

> *It's so much faster. You'd just get so annoyed with [the traditional search interface] because of the amount of effort.* [Participant Number 15, Female]
>
> *When I searched for something, I was able to see whether the results were relevant more immediately than with [the traditional search interface].* [Participant Number 3, Female]
>
> *When all the information came up I was able to scan it quickly and see, well this one is relevant and this one isn't.* [Participant Number 14, Female]

A total of 4 participants described the Web search interface as more intuitive

> *The format of the site means it's more intuitive how to get around it, how to find stuff.* [Participant Number 12, Male]

One of the participants provided the following reason for liking the Web search interface *I could find what I was looking for.* [Participant Number 2, Female]

A total of 7 participants commented on the simplicity of the interface.

> *It's easy because you can just search one comprehensive way rather than spending time*

debating which method you're going to use to actually look for your data [Participant Number 9, Female]

One participant said this could undermine confidence in the interface, however

> *I definitely preferred [the Web search interface], but I know this might sound weird but because it was so easy you worry that what you've done is not right, or it's not reliable* [Participant Number 17, Male]

P7F said that although she preferred the output of the search process in the Web search interface, she preferred using the catalogue search of the traditional search interface to specify search terms:

> *[the traditional search interface] felt a bit more open, whereas this [the Web search interface home page]—everything's hidden behind it. I felt happier with searching with the traditional search interface.* [Participant Number 7, Female]

Another participant, *Participant Number 20, Female*, who said she preferred using the traditional search interface, also cited the catalogue search facility as the reason, saying "it allows you to provide more details and filter the search."

Postsystem Interview

In the postsystem interview, the Web search interface received 35 positive and 17 negative comments, whereas the traditional search interface received 12 positive and 25 negative comments. A total of 6 participants said that they found the Web search interface easy to use, and 6 commented on its speed and simplicity:

It's faster than [the traditional search] interface—you get the same results with fewer clicks [Participant Number 13, Male].

Two described it as "user-friendly":

[the Web search interface] is probably more user-friendly because [the Web search interface] is pretty much like the Google one, so the user may be more familiar with this kind of searching method. [Participant Number 11, Male].

An additional 2 participants compared the search facility favorably with Google and 4 others liked the simple, familiar format.

> *It just seemed so easy, normal—[an] Internet search engine but with a different purpose* [Participant Number 17, Male]

However, some disliked the simplicity of the search box:

I didn't really like the fact that [the main search page] was it. I couldn't automatically do a date or a region filter [Participant Number 7, Female]

As much as 9 participants said they liked the format of the results. According to one participant,

The returns I got were more helpful than with [the traditional search interface]...I think [the Web search interface] has a better grasp of what researchers actually want, so I liked the fact that once you'd got your search returns it said what exactly was the wording of the question, and what category that came under, because sometimes a question will have a different meaning if it's asked under demographics, or asked under some other category, so you might just look and say, "that's not relevant" [Participant Number 7, Female]

A total of 4 participants liked the fact that you could search for, or within, particular surveys and one said she liked the help video.

The negative comments about the Web search interface in the interview were mainly related to the description of particular variables. A total of 4 participants commented that values for some variables did not seem to be available:

I disliked the fact that some of the variables did not seem to have information in—that confused me. I don't know whether that means they're searching datasets they don't have information for? That could be made clearer. [Participant Number 3, Female].

Of the study participants, 2 found the appearance of many variables with the same title confusing, and one felt that the wording of some variable titles was unclear

Some of the questions said things like, "Bus stop, feel, don't know"...I don't know what that question means [Participant Number 8, Female].

However, P8F did recognize this as a potential problem with the survey, rather than with the interface. Also, 2 participants commented that it was not always clear which year variables applied to, and one wondered about the geographical location of the study, which was not apparent just from looking at the variable description.

It was suggested by 2 participants that the Web search interface returned too many variables in the search results, although

that can be managed if you sort them according to which survey they are taken from etc. [Participant Number 13, Male].

One of the participants lacked confidence in the search due to the simplicity of the interface

It's a little less transparent as to what's in the box...I think I'm doing the right thing but I'm not sure. [Participant Number 7, Male]

There were also 2 other participants who found it hard to find the keywords to bring up the required data. One participant commented on the fact the "back" button did not work properly and another did not like the format of the help documentation.

When asked what they liked about the traditional search interface, two key areas came up. A total of 4 participants found the extensive help documentation useful and 5 liked the options provided for filtering results:

It's easier to have a general idea of categorizing topics and areas...you're more likely to exclude something that is not what you want, or include what you want. [Participant Number 11, Male].

Another participant commented,

It looked a lot more professional than [the Web Search interface]. I got the impression it had access to a lot more data. [Participant Number 2, Female].

One participant also mentioned that she found the Nesstar tool helpful (P5F).

When asked what they disliked about the traditional search interface, 3 participants said they found it complicated or hard to use, one described it as less intuitive than the Web search interface and one said it was slower. As much as 4 participants said they found it difficult to get any useful results at all and 8 said their searches returned too much information. One participant mentioned,

You felt that what you got out was quite vague, or not to the point of what you wanted. It just seemed to come up with all sorts of stuff that was completely irrelevant and just wasn't very helpful. Because it would bring up so many items you couldn't really go through them. [Participant Number 17, Male]

Another suggested,

I think it would have been really useful, if they brought up say 200 datasets, if the variables you were actually looking for were highlighted in the small amount of text you've got underneath the heading, because then you can make a judgement. [Participant Number 3, Female]

A total of 5 participants complained about the fact that the results did not give you direct access to the variable data:

I thought I'd worked it out then realized I hadn't. It wasn't easy going from one step to another—it was kind of frustrating. [Participant Number 8, Female]

According to another participant,

There's too much supplementary material before you knew whether that was what you were really looking for or not. [Participant Number 19, Female].

One participant complained that there was no option for sorting the results (Participant Number 13, Male). Another said that it was odd that the variables search was so limited, when compared with the catalogue search:

it just gave you a single box...it didn't give you the ability to search by region and keywords [Participant Number 7, Female]

One participant noted that it was difficult to choose how to look for data: *I'd start searching one way, then I'd think, maybe I should search that way...* [Participant Number 9, Female]

As participants were completing the tasks, they were more likely to make negative comments than positive ones. The Web search interface received 6 positive comments, 4 of which stated that the interface was easy to use. P19F commented on the fact that she had successfully found a variable, and P7F said that she found it helpful to be able to see what a variable contained "without having to go into it." The only 2 positive comments to be made about the traditional search interface also came from P7F, who said that the catalogue search seemed more efficient than the variable search, and after the second task,

> *I found the searching slightly easier this time because I'd got the hang of it.* [Participant Number 7, Female]

Of the 13 negative comments made about the Web search interface while people were completing the tasks, 5 were due to bugs or errors, including the help video being of an inadequate resolution (1 participant), the back button not working correctly (1 participant), and terms in quotation marks that contained white space not being found (3 participants). Besides, 2 participants commented on the lack of an advanced search facility, of the type provided by the traditional search interface, and 2 disliked the fact that a search did not match only complete words (eg, a search for "bus" would return results with "transport" in the variable title). A total of 3 participants commented on the presence of what looked like ghost variables in the results:

> *It's confusing me now. I found what I think are the ones I was looking at before [in the traditional search interface] but when I actually click on it, it's saying that there's no source, no metadata, no value, so I don't really know whether I have found it. I still found the data easily but I've got no idea.* [Participant Number 3, Female]

Of the 16 negative comments participants made while they were using the traditional search interface, 4 were people expressing their dislike for or frustration with the system: *I am actually just getting really annoyed now* [Participant Number 15, Female].

A total of 5 participants commented that it was taking too long to find a variable, and 3 said they were confused or finding the process too difficult. Some participants (n=4) also complained about the format of the search results, including the fact that the variable search returned surveys, rather than taking you directly to the relevant variables (2 participants), and the fact that certain survey years did not seem to appear in the results when they were known to exist (2 participants). P7F also complained about the fact that you could not search within a survey for variables, saying that the variable lists for the datasets were "an awful lot to try and read through."

Discussion

This study has shown that the functionality provided by the Web search interface was preferred to that of the traditional search interface for finding variables in research data archives. Participants were more likely to find a variable that correctly answered the question posed by a task. In addition, they were able to do this more quickly and had more confidence in the results. They found the Web search interface easier to use, and were more satisfied with the overall experience it provided.

We now consider these findings in the context of the wider evidence base: specifically, the merits of leveraging the Web search approach to help users find variables, considering it within the context of HCIR literature.

Query formulation and query reformulation strive to put control of selection and interpretation of results in the user's hands. This is accomplished by allowing the user to quickly formulate and reformulate the query as their understanding of the search domain increases based on the results returned. The Web search interface appears to support this well:

It was simple to use, cause I just used keywords, and I used the same keywords in the other thing and it couldn't find it... [Participant Number 14, Female]; and

> *It's a quick way of finding what variables there are...if you were just looking at say pay, and you just wanted to look at income...I worked on a project looking at minimum wages and things like that, so we mainly use EUROSTAT, but if you could search for something...[the Web search interface] would have been really useful for that kind of thing.* [Participant Number 9, Female]

These examples indicate the broader feeling that traditional search interfaces require a more precise conceptualization of what is required and available for search. By contrast, the Web search interface relies far less on the users' knowledge of the base data and so is better for variable search.

Browsing is generally considered to involve virtually no planning, preparation, or focus. This kind of interaction is common in Web search and is related to query formulation and reformulation, requiring less initial knowledge of the data available. Participants' comments suggest the interface supported this activity:

> *It seemed to be an easier step between the search term and the list of variables* [Participant Number 5, Female]
>
> *When all the information came up I was able to scan it quickly, and see well this one is relevant and this one wasn't* [Participant Number 14, Female]

Browsing for relevance appears to be key to the variable data discovery performed in the study.

Faceted search and navigation enables users to group and interact with information hierarchically, and is becoming both expected and critically important for refining search results. Participants commented that being able to limit results in the Web search interface was useful:

> *I like the different layers of options, so if I search for some different variables and some surveys clearly won't be at all relevant I can select them out.* [Participant Number 3, Female]
>
> *I like that you could click to look for particular surveys or particular variables. So for example if you're looking for something in Northern Ireland*

(NI), I could choose NI surveys and exclude everything else so I don't get swamped with variables, because if you do type in just one thing you get an awful lot of answers coming back and you could get quite lost, so that seemed good. [Participant Number 8, Female]

Surrogates are the titles and abstracts for documents; thumbnails for Web pages, etc, which can be seen interspersed within the search results of modern Web search interfaces. Indeed, amalgamations of surrogates can be seen in many Google searches with documents and information being displayed from Wikipedia and from more general image searches. Surrogates in our Web search interface were limited to the title of the survey. One participant commented that

It helps to know where [the variable] is from. [Participant Number 4, Male]

Relevance feedback modifies an existing query based on available user-based relevance judgments for previously retrieved documents. One participant commented that

in the title of the [variables] they have a lot of information there so it is easy to know when you have found it. [Participant Number 18, Female]

Summarization, analytics, and visual presentation can enable users to better digest the query result, and formulate queries in a familiar interface. Indeed, we received many positive comments on this part of the interaction design: One noted that

Pretty easy [to learn how to use]—it's like a basic search engine and *I like the layout; everyone knows how to use Google so everyone can find the variable they want.* [Participant Number 2, Female]

Another participant commented,

I liked it. It was like a Google search really. It's very familiar—it's like Google search. You just put in all the search terms and it gives up the list, rather than having to go through all the different stages of digging through the literature. [Participant Number 15, Female]

One felt that

It just seemed so easy, normal—[an] Internet search engine, but with a different purpose. [Participant Number 17, Male]

A "Web Search Interface" for Research Variables

The Web Search interface was designed to simplify the process of finding and extracting variables for secondary research. Providing a familiar look, feel, and functionality was a key goal of the design process. It was designed as a Web search engine—set within a scientific social network where users can share methods for relating, extracting, and manipulating data—to take advantage of the fact that the most familiar experience of finding information for its target users will have come from the Web.

Participants were very positive about this approach, stating explicitly that they liked the fact that it resembled a familiar Web search engine; other participants commented on how quick,

user-friendly, simple, and intuitive it felt. Although both interfaces provided a single-box entry system, only the Web search interface provided users with the look, feel, and functionality of a search engine like Google.

It was not only the simplicity of the landing page that was behind the Web search interface's success, but also the format of the results. With the exception of the more detailed catalogue search, the facilities for entering search terms—a single box—were very similar for both interfaces. Whereas the Web search interface presented users with a list of matching variables, the traditional search interface provided a list of surveys, with at least one further click, and possibly some scrolling, required to reach the variable of interest. For some participants, this simply made the task more time consuming. For others, it made it impossible: users expected to see what they were searching for straightaway and when they could not they assumed that something had gone wrong, either with the search process itself or the way they were using it. The traditional search interface was designed for the retrieval of variable data the user already knew to exist; only the Web search interface truly supported the discovery of new data.

This tells us that presenting relevant results immediately after a search is very important: if a user is searching for a variable, and gets back a survey or dataset, they find this confusing. What else can we learn about the format that the results should take? It is not possible to determine from this study whether the Web search interface takes the optimal approach to formatting variable search results, but there is evidence that it uses at least an adequate one, as participants were able to find a variable that mostly or completely matched the search criteria the majority of the time. In fact, participants' comments indicated that they were very happy with the presentation of the results. The majority said that they liked it, and those who specified why focused on the fact that the relevance (or otherwise) of the variable could be seen at a glance. The eye-tracking data show that participants made the decision about variable relevance primarily by looking at the "description" column of the results, and therefore, providing a summary of what the variable contains, and not just its title, appears to be important.

Improving the Web Search Interface

Although participants preferred the Web search interface, it did receive some negative feedback. In several instances this was the result of a bug (eg, the back button not working properly), but it also resulted from the fact that incomplete or seemingly inaccurate variable data were sometimes returned in the results. It is likely that these problems were caused not by an error in the interface itself, but by the original format of the data in question, which was confusing. Nevertheless, ensuring that users fully understand the results that are returned to them, and why they appear as they do, remains a usability challenge for systems providing access to this type of data.

As practicing health science professionals our users have some experience using the traditional interface, but they do not have many years' experience. We also see that they will be more familiar with a Web search interface as this is what they use most days of the week and often multiple times a day. Conforming to this model should be our primary goal as there

is a much more familiarity and there is little refamiliarization required, as there is with a traditional but seldom used, interface.

Study Limitations

As the study was conducted with live websites, it was not completely controlled. The traditional search interface searched a larger data catalogue than the Web search interface, and although this meant that it potentially produced a greater number of correct answers, this did not appear to provide any advantage from the perspective of correctness scores. The time it took a search to complete varied considerably for both interfaces, from less than a second to (occasionally) more than 10. Search times were not deducted from task completion times for a number of reasons: it was not always possible to determine how long a search took (participants often opened a new window over the top to continue with the task when there was a delay); the time to access the server is a property of the system, and as such it may not be appropriate to ignore it; searches rarely took longer than a few seconds. Nevertheless, it should be noted that the task completion times recorded for the Web search interface may rise if the proportion of the data catalogue it provides access to and/or the number of people using it increases.

A second limitation is that the study was conducted with relatively inexperienced researchers. This particular group was used because they are known to have difficulties with data discovery. There may be circumstances where the facilities provided in the traditional search interface are preferable to researchers with more experience, or those currently provided in the Web search interface are not sophisticated enough. Because of the ubiquity of Web search and our participants' constant exposure to this, we might assume that they would be better at Web search.

A third limitation of the study was that it compared only two tools for accessing variable data. This study is one of the first to investigate user preferences for finding and accessing variable data; further work considering other tools or methods is undoubtedly necessary. It would also be useful to examine researchers' preferences in longitudinal, naturalistic settings, as well as controlled, laboratory-based studies.

Study Impact

Previously, researchers working with survey data from the UKDA found it difficult to discover relevant variables for analysis. These difficulties were compounded as single-domain researchers became cross-domain data scientists. To address these difficulties, a Web search interface (MethodBox) was designed as an alternative front end to the archive, enabling users to search through multiple sets of data, supporting documentation and user-contributed metadata in a single process. Since this study, the main UKDA search interface has been significantly overhauled to take account of the findings (Figure 9) and the new-way data are used and searched for.

The new "Discover" Variable and Question Bank interface adopted the Web search interface paradigm described by HCIR and this was shown to be effective for variable search in this study. The interface implements all those features found to be useful to researchers, including faceted search; query reformulation; browsing; surrogates; relevance feedback; summarization, analytics, and visual presentation (Figure 10). The main aim of the study was to empirically support the anecdotal supposition that data scientists share more in common with the "Google Generation" than with their single-domain, single-tool forebears. We studied this with real applications built directly because of this anecdotal supposition; the evaluation of the MethodBox Web search interface provided empirical support for this supposition, which has implications for scientific data search and selection more generally. We have shown that users find the Web search engine approach intuitive and that it helps them to assemble relevant variable data for research. The findings apply not only to MethodBox but also to similar systems that support the need to search for variable data.

The implications of this study for the process of secondary data analysis are substantial. Many researchers, particularly inexperienced ones, or cross disciplinarians, struggle to identify the datasets and variables they should be using to answer a research question. By enabling users to quickly search a data archive at the level of recorded factors/variables, information systems can help users to focus on research rather than on the process of negotiating archives or documents.

A straightforward means of searching provides a greater opportunity for finding relevant factors/variables that the researcher was not previously aware of. This may reduce "investigator bias," whereby research artificially focuses on familiar datasets, but not necessarily those most relevant to the hypothesis.

The simple provision of a Web search interface will not ultimately eliminate the need for researchers to "get to know" a dataset in detail, but it could make the process of data discovery quicker, easier, and far less intimidating. In turn, this may generate "digital crumbs" of metadata about the relationships between variables, users, and research processes. Such metadata may eventually support crowd-sourced secondary research.

Figure 9. New UKDA Interface -- 'Discover' adopting the Web Search Interface and including: Faceted Search (at left); Query Reformulation (at centre); and Browsing (at centre) features.
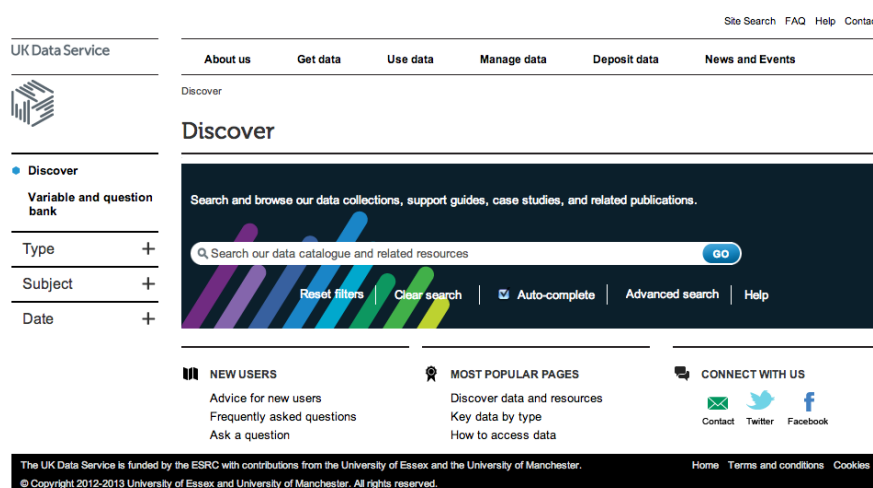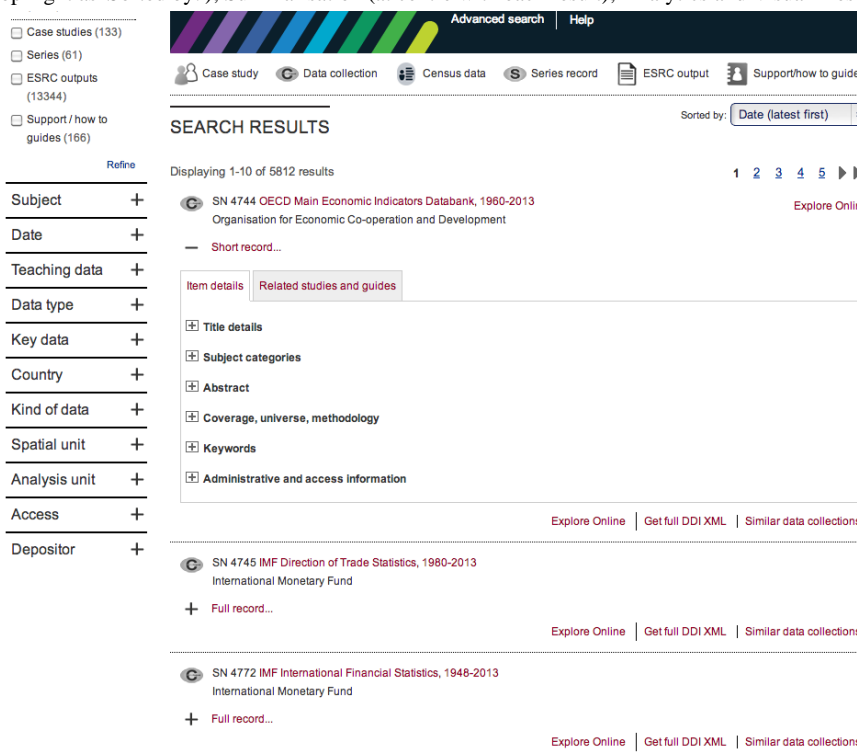

Figure 10. Discover Results -- Search results adopting the Web Search Interface and including: Surrogates (at centre as part of 'Full Record' detail); Relevance Feedback (at top-right as 'Sorted by:'); Summarisation (at centre with each result); Analytics and Visual Presentation (at top-left) features.

Panton Principles and the Science Code Manifesto

Science is based on building on, reusing, and openly criticizing the published body of scientific knowledge. For science to effectively function, and for society to reap the full benefits from scientific endeavors, it is crucial that scientific data be made open. In this case, we support the "Panton Principles" [48]. We further assert that "Code is Method" and likewise support the Science Code Manifesto [49]. In this case, we would like to invite you to access our data and code, and question our analysis and interpretation of that data via the full dataset, experimental protocols, and methodologies [50].

Acknowledgments

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Law M. Reduce, reuse, recycle: Issues in the secondary use of research data. IASSIST Q 2005;29(1):5-10.
2. Fienberg S, Martin M, Straf M. Sharing Research Data. Washington, DC: National Academy Press; 1985.
3. 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]
4. ESRC. The National Data Strategy (2009-2012). London, UK: ESRC; 2009. URL: https://www2.warwick.ac.uk/fac/soc/ier/publications/2009/elias_nds_2009.pdf [accessed 2016-01-10] [WebCite Cache ID 6eRC3vvES]
5. RCUK. UK Data Archive: RCUK. 2015 Jun. URL: http://www.data-archive.ac.uk/ [accessed 2015-06-25] [WebCite Cache ID 6ZY5zJa4l]
6. UKDA. Economic and Social Data Service. 2015 Jun. URL: http://www.esds.ac.uk/ [accessed 2015-06-25] [WebCite Cache ID 6ZY6aoBAf]
7. Thew S, Sufi S, Dunlop I, Hiroeh U, Canoy D, Georgina M. MethodBox: Supporting inter-disciplinary public health research. 2010 Presented at: All Hands Meeting Cardiff; 2010; Cardiff, UK.
8. Freese J. Secondary analysis of large social surveys. In: Research Confidential: Solutions to Problems Most Social Scientists Pretend They Never Have. Ann Arbor, MI: University of Michigan Press; 2009:238.
9. Sanderson M, Croft W. The History of Information Retrieval Research. Proc IEEE 2012 May;100:1444-1451. [doi: 10.1109/JPROC.2012.2189916]
10. Smyth B, Balfe E, Freyne J, Briggs P, Coyle M, Boydell O. Exploiting query repetition and regularity in an adaptive community-based web search engine. User Model User-Adapt Interact 2004;14(5):383-423.
11. Marchionini G. Toward human-computer information retrieval. Bul Am Soc Info Sci Tech 2007 Sep 20;32(5):20-22. [doi: 10.1002/bult.2006.1720320508]
12. UK Government. Office for National Statistics. 2015 Jun. URL: http://www.ons.gov.uk/ons/index.html [accessed 2015-06-25] [WebCite Cache ID 6ZY8TBn6m]
13. European Union. Eurostat. 2015. URL: http://ec.europa.eu/eurostat [accessed 2015-06-25] [WebCite Cache ID 6ZY9ON3YS]
14. US Department of Labour. Bureau of Labor Statistics [About]. 2015. URL: http://www.bls.gov/data/ [accessed 2015-06-25] [WebCite Cache ID 6ZY9hwsm3]
15. University of Michigan. Inter-University Consortium for Political and Social Research. 2015 Jun. URL: http://www.icpsr.umich.edu/icpsrweb/landing.jsp [accessed 2015-06-25] [WebCite Cache ID 6ZYAPTOqm]
16. RAND. Rand Survey Metadata Repository. 2015. URL: https://mmicdata.rand.org/megametadata/ [accessed 2015-06-25] [WebCite Cache ID 6ZYAZZMZl]
17. Torres-Parejo U, Campaña J, Delgado M, Vila M. MTCIR: A multi-term tag cloud information retrieval system. Expert Syst Appl 2013;40:5448-5455.
18. Aknouche R, Asfari O, Bentayeb F, Boussaid O. Integrating query context and user context in an information retrieval model based on expanded language modeling. In: Multidisciplinary Research and Practice for Information Systems. Berlin, Germany: Springer; 2012:244-258.
19. Broder A. A Taxonomy of Web Search. SIGIR Forum Internet New York, NY. New York: ACM; 2002:3-10.
20. Rose D, Levinson D. Understanding user goals in web search. In: Proceedings of the 13th International Conference on World Wide Web Internet. New York: ACM; 2004 Presented at: 13th International Conference on World Wide Web; March 2004; New York p. 13-19. [doi: 10.1145/988672.988675]
21. Catarci T, Kimani S. Human-computer interaction view on information retrieval evaluation. In: Information Retrieval Meets Information Visualization. Berlin: Springer; 2013:48-75.
22. Jansen B, Spink A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Inf Process Manag 2006 Jan;42(1):248-263 [FREE Full text] [doi: 10.1016/j.ipm.2004.10.007]
23. Varian H. Royal Statistical Society Annual Conference. London, UK: Royal Statistical Society; 2013. Statistics at Google URL: http://www.statslife.org.uk/news/227-hal-varian-s-statistics-at-google-talk-now-available-to-view [accessed 2015-12-18] [WebCite Cache ID 6ds1Q5z2d]
24. Gong X, Ke W, Zhang Y, Broussard R. Interactive search result clustering: A study of user behavior and retrieval effectiveness. New York: ACM; 2013 Presented at: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries; 2013; Indianapolis, IN. [doi: 10.1145/2467696.2467726]
25. Kobayashi M, Takeda K. Information retrieval on the web. ACM Comput Surv 2000;32(2):144-173. [doi: 10.1145/358923.358934]
26. Rowlands I, Nicholas D, Williams P, Huntington P, Fieldhouse M, Gunter B, et al. The Google generation: The information behaviour of the researcher of the future. ASLIB Proc 2008 Jul 06;60(4):290-310 [FREE Full text] [doi: 10.1108/00012530810887953]

27.   Zhang Y. Undergraduate students' mental models of the web as an information retrieval system. J Am Soc Inf Sci 2008 Nov;59(13):2087-2098. [doi: 10.1002/asi.20915]

28.   Spring H. Health professionals of the future: Teaching information skills to the google generation. Health Info Libr J 2010;27(2):158-162. [doi: 10.1111/j.1471-1842.2010.00885.x]

29.   Zhang H, Durbin M, Dunn J, Cowan W, Wheeler B. Faceted search for heterogeneous digital collections. New York: ACM; 2012 Presented at: Proceedings of the 12th aCM/iEEE-cS joint conference on digital libraries; 2012; New York p. 425-426. [doi: 10.1145/2232817.2232924]

30.   Wei B, Liu J, Zheng Q, Zhang W, Fu X, Feng B. A survey of faceted search. J Web Eng 2013;12(1-2):41-64 [FREE Full text]

31.   Gong X, Ke W, Khare R. Studying scatter/gather browsing for web search. Proc Am Soc Inform Sci Technol 2013 Jan 24;49(1):1-4. [doi: 10.1002/meet.14504901328]

32.   Kules B, Capra R. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. J Am Soc Inf Sci 2011 Oct 12;63(1):114-138. [doi: 10.1002/asi.21647]

33.   Mislove A, Gummadi K, Druschel P. Exploiting social networks for internet search. 2006 Nov 29 Presented at: 5th Workshop on Hot Topics in Networks (hotNets06); 2006; Irvine, CA URL: http://www.read.cs.ucla.edu/hotnets5/program.pdf#page=79

34.   Chi EH. Information seeking can be social. Computer 2009 Mar;42(3):42-46. [doi: 10.1109/MC.2009.87]

35.   Horowitz D, Kamvar S. The anatomy of a large-scale social search engine. New York: ACM; 2010 Presented at: Proceedings of the 19th International Conference on World Wide Web; 2010; New York p. 431-440. [doi: 10.1145/1772690.1772735]

36.   Zamir O, Etzioni O. Grouper: A dynamic clustering interface to Web search results. Comput Netw 1999 May;31(11-16):1361-1374. [doi: 10.1016/S1389-1286(99)00054-7]

37.   University of Manchester. MethodBox. 2015. URL: https://www.methodbox.org/ [accessed 2015-06-25] [WebCite Cache ID 6ZYAi1HIU]

38.   University of Manchester. Obesity eLab. 2015. URL: http://www.mygrid.org.uk/projects/obesity-e-lab/ [accessed 2015-06-25] [WebCite Cache ID 6ZYAvX06J]

39.   Markey K. Twenty-five years of end-user searching, Part 1: Research findings. J Am Soc Inf Sci 2007 Jun;58(8):1071-1081. [doi: 10.1002/asi.20462]

40.   Robertson S. On the history of evaluation in IR. J Inform Sci 2008 Jun 13;34(4):439-456. [doi: 10.1177/0165551507086989]

41.   Novotny E. I don't think I click: A protocol analysis study of use of a library online catalog in the Internet age. Coll Res Libr 2004 Nov 01;65(6):525-537. [doi: 10.5860/crl.65.6.525]

42.   DDI Alliance. Data Documentation Initiative. 2015. URL: http://www.ddialliance.org [accessed 2015-06-25] [WebCite Cache ID 6ZYB5Oq0G]

43.   Lost Island Labs-University of Manchester. Utopia. 2015. URL: http://getutopia.com [accessed 2015-06-25] [WebCite Cache ID 6ZYBEEodS]

44.   UKDA. Economic and Social Data Service Search. 2015. URL: http://www.esds.ac.uk/search/searchHelpESDS.asp [accessed 2015-06-25] [WebCite Cache ID 6ZYBZq33l]

45.   Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [Medline: 14587545]

46.   Czaja S, Sharit J, Ownby R, Roth D, Nair S. Examining age differences in performance of a complex information search and retrieval task. Psychol Aging 2001;16(4):564-579 [FREE Full text] [doi: 10.1037/0882-7974.16.4.564]

47.   Pak R, Price MM, Thatcher J. Age-sensitive design of online health information: Comparative usability study. J Med Internet Res 2009;11(4):e45 [FREE Full text] [doi: 10.2196/jmir.1220] [Medline: 19917549]

48.   Panton P. Panton Principles: Principles for Open Data in Science. 2015. URL: http://pantonprinciples.org/ [accessed 2015-06-25] [WebCite Cache ID 6ZYCAnEjG]

49.   Climate Code Foundation. Science Code Manifesto. 2015. URL: http://sciencecodemanifesto.org/ [accessed 2015-06-25] [WebCite Cache ID 6ZYCMZjiH]

50.   IAM - University of Manchester. MethodBox Data Set. 2015. URL: http://iam.cs.manchester.ac.uk/projects/methodbox/ [accessed 2015-06-25] [WebCite Cache ID 6ZYC0wmfQ]

Abbreviations

DDI: Data Documentation Initiative
ESDS: Economic and Social Data Service
ESDSG: ESDS government team
EU: European Union
GLM: generalized linear model
HCIR: Human-Computer Information Retrieval
IPCSR: Inter-University Consortium for Political and Social Research
UKDA: UK Data Archive

XSL•FO
RenderX

Original Paper

Providing Doctors With High-Quality Information: An Updated Evaluation of Web-Based Point-of-Care Information Summaries

Koren Hyogene Kwag[1], BSc, MSc; Marien González-Lorenzo[1], MSc Psych, PhD; Rita Banzi[2], PharmD, PhD; Stefanos Bonovas[3], MD, MSc, PhD; Lorenzo Moja[1,4], MD, MSc, PhD

[1]Clinical Epidemiology Unit, IRCCS Galeazzi Orthopedic Institute, Milan, Italy

[2]Laboratory of Drug Regulatory Policies, IRCCS Mario Negri Institute for Pharmacological Research, Milan, Italy

[3]Humanitas Clinical and Research Center, Milan, Italy

[4]Department of Biomedical Sciences for Health, University of Milan, Milan, Italy

Corresponding Author:
Lorenzo Moja, MD, MSc, PhD
Department of Biomedical Sciences for Health
University of Milan
Via Pascal 36
Milan, 20133
Italy
Phone: 39 02503 ext 15097
Fax: 39 0250315102
Email: lorenzo.moja@unimi.it

Abstract

Background: The complexity of modern practice requires health professionals to be active information-seekers.

Objective: Our aim was to review the quality and progress of point-of-care information summaries—Web-based medical compendia that are specifically designed to deliver pre-digested, rapidly accessible, comprehensive, and periodically updated information to health care providers. We aimed to evaluate product claims of being evidence-based.

Methods: We updated our previous evaluations by searching Medline, Google, librarian association websites, and conference proceedings from August 2012 to December 2014. We included Web-based, regularly updated point-of-care information summaries with claims of being evidence-based. We extracted data on the general characteristics and content presentation of products, and we quantitatively assessed their breadth of disease coverage, editorial quality, and evidence-based methodology. We assessed potential relationships between these dimensions and compared them with our 2008 assessment.

Results: We screened 58 products; 26 met our inclusion criteria. Nearly a quarter (6/26, 23%) were newly identified in 2014. We accessed and analyzed 23 products for content presentation and quantitative dimensions. Most summaries were developed by major publishers in the United States and the United Kingdom; no products derived from low- and middle-income countries. The main target audience remained physicians, although nurses and physiotherapists were increasingly represented. Best Practice, Dynamed, and UptoDate scored the highest across all dimensions. The majority of products did not excel across all dimensions: we found only a moderate positive correlation between editorial quality and evidence-based methodology (r=.41, P=.0496). However, all dimensions improved from 2008: editorial quality (P=.01), evidence-based methodology (P=.015), and volume of diseases and medical conditions (P<.001).

Conclusions: Medical and scientific publishers are investing substantial resources towards the development and maintenance of point-of-care summaries. The number of these products has increased since 2008 along with their quality. Best Practice, Dynamed, and UptoDate scored the highest across all dimensions, while others that were marketed as evidence-based were less reliable. Individuals and institutions should regularly assess the value of point-of-care summaries as their quality changes rapidly over time.

(J Med Internet Res 2016;18(1):e15)   doi:10.2196/jmir.5234

KEYWORDS

point-of-care summaries; internet information; evidence-based medicine; information science

XSL•FO
RenderX

Introduction

Pressed for time and obliged to navigate ever-expanding medical literature, doctors are increasingly relying on online information tools to accelerate the search process without compromising the reliability and quality of information retrieved. Point-of-care information summaries offer predigested syntheses of medical research intended to be used when the patient and physician interact (ie, point-of-care) [1]. Web-based point-of-care summaries provide user-friendly interfaces that may improve the retrieval, synthesis, organization, and application of evidence-based content in clinical practice [2,3].

The medical information technology market parallels the efforts by national health systems to streamline clinical workflow and align clinicians' behavior with best practice strategies. Point-of-care summaries play a central role: they increasingly form the knowledge basis of complex information systems, such as computerized physician order entry and computer decision support systems [3-5]. In the United States, the Health Information Technology for Economic and Clinical Health (HITECH) Act requires clinicians and hospitals to integrate electronic health records (EHRs) with clinical decision support rules relevant to a specialty or to high-priority hospital conditions, such as drugs and diagnostic test ordering [6]. In Europe, the integration of point-of-care summaries into the workflow of the prescribers is under scrutiny in several countries [7-10].

As point-of-care information summaries gain ground in the culture of medical practice as stand-alone products or integrated with other systems, their validity must be assessed against marketing claims that they are evidence-based. This review examines the quality of Web-based point-of-care information summaries and their development and progress since 2008.

Methods

Inclusion and Exclusion Criteria

As this is an update of analyses done in 2008 [11] and 2012 [12], detailed methods and operational definitions can be found in the original publication [11]. Briefly, we defined point-of-care information summaries as "Web-based medical compendia specifically designed to deliver predigested, rapidly accessible, comprehensive, periodically updated, and evidence-based information (and possibly also guidance) to clinicians." To be included in this review, a product had to be an online-delivered tertiary publication (summary) that is regularly updated, claims to provide evidence-based information to physicians and other professionals, and is intended for use at the bedside. We considered summaries, regardless of their content development status, number of years on the market, clinical focus or specialty, type of access, or charging agreements. We excluded other online information resources such as guideline databases, meta-lists and search engines, literature surveillance alerting systems, online books, and journal articles (ie, primary and secondary literature). Our analysis was limited to products in the English language.

Search Strategy

To identify the point-of-care information summaries, we re-examined the eligibility of all products that were included or excluded in the 2008 and 2012 analyses. To find new summaries, we searched Medline from August 2012 to December 2014 with the following terms: (("Evidence-Based Medicine"[Mesh]) AND ("Information Storage and Retrieval"[Mesh])) AND (("Online Systems"[Mesh]) OR ("Point-of-Care Systems"[Mesh])). We scanned the references of the papers retrieved and used the Google search engine to identify additional products that may not have been reported in the medical literature. We explored various publisher and librarian association websites (ie, Council of Science Editors, the World Association of Medical Editors, the European Association for Health Information and Libraries, and the American Medical Informatics Association) [13-16], and the 2014 conference proceedings from the Medical Library Association Meeting and Exhibition [17].

Identification of Point-of-Care Information Summaries

One reviewer examined the search results, screened the titles and abstracts of papers identified through Medline, and evaluated the eligibility of products integrating additional information found on product websites. If there was doubt about the inclusion of a product, all authors discussed the eligibility until a consensus was reached. We recorded the reasons for exclusion.

Data Extraction and Analysis

One reviewer extracted information on the general features of each point-of-care information summary. Products that could not be accessed (ie, no subscription available at our institution, no free-trial option, and no response from product representatives to our emails requesting access) were excluded. One reviewer collected data on the general characteristics of products and their content presentation for qualitative (descriptive) evaluation, along with information about the editorial quality, evidence-based methodology, and content volume (breadth of diseases and medical conditions covered) for empirical quantitative analysis. A second reviewer checked the extractions.

Qualitative Evaluation

For each summary included, we collected the following general details: country of development, year of release, vendor or publisher, marketing claims, format (eg, tablets, mobile devices), access and subscription options, annual costs, and targeted audience. Since the 2012 analysis, we have introduced an additional component: ability to be integrated into an EHR system. This entails the capacity to access information from the point-of-care summary directly through the EHR interface. For example, when a physician clicks on a condition written in the patient record, the physician is directed to a new screen detailing disease information and treatment options. A point-of-care summary search tool may be additionally available on the EHR interface to make free-text and *International Classification of Diseases* (ICD)-10 code searches.

Content presentation was analyzed in summaries that we accessed. We examined the different outputs (eg, key point

summary, paragraphs, question and answers, book chapter-like summary, clinical pathway, clinical scenario), use of formal ontology, flexibility, and reporting of references (with or without general or specific citations). We also assessed products' adoption of an intent to recommend, use of a formal strength of recommendation system, as well as the availability of continuing medical education programs or credits, other education materials (eg, lessons on statistical analysis or evidence-based methodology), and patient handouts.

Quantitative Analysis

Two reviewers extracted information about three key dimensions: quality of the editorial process, quality of the evidence-based approach to content development (ie, evidence-based methodology), and volume or breadth of the medical conditions covered. We described products quantitatively using three separate scores that covered components relevant to each dimension. Disagreements were resolved by discussion between the reviewers. A third author was consulted for any unsolved discordances. All Web pages providing useful data were saved and stored in an electronic archive. When information about a particular component (eg, commercial support or critical appraisal) was unclear or could not be found, we contacted publishers by email requesting additional information and clarification of contents. All emails were stored in an electronic archive.

Editorial Quality

We adopted the following indicators of transparency to evaluate the methodological quality of the editorial process: authorship (reporting of authors for each summary), reviewing (implementation of a formal, structured peer-review process), updating (whether or not summaries had been revised or updated in the previous 2 years), conflicts of interest (disclosure of contributing authors' conflict of interest), and commercial support for content development. For this last component, we assigned three points if commercial support was not accepted, one point if commercial support was accepted and reported, and no points if the product developer did not present sufficient information for us to make a judgement. For the remaining items, we assigned three points if the component was judged as "adequate," one point if "unclear," and none if "not adequate" or "not reported." We arbitrarily decided to award three points instead of two for the adequate fulfillment of a criteria in order to give more weight to transparent and accountable reporting, and increase variability within the sample.

In the 2008 and 2012 reviews, we assessed the authorship, authors' conflicts of interest, and updating of products based on the editorial policy statements. If the information provided was insufficient to make an accurate evaluation, we referred to a nonrandom selection of sections (often referred as topics) to assess the dimensions. In the effort to minimize bias between reporting and implementation in the 2014 analysis, we evaluated these dimensions through a random sample of topics. We randomly selected ten blocks or categories of diseases from ICD-10 [18]. If any product did not cover one of the medical conditions identified in a block, we randomly selected another block from ICD-10. In each topic, we checked the reporting of authors as well as any potential conflict of interest. For updating,

topics were considered up-to-date if they had been reviewed or revised within the last 2 years (January 2013 to January 2015). The 2-year time frame was determined based on the average time to changes in evidence that are sufficiently important to require the updating of systematic reviews [19]. Products with eight or more topics updated in the last 2 years were assigned 3 points towards the total editorial quality score. Products with three or less topics updated within that period were assigned no points. Other products with four to seven updated topics were assigned 1 point as well as those that did not consistently provide dates on the articles.

Evidence-Based Methodology

The following components were used to evaluate the strength of the evidence-based methodology for content development: implementation of a literature search or surveillance strategy to identify current information, cumulative versus discretionary approach (prioritization of systematic reviews over other evidence sources), critical appraisal, formal grading of evidence, and citation of expert opinions (separation of expert opinions from other evidence sources in summaries). Three points were assigned if the component was judged "adequate," one if considered "unclear," and none if "not adequate" or "not reported."

Volume (Breadth of Diseases Covered)

As it was not feasible to count the total number of diseases and medical conditions covered in each product, we estimated the comprehensiveness of disease coverage by verifying the presence or absence of a random sample of diseases from the ICD-10 [18]. We randomly selected four chapters: Chapter IV—Endocrine, nutritional and metabolic diseases, VII—Diseases of the eye and adnexa, XII—Diseases of the skin and subcutaneous tissue, and XV—Pregnancy, childbirth and the puerperium. These chapters comprised a total of 35 blocks or categories of diseases or medical conditions. If a point-of-care information summary discussed at least one disease specified within a block, the product was assigned 1 point towards a maximum of 35 total points for volume. We then converted the volume scores into percentages, where 35 points correspond to 100% coverage.

Multimedia Appendix 1 summarizes in a flow diagram the methods used to evaluate products.

Analysis

Volume and quality indicator scores are presented with medians and interquartile ranges (IQR). Point-of-care information summaries were ranked on the basis of (1) editorial quality, (2) the use of an evidence-based approach, and (3) the volume of diseases covered based on a random sample of ICD-10 chapters. Correlations between these three dimensions were assessed by Spearman rank correlation coefficients and their respective P values. Changes in the strength of the products from 2008-2014 were assessed using the matched pairs Wilcoxon signed-rank test. For hypothesis testing, a probability of $<.05$ was considered statistically significant. All statistical tests were two-sided. Stata software was used for statistical analyses.

Results

The search strategy identified 58 products for potential inclusion. After screening, 26 fulfilled our inclusion criteria. Sixteen of these were previously included in the 2008 and 2012 reviews (5 Minute Consult, BestBets, Clin-eGuide, Dynamed, EBM Guidelines, Essential Evidence Topics, eTG Complete, GP Notebook, Map of Medicine, Micromedex, Mosby's Nursing Consult, Nursing Reference Center, PEPID, Rehabilitation Reference Center, UpToDate, and Zynx Evidence). Four products changed into a new product since 2012 (ACP Smart Medicine formerly ACP Pier, Best Practice formerly Clinical Evidence, Clinical Key formerly First Consult, Medscape Drug and Diseases Reference formerly Emedicine). Six products were newly identified in this review (Clinical Access, Cochrane Clinical Answers, Decision Support in Medicine, NICE Pathways, PEMSoft, and Prodigy). Prodigy, which is connected with Clinical Knowledge Summaries (CKS), was considered a new product since CKS was discontinued for some time and only in 2012 was restarted. Figure 1 shows the flow diagram for the selection of point-of-care information summaries in the review.

In order to access the 26 products, we registered for free-trial access online whenever available or contacted the publishers directly requesting temporary access to perform the evaluation. We did not receive a response from the publishers of three products (Clin-eGuide, Mosby's Nursing Consult, and Zynx Evidence), which were prevented from further evaluation. A total of 23 products were included in the content presentation and quantitative analysis.

Figure 1. Flow diagram of point-of-care information summaries included in the review.
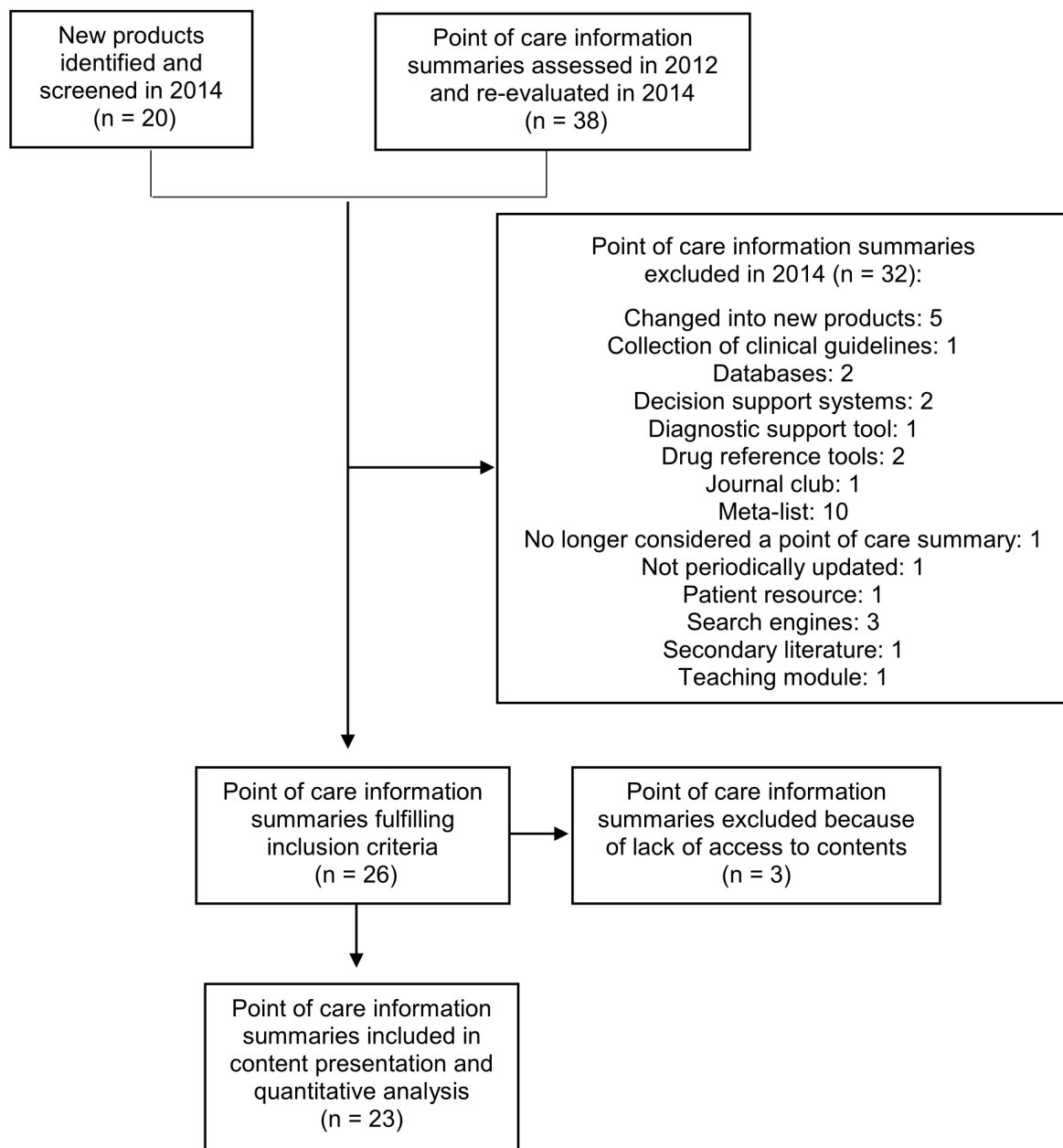
Qualitative Evaluation

General features are summarized in Multimedia Appendix 2. Most of the 26 products were developed by major publishers in the United States (n=12) and United Kingdom (n=8), while others came from the Netherlands (n=4), Finland (n=1), and Australia (n=1). A minority was open access (19%), while most were fee-based (81%) with a median individual subscription price of €244.4 (US$265, £169.52). Regarding their electronic compatibility, over a quarter (7/26, 27%) of products were Web-based only, as others could also be opened on mobile devices. Most products targeted a general audience of health professionals (18/26, 70%), but some were advertised for specific groups such as medical specialists (1/26, 4%), general practitioners (2/26, 8%), nurses (2/26, 8%), emergency medicine doctors (1/26, 4%), pediatricians (1/26, 4%), and rehabilitation professionals (1/26, 4%). Sixteen products out of 26 (62%) could be integrated into EHRs.

Multimedia Appendix 3 presents details of the summary content presentation of the 23 products we could fully evaluate. Products displayed their content in a variety of formats: key point summary, questions and answers, book chapter-like summaries, and clinical pathways (flow charts). Most had a formal ontology for organizing diseases and medical conditions (20/23, 87%) as well as flexible navigation of topic contents (19/23, 83%). Although many products adopted an intent to recommend approach (17/23, 73%), under a third (7/23, 30%) used a formal strength of recommendation system: Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach [20], the Strength of Recommendation taxonomy (SORT) by the American Academy of Family Physicians [21], or individual systems developed for the product. Just under a half (11/23, 48%) of products awarded continuing medical education credits for searches or featured other programs for continuing medical education. Patient education materials and handouts were available in nearly a third (7/23, 30%) of products, and only a few (4/23, 17%) offered additional educational materials for clinicians such as evidence-based medicine and critical appraisal methodology, lessons on cultural competencies, laboratory manuals, and practice resources.

Quantitative Analysis

Figure 2 shows the rank of products based on volume. Disease coverage varied widely: the median volume or coverage of medical conditions was 94% (IQR, 66-100%). The most comprehensive products providing at least one condition per disease category in the four ICD-10 chapters were 5 Minute Consult, Best Practice, Clinical Access, Dynamed, GP Notebook, and UpToDate.

Editorial quality and evidence-based methodology are summarized in Multimedia Appendices 4 and 5; the median scores were 12 (IQR 6-13) and 11 (IQR 4-15), respectively, on a 15-point scale. Five products (ACP Smart Medicine, BMJ Best Practice, Dynamed, Essential Evidence Topics, and UpToDate) received the maximum score for editorial quality. Six (ACP Smart Medicine, BestBets, BMJ Best Practice, Dynamed, EBM Guidelines, and UpToDate) received the maximum score for evidence-based methodology.

The ranking of point-of-care information summaries based on their strength of volume, editorial quality, and evidence-based methodology is shown in Figure 3 (full data reported in Multimedia Appendices 4-6). Best Practice, Dynamed, and UpToDate scored in the highest quartile across all three dimensions. There was a moderate positive correlation between the editorial quality and evidence-based methodology of products (r=.41, P=.0496). No correlations were found between editorial quality and volume (r=.10, P=.64), or between evidence-based methodology and volume (r=.06, P=.80).

Compared to the 2008 evaluation, there were significant improvements in all three dimensions: editorial quality (P=.01), evidence-based methodology (P=.015), and volume (P<.001). Figure 4 shows the evolution of the products in the 2014 assessment that were previously evaluated in 2008.

Figure 2. Estimated volume (breadth) of diseases and medical conditions covered by point-of-care information summaries (based on 4 randomly selected chapters of the ICD-10 classification system).
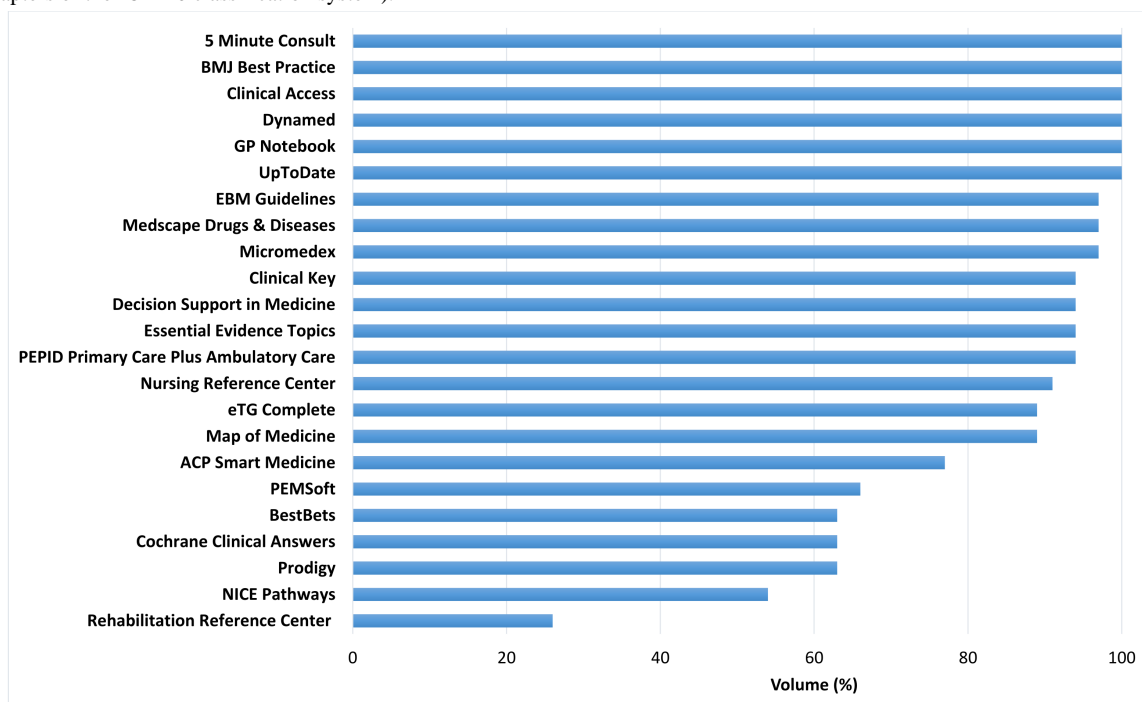

Figure 3. Point-of-care information summary rankings with providers listed in alphabetical order. Quartiles according to 2014 rankings for volume, editorial quality, and evidence-based methodology: black, bottom quartile; dark gray, low intermediate quartile; light gray, high intermediate quartile; white, top quartile (for evidence-based methodology and volume, white represents only the maximum scores of 15 and 100, respectively, as the top quartiles fell on the maximum scores).

Name of Product	Editorial Quality Score	Evidence-Based Methodology Score	Volume (%)
5 Minute Consult	dark gray	dark gray	white
ACP Smart Medicine	white	white	dark gray
BestBets	black	white	black
BMJ Best Practice	white	white	white
Clinical Access	black	black	white
Clinical Key	dark gray	dark gray	dark gray
Cochrane Clinical Answers	dark gray	light gray	black
Decision Support in Medicine	black	black	dark gray
Dynamed	white	white	white
EBM Guidelines	dark gray	white	light gray
Essential Evidence Topics	white	dark gray	dark gray
eTG Complete	light gray	black	dark gray
GP Notebook	black	black	white
Map of Medicine	dark gray	light gray	dark gray
Medscape Drugs & Diseases	light gray	black	light gray
Micromedex	black	dark gray	light gray
NICE Pathways	dark gray	dark gray	black
Nursing Reference Center	dark gray	dark gray	dark gray
PEMSoft	dark gray	dark gray	black
PEPID Primary Care Plus Ambulatory Care	dark gray	dark gray	dark gray
Prodigy	black	dark gray	black
Rehabilitation Reference Center	dark gray	black	black
UpToDate	white	white	white

Figure 4. Evolution of the scores for products evaluated in 2008 and reevaluated in 2014. We wrote "=" when the point-of-care information summary score did not change over time, "↑" when the score improved, and "↓" when the score decreased.

Name of Product	Editorial Quality Score	Evidence-Based Methodology Score	Volume (%)
5 Minute Consult	↑	↑	↑
ACP Smart Medicine	↑	↑	↑
BestBets	=	=	↑
BMJ Best Practice	=	=	↑
Clinical Key	↑	↑	↑
Dynamed	↑	↑	↑
EBM Guidelines	↑	=	↑
eTG Complete	↑	=	↑
GP Notebook	↑	=	↑
Map of Medicine	↑	=	↑
Medscape Drugs & Diseases	=	↑	↑
Micromedex	↓	=	↑
PEPID Primary Care Plus Ambulatory Care	=	↑	↑
UpToDate	=	=	↑

Discussion

Principal Findings

To evaluate products' claims to be evidence-based, we adopted editorial policy, content quality, and coverage of medical knowledge as the key indicators of high-quality point-of-care information summaries. In line with the 2008 and 2012 analyses, the purpose of our study was not to pinpoint the "winning" and "losing" products but to assess the maturity of these tools for clinical decision making and encourage transparent reporting of editorial and content development policies by publishers. We further sought to guide readers in the selection of products for individual or institutional use. Since 2008, there have been improvements in the general features of point-of-care information summaries and the descriptions of their editorial approaches, though suboptimal products are still on the market [11].

Several limitations to our study must be noted, including use of editorial policy statements to determine the implementation of a formal and structured peer-review process and the acceptance of commercial support for content development. We acknowledge that there may have been discrepancies between the reporting and actual implementation of editorial policies. Moreover, although we included quality dimensions informed by evidence in our study, our criteria for assessment may be perceived as arbitrary; users of a given point-of-care summary may have different views or experience. Regardless of potential

differences in opinions, one observation remains clear: publishers have invested notable energy and resources to raise their quality standards in a limited time. Product maturity and the increasing value of reliable information in medical society may sustain the rising popularity of point-of-care summaries among health professionals.

A particular challenge within our study involved the defining of the intervention and execution of the search strategy to identify relevant interventions for inclusion. Since our first evaluation in 2008, there continues to be a discrepancy in the terminology adopted to describe what we identify as "point of care information summaries": Web-based medical compendia that are specifically designed to deliver predigested, rapidly accessible, comprehensive, and periodically updated information to health care providers. These products have been additionally referred to as "evidence-based textbooks" [22], "clinical point-of-care tools" [23], navigators, and services [3]. While we recognize that other terms might be used, we have adopted point-of-care information summaries as the preferred terminology, as it embraces several key content elements. Given the rising interest and adoption of these tools, the development of a common term and definition will facilitate their assessment by researchers as well as by hospitals and health care professionals in search of a compatible tool for use. A common definition might also benefit the PubMed MeSH vocabulary. In fact, the MeSH term "point-of-care systems" comprises a broad range of health care technologies outside of our intervention, such as laboratory and diagnostic instruments [24].

The quality of most products is still moderate, which has also been indicated by the few additional surveys evaluating the quality of point-of-care information summaries [22,25-27]. Clinicians should become familiar with the basic concepts that make an information product a credible source of scientific evidence. Health libraries and local knowledge brokers should endorse and give preference to summaries that are committed to policies to improve editorial and methodological rigor, disclose conflicts of interest [28-30], and ensure complete and accessible reporting of the content development procedure. Users should be skeptical about point-of-care summaries that do not transparently describe how information is found (search strategy), selected (cumulative or discretionary approach), evaluated (critical appraisal), prioritized (grading of evidence and recommendations), and regularly updated (literature surveillance) to maintain their relevance to practice. Publishers may be highly skilled in boosting clinical recommendations through propaganda and legally qualified to sell their products to doctors and hospitals. Moreover, the failure to disclose methods for product development is not in the best interests of the medical community, and might, in fact, draw the line between authoritative and fraudulent therapeutic information.

Point-of-care information summaries largely serve high-income countries. However, information on highly effective medicines and interventions are presumably more valuable in low- and middle-income countries. At the same time, in an increasingly competitive market, publishers cannot make the service "free for everyone" because this would affect their sustainability and might facilitate the opportunistic use of these resources. We encourage publishers to align the prices of their products to the purchasing power of a particular country's physicians through tiered-pricing models and to distribute access through networks active in low- and middle-income countries [31,32]. In addition to their affordability and access, the source of information is critical to the strength and reliability of products.

Dynamed currently has links to over 17,000 guidelines, organized for high or low- to middle-income countries [33]. While the consideration of ready-to-use recommendations is a key first step, more investments in tailoring information to local doctors and other health care providers are needed. For example, information on medicines was never ranked on the basis of the WHO Model List of Essential Medicines, which selects treatments that offer a cure or effective disease management in preference to those that offer only marginal benefit [34]. Doctors are increasingly interested in knowing potential incongruence between investing resources and desired health outcomes [35,36]. In this time of austerity, point-of-care summaries have to do a better job considering the social proven value of medicines.

Future Considerations

It is not easy to predict what directions publishers should take to further improve their services. We propose three approaches. First, as summary providers mature and their contents become broader and more complete (eg, information about medicines, recommendations, and guidelines), information must be re-filtered to meet personal practice needs. Users will need to personalize the product, setting filters to isolate specific information (eg, local hospital guidelines) that is relevant to individual clinical practice. This will prioritize information that can engender changes in health professional behavior [37].

Second, high-quality point-of-care summaries should be integrated into computer decision support systems for EHRs. These computer systems may represent the future of clinical decision making in which evidence-based knowledge from point-of-care summaries is linked with patient information from EHRs to generate case-specific guidance messages through rule- or algorithm-based software [3,38]. Computer decision support systems combined with EHRs might be beneficial for the health care provided to patients, although it is hard to demonstrate their association with benefits on outcomes such as mortality [39].

Third, the potential integration of point-of-care summaries into continuing medical education programs should be recognized [40]. Doubts that are raised during clinical consultation can trigger point-of-care searches that provide health professionals with valuable information that can be directly implemented in the visit. Accreditation systems need to recognize the role of point-of-care summaries as an efficient provider of relevant knowledge.

Conclusion

The maturation of point-of-care summaries can be seen as a virtuous circle [41]. It started with an exogenous factor: technological innovation. As health professionals become increasingly familiar with the summaries, their adoption will become self-reinforcing. In a competitive market, this will probably help lower product prices, leading to more potential users. The last 20 years saw the success of PubMed, The Cochrane Library, and, more recently, WikiProject Medicine, which are now integral parts of medical practice. Publishers and developers of point-of-care summaries need to direct their considerable talents and resources to developing strategies to sustain affordable practice and interventions to improve quality of practice. This change of focus can support their development as indispensable professional tools.

Authors' Contributions

RB and LM conceived and designed the study. KK, MG-L, and RB collected the data. SB provided statistical expertise and analyzed the data. KK, MG-L, and LM drafted the article. All authors contributed to the interpretation of the data, critically revised the article for important intellectual content, and approved the manuscript. LM is the guarantor of the article.

Conflicts of Interest

LM is employed by the IRCCS Galeazzi and Università degli Studi di Milano, which have nonexclusive contracts with commercial publishers to develop or adapt point-of-care services to local or national settings. LM received remuneration for consultancy time. LM is the PI of two randomized controlled trials testing the effectiveness of computer decision support systems based on a point-of-care service (ie, EBM Guidelines) evaluated in this manuscript.

Multimedia Appendix 1

Flow diagram of methods used to evaluate editorial quality, evidence-based methodology, and volume.

[PDF File (Adobe PDF File), 476KB - jmir_v18i1e15_app1.pdf]

Multimedia Appendix 2

General characteristics of point-of-care information summaries.

[PDF File (Adobe PDF File), 266KB - jmir_v18i1e15_app2.pdf]

Multimedia Appendix 3

Content presentation of point-of-care information summaries.

[PDF File (Adobe PDF File), 114KB - jmir_v18i1e15_app3.pdf]

Multimedia Appendix 4

Editorial quality of point-of-care information summaries.

[PDF File (Adobe PDF File), 190KB - jmir_v18i1e15_app4.pdf]

Multimedia Appendix 5

Evidence-based methodology of point-of-care information summaries.

[PDF File (Adobe PDF File), 106KB - jmir_v18i1e15_app5.pdf]

Multimedia Appendix 6

Point-of-care information summary scores for editorial quality, evidence-based methodology, and volume.

[PDF File (Adobe PDF File), 102KB - jmir_v18i1e15_app6.pdf]

References

1. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004 Feb;8(6):iii-iv, 1 [FREE Full text] [Medline: 14960256]
2. Ebell MH. Point-of-care information that changes practice: it's closer than we think. Fam Med 2003 Apr;35(4):261-263. [Medline: 12729310]
3. Moja L, Banzi R. Navigators for medicine: evolution of online point-of-care evidence-based services. Int J Clin Pract 2011 Jan;65(1):6-11. [doi: 10.1111/j.1742-1241.2010.02441.x] [Medline: 21155939]
4. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. Evid Based Med 2006 Dec;11(6):162-164. [doi: 10.1136/ebm.11.6.162-a] [Medline: 17213159]
5. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534 [FREE Full text] [Medline: 11687560]
6. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. J Clin Oncol 2012 Dec 1;30(34):4243-4248. [doi: 10.1200/JCO.2012.42.8011] [Medline: 23071233]

XSL•FO

RenderX

7. Swedish Presidency of the EU. 2009. e-Health for a Healthier Europe! Opportunities for a better use of healthcare resources URL: http://www.medicpen.se/17435-eHealth_for_a_Healthier_Europe.pdf [accessed 2015-10-02] [WebCite Cache ID 6byjmhC4n]

8. Kortteisto T, Raitanen J, Komulainen J, Kunnamo I, Mäkelä M, Rissanen P, EBMeDS (Evidence-Based Medicine electronic Decision Support) Study Group. Patient-specific computer-based decision support in primary healthcare--a randomized trial. Implement Sci 2014;9:15 [FREE Full text] [doi: 10.1186/1748-5908-9-15] [Medline: 24444113]

9. Van de Velde S, Vander SR, Fauquert B, Geens S, Heselmans A, Ramaekers D, et al. EBMPracticeNet: A Bilingual National Electronic Point-Of-Care Project for Retrieval of Evidence-Based Clinical Guideline Information and Decision Support. JMIR Res Protoc 2013;2(2):e23 [FREE Full text] [doi: 10.2196/resprot.2644] [Medline: 23842038]

10. Moja L, Liberati EG, Galuppo L, Gorli M, Maraldi M, Nanni O, et al. Barriers and facilitators to the uptake of computerized clinical decision support systems in specialty hospitals: protocol for a qualitative cross-sectional study. Implement Sci 2014;9:105 [FREE Full text] [doi: 10.1186/s13012-014-0105-0] [Medline: 25163794]

11. Banzi R, Liberati A, Moschetti I, Tagliabue L, Moja L. A review of online evidence-based practice point-of-care information summary providers. J Med Internet Res 2010;12(3):e26 [FREE Full text] [doi: 10.2196/jmir.1288] [Medline: 20610379]

12. Moja L, González LM, Tagliabue L, Banzi R. A review of online evidence-based practice point-of-care information summary providers: an update. 2013 Sep 23 Presented at: 21st Cochrane Colloquium; September 19-23, 2013; Quebec, Canada URL: http://2013.colloquium.cochrane.org/abstracts/review-online-evidence-based-practice-point-care-information-summary-providers-update

13. Council of Science Editors. URL: http://www.councilscienceeditors.org/ [accessed 2015-10-03] [WebCite Cache ID 6c02thAtI]

14. World Association of Medical Editors. URL: http://www.wame.org/ [accessed 2015-10-10] [WebCite Cache ID 6cBDEyBXY]

15. European Association for Health Information and Libraries. URL: http://eahil.eu/ [accessed 2015-10-02] [WebCite Cache ID 6byk4aqA0]

16. American Medical Informatics Association. URL: https://www.amia.org/ [accessed 2015-10-02] [WebCite Cache ID 6bykNpnat]

17. Medical Librarian Association. URL: http://www.mlanet.org/ [accessed 2015-10-03] [WebCite Cache ID 6bzyvr3Cu]

18. World Health Organization. ICD-10 2015. Geneva, Switzerland International Statistical Classification of Diseases and Related Health Problems 10th Revision URL: http://apps.who.int/classifications/icd10/browse/2015/en [accessed 2015-03-01] [WebCite Cache ID 6cBDyLFCC]

19. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 2007 Aug 21;147(4):224-233. [Medline: 17638714]

20. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

21. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. Am Fam Physician 2004 Feb 1;69(3):548-556 [FREE Full text] [Medline: 14971837]

22. Jeffery R, Navarro T, Lokker C, Haynes RB, Wilczynski NL, Farjou G. How current are leading evidence-based medical textbooks? An analytic survey of four online textbooks. J Med Internet Res 2012;14(6):e175 [FREE Full text] [doi: 10.2196/jmir.2105] [Medline: 23220465]

23. Brasure M, Forte M, Butler M, Kane R. Methods Research Reports 14-EHC026-EF. Rockville, MD: Agency for Healthcare Research and Quality (US); 2014. Use of Evidence-Based Practice Center Comparative Effectiveness Reviews by Clinical Point-of-Care Tools Internet URL: http://www.ncbi.nlm.nih.gov/books/NBK206754/pdf/Bookshelf_NBK206754.pdf [accessed 2016-01-08] [WebCite Cache ID 6eOgXbLJc]

24. National Center for Biotechnology Information. MESH term: Point-of-Care Systems URL: http://www.ncbi.nlm.nih.gov/mesh/?term=point-of-care+systems [accessed 2015-11-19] [WebCite Cache ID 6dACECgZ4]

25. Prorok JC, Iserman EC, Wilczynski NL, Haynes RB. The quality, breadth, and timeliness of content updating vary substantially for 10 online medical texts: an analytic survey. J Clin Epidemiol 2012 Dec;65(12):1289-1295. [doi: 10.1016/j.jclinepi.2012.05.003] [Medline: 22974495]

26. Ketchum AM, Saleh AA, Jeong K. Type of evidence behind point-of-care clinical information products: a bibliometric analysis. J Med Internet Res 2011;13(1):e21 [FREE Full text] [doi: 10.2196/jmir.1539] [Medline: 21335319]

27. Banzi R, Cinquini M, Liberati A, Moschetti I, Pecoraro V, Tagliabue L, et al. Speed of updating online evidence based point of care summaries: prospective cohort analysis. BMJ 2011;343:d5856 [FREE Full text] [Medline: 21948588]

28. Shurtz S, Foster MJ. Developing and using a rubric for evaluating evidence-based medicine point-of-care tools. J Med Libr Assoc 2011 Jul;99(3):247-254 [FREE Full text] [doi: 10.3163/1536-5050.99.3.012] [Medline: 21753917]

29. Cook DA, Enders F, Linderbaum JA, Zwart D, Lloyd FJ. Speed and accuracy of a point of care web-based knowledge resource for clinicians: a controlled crossover trial. Interact J Med Res 2014;3(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2811] [Medline: 24566739]

30. Cosgrove L, Shaughnessy AF, Wheeler EE, Krimsky S, Peters SM, Freeman-Coppadge DJ, et al. From caveat emptor to caveat venditor: time to stop the influence of money on practice guideline development. J Eval Clin Pract 2014 Dec;20(6):809-812. [doi: 10.1111/jep.12244] [Medline: 25327453]

31. Publishing TRS. Developing World Access to online journals. URL: http://royalsocietypublishing.org/developing-world [accessed 2015-08-01] [WebCite Cache ID 6bykajoga]

32. Chan L, Arunachalam S, Kirsop B. Open access: a giant leap towards bridging health inequities. Bull World Health Org 2009 Aug 1;87(8):631-635. [doi: 10.2471/BLT.09.064659]

33. Dynamed Plus. URL: http://www.dynamed.com/home/ [accessed 2015-10-02] [WebCite Cache ID 6bynRHgTe]

34. Magrini N, Robertson J, Forte G, Cappello B, Moja LP, de JK, et al. Tough decisions on essential medicines in 2015. Bull World Health Organ 2015 Apr 1;93(4):283-284 [FREE Full text] [doi: 10.2471/BLT.15.154385] [Medline: 26229193]

35. Djulbegovic B. Value-Based Cancer Care and the Excessive Cost of Drugs. JAMA Oncol 2015 Dec 1;1(9):1301-1302. [doi: 10.1001/jamaoncol.2015.3302] [Medline: 26313294]

36. Cherny NI, Sullivan R, Dafni U, Kerst JM, Sobrero A, Zielinski C, et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol 2015;8:1547-1573.

37. Chen C, Haddad D, Selsky J, Hoffman JE, Kravitz RL, Estrin DE, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. J Med Internet Res 2012;14(4):e112 [FREE Full text] [doi: 10.2196/jmir.2152] [Medline: 22875563]

38. Heselmans A, Aertgeerts B, Donceel P, Geens S, Van de Velde S, Ramaekers D. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012 Dec;36(6):3677-3684. [doi: 10.1007/s10916-012-9841-3] [Medline: 22402980]

39. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/AJPH.2014.302164] [Medline: 25322302]

40. Moja L, Kwag KH. Point of care information services: a platform for self-directed continuing medical education for front line decision makers. Postgrad Med J 2015 Feb;91(1072):83-91 [FREE Full text] [doi: 10.1136/postgradmedj-2014-132965] [Medline: 25655251]

41. Farrell D. The real new economy. Harv Bus Rev 2003 Oct;81(10):104-12, 138. [Medline: 14521102]

Abbreviations

EHR: electronic health record
ICD-10: International Classification of Diseases, 10th Revision
IQR: interquartile ranges

XSL•FO
RenderX

Original Paper

Do Therapists Google Their Patients? A Survey Among Psychotherapists

Christiane Eichenberg[1], Dr phil habil; Philipp Y Herzberg[2], Dr rer nat habil

[1]Department of Psychology, Sigmund Freud University, Vienna, Austria

[2]Professorship of Personality Psychology and Psychological Diagnostics, University of the Federal Armed Forces Hamburg, Hamburg, Germany

Corresponding Author:
Christiane Eichenberg, Dr phil habil
Department of Psychology
Sigmund Freud University
Freudplatz 1
Vienna, 1020
Austria
Phone: 43 6601599154
Fax: 43 1 798 40 98 /93
Email: christiane@rz-online.de

Abstract

Background: The increasing use of the Internet and its array of social networks brings new ways for psychotherapists to find out information about their patients, often referred to as patient-targeted googling (PTG). However, this topic has been subject to little empirical research; there has been hardly any attention given to it in Germany and the rest of Europe and it has not been included in ethical guidelines for psychotherapy despite the complex ethical issues it raises.

Objective: This study explored German psychotherapists' behavior and experiences related to PTG, investigated how these vary with sociodemographic factors and therapeutic background, and explored the circumstances in which psychotherapists considered PTG to be appropriate or not.

Methods: A total of 207 psychotherapists responded to a newly developed questionnaire that assessed their experience of and views on PTG. The study sample was a nonrepresentative convenience sample recruited online via several German-speaking professional therapy platforms.

Results: Most therapists (84.5%, 174/207) stated that they had not actively considered the topic of PTG. However, 39.6% (82/207) said that they had already looked for patient information online (eg, when they suspected a patient may have been lying) and 39.3% (81/207) knew colleagues or supervisors who had done so. Only 2.4% (5/207) of therapists had come across PTG during their education and training.

Conclusions: It is essential to provide PTG as a part of therapists' education and training. Furthermore, the complex problems concerning PTG should be introduced into codes of ethics to provide explicit guidance for psychotherapists in practice. This report provides initial suggestions to open up debate on this topic.

Introduction

The Internet has become an essential and frequently used medium for retrieving diverse information about people and organizations. There may be a variety of reasons for this, including private curiosity (some people may wish to look up a friend from school to find out what she is doing now) or there may be job-related reasons for exploring the Web. But what about professional relationships between therapists and patients? Therapists use the Internet for assistance in everyday work tasks; for example, most psychotherapists use email as well as mobile communication to contact their patients as described by Eichenberg and Kienzle [1]. Furthermore, Zur et al [2] have reported an increasing prevalence of "deliberate self-disclosure"

XSL•FO
RenderX

by therapists who maintain a presence on the Internet. For patients, this can make it easier to choose a therapist as Eichenberg et al [3] found in a national survey in Germany: nearly two-thirds (63.5%) of Internet users search the Web for health-related information and 43.7% could imagine seeking help online in cases of emotional distress (eg, obtaining the contact data of a resident psychotherapist).

Could such information interfere with the relationship between therapist and patient? Facebook, Xing, Twitter, etc, open up further possibilities for information use and thereby possibilities for abuse. The Internet represents countless ways of using information without the slightest moral or ethical consideration of any effects this might have on therapeutic relationships. According to Ensher et al [4], just as managers look for background information on the Internet about their potential employees, psychotherapists also look up their potential patients as assumed, but not empirically proven, by Clinton et al [5]. An empirical study revealed that a large majority of psychology graduates (81%) report using online social networks, although approximately 27% of psychotherapists and therapists in education look for online information on their patients [6]. This phenomenon is described by Clinton et al [5] as "patient-targeted googling" (PTG). However, neither the American guidelines for psychotherapists (American Psychological Association [APA]) [7] nor the German guidelines (Berufsverband Deutscher Psychologen [BDP]) [8] give any explicit guidance on this issue.

The APA recommends that psychologists should "...respect the dignity and worth of all people, and the rights of individuals to privacy, confidentiality, and self-determination" (p 4 [7]). This includes keeping intrusions on privacy by the therapist to a minimum. Comparing the US and German guidelines does not lead to any further or more explicit conclusion: the ethical guidelines for psychotherapists in Germany do not explicitly mention the phenomenon of online research of patient information. But its indirect implementation can be found in the professional code of the trade union of German psychologists [8]. It states that psychologists are only allowed to collect, save, and use client or patient data that is in line with the treatment order. The necessity for a bond of trust and the idea that psychologists should inform their patients about all substantial procedures at all steps in the treatment and ask for patients' agreement are well established [8]. But what would be considered a "substantial procedure"? When can the bond of trust be interrupted? When could a therapist satisfy their curiosity instead of following psychotherapeutic orders?

To answer these questions, it seems reasonable to examine PTG. As yet, there have been only 2 articles that have presented empirical material on this [5,6], so more empirical data are needed. Therefore, the aim of this study was to investigate PTG in Germany from the perspective of therapists. Specifically, we intended to answer the following questions:

1. What reasons correspond with what type of information?
2. How much information about PTG do psychotherapists receive in their education or through professional experience?
3. When do psychotherapists consider PTG to be appropriate and inappropriate?
4. What reasons are there for and against PTG in general?
5. Do attitudes differ according to sociodemographic factors?
6. In what ways is the information gained online used for therapeutic reasons?
7. To what extent do therapists know about patients who search for online information about them?
8. Do therapists take precautions to control the information about them available on the Internet?

Methods

Recruitment

The study sample was a nonrepresentative convenience sample recruited online via several professional therapy platforms. (For a discussion of the scientifically proven quality of Web-based studies, see Gosling et al [9].)

Questionnaire

A questionnaire was administered online, generated using the online tool Unipark [10]. It included 36 items organized into 3 sections: (1) sociodemographic data, including age, sex, professional experience, type of psychotherapeutic education, and frequency of Internet use; (2) experience of and attitudes toward PTG structured into open and closed questions with a focus on knowledge of the PTG phenomenon followed by 7 further items to be answered by the participants who had experience with PTG; and (3) online research by patients, including the reactions and experiences of therapists who reported patients searching for information about them via the Internet.

Statistical Analysis

In addition to descriptive statistical methods, inference statistical methods were used for the closed questions (correlation analysis and chi-square test); content analysis was used to analyze the open answers for some of the questions [11]. Inductive categories were designed for single questions, whereas coding entities were defined by its oneness of sense. The data were analyzed with SPSS version 19 and PASW Statistics version 18.

Sample

The sample included 207 psychotherapists (15/207, 7.2% in education), of whom 67.6% (140/207) were female, a similar proportion to the distribution of medical and psychological therapists in Germany [12]. The mean age of participants was 45.00 (SD 10.49) years, which is younger than the mean age of German therapists (mean 53 years). This may be a result of the Web-based study conception, given that younger therapists may be more inclined to use the Internet. More than half (51.2%, 106/207) of the therapists were licensed by health insurance and, on average, they had spent longer than 12 years in professional life. More than two-thirds (70.1%, 145/207) of the therapists worked in their own practice or a group practice with the rest located in clinics or other facilities. The majority of respondents (70.1%, 145/207) worked with adults in a single therapeutic setting; the types of therapy practiced included (multiple answers were allowed) cognitive behavioral therapy (51.2%, 106/207), psychodynamic psychotherapy (44.0%,

XSL·FO

RenderX

91/207), psychoanalysis (25.1%, 52/207), systemic psychotherapy (12.1%, 25/207), and various types of education (18.8%, 39/207). Most therapists (85.0%, 176/207) said that they used the Internet at least once a day for professional reasons; therefore, high Internet affinity can be assumed. In a professional setting, 96.1% (199/207) used the Internet to access factual information and 82.1% (170/207) to exchange information with colleagues.

Results

To What Extent Did the Psychotherapists Carry Out Patient-Targeted Googling?

Most therapists (84.5%, 174/207) stated that they had not actively considered the topic of PTG. Only 2.4% (5/207) had heard about PTG as part of their education or advanced training. Nevertheless, 39.6% (82/207) said that they had already looked for patient information online and 39.3% (81/207) knew colleagues or supervisors who had done likewise. Of the remaining 60.5% (125/207) who claimed that they had not looked for patient information online, 90.4% (113/125) would not do so even if regulations and the law were clarified. The main reasons for this were ethical doubts (36.9%, 42/113) and lack of confidence in the Internet as a source of reliable information (32.8%, 37/113). At the same time, a quarter (24.8%, 28/113) did not even want to know about the online information and another quarter (23.0%, 26/113) claimed that they were not willing to do the extra work involved. Only 13.3% (15/113) supposed that patients would not agree with such behavior. (Respondents were allowed to give multiple answers.)

Did the Psychotherapists With Experience of Patient-Targeted Googling Differ From the Others in Sociodemographic Factors? What Did Patient-Targeted Googling Involve? What Kinds of Information Were Obtained?

Data analysis showed no effect of sex or age on PTG. Nor were the type of therapeutic treatment provided or the age class of patients linked to online information research. Only the frequency of Internet use was significantly correlated with the probability of PTG (r=.18, P<.001).

The more than one-third of therapists (39.6%, 82/207) who looked online for patient information did so for a mean 5.8 cases (SD 8.8). Three-quarters (76%, 62/82) did this without the patients' permission, whereas 21% (17/82) gained permission

from their patients often or all the time to search for the information online. Only 4% (3/82) stated that they only looked for the information together with their patients. Home pages, blogs, and social networks were of greatest interest for obtaining information.

How Much Information About Patient-Targeted Googling Did Psychotherapists Receive in Their Education or Professional Experience?

Only 3 of 207 therapists (1.4%) received information about PTG during their education. They came from different therapeutic backgrounds, so no conclusions could be drawn from this. Only 2 of 207 therapists (1.0%) heard about PTG during advanced training. Overall, only 15.5% (32/207) of therapists had consciously considered the topic of PTG.

When Did Psychotherapists Consider Patient-Targeted Googling to Be Appropriate and Inappropriate? What Reasons Were There For and Against Patient-Targeted Googling in General? Did Attitudes Differ According to Sociodemographic Factors?

Irrespective of whether they themselves had carried out PTG, the therapists were asked about their attitude toward it. More than one-third (38.6%, 80/207) of therapists thought that searching the Web for patient information was unimaginable; for them, there was no possibility of them doing so. Approximately two-thirds felt differently and agreed that certain situations could indicate or allow PTG: 18.1% (38/127) of therapists would agree with using PTG in consensual agreement with the patient, 13.4% (28/127) would allow PTG in circumstances where there was imminent danger, 9.5% (12/127) would agree if there were a reasonable suspicion that the patient was lying, and 8.7% (11/127) allowed for the possibility of PTG at all times (multiple responses were allowed in the questionnaire).

The analysis of the open questions (Textbox 1) illustrated the special justifications for situations when therapists considered PTG to be appropriate. One therapist, for example, tried to gain information from the Internet about a suicidal patient who wanted to kill himself with a gun. To avoid danger, the therapist looked for membership at a shooting or gun club to check whether the patient had access to any kind of firearm and the knowledge of how to use it. Other situations considered legitimate by therapists included finding missing contact details or looking up information with the agreement of the patient.

Textbox 1. When do psychotherapists consider PTG to be appropriate? Response categories and sample quotes to open questions (n=53 therapists with n=59 answers).

---

1. Danger to self and others (n=22)

- "Acute danger to self and others."

- "If a patient tries to endanger others (run amok etc)."

- "Sexual offenders with treatment orders."

- "When I myself as a therapist am clearly threatened."

- "Planned suicides."

2. If discussed/desired (n=9)

- "If my patient desires that I look at his webpage."

- "At the patient's request."

- "If the patients' request is comprehensible for me."

3. Missing extra information (n=8)

- "To get more information."

- "To complete anamnesis."

- "For biographical info."

4. Formalities (n=7)

- "Unpaid bills"

- "If I only had the patient's old phone number and I need the new one."

- "Checking an address or phone number."

5. Patients in public life (n=5)

- "Patient is part of public life and newspaper articles (defamations) are a subject of therapy."

- "Patients who have a public life and assume that you are preinformed when you aren't (eg, you are not informed about footballers if you are not a football fan yourself)."

6. Content of therapy (n=5)

- "Interest in how patients present themselves online."

- "Suspicion of cybermobbing."

- "Young people use the Internet differently to us elder people (often uncritical and uncensored)."

7. Other (n=3)

- "After finishing the therapy I would be all right with it."

- "Pure curiosity."

There were relationships between the therapists' approaches to therapy and their evaluation of PTG's legitimacy. Therapists trained in psychodynamic-oriented therapy (χ^2_1=15.5, P<.001) or psychoanalytical therapy (χ^2_1=17.8, P<.001) responded significantly more often than cognitive behavioral therapists (χ^2_1=13.4, P<.001) that PTG is inappropriate in all situations. These differences may originate in specific aspects of asymmetric therapist-patient relationships in psychodynamic approaches to psychotherapeutic treatment. Analytic reasons (eg, the rule of abstinence) may not only have an impact on

ethics, but also on general techniques in treatment. For other criteria, no associations were found.

Analysis of the therapists' self-written answers on reasons that justify PTG (Textbox 2) showed that the most commonly mentioned reasons in favor of PTG were for a change of perspective, which should lead to a better understanding of the patient and recognition that Internet-based information was freely available. Verification of data and checking for suspected lies were also given as reasons to search for patient information, as was curiosity.

Textbox 2. Justifications for researching a patient's information via the Internet. Response categories and sample quotes to open question (n=132 therapists with n=149 answers).

---

1. Better understanding because of more information and a change of perspective (n=34)

- "Better understanding of the patient's social environment."

- "External, more widespread information that is not controlled directly by the patient."

- "eg, Patient is a refugee and I can imagine better his home and be more empathic."

2. Therapy-relevant information is on the Internet (n=17)

- "To be authentic to patients who attach importance to their Web presence."

- "How do patients present themselves on the Web?"

- "To gather information about how patients present themselves or so that they don't overlook that their self-expression can be seen by others."

3. Online information is public (n=8)

- "Anyone who provides their personal data on the Internet implicitly gives permission for this to be seen by others. That's why I don't need to ask for the patient's permission."

- "It is about information which belongs to patients, normally provided by them; and if not it is still part of patients' expression of personality."

- "Anyone who provides online information needs to expect that it will be read."

4. On the request of the patient (n=8)

- "Permission of patient after agreement or request."

- "After the patient's explicit request."

5. Curiosity (n=7)

- "If it is an interesting patient and you want to get to know more about him."

- "Sometimes, once in a blue moon, I do it out of curiosity. But I don't think it is essential or reasonable. In the end it is only one option of investigation: I want to gather information about a patient, eg by doing a third-party review of the patient's case history without consent."

6. Controlling patients' statements (n=6)

- "A kind of reality check. Is the patient really as famous as he says?"

- "Verifying patients' information about their activities and occupation. I have only done that in the case of narcissistic male patients and got a feeling of greater objectivity later on."

7. Suspicion of lying and concealment (n=5)

- "Suppressing facts such as criminal proceedings."

- "Trying to gain secretive factual information, to clear up discrepancies."

8. Nothing (n=64)

- "Currently I cannot imagine any situation where Internet research could be helpful for the therapeutic process."

- "Under the aspect of a relation of trust: nothing."

- "I do not know any reason!"

In their arguments against PTG (Textbox 3), therapists stated that the relationship of trust could be damaged and that patients should also have the right to decide for themselves what information they wanted to share. Protection of privacy and doubts about the real advantage and usefulness of the information were mentioned by many of the therapists who were against PTG. The risk to countertransference in therapeutic work was also an issue raised.

Textbox 3. Arguments against researching patients' information on the Internet. Response categories and sample quotes to open question (n=103 therapists with n=128 answers).

---

1. Disturbance in the trust relationship (n=39)

- "The open relationship of trust with patients. If relatives of patient provide written or oral information, for example, I would handle it the same open way as if it was obtained from Internet research."

- "It harms the bond of trust; patients don't "lie" to me—they deceive themselves."

2. Patients' self-determined information control (n=20)

- "Patients need to be in control of what they say."

- "Personal rights, privacy."

- "The right to lie."

3. Border violation/ensuring privacy (n=17)

- "Ensuring privacy."

- "The right of patients to 'privacy'—to appear in therapy the way they want to and need to."

4. Rule of abstinence and curiosity (n=13)

- "The rule of abstinence for psychotherapists as example."

- "Personal curiosity."

5. Manipulation/lack of impartiality (n=12)

- "My principle: all I learn about my patients is what I am told by them, not information obtained behind their back. This influences the unconscious therapeutic relationship."

- "Corruption of therapeutic neutrality in front of patients by having information they might not have wanted to give to me—concealed information could have a special function."

6. Doubtful reliability or usefulness of the information (n=11)

- "Not objective, only parts of the whole, not possible to demonstrate validity."

- "Lots of trash on the Internet"

- "You don't get the information you really needed for therapy."

7. Acting of countertransference (n=7)

- "Substantial disturbance in the relationship of trust as well as in transference and countertransference."

- "I consider the research of such data to be a professional and ethical problem. Professional, because instead of analyzing the countertransference you start acting; and ethical because of violation of abstinence and destruction of the relationship of trust/protected area."

8. Nothing (n=9)

- "Precisely nothing."

The summarized answers to the open questions show an ambivalent attitude toward PTG in the therapists' behavior and thought. Ignoring the answers in the questionnaire that were neither for nor against PTG, there were 85 responses in total justifying PTG and 119 arguments against it.

A similar result was also found with the answers to closed questions (Table 1). Of the arguments to justify PTG, the 2 that received the most agreement were that therapists should have access to freely available Internet information and that therapists should be in a position to separate their curiosity from necessity. More support can be found for the arguments against PTG, with most therapists agreeing that there were risks of curiosity being the motivation and there being the potential of harming the relationship of trust with the patient and of acting out countertransference. Table 1 sheds more light on therapists' ambivalent opinions.

XSL•FO
RenderX

Table 1. The proportion and number of therapists who agreed or disagreed with patient-targeted googling (multiple responses were allowed; N=207).

Specific statements about PTG	n (%)
Justifying PTG	
Information is provided online for all people. Online information is information shared with the therapist as well.	58 (28.0)
A good therapist can differentiate between curiosity and therapeutic need and does not run the risk of doing PTG with intrinsic motivation.	39 (18.8)
The Internet is in cases of emergency the quickest available resource to use.	32 (15.5)
Decisions have to be made on the basis of patient benefit. This even includes seeking additional information a patient does not want or is unable to give but which might accelerate help.	24 (11.6)
None of the above (positive) statements.	77 (37.2)
Against PTG	
Personal curiosity is certainly a motivation for PTG (perhaps unconsciously).	125 (60.4)
The bond of trust between patient and therapist collapses because of PTG.	97 (46.9)
Internet information is not reliable in the case of patient's inquiry.	80 (38.6)
Therapists are not allowed to gain information they were not officially provided with; this includes information from the Internet.	65 (31.4)
None of the above (negative) statements.	21 (10.1)

How Was the Online Information Used Therapeutically?

Nearly two-thirds (65%, 53/82) of therapists who used PTG did not annotate their findings in the patient's record because they considered the information to lack importance or saw it as mirroring information that was already known. The rest of the therapists referred to fulfilling obligatory documentation requirements (10%, 8/82) or to sporadic documentation (26%, 21/82). Patients were not consulted about online research in 38% (31/82) of cases for these reasons. Some emotional reasons of therapists also came across; for example, the pursuit of patients or curiosity should not be part of the therapy. Online results were discussed within therapy sessions (always: 31%, 25/82; sometimes: 32%, 26/82) to clarify mismatches mostly in cases where the online information was relevant to the therapy.

In the opinion of nearly one-third (32%, 26/82) of the therapists, no important or interesting details about their patients were found on the Web. PTG is seen by many as having a potential therapeutic use in allowing a better understanding to be gained of the public roles of some patients as well as providing an interesting focus on patients' self-expression. Furthermore, it has been understood as providing "certification for issues discussed in therapy" to make sure the patient is being understood in the correct way. A few therapists who did PTG said that personal interest certainly provided high motivation for online investigation.

To What Extent Do Therapists Know About Patients Who Search for Online Information About Them?

Therapists raised more concerns about being "googled" than they did about PTG; 91.3% (189/207) said that they had already thought about this issue. In the answers given to open questions, views were expressed that patients' curiosity is justified, whereas the negative perceptions about being googled focused on the violation of privacy, concerns about the control of information provided online, and worries about potential negative rating of therapists online. Of the therapists who had never considered the topic before (8.7%, 18/207), some said that they had no online information so the topic did not concern them. More than half of the therapists (54.6%, 113/207) were researched online by a patient at least once or were content with patients researching them. There was a significant difference between the sexes; male therapists were more often the subject of research than women (χ^2_1=6.8, P=.009) and correlation analysis showed that the frequency of being researched increased with the therapists' length of time in therapeutic practice (r=.31, P<.001). There was also a significant correlation between the amount of time therapists spent using the Internet and the number of times they were targeted for online research by patients (r=.22, P=.002). Results also suggest that a patient's interest in knowing his or her therapist seems to increase with treatment in long-term therapy.

Did the Therapists Take Precautions to Control the Information Available About Them on the Internet?

Most therapists (58.9%, 122/207) controlled the information available about them on the Internet for security reasons by uploading only carefully selected information. Almost one-third (29.5%, 61/207) preferred not to post any personal information online and 46.4% (96/207) did use search engines to check for the online information available about them. Nearly 10% (21/207) did not think it was necessary to protect themselves or had never thought about this. Only 4.3% (9/207) used Google alerts (a tool for online Web monitoring of new content that can also be used for names, etc) for searching their own name (to see when any new entries became available online), whereas 7.7% (16/207) employed other methods to keep updated. Only 40.5% (84/207) declared their membership to social networks and only 9 of 207 (4.3%) therapists allowed unrestricted access to their social network accounts, with most members of social

networks sharing their details only with friends or using false names or nicknames.

Discussion

Principal Results

This study investigated PTG by German psychotherapists, focusing on the experience of and attitudes toward PTG for a sample of therapists. All health care professionals have the option of using the Internet for looking up information about their patients, but this has particular relevance for psychotherapeutic relationships. Not only must psychotherapists keep in mind the ethical aspects of PTG (which would be of concern to all health care professionals), but they must also consider their therapeutic relationship with their patients, which raises many more aspects of concern. For example, what might be the effects of PTG on the therapeutic relationship, such as in countertransference that is not reflected but acted out? For certain, PTG influences the therapeutic relationship on a very particular individual basis, and often with profound consequences, such as harm to the bond of trust. When could PTG be seen as a symptom of a failed therapeutic relationship? Can PTG be understood as a violation of borders when it comes to the terms of the rule of abstinence on the basis of the code of ethics of psychological psychotherapists in Germany? Stellpflug and Berns [13] state that the relationship of trust between therapist and patient should not be abused for the satisfaction of the therapist's own interests and needs; this would mean that there has been a clear violation of guidelines when, for instance, a therapist searches on the Internet for information on their patients out of curiosity. Conversely, are there any reasons or situations that legitimate PTG? If its use is legitimate, how should the therapist proceed with information found on the Internet?

Results of the current study show that the majority (84.5%) of therapists who responded had not actively engaged with the topic of PTG. Yet 39.6% said that they had already searched for patients' online information, which proves that there has been use of the Internet as a source of information about patients without full consideration. The correlations found between research activities and the general use of Internet are not surprising: frequent use of the Internet and the consequent integration of this medium into daily life make its use in other contexts more likely. In contrast, the relationship found between the psychotherapy orientation of the therapist and attitudes toward PTG is more notable: psychodynamic-oriented therapists were much more often of the opinion that there were no justified reasons for PTG than were their behaviorally trained colleagues. Given that PTG is almost never discussed during their therapeutic education, these differences in attitude must be due to broader aspects of their therapeutic positioning, such as their conception of the working alliance, the therapeutic relationship,

the rule of abstinence, or privacy. Further studies regarding this are needed. In addition to studies that focus on the psychotherapist taking into account the bidirectional bond between psychotherapist and the patient, there is also a need to focus on the patient's perspective: the possibility of patients using the Internet to gain information about psychotherapists and how to respond to this. As well as describing the use of information gained from the Internet about therapists and patients, there would be value in discussing the clinical utility of this information-seeking behavior.

Limitations

In general, PTG is seldom discussed and has not been the object of empirical scientific analysis. Therefore, this survey should only be thought of as a first explorative study to improve understanding of PTG. Due to the data collection procedure used, this study did not involve a representative sample of Internet-using psychotherapists in Germany (let alone worldwide). We were not able to test whether PTG was over- or underestimated in our sample. However, methodical studies have shown that Web-based surveys can achieve comparable response rates to questionnaires delivered by mail [14]. A theoretical bias also cannot be excluded; therapists who are interested in ethics may be overrepresented in the study and their interest in ethics may be driven by the idea of a therapeutic use of the Internet. However, there is no evidence of any self-selection of participants of this kind. A further limitation was that the therapist-patient dyad was not investigated. This will clearly need to be looked at in future studies.

Implications

In future, discussion of PTG should become part of therapists' education and training. For instance, as well as giving information about the prevalence and circumstances of PTG gained from empirical studies such as this one, emerging therapists should be pointed toward the influence of PTG on the therapist-patient relationship. This could be accomplished through a discussion of the pros and cons of PTG, and augmented by case studies and analyses of the feelings of countertransference and transference of self. Furthermore, the complex nature of problems related to PTG should be introduced into codes of ethics to provide explicit guidance for therapists in practice. In the first therapy session, the role of modern media in the therapeutic process should be discussed (eg, whether the therapist can be contacted via email or text messaging or whether mental health programs should be a part of therapy); in this context, the need to search online for information about each other can be adressed. Potential implicit expectations of patients (eg, in searching for the therapist on Facebook and requesting to be accepted as a friend) open up new and wide fields that need to be understood to maintain quality in patient treatment.

XSL•FO

RenderX

Conflicts of Interest

None declared.

References

1. Eichenberg C, Kienzle K. Psychotherapeuten und Internet. Psychotherapeut 2011 Jun 11;58(5):485-493. [doi: 10.1007/s00278-011-0833-4]
2. Zur O, Williams MH, Lehavot K, Knapp S. Psychotherapist self-disclosure and transparency in the Internet age. Prof Psychol-Res Pr 2009;40(1):22-30. [doi: 10.1037/a0014745]
3. Eichenberg C, Wolters C, Brähler E. The internet as a mental health advisor in Germany--results of a national survey. PLoS One 2013 Nov;8(11):e79206 [FREE Full text] [doi: 10.1371/journal.pone.0079206] [Medline: 24278121]
4. Ensher EA, Nielson TR, Grant-Vallone E. Tales from the hiring line: effects of the Internet and technology on HR processes. Organ Dyn 2002 Dec;31(3):224-244. [doi: 10.1016/S0090-2616(02)00111-0]
5. Clinton BK, Silverman BC, Brendel DH. Patient-targeted googling: the ethics of searching online for patient information. Harv Rev Psychiatry 2010;18(2):103-112. [doi: 10.3109/10673221003683861] [Medline: 20235775]
6. Lehavot K, Barnett JE, Powers D. Psychotherapy, professional relationships, and ethical considerations in the myspace generation. Prof Psychol-Res Pr 2010;41(2):160-166. [doi: 10.1037/a0018709]
7. American Psychological Association. Ethical Principles of Psychologists and Code of Conduct. 2010. Privacy and confidentiality URL: http://www.apa.org/ethics/code/principles.pdf [accessed 2015-12-17] [WebCite Cache ID 6dqxCxwSD]
8. Berufsverband Deutscher Psychologen. Ethische Richtlinie der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbandes Deutscher Psychologinnen und Psychologen. 2005. URL: http://www.bdp-verband.org/bdp/verband/clips/BDP_Ethische_Richtlinien_2005.pdf [accessed 2015-12-17] [WebCite Cache ID 6dqxEgWvj]
9. 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;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
10. Unipark Survey Software. URL: http://www.unipark.com/en/ [accessed 2015-12-17] [WebCite Cache ID 6dqxNbDzp]
11. Mayring P. Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Klagenfurt, Austria: GESIS–Leibniz Institute for the Social Sciences; 2014. URL: http://www.ssoar.info/ssoar/handle/document/39517 [accessed 2015-12-17] [WebCite Cache ID 6dqxR8a4Z]
12. Bundesarztregister. 2016. Statistische Informationen aus dem Bundesarztregister, KBV URL: http://gesundheitsdaten.kbv.de/cms/html/16396.php [accessed 2015-12-17] [WebCite Cache ID 6eHetzfFj]
13. Stellpflug MH, Berns I. Text und Kommentierung. In: Musterberufsordnung für die Psychologischen Psychotherapeuten und Kinder-und Jugendlichenpsychotherapeuten. Berlin/Heidelberg: Psychotherapeutenverlag; 2015:978-973.
14. Kaplowitz MD, Hadlock TD, Levine R. A comparison of Web and mail survey response rates. Public Opin Quart 2004 Apr 22;68(1):94-101. [doi: 10.1093/poq/nfh006]

Abbreviations

PTG: patient-targeted googling

<u>Original Paper</u>

Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment

Timothy W Bickmore[1*], PhD; Dina Utami[1*], MS; Robin Matsuyama[2*], PhD; Michael K Paasche-Orlow[3*], MD, MPH

[1]Northeastern University, College of Computer and Information Science, Boston, MA, United States

[2]Virginia Commonwealth University, Richmond, VA, United States

[3]Boston Medical Center, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Timothy W Bickmore, PhD
Northeastern University
College of Computer and Information Science
360 Huntington Ave, 910-177
Boston, MA, 02115
United States
Phone: 1 6173735477
Fax: 1 617 812 2589
Email: bickmore@ccs.neu.edu

Abstract

Background: Conventional Web-based search engines may be unusable by individuals with low health literacy for finding health-related information, thus precluding their use by this population.

Objective: We describe a conversational search engine interface designed to allow individuals with low health and computer literacy identify and learn about clinical trials on the Internet.

Methods: A randomized trial involving 89 participants compared the conversational search engine interface (n=43) to the existing conventional keyword- and facet-based search engine interface (n=46) for the National Cancer Institute Clinical Trials database. Each participant performed 2 tasks: finding a clinical trial for themselves and finding a trial that met prespecified criteria.

Results: Results indicated that all participants were more satisfied with the conversational interface based on 7-point self-reported satisfaction ratings (task 1: mean 4.9, SD 1.8 vs mean 3.2, SD 1.8, $P<.001$; task 2: mean 4.8, SD 1.9 vs mean 3.2, SD 1.7, $P<.001$) compared to the conventional Web form-based interface. All participants also rated the trials they found as better meeting their search criteria, based on 7-point self-reported scales (task 1: mean 3.7, SD 1.6 vs mean 2.7, SD 1.8, $P=.01$; task 2: mean 4.8, SD 1.7 vs mean 3.4, SD 1.9, $P<.01$). Participants with low health literacy failed to find any trials that satisfied the prespecified criteria for task 2 using the conventional search engine interface, whereas 36% (5/14) were successful at this task using the conversational interface ($P=.05$).

Conclusions: Conversational agents can be used to improve accessibility to Web-based searches in general and clinical trials in particular, and can help decrease recruitment bias against disadvantaged populations.

Introduction

The majority of US adults look online for health information [1,2]. However, disparities in the use of the Internet for finding health information remain [3,4]. One specific cause of these disparities may be that keyword-based search engines such as Google—although the primary search portals for most users—may actually represent a significant barrier for many disadvantaged individuals. Prior research has demonstrated that people with low health literacy, the ability to acquire and act

on information related to health care [5,6], have particular difficulty using keyword-based search interfaces. Agree et al [3] demonstrated that individuals with low health literacy had lower success rates when using these interfaces to search for general health information on the Web. Usability by people with low health literacy is important because this population comprises 36% of US adults [5].

In addition to general-purpose search engines, many search engines and interfaces have been developed for specific kinds of health care information. One example is the clinical trial search engine, which retrieves descriptions of clinical trials from a repository or database [7]. Several of these search engines are available on the Web, developed by both commercial firms and the US government (eg, the National Cancer Institute [8]). Individuals use these search engines to find trials for which they may be eligible and in which they may be interested in participating. Utami et al [9] found that individuals with low health literacy found fewer clinical trials and took longer to complete standardized search tasks using a Web-based clinical trial search engine compared to those with adequate health literacy. Usability of clinical trial search engines by people with low health literacy is especially important because there is a disproportionate representation of minorities in this group [10-12] leading to reduced access for disadvantaged populations to information about clinical trials. Although Web-based clinical trial search engines hold the promise of providing universal access to information, conventional search systems may further promote disparities in clinical trial recruitment by catering primarily to populations of well-educated individuals with high levels of health and computer literacy.

Conventional Web form-based search engine user interfaces (eg, Google) typically make exclusive use of user-supplied keywords, whereas others combine keyword input with multiple-choice options, referred to as "facet-based" search interfaces [13]. Several prior studies have investigated the use of these search interfaces for users with low domain knowledge [14], who speak a language that is different than that of the Web form [15], who are children [16], or older adults [17], all who share characteristics with our task and population. These studies have demonstrated that even the simplest keyword-based search interfaces are unusable for many users and that special design considerations—such as simplifying results [17] and providing language and interaction support [16]—are important for disadvantaged users. Users may also be influenced by contextual cues when evaluating results from search engines [18] and those with low health literacy may be particularly susceptible to these cues when evaluating search results, relying on such features as position in search results, quality of pictures, and celebrity endorsements [19].

In this paper, we describe the design of a Web-based clinical trial search engine that we designed to mitigate barriers associated with low health literacy. The search task is framed as a conversation with an animated character to make it as familiar and approachable as possible, and a number of additional features and simplifications were made to help users with low health literacy navigate the overall clinical trial search process. We conducted a randomized trial, comparing the conversational search engine to an existing conventional Web-based search engine.

Design of the Conversational Search Engine

The overall task the conversational search engine supports is finding one or more cancer-related clinical trials for which the user is eligible, based on initial demographic criteria and in which the user is provisionally interested, using publically available information. The search engine indexes trials from the National Cancer Institute's (NCI) database of more than 10,000 active trials (at the time this work was conducted) [8].

Based on our experience in developing several health counseling dialog systems for patients with low health literacy [20], previous studies demonstrating greater user recall with audiovisual information combined with conversational style [21], and with animation combined with speech compared to text [22], we designed the overall interaction as a dialog with an embodied conversational agent [23]. The agent speaks using synthetic speech, generated from an augmented transition network-based dialog engine [24], template-based text generation [25], and a dynamically updated user model accompanied by conversational nonverbal behavior (eg, hand gestures, facial displays, gaze) animated in synchrony with the speech [26] (Figure 1). The agent also manipulates artifacts it is discussing with the user; in this case, documents that represent aspects of the clinical trials being discussed. User inputs to the conversation are restricted to multiple-choice selection of utterances from a list that is dynamically updated during each turn of the conversation. Thus, the interaction is system-initiated at the dialog adjacency-pair level (eg, agent question / user response), but user initiative is provided by allowing the user to select topics of conversation and ask questions at predefined points in the dialog by selecting from predefined lists. We have successfully used this interface modality with more than a thousand patients in clinical trials, including hundreds who have low health literacy and many who have never touched a computer before [20,27]. The resulting system could be characterized as a "fully faceted" search interface in which users are never asked to recall and type text, but are always scaffolded with the range of possible inputs they can make [13].

The overall search experience is framed as an extended conversation, in which the user is first interviewed about their requirements and preferences and then shown candidate trials with the agent providing as much scaffolding—through tutorials, explanations, and suggestions—along the way as possible. Given that clinical trial descriptions can be very complex and tedious for users to read, we err on the side of eliciting as much information as possible from users before the search in an attempt to identify trials that are most fitting. In addition, we designed the system to display information about a trial in stages, revealing only the details a user needs at each point of their evaluation. The overall flow of a typical conversation is shown in Figure 2.

To define the search criteria the agent elicits from the user, we leveraged qualitative findings from our usability study [9]. Participants in this study were asked to choose between pairs of clinical trial descriptions and then asked to explain their rationale. Analysis of explanations using grounded theory [28]

revealed information-seeking practices and deliberation themes. We cross-referenced the resulting list of search criteria preferences elicited from users in this study with the clinical trial schema in the NCI database. We found that some user criteria already existed as database indexes, including participant age, sex, cancer type, study geographic location, trial type and phase, and the use of an investigational drug. Additional user criteria did not exist as database indexes, but could be inferred through text classification of text fields in the database, including subjective assessment of the likelihood that study participation would involve painful procedures, subjective assessment of protocol invasiveness (eg, survey vs diagnostic vs treatment), and overall time commitment. These inferred criteria (pain, invasiveness, and time burden) were computed for each trial based on a decision tree algorithm (ID3 [29]) that used word occurrence features in the trial description text trained on hand-rated examples. In our runtime clinical trial search algorithm, the search criteria that could be mapped to existing database indexes are used to search the database, whereas the criteria inferred via text classification are used to sort results.

In addition to the overall structure of the interaction shown in Figure 2, we included several features in the search user interface to assist users with low literacy in their search:

1. Dictionary: the NCI website provides users with a dictionary of medical terms; however, this dictionary is available on the NCI site as a separate module from the search engine. In our user studies of this website, we observed that finding definitions often distracted users from their main search task. We integrated a dictionary with our search functionality; although the agent explains search results to users, the character automatically extracts difficult terms from the text and offers to explain them.

2. Read aloud: users are able to ask the agent to read aloud, and repeat if necessary, any clinical trial text or definition. While speaking aloud, the agent holds up a visualization of the text, enabling users to read along.

3. Simplified title: clinical trial titles can be very long, complex, and hard to remember. We simplified display titles using the phase and type of the trial, and cancer type the trial is for (eg, replacing "A Study of BFTZ2252 Given With Patanobib or Dopepaxel in Subjects With Non-Small Cell Lung Cancer" with "Phase II Treatment Trial for Lung Cancer").

4. Education modules: the agent offers to explain several concepts underlying clinical trials, such as voluntariness and risk, at appropriate points in the search.

5. User search criteria confirmation: the agent periodically displays the user's search criteria and allows them to revise their choices. This is done after the initial interview, after each matching clinical trial is discussed, and whenever the agent runs out of trials to show the user.

6. Query refinement: if a search returns no results or the user exhausts the list of indexed trials, the agent suggests ways to modify the search criteria.

7. Bookmarking: users can save references to clinical trials for review in subsequent search sessions.

8. Summary of views: the agent periodically displays a list of trials viewed during the search session with bookmarked trials highlighted.

9. Levels of detail: each clinical trial description is presented in 3 levels of detail with associated displays. Users are initially only shown the title and eligibility criteria for a trial (see Figure 1), but are then given the option to view the trial research purpose, after which they are given the option of viewing details of the trial procedures.

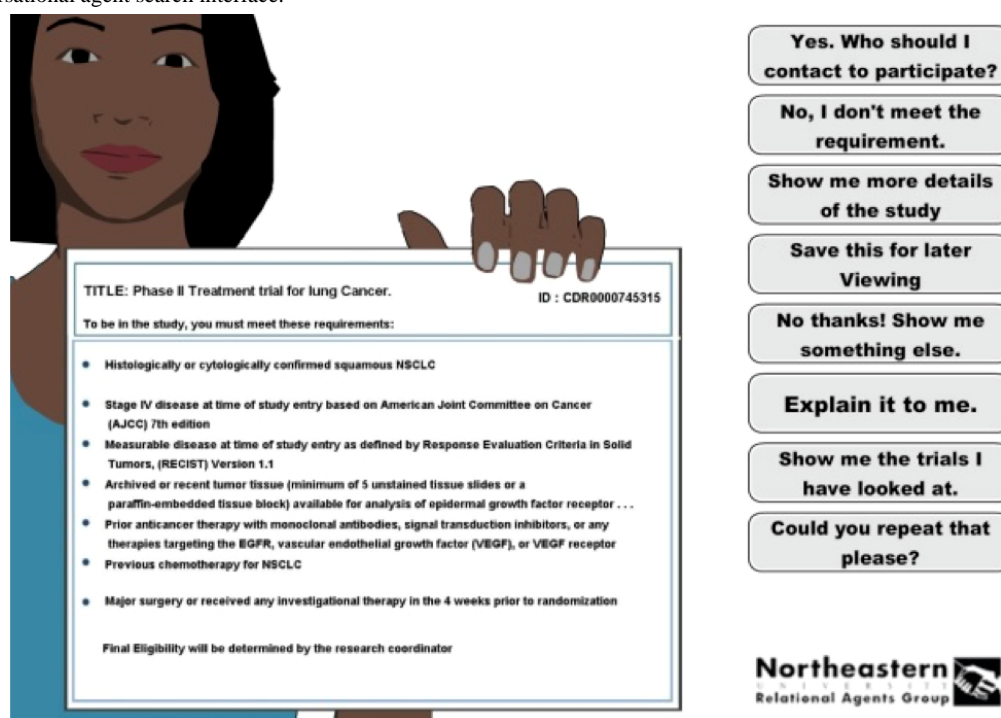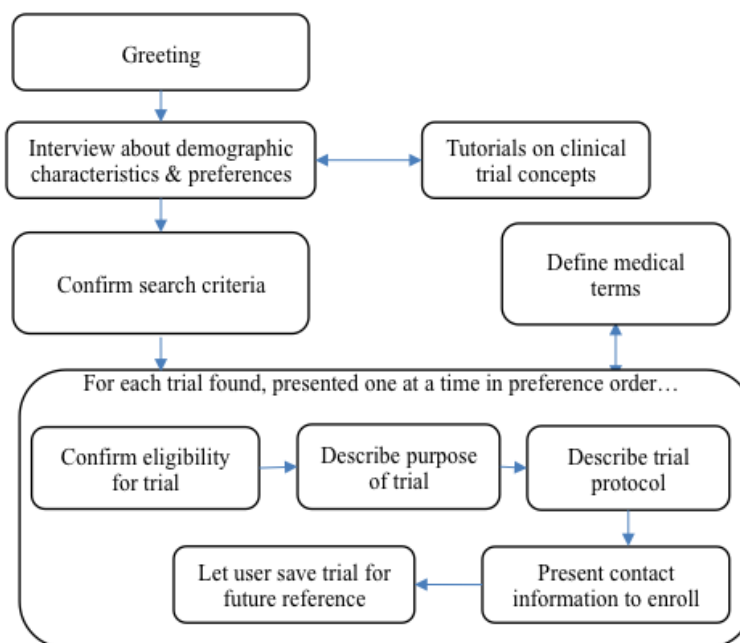**Figure 1.** Conversational agent search interface.

Figure 2. Typical dialog flow in a search session.

Study Aims

The overall aim of this work is to develop a Web-based search interface that is more usable by individuals with low health and computer literacy. We hypothesize that the conversational search interface will lead to greater search successes and higher levels of satisfaction compared to conventional keyword- and facet-based search engines for all users, but that the differences will be especially pronounced for individuals with low health literacy.

Methods

To evaluate our system, we conducted a between-subjects randomized trial comparing our conversational agent search engine ("agent") to the conventional facet- and keyword-based search engine ("control") developed by the NCI ([7]) with both search engine interfaces indexing the same set of clinical trials. Participants were recruited from a pool of adult English-speaking cancer patients from across the literacy spectrum. Participants already had sociodemographic measures recorded and health literacy assessments completed for a prior study at Virginia Commonwealth University. The study protocol was approved by the Boston Medical Center and Virginia Commonwealth University IRBs and informed consent was obtained from all study participants. Health literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine (REALM) [30]. Participants were split into adequate and inadequate health literacy groups using a REALM score of 9th grade as a cut-off as other authors have done [31-34].

We asked participants to perform 2 search tasks. In task 1, participants were asked to search for a clinical trial for which they would be provisionally eligible and in which they would be interested. In task 2, they were asked to search for a trial for someone else with specified eligibility criteria (ie, age, cancer type, trial type, geographic location) as a standardized test so that we could determine whether any trials they found actually matched the specified criteria (Textbox 1).

To ensure accessibility for participants with low computer and health literacy, we designed the experiment so that they were able to do the study either in the laboratory (for those without access to computers) or at home. The experiment software first gave participants a short tutorial on using the system, including a practice task. The first search task was then displayed. To complete each task, participants were redirected onto another Web page that had the agent or the NCI search engine (Figure 3). If the participant found a trial, they entered the clinical trial ID number into a text field and clicked an "I found a trial" button. If they could not find a trial, they clicked on an "I cannot find a trial" button. As soon as users completed each task, they were prompted to fill in a Web form questionnaire that captured the study measures. At the completion of each task, the first 7 questions in Table 1 were automatically administered via Web forms. Participants were also asked to recall the number of trials they examined and the number of these that met their criteria after each task. The Web server also captured the clinical trial ID that the participants found (if any) and the time needed to complete each task. At the completion of both tasks, the remaining 5 questions (questions 8-12 in Table 1) were administered verbally by a research assistant (for those conducting the study in person) or over the phone (for those conducting the study remotely). Participants completing the study at home did so on their own, without online assistance, and were called within 36 hours of completing the online tasks to obtain final outcome measures. Nonparametric statistics (chi-square tests for frequencies and Mann-Whitney U tests for all other measures) were used given the nature of the data and nonnormality of most distributions.

Textbox 1. Standardized search task (task 2).

Now here is your second task. Please write it down.

This is Rosa. She is a cancer patient.

<IMAGE OF ROSA>

Here is some information about Rosa:

Age: 70 years old.

Cancer type: Breast cancer.

Trial type she would like: Treatment trial.

Location of trial she would like: Can be anywhere.

We would like you to use the information above to find a clinical trial for Rosa.

Once you find a trial, please enter its ID number into the box on the bottom right of the screen where it says "TRIAL ID" and click the button that says "I found a trial." Also, please write down the protocol ID number on a piece of paper.

If you have spent some time looking but do not think you can find a trial, then click on the button that says "I cannot find a trial" at the bottom left corner of the screen.

If you are ready to begin, click the "I am ready" button.

Figure 3. Experimental setup for control condition.
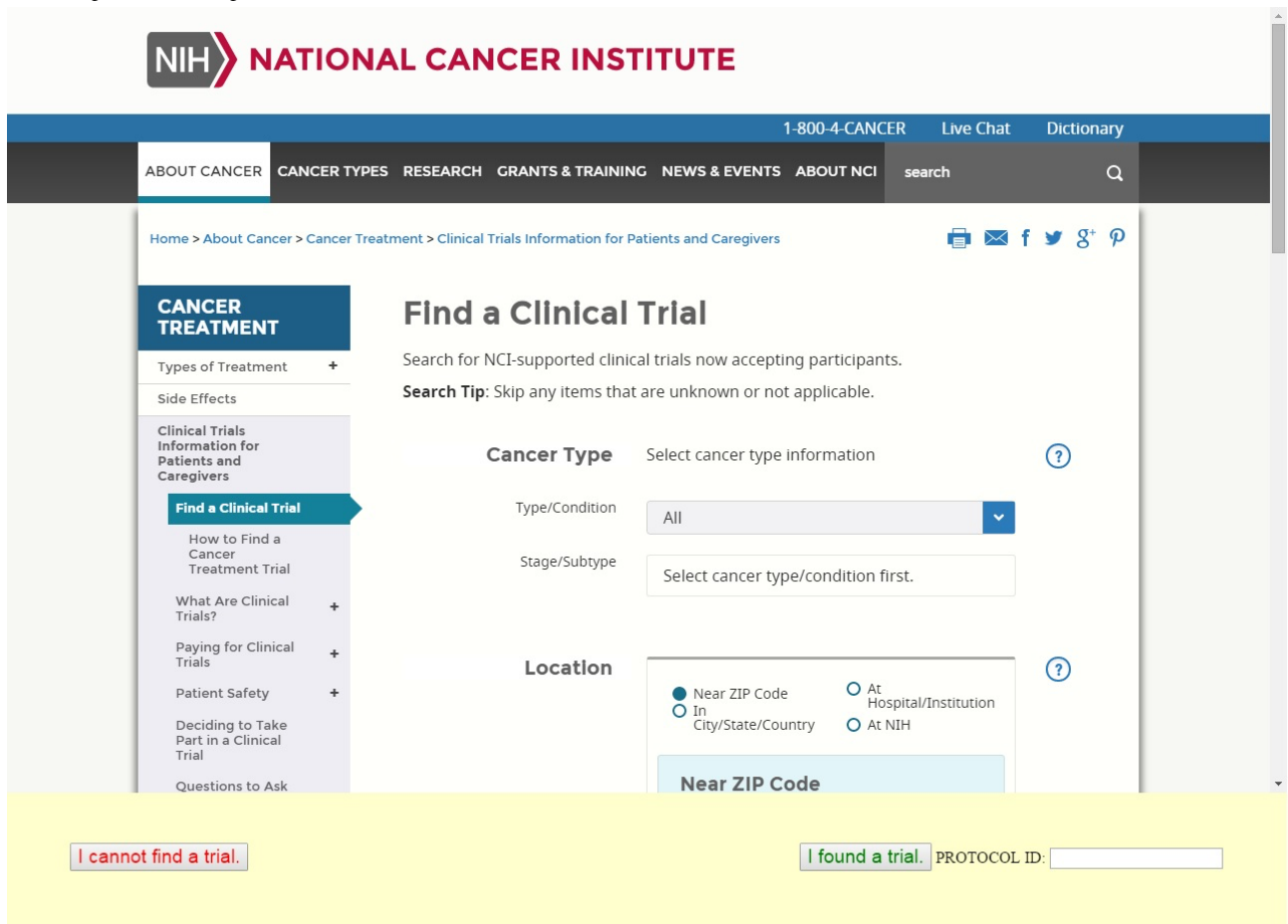
Table 1. Self-report measures used in the study.

Scale self-report item	Anchor 1	Anchor 7
1. To what degree did you know what you wanted in a trial? (task 1)	Not at all	Exactly
2. To what degree did the trial match what you were looking for?	Not at all	Exactly
3. How likely are you to sign up for the trial that you found? (task 1)	Not likely	Very likely
4. How much time do you feel it took to use the system?	Too little	Too much
5. How satisfied were you with the clinical trial search system?	Not at all	Very
6. How frustrated do you feel right now?	Not at all	Very
7. How pleased do you feel right now?	Not at all	Very
8. How much pressure did you feel to volunteer for a trial?	No pressure	A lot of pressure
9. How much information do you feel was presented by the system?	Not enough	Too much
10. How likely would you be to use the system again, if you wanted to find another trial?	Not likely	Very likely
11. How likely would you be to recommend the system to someone else who was looking for a trial?	Not likely	Very likely
12. How much do you trust the information you received from the system?	Not at all	Very much

Results

Participant sociodemographic information is shown in Table 2. A total of 89 individuals participated; mean age was 59.2 (SD 9.8) years, 46% (48/89) were female, and 27% (23/89) had low health literacy. A current cancer diagnosis was reported by 98% (87/89) of participants: 32% (28/89) with hematologic cancer, 14% (12/89) with breast cancer, 14% (12/89) with genitourinary cancer, 14% (12/89) with head and neck cancer, and 10% (9/89) with lung cancer. Most (70%, 62/89) reported regular computer use and regular use of Web-based search engines (52%, 46/89). Although only 21% (19/89) reported previous participation in a cancer-related clinical trial, 52% (46/89) expressed interest in participating in one. Approximately half of the participants (48%, 43/89) were randomized to the agent condition. Of the 89 participants, 53 (60%) conducted the study in the laboratory and 36 (40%) conducted the study over the Web at home. A few participants could not complete some of the tasks (17%, 14/89) due to technical or other problems. The primary study results are shown in Table 3.

Task 1 Results for All Participants

In the initial task, participants were asked to find a clinical trial for themselves. Most participants started this task without a clear idea of what they were looking for and rated the degree they knew what they wanted in a trial a mean 2.8 (SD 1.9) on a scale of 1 to 7. Nevertheless, 45% (19/42) in the agent group and 31% (14/45) in the control group successfully found a trial; there was no significant difference between groups ($\chi^2_1=1.8$, $P=.52$). The degree to which participants felt these trials matched what they were looking for was significantly greater in the agent condition compared to the control condition (mean 3.7, SD 1.6 vs mean 2.7, SD 1.8, $U=465$, $P=.01$). Participants were significantly more satisfied with the agent compared to the conventional interface (rating mean 4.9, SD 1.8 vs mean 3.2, SD 1.8, $U=363$, $P<.001$) and felt significantly less frustrated (rating mean 2.1, SD 1.7 vs mean 3.7, SD 2.2, $U=405$, $P<.001$) and more pleased (rating mean 5.1, SD 2.1 vs mean 3.4, SD 1.9, $U=380$, $P<.001$) with the agent after completing the task compared to those in the control condition.

There were no significant differences between results for those who completed the study in person versus at home.

Task 2 Results for All Participants

In the second task, participants were asked to find a clinical trial that satisfied a prespecified set of criteria as a standardized task. Although 48% (20/42) in the agent group and 40% (18/45) in the control group claimed to find trials that met the criteria ($\chi^2_1=0.5$, $P=.52$), only 43% (18/42) and 31% (14/45), respectively, actually found a correct trial ($\chi^2_1=1.3$, $P=.28$). However, participants in the agent group felt that the trials they found matched the criteria to a greater degree compared to those in the control group (mean 4.8, SD 1.7, vs 3.4, SD 1.9, $U=381$, $P<.001$). As with task 1, participants in the agent group were significantly more satisfied (rating mean 4.8, SD 1.9 vs mean 3.2, SD 1.7, $U=336$, $P<.001$) and pleased (rating mean 4.6, SD 1.8 vs mean 3.1, SD 1.7, $U=358$, $P<.001$), and significantly less frustrated (rating mean 2.6, SD 1.9 vs mean 3.8, SD 2.2, $U=429$, $P=.01$) after completing their task compared to those in the control group.

As with task 1, searching with the agent tended to take longer compared to the conventional interface (mean 8.2, SD 5.3 minutes vs mean 6.4, SD 4.3 minutes), but this did not meet statistical significance ($U=507$, $P=.06$). However, participants felt the agent took significantly less time compared to the conventional interface (mean 4.2, SD 1.1 vs 5.1, SD 1.7, $U=466$, $P=.03$).

Table 2. Participant sociodemographics.

Variable	All N=89	Agent n=43	Control n=46	P
Sex (female), n (%)	48 (54)	18 (42)	30 (65)	.03
Age (years), mean (SD)	59.2 (9.8)	58.6	59.7	.59
Race, n (%)				.45
Black	41 (46)	19 (44)	22 (48)	
White	48 (54)	24 (56)	24 (52)	
Other	0 (0)	0 (0)	0 (0)	
Hispanic or Latino, n (%)	0 (0)	0 (0)	0 (0)	N/A
Education, n (%)				.38
<High school	15 (17)	9 (21)	6 (13)	
High school	10 (11)	6 (14)	4 (9)	
>High school	62 (71)	27 (63)	35 (78)	
Married, n (%)	40 (45)	16 (32)	24 (52)	.20
Health literacy (REALM Score)				
Mean (SD)	57 (15)	56 (16)	59 (15)	.35
Adequate (≥60), n (%)	65 (73)	27 (64)	38 (83)	.05
Study location (in person), n (%)	53 (60)	27 (63)	26 (57)	.67
Computer experience, n (%)				.64
Never used one	7 (8)	5 (12)	2 (4)	
Tried one	21 (24)	10 (24)	11 (24)	
Use regularly	53 (61)	24 (57)	29 (64)	
Expert	6 (7)	3 (7)	3 (7)	
Search engine experience, n (%)				.99
Never used one	15 (17)	7 (17)	8 (18)	
Tried one	16 (18)	8 (19)	8 (18)	
Use regularly	45 (52)	22 (52)	23 (51)	
Expert	11 (13)	5 (12)	6 (13)	
Clinical trials knowledge, n (%)				.12
None	16 (18)	12 (29)	4 (9)	
A little	40 (46)	16 (38)	24 (53)	
Fair amount	29 (33)	13 (31)	16 (36)	
Expert	2 (2)	1 (2)	1 (2)	
Participated in cancer clinical trial before (yes), n (%)	18 (21)	7 (17)	11 (24)	.43
Actually interested in participating in a trial now? (yes), n (%)	44 (52)	23 (56)	21 (48)	.52
To what degree do you know what you want in a trial?[a] mean (SD)	2.8 (1.9)	2.7 (1.7)	3.0 (2.1)	.60

[a] Anchor 1=I didn't know at all; anchor 7=I knew exactly.

Results for Low Health Literacy Participants

Table 4 shows the results by study condition for the 24 participants with low health literacy. The results are very similar to those for all study participants (Table 3), with one notable exception: in the standardized task (task 2), none of the low literacy participants were able to find a clinical trial that met the given criteria using the conventional interface. However, 36% (5/14) of low literacy participants were able to find a correct clinical trial using the agent. This difference was near significant (χ^2_1=3.7, P=.05).

Table 3. Primary study results.

Measure	Task 1			Task 2		
	Agent (n=43)	Control (n=46)	P	Agent (n=43)	Control (n=46)	P
Completed task, n (%)[a]	37 (86)	37 (80)	.58	36 (84)	37 (80)	.79
Declared found a trial, n (%)[a]	19/42 (45)	14/45 (31)	.19	20/42 (48)	14/45 (40)	.52
Found a correct trial, n (%)[a]	__[b]	__[b]	__[b]	18/42 (43)	14/45 (31)	.28
Elapsed time (minutes), mean (SD)[c]	12.6 (9.2)	9.0 (8.4)	.06	8.15 (5.3)	6.4 (4.9)	.06
Number of trials examined (self-report), mean (SD)[c]	2.8 (3.0)	3.8 (6.7)	.56	3.0 (3.1)	4.9 (9.4)	.54
Trials examined that meet criteria (self-report; % of trials examined), mean (SD)[c,d]	56 (39)	34 (35)	.06	64 (37)	44 (41)	.09
To what degree did the trial match what you were looking for? (range 1-7),[c] mean (SD)	3.7 (1.8)	2.7 (1.6)	.01	4.8 (1.7)	3.4 (1.9)	.00
How likely are you to sign up for the trial that you found? (range 1-7), mean (SD)[c]	3.3 (1.7)	2.9 (1.7)	.21	__[e]	__[e]	__[e]
How much time do you feel it took to use the system? (range 1-7), mean (SD)[c]	4.3 (1.3)	4.6 (1.8)	.61	4.2 (1.1)	5.1 (1.7)	.03
How satisfied were you with the clinical trial search system? (range 1-7), mean (SD)[c]	4.9 (1.8)	3.2 (1.8)	<.001	4.8 (1.9)	3.2 (1.7)	<.001
How frustrated do you feel right now? (range 1-7), mean (SD)[c]	2.1 (1.7)	3.7 (2.2)	.001	2.6 (1.9)	3.8 (2.2)	.01
How pleased do you feel right now? (range 1-7), mean (SD)[c]	5.1 (2.1)	3.4 (1.9)	.001	4.6 (1.8)	3.1 (1.7)	.001
How much pressure did you feel to volunteer for a trial? (range 1-7), mean (SD)[c]	1.2 (0.6)	1.4 (1.1)	.99	__[e]	__[e]	__[e]
How much information do you feel was presented by the system? (range 1-7), mean (SD)[c]	4.2 (1.7)	4.3 (1.8)	.98	__[e]	__[e]	__[e]
How likely would you be to use the system again, if you wanted to find another trial? (range 1-7), mean (SD)[c]	5.0 (2.1)	4.1 (2.4)	.07	__[e]	__[e]	__[e]
How likely would you be to recommend the system to someone else who was looking for a trial? (range 1-7), mean (SD)[c]	5.1 (2.2)	4.5 (2.5)	.25	__[e]	__[e]	__[e]
How much do you trust the information you received from the system? (range 1-7), mean (SD)[c]	5.7 (1.6)	5.1 (1.9)	.13	__[e]	__[e]	__[e]

[a] Chi-square test.

[b] Task 1 involved participants finding trials they were interested in, so there was no way to objectively assess whether the trials they found were "correct".

[c] Mann-Whitney U test.

[d] Trials examined that meet criteria was a subjective self-report measure.

[e] Task 2 involved participants finding trials to satisfy criteria for a hypothetical patient, so it did not make sense to ask questions related to their own participation.

Table 4. Study results for low health literacy participants.

Measure	Task 1			Task 2		
	Agent (n=15)	Control (n=8)	P	Agent (n=15)	Control (n=8)	P
Completed task, n (%)[a]	12 (80)	8 (100)	.18	12 (80)	8 (100)	.18
Declared found a trial, n (%)[a]	6/14 (43)	1/8 (13)	.14	5/14 (36)	1/8 (13)	.24
Found a correct trial, n (%)[a]	__[b]	__[b]	__[b]	5/14 (36)	0/8 (0)	.05
Elapsed time (minutes), mean (SD)[c]	13.3 (11.4)	8.2 (6.3)	.47	6.8 (4.3)	4.6 (4.3)	.25
Number of trials examined (self-report), mean (SD)[c]	3.0 (3.4)	0.9 (1.1)	.16	2.8 (3.2)	1.5 (2.1)	.36
Trials examined that meet criteria (self-report; % of trials examined), mean (SD)[c,d]	74 (30)	75 (35)	.89	80 (31)	67 (58)	.92
To what degree did the trial match what you were looking for? (range 1-7), mean (SD)[c]	4.1 (1.9)	2.4 (1.4)	.06	5.3 (2.1)	3.3 (2.0)	.04
How likely are you to sign up for the trial that you found? (range 1-7), mean (SD)[c]	3.8 (1.1)	3.6 (1.9)	.81	__[e]	__[e]	__[e]
How much time do you feel it took to use the system? (range 1-7), mean (SD)[c]	3.8 (0.9)	4.1 (2.5)	.83	3.9 (0.3)	4.5 (2.3)	.86
How satisfied were you with the clinical trial search system? (range 1-7), mean (SD)[c]	5.3 (1.6)	2.9 (1.7)	.01	5.7 (1.6)	2.9 (1.4)	.002
How frustrated do you feel right now? (range 1-7), mean (SD)[c]	2.9 (2.0)	4.8 (2.1)	.01	2.3 (2.1)	3.1 (1.9)	.15
How pleased do you feel right now? (range 1-7), mean (SD)[c]	5.5 (1.6)	3.3 (2.7)	.04	5.7 (1.7)	2.6 (1.1)	.001
How much pressure did you feel to volunteer for a trial? (range 1-7), mean (SD)[c]	1.1 (0.3)	2.3 (2.2)	.10	__[e]	__[e]	__[e]
How much information do you feel was presented by the system? (range1-7), mean (SD)[c]	4.8 (1.5)	4.3 (1.5)	.21	__[e]	__[e]	__[e]
How likely would you be to use the system again, if you wanted to find another trial? (range 1-7), mean (SD)[c]	5.5 (1.8)	5.0 (2.4)	.74	__[e]	__[e]	__[e]
How likely would you be to recommend the system to someone else who was looking for a trial? (range 1-7), mean (SD)[c]	6.1 (1.9)	5.1 (2.7)	.45	__[e]	__[e]	__[e]
How much do you trust the information you received from the system? (range 1-7), mean (SD)[c]	6.3 (1.0)	5.1 (2.1)	.71	__[e]	__[e]	__[e]

[a] Chi-square test.

[b] Task 1 involved participants finding trials they were interested in, so there was no way to objectively assess whether the trials they found were "correct".

[c] Mann-Whitney *U* test.

[d] Trials examined that meet criteria was a subjective self-report measure.

[e] Task 2 involved participants finding trials to satisfy criteria for a hypothetical patient, so it did not make sense to ask questions related to their own participation.

Analyzing differences between low and high health literacy participants across all study conditions indicated a few significant differences. Participants with low health literacy were more likely overall to state that the trials they read satisfied their criteria compared to those with high health literacy (task 1: 74% vs 37%, *U*=82, *P*<.001; task 2: 77% vs 49%, *U*=148, *P*=.02). Participants with low health literacy were more likely to say they would sign up for the trial they found in task 1 (rating mean 3.8, SD 1.4 vs mean 2.8, SD 1.8, *U*=380, *P*=.048) and

that they would recommend the system to a friend (rating mean 5.7, SD 2.3 vs mean 4.5, SD 2.3, *U*=379, *P*=.01) compared to those with adequate health literacy.

Discussion

Principal Results

In our comparison of a conversational agent-based search user interface to a conventional keyword- and facet-based search

XSL•FO

RenderX

engine interface, participants were more satisfied with the agent and felt the agent was better at finding trials that matched their criteria compared to the conventional interface. Participants also felt more pleased and less frustrated after interacting with the agent compared to the conventional interface.

In our standardized task (task 2), it is notable that none of the low health literacy participants were able to find a correct clinical trial using the conventional search engine interface, whereas 36% (5/14) were able to do so with the conversational agent. These results reinforce our earlier findings that conventional search interfaces are unusable by individuals with low health or computer literacy [9]. It is encouraging that the conversational interface was able to provide accessibility to at least a third of these users, while being rated more highly on satisfaction by all users, including those with high health literacy. Nonetheless, it appears that research on additional adaptations is warranted in order to succeed with an even broader portion of the population.

The conversational interface does take more time to use compared to the conventional interface: 40% longer in task 1 and 27% longer in task 2 (although these differences were not statistically significant). There are several reasons for this: the time required to hear spoken prompts rather than reading them, the interview by the agent to obtain search criteria, and social dialog, tutorials, and other "off-task talk" used by the agent to improve approachability, engagement, and comprehension. However, users in our target demographic are clearly happy to spend the extra time with the conversational user interface to obtain better results; in the browsing task, they chose to spend almost twice as long finding a trial compared to those using the conventional interface. In addition, their subjective impression of time taken in task 2 indicates that users felt the conversational agent interface actually took significantly less time to use compared to the conventional interface.

Limitations

Our study had several limitations beyond the small number of participants involved. Some (21%) of the users in our study had previously been involved in clinical trials and thus are not representative of the general population of people with cancer because they may have had higher than average background knowledge about clinical trials. However, people with prior trial experience were randomly assigned to the 2 trial arms, so both groups should be equivalent in this regard. Another limitation relates to the use of the REALM as a measure of health literacy. Although this tool successfully differentiated among participants regarding the impact of our intervention, other measures could potentially have provided a more refined capacity to delve deeper within specific dimensions of electronic and computer literacy. Although the eHealth Literacy Scale (eHEALS) was designed for such a purpose, it is a subjective self-report measure [35]. We opted for an objectively scored measure in the current study; future research is warranted to further differentiate how a conversational search interface may ameliorate various dimensions of low health and computer literacy.

Comparison With Prior Work

Several studies have investigated the use of standard keyword-based search interfaces for users with low domain knowledge [14], who speak a foreign language [15], who are children [16] or older adults [17], which all share characteristics with our task and population. These studies have demonstrated that even the simplest keyword-based search interfaces are unusable for many users and that special design considerations—such as simplifying results [17] and providing language and interaction support [16]—are important for users, especially those with low health or computer literacy

Other studies have investigated the use of conversational agents to communicate health information to individuals with low health literacy. Bickmore et al [20,36] and Wang et al [37] have developed conversational agents for physical activity promotion, hospital discharge instruction, explanation of medical documents, and family health history-taking to individuals with low health literacy. Most of these studies have demonstrated that participants with low health literacy have significantly higher levels of satisfaction with conversational interfaces compared to participants with adequate health literacy.

Conclusions

Our findings suggest that conversational agent-based search engine interfaces could be a good alternative to conventional Web form-based interfaces for many kinds of applications, but especially for those intended for low health literacy users or those with limited computer experience or skills.

Conflicts of Interest

None declared.

References

1. National Cancer Institute. Health Information National Trends Survey URL: http://hints.cancer.gov/question-details. aspx?dataset=41&method=mail&qid=757 [accessed 2015-12-23] [WebCite Cache ID 6cENkpafC]

2. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/ health-online-2013/ [accessed 2015-12-23] [WebCite Cache ID 6e0HETiK0]

3.   Agree EM, King AC, Castro CM, Wiley A, Borzekowski DL. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]

4.   van Dijk J. Digital divide research, achievements and shortcomings. Poetics 2006;34(4-5):221-235.

5.   Nielsen-Bohlman L, Panzer A, Kindig D. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004. URL: http://www.ncbi.nlm.nih.gov/books/NBK216032/ [accessed 2015-12-23] [WebCite Cache ID 6e0HlhGII]

6.   Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]

7.   Atkinson NL, Saperstein SL, Massett HA, Leonard CR, Grama L, Manrow R. Using the Internet to search for cancer clinical trials: a comparative audit of clinical trial search tools. Contemp Clin Trials 2008 Jul;29(4):555-564 [FREE Full text] [doi: 10.1016/j.cct.2008.01.007] [Medline: 18346942]

8.   National Cancer Institute. 2014. Find NCI-supported clinical trials URL: http://www.cancer.gov/about-cancer/treatment/clinical-trials/search [accessed 2015-11-29] [WebCite Cache ID 6dO9qQfUX]

9.   Utami D, Bickmore TW, Barry B, Paasche-Orlow MK. Health literacy and usability of clinical trial search engines. J Health Commun 2014;19 Suppl 2:190-204. [doi: 10.1080/10810730.2014.938842] [Medline: 25315593]

10.   Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]

11.   Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]

12.   Volandes AE, Paasche-Orlow M, Gillick MR, Cook EF, Shaykevich S, Abbo ED, et al. Health literacy not race predicts end-of-life care preferences. J Palliat Med 2008 Jun;11(5):754-762. [doi: 10.1089/jpm.2007.0224] [Medline: 18588408]

13.   Hearst M. Search User Interfaces. Cambridge: Cambridge University Press; 2009.

14.   Zhang X, Liu J, Yuan X, Cole M, Belkin N, Liu C. Do users with different domain knowledge select different sets of documents? In: Proceedings of the Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR). 2012 Presented at: Symposium on Human-Computer Interaction and Information Retrieval (HCIR); Oct 4-5 2012; Cambridge, MA.

15.   Komlodi A, Jozsa E, Koles M, Hercegfi K. Search quality differences in native and foreign language searching. 2011 Presented at: Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR); Oct 20, 2011; Mountain View, CA.

16.   Gossen T, Nitsche M, Nurnberger A. Knowledge journey: a Web search interface for young users. 2012 Presented at: Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR); Oct 4-5, 2012; Cambridge, MA. [doi: 10.1145/2391224.2391225]

17.   Aula A, Käki M. Less is more in Web search interfaces for older adults. First Monday 2005;10(5):100-101 [FREE Full text]

18.   Allam A, Schulz PJ, Nakamoto K. The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating Google output. J Med Internet Res 2014;16(4):e100 [FREE Full text] [doi: 10.2196/jmir.2642] [Medline: 24694866]

19.   Mackert M, Kahlor L, Tyler D, Gustafson J. Designing e-health interventions for low-health-literate culturally diverse parents: addressing the obesity epidemic. Telemed J E Health 2009 Sep;15(7):672-677 [FREE Full text] [doi: 10.1089/tmj.2009.0012] [Medline: 19694596]

20.   Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]

21.   Bol N, van Weert JC, de Haes HC, Loos EF, Smets EM. The effect of modality and narration style on recall of online health information: results from a Web-based experiment. J Med Internet Res 2015;17(4):e104 [FREE Full text] [doi: 10.2196/jmir.4164] [Medline: 25910416]

22.   Meppelink CS, van Weert JC, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]

23.   Cassell J. Embodied Conversational Agents. Cambridge, MA: MIT Press; 2000.

24.   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]

25.   Reiter E, Dale R. Building Natural Language Generation Systems. Cambridge: Cambridge University Press; 2000.

26.   Bickmore T, Schulman D, Shaw G. DTask and LiteBody: open source, standards-based tools for building Web-deployed embodied conversational agents. In: Intelligent Virtual Agents. 2009 Presented at: 9th International Conference, IVA 2009; Sep 14-16, 2009; Amsterdam p. 425-431. [doi: 10.1007/978-3-642-04380-2_46]

XSL•FO
RenderX

27.   Bickmore T, Caruso L, Clough-Gorr K, Heeren T. "It's just like you talk to a friend"-relational agents for older adults. Interact Comput 2005;17(6):711-735.

28.   Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage; 2014.

29.   Quinlan J. Induction of decision trees. Mach Learn 1986;1(1):81-106.

30.   Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]

31.   Lindau ST, Basu A, Leitsch SA. Health literacy as a predictor of follow-up after an abnormal Pap smear: a prospective study. J Gen Intern Med 2006 Aug;21(8):829-834 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00534.x] [Medline: 16881942]

32.   Mancuso CA, Rincon M. Impact of health literacy on longitudinal asthma outcomes. J Gen Intern Med 2006 Aug;21(8):813-817 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00528.x] [Medline: 16881939]

33.   Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 2006 Aug;21(8):806-812 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00539.x] [Medline: 16881938]

34.   Lincoln A, Paasche-Orlow MK, Cheng DM, Lloyd-Travaglini C, Caruso C, Saitz R, et al. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J Gen Intern Med 2006 Aug;21(8):818-822 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00533.x] [Medline: 16881940]

35.   Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]

36.   Bickmore TW, Pfeifer LM, Paasche-Orlow MK. Using computer agents to explain medical documents to patients with low health literacy. Patient Educ Couns 2009 Jun;75(3):315-320 [FREE Full text] [doi: 10.1016/j.pec.2009.02.007] [Medline: 19297116]

37.   Wang C, Bickmore T, Bowen DJ, Norkunas T, Campion M, Cabral H, et al. Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories. Genet Med 2015 Oct;17(10):822-830. [doi: 10.1038/gim.2014.198] [Medline: 25590980]

Abbreviations

NCI: National Cancer Institute
REALM: Rapid Estimation of Adult Literacy in Medicine

XSL•FO
RenderX

Original Paper

Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed

Ashley N.D Meyer[1,2], PhD; Christopher A Longhurst[3], MD, MS; Hardeep Singh[1,2], MD, MPH

[1]Houston Veterans Affairs Center for Innovations in Quality, Effectiveness and Safety, Health Services Research and Development, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, United States

[2]Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, United States

[3]Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Hardeep Singh, MD, MPH
Houston Veterans Affairs Center for Innovations in Quality, Effectiveness and Safety
Michael E. DeBakey Veterans Affairs Medical Center
VA HSR&D Center of Innovation (152)
2002 Holcombe Boulevard
Houston, TX, 77030
United States
Phone: 1 713 794 8601
Fax: 1 713 748 7359
Email: hardeeps@bcm.edu

Abstract

Background: Despite visits to multiple physicians, many patients remain undiagnosed. A new online program, CrowdMed, aims to leverage the "wisdom of the crowd" by giving patients an opportunity to submit their cases and interact with case solvers to obtain diagnostic possibilities.

Objective: To describe CrowdMed and provide an independent assessment of its impact.

Methods: Patients submit their cases online to CrowdMed and case solvers sign up to help diagnose patients. Case solvers attempt to solve patients' diagnostic dilemmas and often have an interactive online discussion with patients, including an exchange of additional diagnostic details. At the end, patients receive detailed reports containing diagnostic suggestions to discuss with their physicians and fill out surveys about their outcomes. We independently analyzed data collected from cases between May 2013 and April 2015 to determine patient and case solver characteristics and case outcomes.

Results: During the study period, 397 cases were completed. These patients previously visited a median of 5 physicians, incurred a median of US $10,000 in medical expenses, spent a median of 50 hours researching their illnesses online, and had symptoms for a median of 2.6 years. During this period, 357 active case solvers participated, of which 37.9% (132/348) were male and 58.3% (208/357) worked or studied in the medical industry. About half (50.9%, 202/397) of patients were likely to recommend CrowdMed to a friend, 59.6% (233/391) reported that the process gave insights that led them closer to the correct diagnoses, 57% (52/92) reported estimated decreases in medical expenses, and 38% (29/77) reported estimated improvement in school or work productivity.

Conclusions: Some patients with undiagnosed illnesses reported receiving helpful guidance from crowdsourcing their diagnoses during their difficult diagnostic journeys. However, further development and use of crowdsourcing methods to facilitate diagnosis requires long-term evaluation as well as validation to account for patients' ultimate correct diagnoses.

XSL•FO
RenderX

Introduction

Errors of clinical diagnosis affect at least 5% of US adults every year and approximately half of these errors could result in serious harm to the patients [1]. To address the extent and severity of this problem, both systems and cognitive solutions have been proposed. However, only a few of these have been tested and only a fraction of those tested have been shown to improve diagnostic outcomes [2-4]. Patients with difficult-to-diagnose conditions often seek care from several physicians and institutions before obtaining a diagnosis. One intervention that could benefit patients is the use of second opinions [5-7], and this has been shown to catch previously missed diagnoses, at least in the realms of radiology and pathology [6]. Several formal programs currently exist to provide second opinions to patients. [7] For example, in the NIH Undiagnosed Diseases Network based at several centers across the US [8], medical experts diagnose undiagnosed individuals or those with rare diseases. The program, however, has strict eligibility requirements for patients and requires a clinician referral. Additional programs include Best Doctors' second-opinion program that is open to employee beneficiaries only and Cleveland Clinic's MyConsult program [5,9], both of which involve comprehensive review of patients' medical records, but no dynamic interactions with the patients.

A recently developed software platform, CrowdMed [10], aims to overcome some limitations of the aforementioned programs, namely; strict eligibility requirements, needed referrals, and limited interaction with patients; by leveraging the "wisdom of the crowd" or crowdsourcing to help undiagnosed or misdiagnosed patients. Crowdsourcing is a "participative online activity" in which a group of individuals of varying knowledge, heterogeneity, and number comes together to solve a problem [11]. It has been used for a variety of problems in different fields ranging from simple text translation to more complicated tasks, such as solving the BP oil spill disaster in the Gulf of Mexico [12]. In medicine, it has been utilized for health and medical research, such as estimating flu prevalence [13]; for informatics solutions, including establishing problem-related medication pairs [14], and for examining specific diseases through image analysis. In the latter situation, crowdsourcing has been used to inspect blood samples to determine the presence or absence of malarial infections [15-17] and to categorize colorectal polyps [18,19] or diabetic retinopathy [20]. However, until now crowdsourcing had not been used to come up with a diagnosis from all possible diagnoses a patient might have. Of note, this platform allows laypersons without health care training or experience to participate. Although patients have been "googling for a diagnosis" for more than a decade and even using online symptom checkers [21,22], this is the first description of a crowd of people working together online towards a more accurate diagnosis. We conducted an independent evaluation of this untested approach to determine whether this could be beneficial to patient care.

Methods

A Description of CrowdMed

For a small fee, the CrowdMed website allows undiagnosed patients to submit their clinical information and obtain potential diagnoses expeditiously. Patients anonymously answer a comprehensive set of medical questions and upload relevant test results and images related to their cases (Figure 1).

Patients also decide how long they want their cases open and whether they wish to compensate the case solvers. Anyone (including nonmedical persons) can sign up to be a case solver and select cases they think they can help solve (Figure 2).

While the cases are open, patients and case solvers can discuss details online about potential diagnoses, further work-up that should be done, and newly obtained test results and/or appointments completed with the patients' physicians. Thus, case details can unfold online while the case is still open. All diagnostic suggestions and all case discussions are available to all case solvers as they are suggested and discussed throughout the open period. This enables the entire group of case solvers to work in concert to solve each case.

When a patient's case is closed, the patient receives a detailed report containing the entire list of diagnostic suggestions made by the case solvers and suggested next steps, so that they can discuss them with their physicians. Diagnoses are ranked in decreasing order of "relative popularity." The relative popularity of diagnoses is determined by case solvers' "bets" on each diagnosis in terms of their beliefs that the diagnosis is the most specific, accurate, root cause of the symptoms presented. CrowdMed takes these bets and assigns points to each diagnosis using a prediction market algorithm, thereby determining the "relative popularity" of each diagnosis suggested. Finally, patients are provided with case solvers' reasoning for choosing particular diagnoses. Patients choose which case solver(s) to compensate based on whose answers they found helpful. If the patient decides to reward multiple solvers, they also decide how to divvy up the compensation. Afterward, patients are invited to fill out surveys about their outcomes.

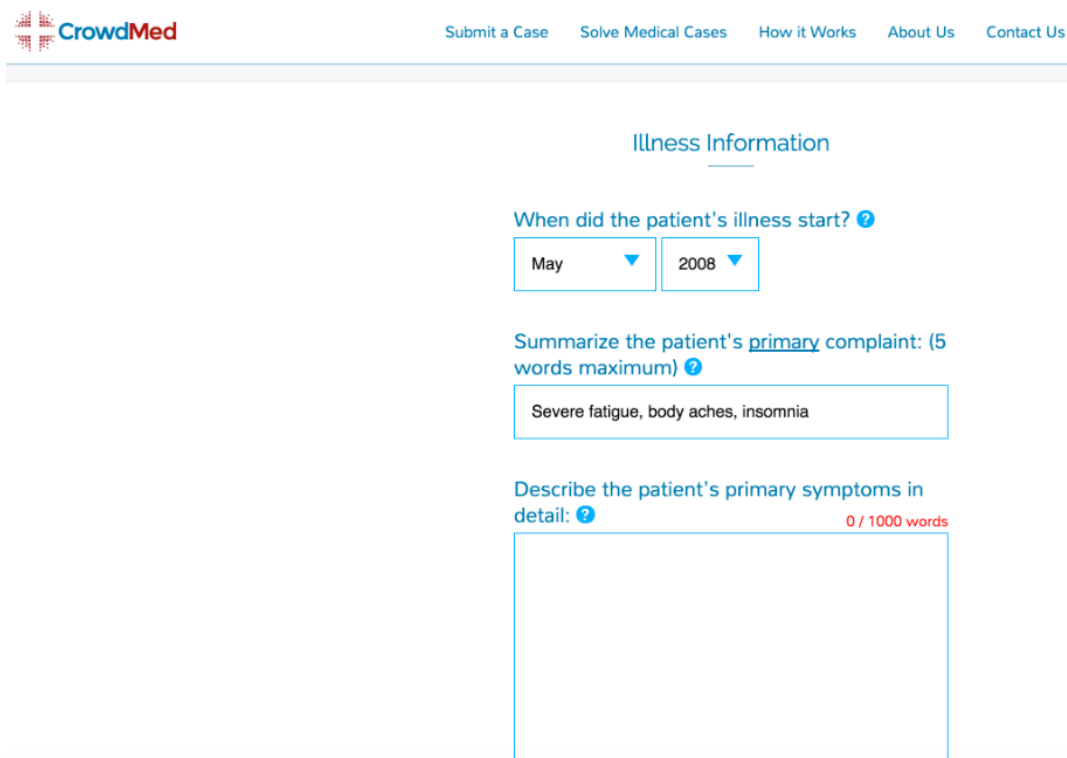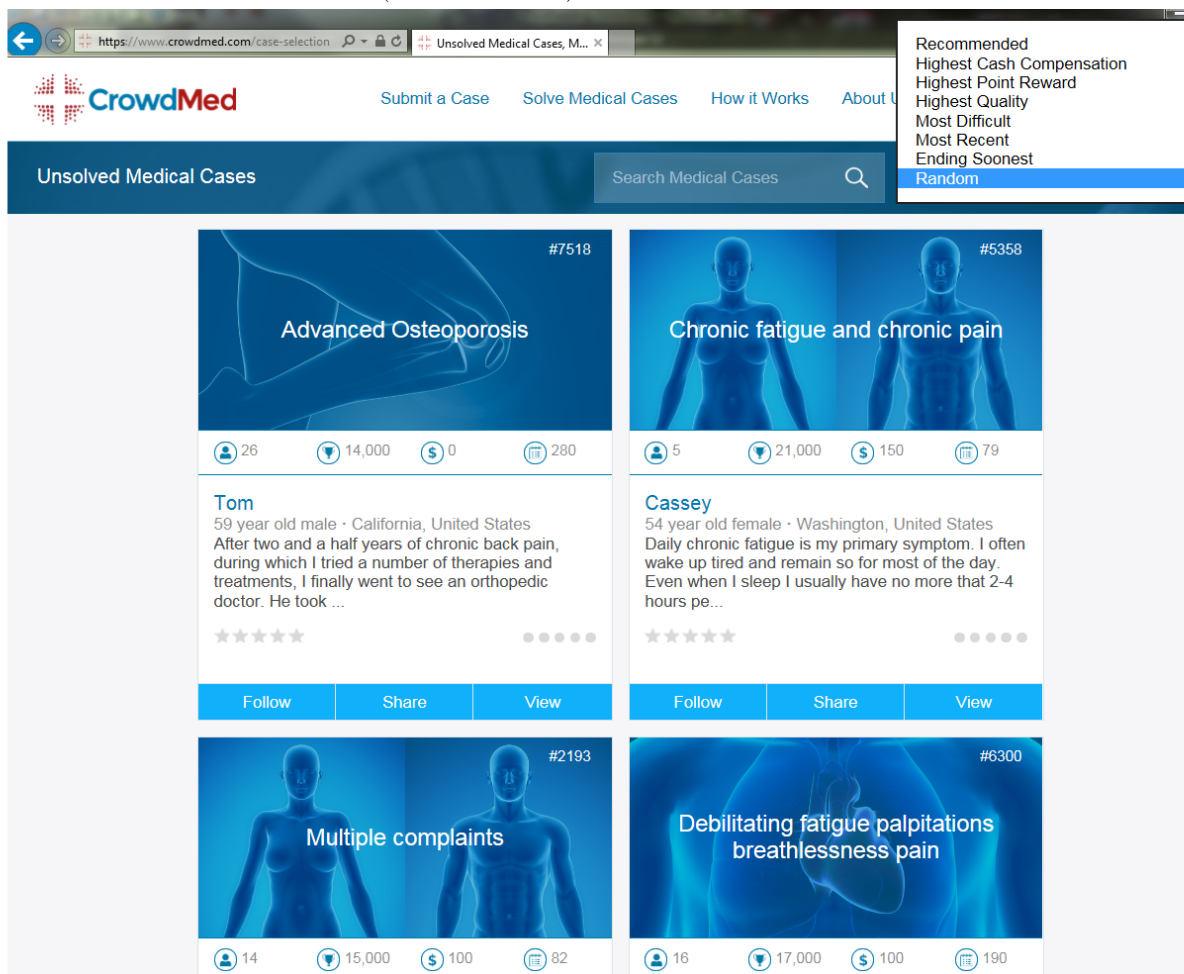**Figure 1.** Screenshot of case submission.

Figure 2. Screenshot of case selection for solvers (names are fictitious).

Independent Evaluation

We independently analyzed all CrowdMed data collected from May 2013 to April 2015. Specifically, we analyzed data on patients' demographic and case characteristics; case solvers' demographic and performance characteristics; and preliminary case outcomes. Outcomes included whether patients would recommend CrowdMed, if the program provided insights leading them closer to correct diagnoses, and estimated improvements in patients' productivity and medical expenses. Data were summarized using descriptive statistics and independent samples *t* tests using IBM SPSS Statistics 22.

Results

Patients and Cases

During the study period, 397 cases were completed (350 from the United States). Patients' self-reported mean (SD) age was 47.8 (18.8) years (age range 2-90) and 182 were males (45.8%).

Before case submission, patients reported visiting a median of 5 physicians (interquartile range [IQR] 3-10; range 0-99), incurred a median of US $10,000 in medical expenses (IQR US $2500-US $50,000; range US $0-US $5,000,000) including payments by both patients and payers, spent a median of 50 hours (IQR 15-150; range 0-12,000) researching their illnesses online, and had symptoms for a median of 2.6 years (IQR 1.1-6.9; range 0.0-70.6). Online case activity lasted a median of 60 days (IQR 30-90; range 2-150) and case solvers were offered a median of US $100 in compensation (IQR US $0-US $200; range US $0-US $2700) for diagnostic suggestions. A total of 59.7% (237/397) of the cases were compensated with a median compensation of US $200 (IQR US $100-US $300; range US $15-US $2700).

Case Solvers

During the study period, CrowdMed had 357 active case solvers; of which 37.9% (132/348) were male, 76.7% (264/344) were from the US, and 58.3% (208/357) worked or studied in the medical industry; including 36 physicians and 56 medical students. Mean (SD) age was 39.6 (13.8) years (range 17-77 years).

Solvers participated in a median of 3 cases (IQR 1.0-12.8; range 0-415), earned a median of US $0 (IQR US $0-US $1.18; range US $0-US $3952) and a mean (SD) of US $93.97 (US $364.72; the majority earned US $0). Median solver rating was 3 (out of 10; IQR 3-6; range 1-10) and significantly higher ($P=.006$) for medical industry-based solvers (mean [SD] 4.8 [2.5]; range 1-10) than for others (mean [SD] 4.1 [2.2]; range 1-10).

Outcomes

At completion, 50.9% (202/397) of patients were likely to recommend CrowdMed to a friend, 59.6% (233/391) reported that the process gave insights leading them closer to correct diagnoses, 57% (52/92) reported estimated decreases in medical expenses, and 38% (29/77) reported estimated improvements in school or work productivity (Table 1).

Table 1. Case outcomes as assessed in a postcase survey.

Case outcomes		n (%)
On a scale of 1-5, How likely are you to recommend CrowdMed to a friend (with 5 being most likely)? (391/397 surveyed answered; 98.5% response rate)		
	1	39 (10.0)
	2	43 (11.0)
	3	107 (27.4)
	4	76 (19.4)
	5	126 (32.2)
Did CrowdMed Medical Detective community provide insights that lead you closer to a correct diagnosis or cure? (391/397 surveyed answered; 98.5% response rate)		
	No	158 (40.4)
	Yes	233 (59.6)
How much do you estimate that your CrowdMed results will reduce the cost of your medical case going forward? (92/147 surveyed answered; 62.6% response rate) [a]		
	1-20%	25 (27.2)
	21-50%	15 (16.3)
	51-80%	10 (10.9)
	>80%	2 (2.2)
	Not at all	40 (43.5)
How much lost work or school productivity do you estimate that your CrowdMed results will help you regain going forward? (77/147 surveyed answered; 52.4% response rate) [a]		
	1-20%	12 (15.6)
	21-50%	8 (10.4)
	51-80%	7 (9.1)
	81-99%	1 (1.3)
	All	1 (1.3)
	None	48 (62.3)

[a]These questions were added to the postcase survey later.

Patients reporting helpful insights from CrowdMed saw fewer doctors (mean [SD] 7.2 [7.3]; range 0-99) before participating than those who did not report receiving helpful insights (mean [SD] 9.2 [10.7]; range 0-50), $P=.047$. The 14 most common diagnoses suggested as the most popular diagnosis for a case are presented in Table 2.

XSL•FO
RenderX

Table 2. The 14 most common diagnoses suggested as the most popular diagnosis across 397 cases.

Diagnosis	n (%)
Lyme disease	8 (2.0)
Dysautonomia	7 (1.8)
Chronic fatigue syndrome	6 (1.5)
Irritable bowel syndrome	6 (1.5)
Mast cell activation disorder	6 (1.5)
Postural orthostatic tachycardia syndrome	5 (1.3)
Ehlers-Danlos syndrome	4 (1.0)
Sjögren's syndrome	4 (1.0)
Abdominal cutaneous nerve entrapment syndrome	3 (0.8)
Gastroesophageal reflux disease	3 (0.8)
Hypothyroidism	3 (0.8)
Multiple sclerosis	3 (0.8)
Myasthenia gravis	3 (0.8)

In addition, some patients informally reported to CrowdMed that the program helped them find diagnoses that their physicians previously were unable to determine, including Sjögren's syndrome and chorda tympani dysfunction.

Discussion

Main Findings

Our independent evaluation suggests that at least some patients with undiagnosed illnesses reported receiving helpful guidance from crowdsourcing their diagnoses during their difficult diagnostic journeys. Several of the conditions most commonly suggested by case solvers are conditions well known to represent diagnostic challenges. The crowdsourcing strategy enabled dynamic interaction between patients and case solvers as more case details unfolded over time.

Novel approaches are needed to help patients who experience difficulties in obtaining a correct and timely diagnosis. In that regard, advantages of using "wisdom of the crowd" could include low cost, increased program accessibility for patients, and relatively quick opinions. Although the data we obtained were useful for understanding this program, there were several limitations of our study. The postparticipation survey was rather limited in scope as it was designed for business purposes and not for research. In addition, there was no way to verify patient-reported data and some patient-reported data might be outside of realistic boundaries (eg, 1 patient reported spending 12,000 hours researching illnesses online). Furthermore, downstream outcomes of patients were not systematically collected, so it is not known what their eventual diagnoses were or if the program identified them accurately. Further development and use of crowdsourcing methods to facilitate diagnosis requires long-term evaluation as well as validation to account for patients' ultimate correct diagnoses.

Although crowdsourcing appears to have potential, it is important to identify factors that lead to successful crowdsourcing to improve the process and help improve patient care. Multidisciplinary research is needed to gain both technical and nontechnical insights into how this can be done. For example, previous researchers have identified the importance of both finding crowd members with the appropriate skills to the relevant problem and providing adequate motivation to the crowd for the successful use of crowdsourcing for problem solving [23]. Finally, the potential legal ramifications of giving individuals without medical degrees (who make up a substantial portion of the case solvers) the ability to render diagnostic opinions would need to be considered [24].

Conclusions

In conclusion, our independent evaluation suggests that some patients with undiagnosed illnesses report receiving helpful guidance from crowdsourcing their diagnosis. Further development and use of crowdsourcing methods to facilitate diagnosis require multidisciplinary research and long-term evaluation that includes validation to account for patients' ultimate correct diagnoses.

XSL·FO
RenderX

conclusions reached, or manuscript preparation; and did not commission this report or provide funding for it. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or any other funding agency.

Conflicts of Interest

None declared.

References

1. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
2. Graber ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, et al. Cognitive interventions to reduce diagnostic error: A narrative review. BMJ Qual Saf 2012 Jul;21(7):535-557. [doi: 10.1136/bmjqs-2011-000149] [Medline: 22543420]
3. Singh H, Graber ML, Kissam SM, Sorensen AV, Lenfestey NF, Tant EM, et al. System-related interventions to reduce diagnostic errors: A narrative review. BMJ Qual Saf 2012 Feb;21(2):160-170 [FREE Full text] [doi: 10.1136/bmjqs-2011-000150] [Medline: 22129930]
4. National Academies of Sciences Engineering and Medicine. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; Sep 22, 2015.
5. Meyer AN, Singh H, Graber ML. Evaluation of outcomes from a national patient-initiated second-opinion program. Am J Med 2015 Oct;128(10):1138.e25-1138.e33. [doi: 10.1016/j.amjmed.2015.04.020] [Medline: 25913850]
6. Payne VL, Singh H, Meyer AN, Levy L, Harrison D, Graber ML. Patient-initiated second opinions: Systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. Mayo Clin Proc 2014 May;89(5):687-696. [doi: 10.1016/j.mayocp.2014.02.015] [Medline: 24797646]
7. Reddy S. The Wall Street Journal. 2015 Aug 24. New ways for patients to get a second opinion: Online services from established medical centers and independent businesses URL: http://www.wsj.com/articles/new-ways-to-get-a-second-opinion-1440437584 [accessed 2015-10-15] [WebCite Cache ID 6cJ0Szema]
8. Undiagnosed Diseases Network. The Undiagnosed Diseases Network. 2015. URL: http://undiagnosed.hms.harvard.edu/ [accessed 2015-12-22] [WebCite Cache ID 6dy5TLYOj]
9. MyConsult Online Expert Opinion. Cleveland Clinic. 2015. URL: http://www.eclevelandclinic.org/aboutMyConsultHome [accessed 2015-10-15] [WebCite Cache ID 6cJ20vGhc]
10. CrowdMed. CrowdMed Inc. 2015. URL: https://www.crowdmed.com/ [accessed 2015-06-26] [WebCite Cache ID 6ZaFhaPIS]
11. Estellés-Arolas E, González-Ladrón-de-Guevara F. Towards an integrated crowdsourcing definition. J Infor Sci 2012 Mar 09;38(2):189-200. [doi: 10.1177/0165551512437638]
12. Schenk E, Guittard C. Towards a characterization of crowdsourcing practices. J Innov Econ Manage 2011;7(1):93-107 [FREE Full text] [doi: 10.3917/jie.007.0093]
13. Ranard BL, Ha YP, Meisel ZF, Asch DA, Hill SS, Becker LB, et al. Crowdsourcing—Harnessing the masses to advance health and medicine, a systematic review. J Gen Intern Med 2014 Jan;29(1):187-203 [FREE Full text] [doi: 10.1007/s11606-013-2536-8] [Medline: 23843021]
14. McCoy AB, Wright A, Krousel-Wood M, Thomas EJ, McCoy JA, Sittig DF. Validation of a crowdsourcing methodology for developing a knowledge base of related problem-medication pairs. Appl Clin Inform 2015;6(2):334-344. [doi: 10.4338/ACI-2015-01-RA-0010] [Medline: 26171079]
15. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: An online game for analyzing images of infected thick blood smears. J Med Internet Res 2012;14(6):e167 [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]
16. Mavandadi S, Dimitrov S, Feng S, Yu F, Sikora U, Yaglidere O, et al. Distributed medical image analysis and diagnosis through crowd-sourced games: A malaria case study. PLoS One 2012;7(5):e37245 [FREE Full text] [doi: 10.1371/journal.pone.0037245] [Medline: 22606353]
17. Mavandadi S, Dimitrov S, Feng S, Yu F, Yu R, Sikora U, et al. Crowd-sourced BioGames: Managing the big data problem for next-generation lab-on-a-chip platforms. Lab Chip 2012 Oct 21;12(20):4102-4106 [FREE Full text] [doi: 10.1039/c2lc40614d] [Medline: 22918378]
18. McKenna MT, Wang S, Nguyen TB, Burns JE, Petrick N, Summers RM. Strategies for improved interpretation of computer-aided detections for CT colonography utilizing distributed human intelligence. Med Image Anal 2012 Aug;16(6):1280-1292 [FREE Full text] [doi: 10.1016/j.media.2012.04.007] [Medline: 22705287]
19. Nguyen TB, Wang S, Anugu V, Rose N, McKenna M, Petrick N, et al. Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography. Radiology 2012 Mar;262(3):824-833 [FREE Full text] [doi: 10.1148/radiol.11110938] [Medline: 22274839]

20. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res 2014;16(10):e233 [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]

21. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: Audit study. BMJ 2015;351:h3480 [FREE Full text] [Medline: 26157077]

22. Tang H, Ng JH. Googling for a diagnosis—Use of Google as a diagnostic aid: Internet based study. BMJ 2006 Dec 2;333(7579):1143-1145 [FREE Full text] [doi: 10.1136/bmj.39003.640567.AE] [Medline: 17098763]

23. Parvanta C, Roth Y, Keller H. Crowdsourcing 101: A few basics to make you the leader of the pack. Health Promot Pract 2013 Mar;14(2):163-167. [doi: 10.1177/1524839912470654] [Medline: 23299912]

24. Zettler P. The Health Care Blog. 2015 Jun 19. Do you need a medical degree to crowdsource medicine? URL: http://thehealthcareblog.com/blog/2015/06/19/do-you-need-a-medical-degree-to-crowdsource-medicine/ [accessed 2015-06-30] [WebCite Cache ID 6Zg1GBjXD]

XSL•FO

RenderX

Original Paper

Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England

Matthew Sperrin[1], PhD (Statistics); Helen Rushton[1], MSc (Statistics); William G Dixon[1,2,3], MRCP, PhD; Alexis Normand[4], MSc; Joffrey Villard[4], PhD; Angela Chieh[4], MSc; Iain Buchan[1], MD

[1]Health e-Research Centre, Farr Institute, Institute of Population Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

[2]Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom

[3]Rheumatology Department, Salford Royal NHS Foundation Trust, Salford, United Kingdom

[4]Withings, Issy-les-Moulineaux, France

Corresponding Author:
Matthew Sperrin, PhD (Statistics)
Health e-Research Centre, Farr Institute
Institute of Population Health, Manchester Academic Health Science Centre
University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 0161 3067629
Fax: 44 0161 3067629
Email: matthew.sperrin@manchester.ac.uk

Abstract

Background: Digital self-monitoring, particularly of weight, is increasingly prevalent. The associated data could be reused for clinical and research purposes.

Objective: The aim was to compare participants who use connected smart scale technologies with the general population and explore how use of smart scale technology affects, or is affected by, weight change.

Methods: This was a retrospective study comparing 2 databases: (1) the longitudinal height and weight measurement database of smart scale users and (2) the Health Survey for England, a cross-sectional survey of the general population in England. Baseline comparison was of body mass index (BMI) in the 2 databases via a regression model. For exploring engagement with the technology, two analyses were performed: (1) a regression model of BMI change predicted by measures of engagement and (2) a recurrent event survival analysis with instantaneous probability of a subsequent self-weighing predicted by previous BMI change.

Results: Among women, users of self-weighing technology had a mean BMI of 1.62 kg/m^2 (95% CI 1.03-2.22) lower than the general population (of the same age and height) (P<.001). Among men, users had a mean BMI of 1.26 kg/m^2 (95% CI 0.84-1.69) greater than the general population (of the same age and height) (P<.001). Reduction in BMI was independently associated with greater engagement with self-weighing. Self-weighing events were more likely when users had recently reduced their BMI.

Conclusions: Users of self-weighing technology are a selected sample of the general population and this must be accounted for in studies that employ these data. Engagement with self-weighing is associated with recent weight change; more research is needed to understand the extent to which weight change encourages closer monitoring versus closer monitoring driving the weight change. The concept of isolated measures needs to give way to one of connected health metrics.

(*J Med Internet Res 2016;18(1):e17*)   doi:10.2196/jmir.4767

KEYWORDS

XSL•FO
RenderX

Introduction

Self-monitoring of weight has a long history rooted in consumer demand for weight control, reinforced in recent decades by public concern over rising obesity levels [1,2]. Frequent self-weighing is associated with weight loss [3,4] and there is evidence of a dose-response relationship with more frequent weighing associated with higher weight loss [5-7]. Technologies that enable weight to be captured digitally and fed automatically into consumer health records can enhance both the utilization and effects of self-weighing [8,9]. This is an example of *connected health technology*: the application of technology to help individuals and their health care providers monitor and maintain health [10].

Data from connected health technologies have potential for adoption in clinical practice and research; however, there are at least 2 concerns with their use. The first concern is that, on an individual level, the accuracy of the data may be considered inferior to that recorded by a health professional. Generally, self-measured height is overestimated and weight is underestimated [11-13]; however, the extent of this is minor and use of self-reported height and weight is considered valid [14-16]. Recall bias may also apply to historical weight measures. A connected health approach may overcome this concern as automatic transfer of data from the weighing device to the consumer health record bypasses reporting and recall bias. The second concern applies on an epidemiological level: there is inherent selection bias in the individuals who choose to self-monitor, so it is difficult to draw population-wide conclusions. Existing literature focuses on participants who volunteered for self-weighing; therefore, the organic uptake of self-weighing remains relatively unexplored.

The aim of this study was to explore the possibility of using data collected from contemporary self-weighing smart scales for epidemiological research. Our first objective was to compare the population of people using smart scales in England with the wider population to get an idea of the selection bias. Our second objective was to understand how engagement with the smart scales varies between participants and how this engagement affects (or is affected by) weight change.
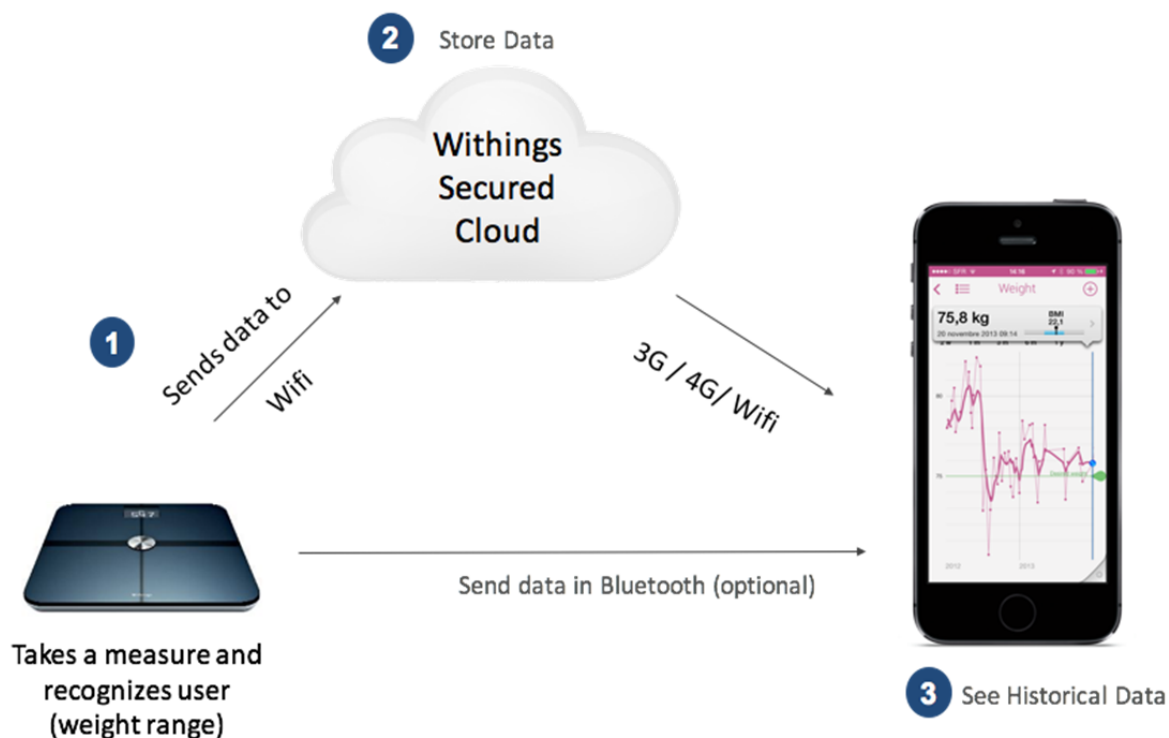
Methods

Data

There were 2 sources of data used in this study. The first dataset was the 2011 wave of the Health Survey for England (HSE), used to obtain a representation of the distribution of height, weight, and body mass index (BMI) in England. The HSE is a series of annual cross-sectional surveys carried out in England. First piloted in 1991, it has been fully running since 1992. Weight is measured by a nurse to the nearest 100 g using an electronic scale after removal of shoes or bulky clothing (participants were not weighed if they were pregnant, unsteady on their feet, or chair-bound). Height, to the nearest millimeter, is measured by a nurse using a portable stadiometer. Previous surveys reported, on average, 70% of households agreed to an interview and BMI was available from approximately 90% of those interviewed (with some variation by year and region) [17].

The second data source was a random sample of Withings Smart Scale users based in England, representing the population engaged with self-weighing. A *user* is defined as someone who obtained a Withings Smart Scale and created an account under which their measurements are stored. Scales were self-purchased by potential users from retail stores or from the Withings website. The process of a self-weighing and the data being stored is described in Figure 1. A random sample was generated from all users with at least one self-weighing; the full dataset of English Withings Smart Scale users could not be used due to commercial sensitivity; however, the random sample was large enough to afford reasonable contrasts in demographic characteristics and BMI. The follow-up time for a Withings Smart Scale user was defined as the time interval between the first and last available measurement.

The anonymized HSE is publicly available for research purposes. The Withings Smart Scale users consented to their data being used for research purposes as part of the Terms and Conditions when setting up a user account (see [18]).

We restricted analysis a priori to persons aged 16 or older. BMI measurements below 15 and above 70 were assumed to be erroneous and were removed. BMI was used as a continuous variable as well as a categorical variable using the World Health Organization cut-offs [19].

Figure 1. Description of the self-weighing process and data storage for Withings Smart Scale.

Statistical Methods

Descriptive statistics were produced using standard methods. We compared these between the 2 datasets (HSE and Withings Smart Scale) and additionally stratified this comparison by gender. Continuous variables that were not expected a priori to have substantial skew (age, height, weight, and BMI) were summarized using means and standard deviations, and compared using *t* tests. Continuous variables that were expected to be skewed (measurements per person, follow-up days, and measurements per person per month) were summarized using medians and the interquartile range. Categorical variables were summarized by the counts and proportions of participants in each group and compared using Fisher exact tests.

We compared the BMI of the smart scale users with the HSE participants using linear regression, with BMI as the response and an indicator of smart scale user as the predictor of primary interest.

Withings Smart Scale data were investigated in more detail to explore the association between engagement with self-weighing and BMI. First, determinants of BMI change over the follow-up period were examined using linear regression. BMI change (the response) was calculated as a single measurement for each individual as the difference between the first and last BMI measures reported divided by the time (in months) between them, with negative change representing overall BMI loss. Individuals required at least 2 measurements to be included in this model. Primary predictors of interest were number of measurements per month, total follow-up time, and initial weight. Second, a multilevel Cox proportional hazard model was used to assess determinants of a weighing event occurring. This was treated as a recurrent event with frailty terms used to account for within-person correlation. The primary covariates

of interest for this model were BMI at the previous reading and a measure of the recent change in BMI. Recent change in BMI was considered in 2 ways in 2 separate analyses. The "current" incremental change was defined as the difference in BMI between the previous weighing and the current weighing. This may represent an individual's perception of recent weight change when making the current weighing. The "previous" incremental change was defined as the difference in BMI between the 2 previous weighings. Therefore, this represents a BMI change that has already been observed before the current weighing. For both measures of change, we recorded whether this was a gain or loss and this was represented in 2 separate variables. For example, if BMI at weighing *t* minus BMI at time *t–1* equalled –0.3, this was recorded as a BMI loss of 0.3 (and the variable for BMI gain was set to zero). We also included an indicator variable of whether BMI was lost or gained. This allowed for some flexibility in modeling the BMI change: a discontinuity at a BMI change of zero represented by the indicator variable and different slopes depending on whether BMI was gained or lost. For these models, the time interval between the first 2 BMI measurements was excluded because the previous change variable was not available; all other time intervals were included. Consequently, individuals required at least 3 measurements to contribute to this model.

For all the preceding models, height, age, and age squared (age^2) were included as confounders because they are all known to be associated with BMI [1,20]. Separate models were fitted for men and women because it was known a priori that BMI should be interpreted differently for each gender [20]. All analyses were carried out using Stata version 13 software.

Results

For the Withings Smart Scale data, there were 975 users in the sample; for the HSE data there were 7035 individuals. A data exclusion flowchart is given in Figure 2.

The baseline characteristics of the 2 populations are given in Table 1; this used the first recorded height and weight measurement for each individual in the Withings Smart Scale data. The Withings Smart Scale data contained more men and a younger population than the HSE. For the Withings Smart Scale data, the median follow-up was 377 (IQR 187-700) days for men and 351 (IQR 143-655) days for women, with a median of 87 (IQR 30-188) weighings per man over the entire follow-up period (median 7.6, IQR 3.7-16.1 per month) and median 50 (IQR 15-123) per woman over the entire follow-up period (median 5.5, IQR 2.2-14.1 per month). Example trajectories from the Withings Smart Scale data are also visualized in Figure 3.

Figure 2. Data exclusion flowchart for Health Survey for England data (left) and Withings Smart Scale data (right).
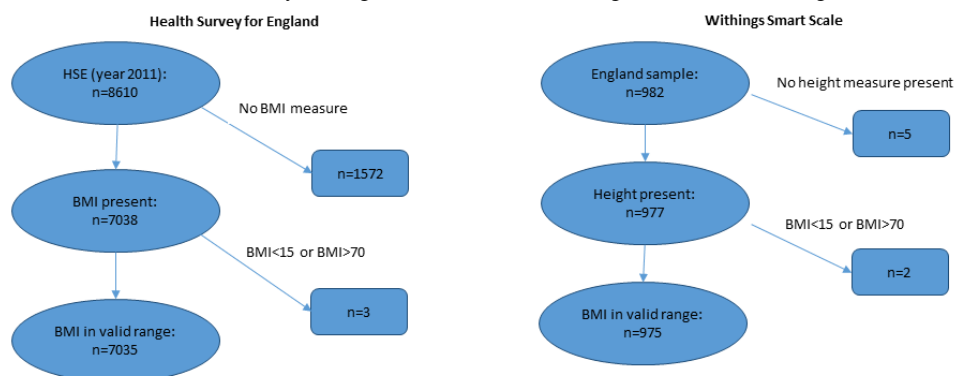
Table 1. Comparison of baseline characteristics between Withings Smart Scale and Health Survey for England (HSE) participants (N=8010).

Variable	Smart scale n=975			HSE n=7035			P [a]		
	Men	Women	Overall	Men	Women	Overall	Men	Women	Overall
Participants, n (%)	591 (60.6)	384 (39.4)		3164 (44.98)	3871 (55.02)				<.001
Age (years), mean (SD)	39.00 (10.52)	39.34 (12.55)	39.13 (11.36)	49.30 (18.11)	48.86 (18.36)	49.05 (18.25)	<.001	<.001	<.001
Measurements per person, median (IQR)	87 (30-188)	50 (15-123)							
Follow-up days, median (IQR)	377 (187-700)	351 (143-655)							
Measurements per person per month, median (IQR)	7.6 (3.7-16.1)	5.5 (2.2-14.1)							
BMI at first measurement (kg/m^2), mean (SD)	28.32 (5.42)	25.17 (5.34)	27.08 (5.60)	27.51 (4.79)	27.30 (5.77)	27.39 (5.35)	<.001	<.001	.09
Height (cm), mean (SD)	178.91 (7.77)	165.19 (6.47)	173.51 (9.90)	175.05 (7.42)	161.62 (6.81)	167.66 (9.74)	<.001	<.001	<.001
Weight (kg), mean (SD)	90.65 (18.27)	68.77 (15.62)	82.03 (20.31)	84.37 (15.80)	71.29 (15.58)	77.17 (16.98)	<.001	.003	<.001
BMI (kg/m^2), n (%)									
Underweight (<18.5)	4 (0.7)	14 (3.6)	18 (1.8)	31 (0.98)	77 (1.99)	108 (1.54)	.01	<.001	.12
Normal (18.5-24.9)	160 (27.1)	213 (55.5)	373 (38.3)	966 (30.53)	1474 (38.08)	2440 (34.68)			
Overweight (25.0-29.9)	241 (40.8)	99 (25.8)	340 (34.9)	1373 (43.39)	1286 (33.22)	2659 (37.80)			
Obese (≥30)	186 (31.5)	58 (15.1)	244 (25.0)	794 (25.09)	1034 (26.71)	1828 (25.98)			

[a] Based on Fisher exact test or *t* test.

The regression model for Withings Smart Scale user status on BMI is given in Table 2. Among women, after correction for potential confounders, Withings Smart Scale users had a mean BMI of 1.62 (95% CI 1.03-2.22) lower than the general population (of the same age and height) (*P*<.001). The opposite pattern was seen among male Withings Smart Scale users, who had a mean BMI of 1.26 (95% CI 0.84-1.69) greater than the general population (of the same age and height) (*P*<.001). The results from both samples also corroborated that shorter men and women tend to have higher BMI (reflected in the negative coefficient for height in Table 2). There is a quadratic relationship between BMI and age with BMI generally increasing up to age 60 years then declining (see Figure 4).

Figure 3. Example BMI trajectories of the first 100 men and 100 women in the Withings Smart Scale data over time (January 1, 2010 to January 1, 2014).
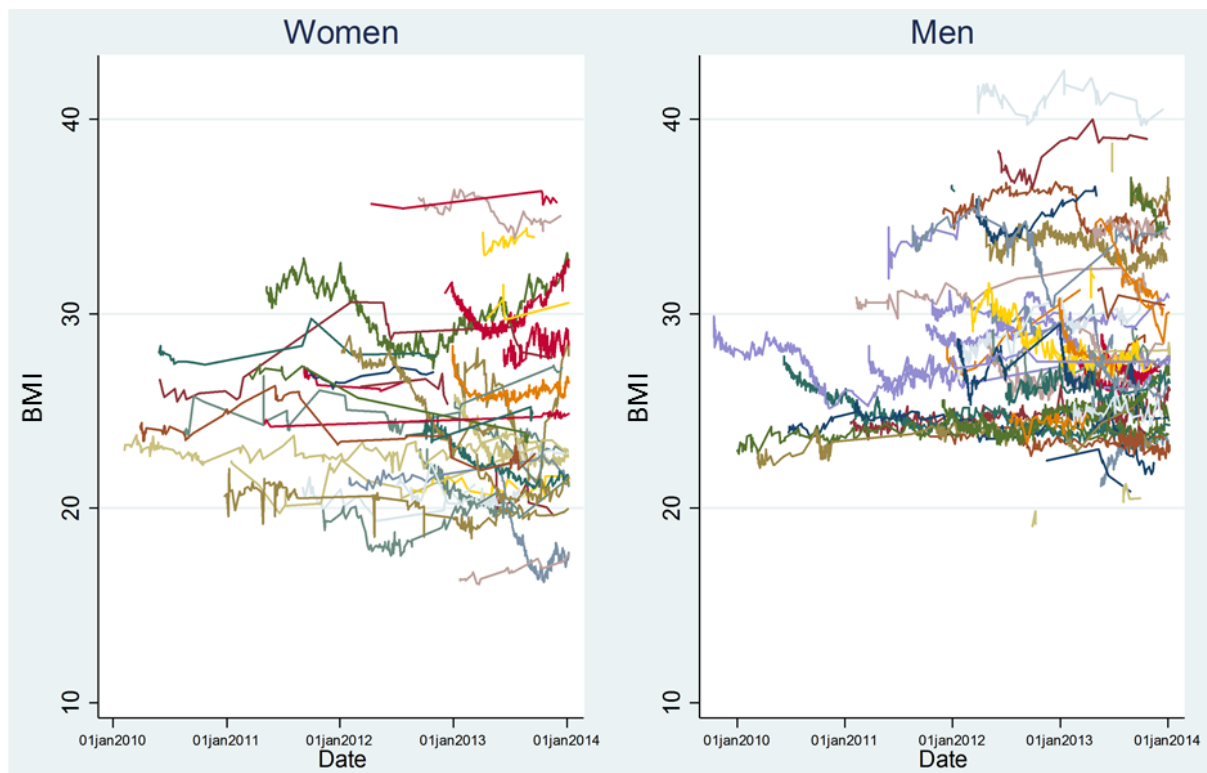
Table 2. Results of regression model comparing BMI between Withings Smart Scale and Health Survey for England (HSE) data.

Variable	Men, n=3755		Women, n=4255	
	Coef (95% CI)	*P*	Coef (95% CI)	*P*
Smart scale cohort indicator	1.26 (0.84, 1.69)	<.001	−1.62 (−2.22, 1.03)	<.001
Age	0.34 (0.29, 0.39)	<.001	0.27 (0.22, 0.32)	<.001
Age^2	−0.0028 (−0.0033, −0.0023)	<.001	−0.0022 (−0.0027, −0.0017)	<.001
Height	−0.03 (−0.05, −0.01)	<.001	−0.07 (−0.094, −0.04)	<.001
Intercept	23.42 (19.60, 27.65)	<.001	30.94 (26.55, 35.34)	<.001

Figure 4. Model-estimated BMI for Health Survey for England data for men of average height (175 cm; blue line) and women of average height (162 cm; pink line).
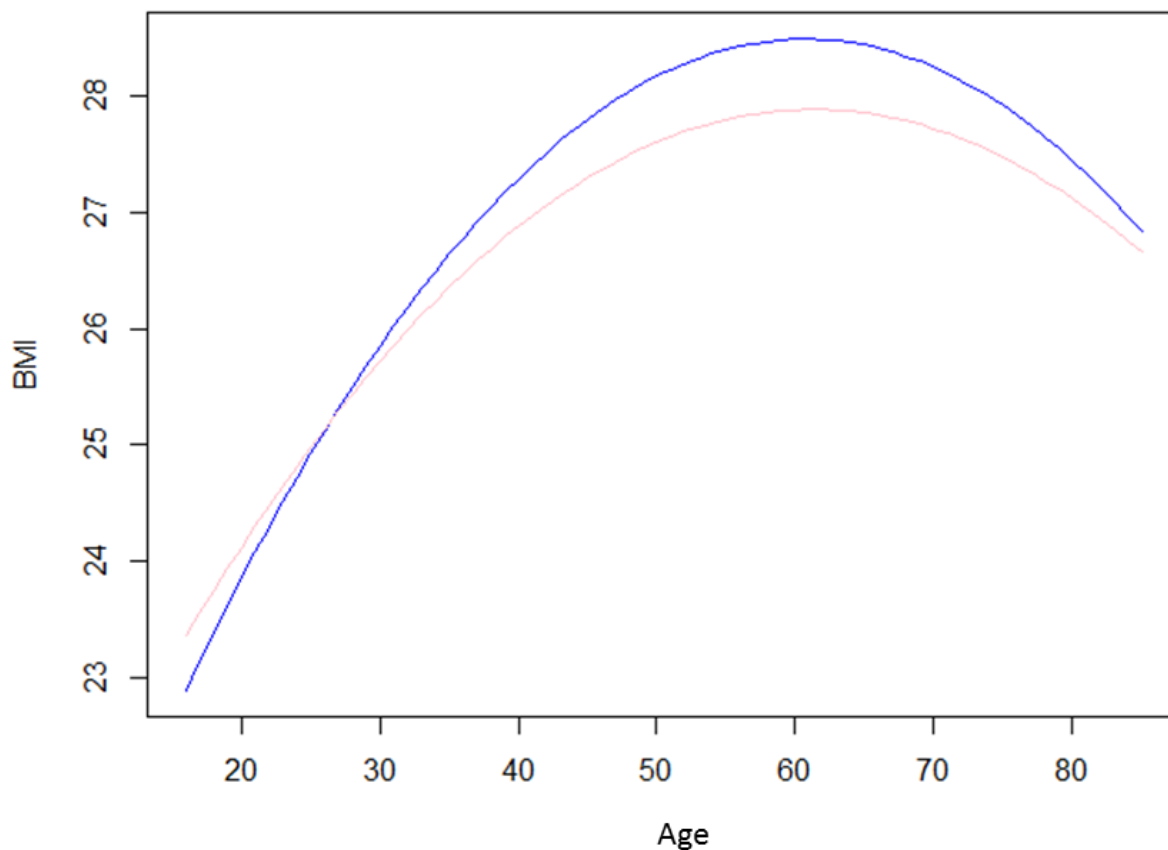
We then looked in more detail at the Withings Smart Scale data to understand how engagement with the smart scale technology related to BMI change over time. First, in the regression of BMI change against measurement intensity, we found that more frequent measurement over the entire period was associated with greater weight loss per month in both women (regression coefficient 0.03, 95% CI 0.02-0.05 kg/m^2 per measurement per month, P=.01) and men (regression coefficient 0.03, 95% CI 0.01-0.05 kg/m^2 per measurement per month, P<.001). To put this in context, a man with the median follow-up of 377 days (12.4 months, from Table 1) would be expected to lose 0.37 kg/m^2 more over the follow-up period than a man with one fewer measurement per month; this is equivalent to 1.13 kg for a man of average height (175 cm). Similarly, a woman with median follow-up of 355 days (11.7 months, from Table 1) would be expected to lose 0.35 kg/m^2 more over the follow-up than a woman with one fewer measurement per month; this is equivalent to 0.92 kg for a woman of average height (162cm). Higher initial BMI led to a greater reduction per month. See Table 3 for the full regression results.

Table 3. Results of regression model for weight loss versus measurement intensity.

Variable	Men, n=586		Women, n=376	
	Coef (95% CI)	P	Coef (95% CI)	P
Measurements per month	−0.03 (−0.05, −0.02)	<.001	−0.03 (−0.05, −0.01)	.01
Time observed (months)	0.006 (−0.006, 0.018)	.30	0.01 (−0.004, 0.031)	.12
BMI at start	−0.12 (−0.15, −0.09)	<.001	−0.05 (−0.09, −0.02)	.005
Intercept	−5.71 (−9.89, −1.54)	.007	−3.07 (−8.65, 2.50)	.28
Age	0.02 (−0.07, 0.11)	.65	0.08 (−0.004, 0.17)	.06
Age^2	−0.0002 (−0.0012, 0.0009)	.76	−0.0010 (−0.0020, 0.0000)	.049
Height (m)	2.75 (0.66, 4.85)	.01	0.77 (−2.43, 3.96)	.64

We then considered longitudinal patterns of subsequent weighings based on recent weight change. The results of these analyses are summarized in Table 4. Because at least 3 measurements were required for these models, 9 men and 13 women who had only 1 or 2 measurements each were removed. For the current measure of weight change (see Methods), we

XSL•FO
RenderX

found recent weight loss encouraged a subsequent measurement, with hazard ratio (HR) 7.38 (95% CI 7.03-7.75) per unit BMI in men (ie, propensity for weighing is 7.38 times higher for a man whose BMI has dropped 1 kg/m^2 compared with a man who has remained the same weight) and HR 5.86 (95% CI 5.50-6.25) per unit BMI in women. For the previous measure of weight change (see Methods), weight loss encouraged subsequent measurements but to a lesser extent, with HR 2.88 (95% CI 2.74-3.02) in men and HR 2.44 (95% CI 2.28-2.60) in women. On the other hand, recent weight gain discouraged

subsequent measurements in both men and women. Under the current measure of weight gain, HR 0.09 (95% CI 0.09-0.10) was observed for men and HR 0.10 (95% CI 0.09-0.10) was observed for women. For the recent measure of weight gain, smaller effects were observed, but in the same direction with HR 0.41 (95% CI 0.40-0.43) in men and HR 0.40 (95% CI 0.38-0.42) in women. The fact that in all cases the effect was more pronounced for the current incremental change suggests that perceived recent weight change is more important than measured historical weight change as a predictor of further weighing.

Table 4. Hazard ratios (HR) calculated from the Cox proportional hazards model.

Variable	Men (88,769 observations on 575 participants)		Women (41,894 observations on 363 participants)	
	HR (95% CI)	P	HR (95% CI)	P
Current change				
BMI	0.99 (0.98-0.99)	<.001	1.02 (1.01-1.02)	<.001
Time since first weighing (months)	0.98 (0.97-0.98)	<.001	0.98 (0.98-0.98)	<.001
Indicates BMI lost	1.20 (1.18-1.22)	<.001	1.06 (1.03-1.09)	<.001
BMI change (gain)	0.09 (0.09-0.10)	<.001	0.10 (0.09-0.10)	<.001
BMI change (loss)	7.38 (7.03-7.75)	<.001	5.86 (5.50-6.25)	<.001
Age	1.03 (1.03-1.04)	<.001	1.01 (1.01-1.02)	<.001
Age^2	0.9997 (0.9996-0.9997)	<.001	0.9999 (0.9999-1.0000)	.28
Height (m)	1.79 (1.63-1.96)	<.001	1.08 (0.93-1.26)	.30
Previous change				
BMI	0.97 (0.97-0.97)	<.001	1.00 (1.00-1.00)	.30
Time since first weighing (months)	0.97 (0.97-0.97)	<.001	0.97 (0.97-0.97)	.002
Indicates BMI lost	1.15 (1.12-1.17)	<.001	0.98 (0.95-1.001)	.12
BMI change where BMI gained	0.41 (0.40-0.43)	<.001	0.40 (0.38-0.42)	<.001
BMI change where BMI lost	2.88 (2.74-3.02)	<.001	2.44 (2.28-2.60)	<.001
Age	1.05 (1.04-1.05)	<.001	1.02 (1.02-1.03)	<.001
Age^2	0.9996 (0.9995-0.9996)	<.001	0.9999 (0.9998-0.9999)	<.001
Height (m)	1.47 (1.34-1.61)	<.001	1.18 (1.01-1.37)	.04

Discussion

Summary

This study compared English users of Withings smart scales connected to consumer health records to the general population in England. We found that Withings Smart Scale users are younger and more likely to be male than the general population in England. Among women, we found Withings Smart Scale users had, after correction for confounding, a BMI 1.62 kg/m^2 lower than the general population; for a woman of average height (162 cm), this is a weight difference of 4.25 kg. Among men, we found Withings Smart Scale users had, after correction, a BMI 1.26 kg/m^2 higher than the general population; for a man of average height (175 cm), this is a weight difference of 3.86 kg. Looking in more detail at Withings Smart Scale users, we found that more frequent measurement was associated with

greater weight loss; again considering average height, each additional weighing per month was associated with further weight loss over the entire follow-up period of 1.13 kg for men and 0.92 kg for women. A positive feedback loop was identified in which a recent observed decrease in weight encourages further weighing.

Strengths and Limitations

A strength of the study is that we used data from large, robust sources for both the general population and the randomly selected population of individuals who use a popular brand of smart scales to monitor their weight. We employed advanced modeling techniques, including multilevel Cox regression, to exploit the longitudinal richness of the data.

A limitation is that the BMI comparison is based on standardized measurement in HSE, whereas readings in the Withings Smart

Scale data were not standardized to such things as the amount of clothing worn. However, even self-reported height and weight without automated data capture from one type of instrument are generally accepted to be sufficiently accurate for such comparisons to be made [14]. However, the height data are nonstandardized self-reports recorded into the consumer's online health record.

The HSE is a cross-sectional study and the Withings Smart Scale data are longitudinal. Therefore, there is a difference in timeframe, although this was minimized by using the 2011 wave of HSE, which is within the Withings Smart Scale data timeframe. Although changes in BMI in the English population are likely to be small over the Withings Smart Scale data timeframe (2010-2013) [1], changes over time in the Withings Smart Scale data could be larger, especially because the use of smart scales has become more widespread over the period. A future study will consider the emergence of use of smart scales over time and test the hypothesis that the smart scale user population converges to the general population over time.

A further limitation is that this is an observational study, so propensity to use self-weighing technology is subject to confounding. We have mitigated this by correcting our comparative models for age, gender, and height. However, we could not consider unmeasured potential confounding factors. An important unmeasured confounder is baseline engagement with weight or BMI; it is likely that individuals with more interest in BMI monitoring are more likely to purchase self-weighing technology, which would amplify the association of smart scale use with BMI control. Therefore, the results of our study should not be interpreted causally and further studies are needed to isolate the causal effect of self-weighing.

Comparison With Existing Literature

Our findings reinforce those of others that found increased engagement with self-weighing is associated with greater weight loss or reduced weight gain [3-7,21-23]. However, to the best of our knowledge, all existing studies concern participants in weight control programs. Therefore, our study adds to the literature because it demonstrates this effect in a population of smart scale users who may or may not be engaging in weight control programs. In addition, we have uncovered a positive feedback loop in which a weighing showing a decrease in weight encourages a further weighing in the near future.

Unlike other studies, our observations suggest that women engaging with self-weighing technology tend to be lighter than average, whereas men tend to be heavier. A possible hypothesis for this finding could be that men who engage may be fit with high muscle mass.

Implications for Research/Practice

Users of Withings Smart Scale devices are not representative of the general population. Any inferences about the general population should be corrected for at least age and gender by regression analysis or reweighting. In addition, even after correction for age and gender, BMI measures differ between the smart scales and the general population. Because this difference is in the opposite direction for men and women, there may be complementary reasons for engagement with smart scales between the genders. Further qualitative research into these drivers may allow for transfer across the genders and improve uptake of such devices.

Connected health technologies incorporating self-weighing can provide richer data than those from infrequent contact with health professionals. In particular, much higher longitudinal resolution of BMI can be captured for individuals and populations. However, these data are complex: the relation between the frequency of self-weighing and the underlying level and change in the weight itself needs careful consideration. Usefully, self-weighing is associated with better weight control; however, more research is needed to examine potential mediators and confounders of this relationship.

As personal health records start to gather data from a wider ecosystem of frequent measurement, the links between health observations and behaviors will become more tightly coupled. For example, physical activity monitoring from smart watches linked to weight measures from smart scales brings together information on weight control interventions and outcomes in a potentially persuasive ensemble. The statistical challenges of harnessing linked observation, intervention, and outcome processes should not be underestimated.

Connected health ecosystems are being driven by the consumer health/wellness market, but they also have the potential to support clinical interventions and research [24,25]. At present, such technologies are not ubiquitous; therefore, the selection biases due to the characteristics of those who opt to buy and use them must be considered.

The use of connected health technologies is a promising area for clinical research and practice as well as consumer health markets. Their real potential may be realized through their linkage with each other and with more conventional sources such as electronic health records.

Conclusion

In this paper, we have demonstrated that current engagement with smart scale technology involves a selected population. Therefore, use of the associated data needs to correct for this selection. We have also demonstrated an opposing selection effect between men and women, with male users being heavier than average and female users being lighter, as well as a positive feedback loop with more frequent weighings following greater weight loss. The drivers behind these findings need to be explored in more detail to understand how engagement with smart scale technology drives, and is driven by, healthy behavior.

Acknowledgments

Authors' Contributions

MS, WGD, AN, and IB conceived and designed the study. MS and HR carried out the statistical analysis. All authors contributed to the drafting and revision of the manuscript, and approved the final version for submission.

Conflicts of Interest

AN, JV, and AC are employees of Withings, who develop self-weighing and other self-monitoring equipment. MS, HR, WGD, and IB have no conflicts to declare.

References

1. Sperrin M, Marshall AD, Higgins V, Buchan IE, Renehan AG. Slowing down of adult body mass index trend increases in England: a latent class analysis of cross-sectional surveys (1992-2010). Int J Obes (Lond) 2014 Jun;38(6):818-824. [doi: 10.1038/ijo.2013.161] [Medline: 23995474]

2. Tufano J, Karras B. Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends. J Med Internet Res 2005;7(5):e58 [FREE Full text] [doi: 10.2196/jmir.7.5.e58] [Medline: 16403722]

3. Burke L, Wang J, Sevick M. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]

4. Butryn ML, Phelan S, Hill JO, Wing RR. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity (Silver Spring) 2007 Dec;15(12):3091-3096. [doi: 10.1038/oby.2007.368] [Medline: 18198319]

5. VanWormer J, Martinez A, Martinson B, Crain A, Benson G, Cosentino D, et al. Self-weighing promotes weight loss for obese adults. Am J Prev Med 2009 Jan;36(1):70-73. [doi: 10.1016/j.amepre.2008.09.022] [Medline: 18976879]

6. VanWormer J, Linde J, Harnack L, Stovitz S, Jeffery R. Self-weighing frequency is associated with weight gain prevention over 2 years among working adults. Int J Behav Med 2012 Sep;19(3):351-358 [FREE Full text] [doi: 10.1007/s12529-011-9178-1] [Medline: 21732212]

7. Vanwormer J, French S, Pereira M, Welsh E. The impact of regular self-weighing on weight management: a systematic literature review. Int J Behav Nutr Phys Act 2008;5:54 [FREE Full text] [doi: 10.1186/1479-5868-5-54] [Medline: 18983667]

8. Burke L, Styn M, Sereika S, Conroy M, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]

9. Spring B, Duncan J, Janke E, Kozak A, McFadden H, DeMott A, et al. Integrating technology into standard weight loss treatment: a randomized controlled trial. JAMA Intern Med 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1221] [Medline: 23229890]

10. Kvedar J, Coye M, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]

11. Connor GS, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obes Rev 2007 Jul;8(4):307-326. [doi: 10.1111/j.1467-789X.2007.00347.x] [Medline: 17578381]

12. Jerome G, Dalcin A, Coughlin J, Fitzpatrick S, Wang N, Durkin N, et al. Longitudinal accuracy of web-based self-reported weights: results from the Hopkins POWER Trial. J Med Internet Res 2014;16(7):e173 [FREE Full text] [doi: 10.2196/jmir.3332] [Medline: 25042773]

13. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? J Med Internet Res 2014;16(1):e4 [FREE Full text] [doi: 10.2196/jmir.2909] [Medline: 24398335]

14. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. Public Health Nutr 2002 Aug;5(4):561-565 [FREE Full text] [doi: 10.1079/PHN2001322] [Medline: 12186665]

15. Bonn SE, Trolle LY, Bälter K. How valid are Web-based self-reports of weight? J Med Internet Res 2013;15(4):e52 [FREE Full text] [doi: 10.2196/jmir.2393] [Medline: 23570956]

16. Lassale C, Péneau S, Touvier M, Julia C, Galan P, Hercberg S, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Santé study. J Med Internet Res 2013;15(8):e152 [FREE Full text] [doi: 10.2196/jmir.2575] [Medline: 23928492]

17. Mindell J, Biddulph JP, Hirani V, Stamatakis E, Craig R, Nunn S, et al. Cohort profile: the health survey for England. Int J Epidemiol 2012 Dec;41(6):1585-1593 [FREE Full text] [doi: 10.1093/ije/dyr199] [Medline: 22253315]

18. Withings. 2015 Apr 26. Withings services terms and conditions URL: http://www.withings.com/jp/en/legal/withings-services-terms-and-conditions [accessed 2016-01-09] [WebCite Cache ID 6ePk32abR]

19.   World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation (WHO Technical Report Series 894). Geneva: WHO; 2000.

20.   Sperrin M, Marshall A, Higgins V, Renehan A, Buchan I. Body mass index relates weight to height differently in women and older adults: serial cross-sectional surveys in England (1992-2011). J Public Health (Oxf) 2015 Jun 1:1-7 [FREE Full text] [doi: 10.1093/pubmed/fdv067] [Medline: 26036702]

21.   Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265. [doi: 10.1002/oby.20946] [Medline: 25521523]

22.   Madigan C, Aveyard P, Jolly K, Denley J, Lewis A, Daley A. Regular self-weighing to promote weight maintenance after intentional weight loss: a quasi-randomized controlled trial. J Public Health (Oxf) 2014 Jun;36(2):259-267 [FREE Full text] [doi: 10.1093/pubmed/fdt061] [Medline: 23753256]

23.   Madigan C, Daley A, Lewis A, Aveyard P, Jolly K. Is self-weighing an effective tool for weight loss: a systematic literature review and meta-analysis. Int J Behav Nutr Phys Act 2015;12(1):104 [FREE Full text] [doi: 10.1186/s12966-015-0267-4] [Medline: 26293454]

24.   Oldenburg B, Taylor C, O'Neil A, Cocker F, Cameron L. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]

25.   Taylor J, Oguntuase O, Gorst S. Does telehealth promote self-care? A qualitative study examining the experience of older people using telehealth for long-term condition management. Int J Integ Care 2014:e1 [FREE Full text]

Abbreviations

BMI: body mass index
HR: hazard ratio
HSE: Health Survey for England

Original Paper

Establishing and Governing e-Mental Health Care in Australia: A Systematic Review of Challenges and A Call For Policy-Focussed Research

Carla Meurk[1,2], PhD; Janni Leung[1], BHealth Sc; Wayne Hall[2], PhD; Brian W Head[3], PhD; Harvey Whiteford[1,4], MBBS, MPH, PhD

[1]School of Public Health, Faculty of Medicine and Biomedical Sciences, The University of Queensland, Herston, Australia

[2]Centre for Youth Substance Abuse Research, The University of Queensland, Herston, Australia

[3]Institute of Social Science Research, The University of Queensland, St Lucia, Australia

[4]Policy and Epidemiology Group, Queensland Centre for Mental Health Research, Archerfield, Australia

Corresponding Author:
Carla Meurk, PhD
School of Public Health
Faculty of Medicine and Biomedical Sciences
The University of Queensland
Level 2, Public Health Building
Herston, 4006
Australia
Phone: 61 733655345
Fax: 61 733655442
Email: c.meurk@uq.edu.au

Abstract

Background: Growing evidence attests to the efficacy of e-mental health services. There is less evidence on how to facilitate the safe, effective, and sustainable implementation of these services.

Objective: We conducted a systematic review on e-mental health service use for depressive and anxiety disorders to inform policy development and identify policy-relevant gaps in the evidence base.

Methods: Following the PRISMA protocol, we identified research (1) conducted in Australia, (2) on e-mental health services, (3) for depressive or anxiety disorders, and (4) on e-mental health usage, such as barriers and facilitators to use. Databases searched included Cochrane, PubMed, PsycINFO, CINAHL, Embase, ProQuest Social Science, and Google Scholar. Sources were assessed according to area and level of policy relevance.

Results: The search yielded 1081 studies; 30 studies were included for analysis. Most reported on self-selected samples and samples of online help-seekers. Studies indicate that e-mental health services are predominantly used by females, and those who are more educated and socioeconomically advantaged. Ethnicity was infrequently reported on. Studies examining consumer preferences found a preference for face-to-face therapy over e-therapies, but not an aversion to e-therapy. Content relevant to governance was predominantly related to the organizational dimensions of e-mental health services, followed by implications for community education. Financing and payment for e-services and governance of the information communication technology were least commonly discussed.

Conclusions: Little research focuses explicitly on policy development and implementation planning; most research provides an e-services perspective. Research is needed to provide community and policy-maker perspectives. General population studies of prospective treatment seekers that include ethnicity and socioeconomic status and quantify relative preferences for all treatment modalities are necessary.

XSL·FO
RenderX

Introduction

Growing evidence attests to the efficacy of Internet-assisted therapeutic tools, particularly in the treatment of common mental disorders such as mild to moderate depressive and anxiety disorders [1]. Prominent mental health researchers and advocates are optimistic about the potential for e-mental health care to enhance accessibility and increase cost efficiency of services, promote consumer empowerment, and overcome geographic obstacles to service utilization [2-6]. They have called on government to support and promote the development and implementation of these tools [7,8]. Recent translational research has detailed the organizational infrastructure that exists in Australia and called for further translational research focused on closing the evidence-practice gap, ensuring the viability of e-services through financing and enhancing the reach of, and adherence to, e-therapies especially through health promotion [7,9].

Realizing the potential of these technologies, however, will require that these treatments are embedded within the existing health system as part of a continuum of mental health care and alongside other modalities such as face-to-face psychological treatment and pharmacotherapies.

While meta-analyses show that Internet-based and Internet-assisted therapies are effective and have an important role in the Australian health system, evidence that these interventions *can* work under experimental conditions is not sufficient to show that an intervention *should* be upscaled and implemented from economic, social, and ethical perspectives [10,11]. Nor does it precisely describe how these services will operate within a health system [12-14]. More to the point, evidence on what works in achieving positive treatment outcomes in controlled trials does not necessarily provide information about *how* health policy makers and health professionals might act to implement these new technologies at scale using conventional policy mechanisms and changing established clinical practices [13].

We can think about the informational requirements for introducing a new technology into the health care system in terms of a hierarchy of policy-relevant information (see Figure 1). This is a hierarchy based on information type rather than methodological rigor. Under this view, efficacy and effectiveness studies—randomized controlled trials, systematic reviews, and meta-analyses—which constitute the pinnacle of a hierarchy of evidence types within the biomedical sciences, form the bedrock for subsequent investigations of the cost-effectiveness, acceptability, and logistics of implementing efficacious technologies. The hierarchy of policy relevance is not immutable, and stages of development are interrelated to some degree. For example, the acceptability of technologies can be optimized through incorporating user preferences into the development of technologies as well as through promotion of fully developed ones.

These informational requirements apply, in different ways, to multiple domains: clinical settings, research settings, communities, and within government. Achieving successful implementation depends on harmonizing interacting processes that are initiated in each domain. Thus, a pluralist approach needs to be taken as to what constitutes relevant and useful information to facilitate implementation in different contexts. Clinician, research, community, and policy-maker perspectives all need to be carefully enumerated to ascertain how particular issues are framed, identify mechanisms for action, and describe the scope and limits of what can feasibly and ethically be changed in, and through, each domain in order to facilitate uptake.

The objective of this systematic review was to take stock of what is currently known about the utilization of e-mental health, interpreted from a policy-making perspective on implementation. Our aims were to (1) identify current knowledge about e-mental health service utilization in Australia for depressive and anxiety disorders, (2) synthesize evidence relevant to e-mental health policy development, and (3) identify future directions for policy-focused research.

Figure 1. Hierarchy of policy-relevant information.

Methods

This systematic review employed an a priori protocol based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol (PRISMA-P) 2015 guidelines [15]. The research questions and inclusion criteria were established before the review started through meetings, written proposals, and discussions between the authors. CM and JL were responsible for identifying and reviewing literature. Disagreements in screening and data extraction were resolved through consensus meetings between CM and JL. Data were stored in Endnote and Excel. Sources were appraised according to their study design, using standard quality criteria. Data were synthesized by area and level of policy relevance. We report the results in accordance with the PRISMA statement [16].

Eligibility Criteria

The research question and eligibility criteria were formulated based on a PICO model (Population, Intervention or factors, Comparison, and Outcome). The population of interest was Australia. Literature from New Zealand, United Kingdom, Norway, Sweden, Finland, and Canada were included in the initial search in case an insufficient number of studies were found in Australia. However, sufficient Australian studies were found, and thus international studies were excluded at the screening stage. Our outcome of interest was the use of e-mental health services for depressive (affective) disorders or anxiety disorders (International Classification of Diseases, 10th revision, codes F30 to F44). E-mental health is a relatively new and rapidly evolving field. Thus, only published literature and gray literature from 2005 were included.

Search and Study Selection

The search was conducted during February 2015. A research librarian was consulted regarding the search strategy. Limits used were English language, human subjects, and dates from 2005-2015. Gray literature and peer-reviewed publications were included in our search. Databases searched included Cochrane, PubMed, PsycINFO, CINAHL, Embase, ProQuest Social Science, and Google Scholar. We reviewed the references in the final included studies to find additional research studies, as part of our supplementary search.

Search terms used were MeSH (Medical Subject Headings) for Cochrane and PubMed, Thesaurus of Psychological Index Terms for PsycINFO, CINAHL headings for CINAHL, EMTREE for Embase, ProQuest Social Science, and Google Scholar. For each of these databases, the general search strategy took the following form:

(((e-health & (affective disorder OR anxiety disorder)) OR e-mental health) & (<list of factors associated with use, eg service use OR barriers OR attitude OR socio-economic OR preference>)). Using PubMed as an example, the search strategy was ((((Telemedicine) OR Therapy, Computer-Assisted)) AND ((((((((mental health) OR mood disorders) OR depression) OR anxiety disorders) OR anxiety) OR psychotherapy) OR mental health service) OR community mental health services))) AND (((((((((Australia) OR New Zealand) OR United Kingdom) OR UK) OR Norway) OR Sweden) OR Finland) OR Canada)) AND (((((((((((((Epidemiological factors) OR Health services accessibility) OR Health care disparities) OR Attitude to health) OR Health services research) OR Socioeconomic Factors) OR Demography) OR Social determinants of health) OR Health literacy) OR Patient satisfaction)) OR (Prefer* OR Challeng* OR Barrier* OR Facilitat*))

From the search, all studies were compiled and duplicates were removed. The titles and abstracts of the studies were screened to remove irrelevant studies. The full texts of the studies were then screened by JL and CM on the eligibility criteria for inclusion in the systematic review.

Data Extraction

Data extraction was conducted by JL and CM who compared the extracted data to ensure consistency in data collection methods. Study characteristics extracted included the study aims and information on the sample. The factors of interest included any variables that could be a facilitator or barrier for e-mental health service usage for help-seekers, for example, knowledge and attitudes, sociodemographic, psychological, technological, and environmental factors. We were also interested in institutional and organizational factors that might facilitate or impede the use of e-mental health via service provision. Finally, we were interested in assessing the character of studies conducted in this area, including study design and methods of analysis.

Quality and Bias Assessment

This review differs from the usual aims of systematic reviews in the biomedical sciences in that we wished to analyze past studies in terms of how they might be used to inform government policy. Thus, while we appraised study quality, policy relevance was our key concern. In line with standard protocols, we undertook a quality assessment based on the levels of evidence of the National Health and Medical Research Council (NHMRC) Evidence Hierarchy [17]. Level I evidence included systematic reviews. There was no existing Level I evidence on this topic (ie, on e-mental health service use as opposed to systematic reviews on efficacy of e-therapies, of which there are several [1,18,19]). Level II evidence included randomized controlled trials, observational studies, or case-control studies. Level III evidence included qualitative interviews or focus groups, and Level IV included commentaries and expert opinions. Category IV articles were excluded from evaluation in our findings, as they did not present new empirical evidence. They were instead reviewed and referred to where relevant in our introduction and discussion.

The distinction between studies on samples based on service users and/or online help-seekers versus samples of prospective service users is an important consideration for this review and therefore the sample source was incorporated into the assessment criteria. Level of evidence ratings were labeled "EU" for studies on existing e-mental health service users and/or self-selected e-mental health help-seekers, "PU" for studies drawn from community/general population samples including prospective users, and "SP" for studies that sampled service providers.

Synthesis of Results

Data were synthesized for analysis according to their level of policy relevance and area of policy relevance. These categories are defined below.

Levels of Policy Relevance

We assessed papers on a 3-point scale (Low, Mod, High) of policy relevance. Our intention was to qualitatively score items against the hierarchy of policy-relevant evidence shown in Figure 1. Policy relevance="Low" were studies focused on showing that a treatment or intervention is clinically effective. Policy relevance="Mod" were studies that justified implementation of an intervention and defined the parameters for an intervention's usefulness. These include cost-effectiveness and prioritization studies, as well as analyses of the ethical and social acceptability of the broad-scale implementation of a particular treatment. Acceptability, from a policy perspective, has a different meaning to the way it is commonly used in clinical trials—although there is overlap. In clinical trials, acceptability refers to elements such as satisfaction with treatment and treatment compliance. For the purposes of this review, acceptability refers to the "attractiveness" and appeal of an intervention among a significant sector of society. Acceptability includes both a disposition to use an intervention oneself and support for the idea of the intervention, for example, that key sectors of the public believe that e-mental health is a good idea and that it is appropriate for the government to deliver some mental health services in this way. Policy relevance="High" refers to studies that provide explicit, empirical, or analytical evidence to support particular approaches to facilitating and governing the delivery of e-mental health care.

Area of Policy Relevance

For each study, we identified how it contributed to an area of policy relevance. We labeled these Target Demographic (T), Facilitating Uptake (F), and Governing Mechanisms (G). These areas are not independent from one another, and each study could potentially contribute to more than one area of policy relevance.

Target Demographic (T) refers to findings relevant to understanding e-mental health service use among specific sectors of the population. To examine this aspect of the literature, we analyzed information about sample characteristics, study inclusion and exclusion criteria, means of sample recruitment, mental health disorder targeted, phase of intervention (prevention or treatment), and platform or mode of e-mental health service.

Facilitating Uptake (F) refers to findings that are useful in understanding what characteristics explain willingness to use e-mental health care and under what conditions e-mental health will be attractive to different groups of people. To examine this aspect of the literature, we extracted data on what outcomes, relevant to facilitating uptake, were measured and reported, including individual level facilitators and barriers of use.

Governing Mechanisms (G) refers to findings that provide information on governance arrangements and policy settings needed to facilitate the establishment of e-mental health services within the health care system. We provided details about the policy implications of papers, classified according to a typology of policy mechanisms relevant to health governance: Organization, Regulation, Community Education, Finance, and Payment [20]. We added Information Communication Technology as a category, as this is a rapidly evolving area of health policy that may or may not be adequately encompassed by existing typologies for classifying policy mechanisms.

Results

Study Selection

As shown in Figure 2, the database search yielded 1081 records, comprising 17% from Cochrane, 38% from PubMed, 7% from PsycINFO, 4% from CINAHL, 25% from Embase, and 9% from ProQuest Social Science. The supplementary search yielded an additional 20 records for consideration of which four were included in full-text screening. After duplicates were removed, 1035 records went through the title screening stage to exclude studies that were not on e-mental health (eg, studies on stroke, dementia, chronic pain, or weight management), from which 784 records were excluded, leaving 251 records for abstract screening. From screening the abstracts, 159 records were excluded, which left 92 records for full-text assessment for eligibility. A further 62 records were excluded due to the following reasons: not Australia-focused (12/62, 19%), not e-mental health for consumers (14/62, 23%), not for anxiety or depressive disorders (3/62, 5%), and not on e-mental health usage (33/62, 53%). A list of the excluded studies along with the reasons for exclusion is presented in Multimedia Appendix 1. A total of 30 studies were included in the analysis.

XSL•FO

RenderX

Figure 2. PRISMA flow diagram for study inclusion.
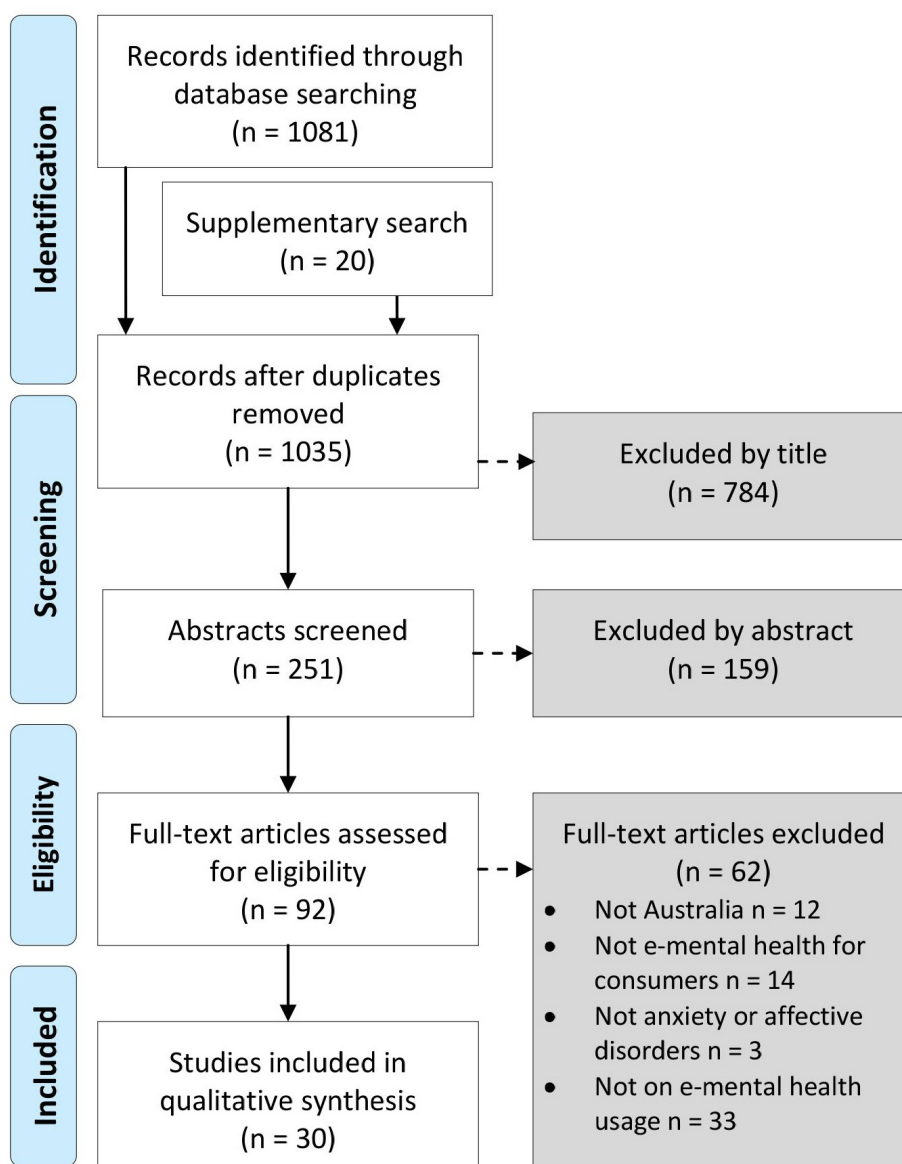
Study Characteristics

Our findings show that the majority of research to date has been conducted on clinical and self-selected online help-seeking populations (see Table 1 [21-50]). From the included studies, 63% (19/30) of studies were conducted on existing or recruited e-mental health users and participants (EU), 30% (9/30) were conducted on general populations or prospective users (PU), and 10% (3/30) of studies were conducted on service providers (SP). Most empirical research (24/30, 80%) has been undertaken by the developers of the interventions being discussed. Our key focus was to draw together literature on e-mental health service use, including facilitators and barriers; however, only 60% (18/30) of the studies included e-mental health utilization as a research question. Of the included studies, 93% (28/30) were trials and online surveys (level II evidence) and the other two were qualitative interview studies (level III evidence). Sample sizes varied markedly across studies, with sample sizes ranging from 10 to 110,825. Fewer than half (13/30, 43%) of the studies were given a policy relevance rating of Low; 47% (14/30) had a rating of Mod. A small minority of studies (3/30, 10%) provided empirical evidence focused on implementation, that is, had a policy relevance rating of High. In terms of area of policy relevance, we classified 97% (29/30) of studies as relevant to understanding e-mental health Target Demographic (T), 57% (17/30) as relevant to Facilitating Uptake (F), and 77% (23/30) as relevant to Governing Mechanisms (G) of e-mental health within the health system.

Table 1. Study characteristics (N=30).

Reference	Study aims (Aims type, Level of evidence – Sample type)[a]	Sample size	Policy relevance: level[b]; area[c]
Anderson et al, 2012 [21]	Examine the quality of the working alliance in online cognitive behavioral therapy (CBT) for anxiety disorders in youth and the role of working alliance and compliance in predicting treatment outcome (B, II-EU)	132 children and adolescents, and their parents	Low; T
Bennett et al, 2010 [22]	Describe ehub and populations for whom it may be suited (B, II-EU)	110,825 website visitors to e-mental health site	Mod; TG
Bradford, Rickwood, 2014 [23]	Determine whether adolescents prefer online over more traditional types of mental health service delivery, what their help-seeking intentions are for a commonly experienced mood disorder and the factors that affect these intentions (A, II-PU)	231 school students	Mod; TFG
Casey et al, 2013 [24]	Assess the impact of providing e-mental health information on attitudes toward e-mental health services (A, II-PU)	217 general convenience sample	High; FG
Christensen et al, 2006 [25]	Examine predictors of depression and anxiety scores on the MoodGYM website as a function of user characteristics, and to compare the compliance rates of the original site with the new public version of the site (B, II-EU)	58,398 public registrants to the MoodGYM site	Low; T
Crisp, Griffiths, 2014 [26]	Examine the characteristics associated with interests and preferences in using online mental health interventions (A, II-PU)	4761 Well-Being Project participants	Mod; TFG
Dear et al, 2013 [27]	Evaluate the efficacy, acceptability, and feasibility of a brief iCBT[d] program, Managing Your Mood Program, to treat depression among older adults aged 60 years and older (A, II-EU)	20 older adults	Low; TFG
Dear et al, 2015 [28]	Examine acceptability, efficacy, and health economic impact of two self-guided iCBT programs for adults over 60 years of age with anxiety and depression (A, II-EU)	47 older adults	Mod; TFG
Dingwall et al, 2015 [29]	Examine the acceptability, feasibility, and appropriateness of e-mental health resource app for use by service providers with Aboriginal and Torres Strait Islander communities (A, III-SP)	15 Aboriginal and Torres Strait Islander service providers	Mod; TFG
Donker et al, 2013 [30]	Predict treatment outcomes of new e-couch Internet-delivered Interpersonal Psychotherapy (iIPT) and CBT against MoodGYM CBT (B, II-EU)	1843 spontaneous website visitors	Low; T
Ellis et al, 2012 [31]	Explore young people's attitudes and behaviors in relation to mental health and technology use (A, II-PU)	1038 young people	Mod; TFG
Ellis et al, 2013 [32]	Explore young men's knowledge, attitudes, and behavior towards mental health and technology use (A, II-PU)	486 young men from online surveys and 118 from focus groups	Mod; TFG
Gun, Titov, Andrews, 2011 [33]	Explore levels of acceptability of Internet-based treatment programs for anxiety and depression (A, II-EU & SP)	1543 health professionals and lay people	High; TFG
Johnston et al, 2014 [34]	Explore the efficacy and acceptability of iCBT for young adults with anxiety and depression (A, II-EU)	18 young adults	Low; TFG
Keane et al, 2013 [35]	Examine the use of the Internet to access mental health information by demographic characteristics (A, II-PU)	2996 general population	High; TFG
Kirkpatrick et al, 2013 [36]	Report acceptability, feasibility, and preliminary efficacy of established iCBT course (Well-being Course) being administered by nongovernmental organization for anxiety (A, II-EU)	10 adult callers or website visitors of Mental Health Australia	Mod; TFG
Kiropoulos et al, 2008 [37]	Compare the effectiveness of iCBT versus face-to-face CBT for panic disorder and agoraphobia (B, II-EU)	86 people with panic disorder	Low; T
Klein et al, 2011 [38]	Evaluate the Anxiety Online programs (B, II-EU)	225 people self-selected for e-therapy programs	Low; T
Klein et al, 2010 [39]	Open trial to evaluate posttraumatic stress disorder online (B, II-EU)	22 adults with posttraumatic stress disorder	Low; T
Klein, Richards, Austin, 2006 [40]	Compare the efficacy of Internet-based self-help and self-help manual for treating panic disorders (B, II-EU)	55 people with panic disorder	Low; T

Reference	Study aims (Aims type, Level of evidence – Sample type)[a]	Sample size	Policy relevance: level[b]; area[c]
Morgan, Jorm, Mackinnon, 2012 [41]	Test the effectiveness of an automated email-based campaign promoting self-help behaviors (B, II-EU)	1326 adults with depression	Low; TG
Neil et al, 2009 [42]	Investigate adherence rates to a CBT website in adolescent samples from a school-based or community setting (A, II-EU)	1000 school-based and 7207 community-based adolescents	Mod; TFG
O'Kearney et al, 2009 [43]	Evaluate the benefits of MoodGYM compared to a usual high school curriculum (B, II-EU)	157 girls	Low; TG
Pier et al, 2008 [44]	Evaluate the efficacy of an iCBT intervention (Panic Online) for the treatment of panic disorder (B, II-EU)	65 people with panic disorder	Low; TG
Proudfoot et al, 2010 [45]	Explore community attitudes toward the appropriation of mobile phones for mental health monitoring and management (A, II-PU)	525 from online survey; 47 from focus groups; 20 interviews	Mod; TFG
Robertson et al, 2006 [46]	Test the feasibility of implementing an e-mental health system for the treatment for depression (A, II-EU)	144 depressed adults	Low; TG
Sinclair et al, 2013 [47]	Understand rural clinicians' attitudes towards the acceptability of online mental health resources as a treatment option in the rural context (A, III-SP)	21 rural clinicians	Mod; TFG
Titov et al, 2010 [48]	Examine characteristics of adults with anxiety and depression treated at an Internet clinic with national survey data and outpatient clinic data (B, II-PU)	774 volunteers to an Internet Clinic, 454 patients in an anxiety disorders outpatient clinic, 627 National survey cases	Mod; TG
Wootton et al, 2011 [49]	Establish the acceptability of iCBT treatments for adults with obsessive compulsive disorder (A, II-PU)	129 volunteers to an online survey, 135 in an anxiety disorders outpatient clinic, 297 National survey cases	Mod; TFG
Zou et al, 2012 [50]	Perform feasibility study for iCBT for anxiety in older adults (A, II-EU)	22 older adults with anxiety	Mod; TFG

[a]Study aims type: A=includes investigation of barriers and facilitators of e-mental health use as part of the research aim; B=provides information about e-mental health use, including barriers and facilitators, even though this was not part of the research aim.
Study aims level of evidence: II=randomized controlled trials, observational studies, or case-control studies; III=case series, focus groups; EU=study of existing e-mental health service users or self-selected sample; PU=study was on prospective e-mental health users; SP=study of service providers.

[b]Policy relevance level: Low=minimal policy relevance, Mod=some policy relevance, High=direct policy relevance/policy-focused.

[c]Policy relevance area: T=Target Demographic, F=Facilitating Uptake, G=Governing Mechanisms

[d]iCBT=Internet-based cognitive behavioral therapy.

Target Demographic

The 29 studies that provided information on target demographics provided variable detail on their study samples and the characteristics of e-mental health users (see Multimedia Appendix 2). As mentioned already, there was a bias towards online recruitment and self-selected e-mental health service users. Study samples tended to be biased towards females. Program development has targeted different age groups with tailored programs, and young people have received particular attention to date (6/29, 21%). Half of the studies (15/29, 52%) provided information about socioeconomic status (mostly employment status) of e-mental health care users. Where relevant information was provided, it appeared there was a bias towards middle- to high-income earners. Nearly half of studies (13/29, 45%) provided information about educational attainment and in these there was a bias towards more highly educated members of the public. Only 14% of studies (4/29) asked or provided information about ethnicity, directly or indirectly. Inclusion/exclusion criteria for studies requiring fluency in written English, as well as technological requirements (eg, access to a computer, Internet, and printer), reinforce these biases. Only 14% of studies (4/29) cited statistics on the geographic location of participants (eg, urban versus rural), and few studies provided information on relationship status of participants.

Facilitating Uptake

Synthesized results of the 17 studies that provided information on measurements related to facilitating uptake of e-mental health are presented in Table 2 [23,24,26-29,31-36,42,45,47,49,50] and Multimedia Appendix 3. As detailed earlier, the majority of research included in this review was designed to justify, or enhance, the efficacy and effectiveness of online interventions rather than to investigate the appeal of currently available online therapies as a possible course of treatment for prospective help-seekers (see Multimedia Appendix 3). Consequently, many of these studies were focused on enhancing uptake through program development rather than investigating how systems-wide implementation could be achieved via policy and planning. Having said this, satisfaction with treatment was linked to likelihood of recommending e-mental health to others in a number of studies [27,28,34,36,50].

XSL•FO
RenderX

Table 2. Facilitators and barriers for e-mental health utilization (N=17).

Reference	E-therapy utilization: Facilitators	E-therapy utilization: Barriers	Non-significant factors
[23]	Motivated to seek face-to-face help rather than receive no help	Not preferring online treatment	Self-reliance
		Females prefer face-to-face help	Shyness
	Males who would have otherwise chosen no help	Lower mental health literacy	Stigma
	Higher mental health literacy	Viewing e-therapy as impersonal	
	Anonymity of the Internet	Lack of trust	
	Accessibility of information	Not knowing who you are talking to	
	Connecting with others who have been through the same thing	Lack of customized feedback	
[24]	Knowledge about e-mental health through provision of textual information	Lack of knowledge about e-mental health	Type of e-mental health service
		Attitude that online programs without therapist assistance are not helpful	
[26]	Female	Male	—
	Higher education	Low education	
	Not married	Young age	
	History of depression	Lack of interest	
	Higher depressive symptoms	Stigma	
	More free time	Too busy	
		Prefer to deal alone	
[27]	High adherence	—	—
	High satisfaction linked to likelihood of recommending to others		
[28]	High satisfaction linked to likelihood of recommending to others	—	—
[29]	Attractive visual appeal	Technology issues	Individual mental health issues
	Ease of use	Time constraints for service providers	Age
	Culturally appropriate	Concern for job security	Sex
	Enjoyable / fun	Translation into Indigenous languages	
	Appropriate training for service providers		
[31]	Positive attitudes towards e-mental health in general	Male	—
		Interactive games were not preferred	
[32]	Privacy and anonymity	Ideas about masculinity	—
		Preference for reliance on informal networks	
		Preference for self-help	
		Generalized scepticism of "interventions"	

Reference	E-therapy utilization: Facilitators	E-therapy utilization: Barriers	Non-significant factors
[33]	Low severity of mental health symptoms	Lack of information about effectiveness of e-mental health	The need for reliable Internet
		Lack of knowledge about treatments available	Lack of computer skills
		Lack of established guidelines	IT support
		Unclear about legal issues involved or liabilities of recommending e-therapies	
		Lack of training for health professionals	
		Preference for not seeking help at all over using e-mental health	
		Lack of experience in using e-mental health treatments	
[34]	Good adherence	—	Low acceptability
	High satisfaction, linked to likelihood of recommending e-mental health		
[35]	Female	Male	Metropolitan versus rural location of residence
	Younger age (15-54)	Older age	
		Low overall usage	
[36]	High satisfaction	Therapist initial scepticism	—
[42]	Monitored settings, such as school-based settings	Unmonitored-settings	History of depression
	Female	Male	
	Living in rural areas		
[45]	Symptoms of depression, anxiety, or stress were more likely to be interested in mobile mental health	Perceived as not helpful	Sex
		Negative attitudes towards technology	Age
	Speed and convenience	Privacy concerns	Employment
	Ease of access	Lack of Internet access on mobile phone	Marital status
	Positive attitude towards self-help	Small screen of mobile phone	
	At least some access		
	Less confronting than face-to-face-consultation		
[47]	Usability, privacy	Inadequate (private) Internet access in some rural settings	—
	Provides some services to rural areas where there is a lack of service	Reading difficulties among consumers	
	Training for clinicians	Computer literacy	
	Provision of informational materials for providers and consumers	Difficulty accessing training in the rural environment	
	Ability for e-mental health to be integrated with existing care	Practitioner concerns about lack of feedback from clients, rumination or social isolation	
	Promotion of e-mental health as an effective treatment	Scepticism about the effectiveness of e-mental health treatments	
		Lack of time to explore resources	

Reference	E-therapy utilization: Facilitators	E-therapy utilization: Barriers	Non-significant factors
[49]	Embarrassment of face-to-face	Prefer face-to-face	Got told not to use
	Believed that e-mental health would be useful	Embarrassment	Lack of access to computer/Internet
	Privacy and anonymity	Perceived as not effective	
	Convenience	Cannot see a person	
	Bridges travel issues	Inferior to communication with therapist	
	Reduced costs	Do not know what e-mental health care is	
	Willingness to try	Prefer self-management	
	Useful for mild symptoms	Too confronting	
		Problems not severe enough	
		Prefer medications	
		Sounds too risky	
		Lack of time	
[50]	High level of satisfaction, related to likelihood of recommending treatment to a friend.	—	—

Six of 30 studies (20%) sought to understand treatment preferences for online therapies compared to face-to-face psychological therapies, including "interest" or "willingness to try" online therapies [23,31-33,45,49]. We did not find studies that directly compared preferences for online therapies, face-to-face therapies, and pharmacotherapies. Two studies quantified relative preferences [23,33] and found a preference for face-to-face therapies over online therapies. In a sample of adults who visited a website for depressive and anxiety disorders, 63% of participants preferred face-to-face, compared to 7% who preferred e-mental health services [33]. Similarly, in a non-clinical sample of students in grades 10-12 recruited from schools, 58% preferred face-to-face, compared to 16% who preferred e-mental health services [23]. There was some indication that online therapies with practitioner support were preferred to online-only therapies [23,47]. The exception to this rule was that young men preferred informational websites to treatment-oriented websites [31].

Facilitators and barriers for e-mental health utilization are presented in Table 2. Stigma, broadly defined, was highlighted as both facilitating the use of e-therapies (including, "embarrassment" of seeking face-to-face help), as a barrier to use, and as non-significant [23,26,49]. Mental health literacy was highlighted as a facilitator in one study [23], and awareness (or lack thereof) of e-mental health was identified as important in four studies [24,32,47,49]. Being a rural resident was identified as a facilitator [42], a barrier [47] and as non-significant [35]. Some perceived qualities of e-mental health care were both facilitators and barriers, depending on whether different individuals interpreted them positively or negatively. For example, some studies identified "anonymity" as a facilitator of e-mental health use [23,47,49], but anonymity was arguably also a barrier when e-mental health services were seen as

depersonalized [23,45]. Assessments about using e-mental health care differed depending on different beliefs as to whether important requirements, such as the need for privacy, were met. For example, concerns with privacy could be a facilitator of use, if e-mental health care was perceived as private [47,49]. However, "concern with privacy" was also deemed a barrier to use [45], indicating that some people do not perceive e-mental health care as protecting privacy. A preference for "self-help" was also reported as being a barrier or facilitator to the use of e-mental health [23,26,32,49], depending on whether e-mental health was viewed as consistent with self-help or not. Both lower symptom severity [33] and higher symptom severity [26,45] have been identified as facilitators of use.

Governing Mechanisms

Over three-quarters (23/30, 77%) of studies examined factors from which we could draw inferences about policy mechanisms needed to establish e-mental health within the health system (see Table 3 [22-24,26-29,31-36,41-50]). However, none of these studies characterized the policy settings required to implement e-therapies. Nineteen of these studies (83%) provided insight into the organizational requirements for establishing e-mental health. These described settings in which e-mental health could justifiably be implemented, namely, schools, general practice, non-governmental mental health organization websites, and through direct-to-public online delivery. These studies also described configurations of e-mental health care delivery (eg, informational websites, peer support websites, Internet-only therapy or clinician-moderated e-mental health care) that may be best accepted by different sectors of the population. However, these studies did not provide details on the relative merits of implementation of different organizational types at scale, nor how implementation in different settings might occur.

Table 3. Governing mechanisms (N=23).

Reference	Implications for governing mechanisms	Details related to governing mechanisms[a]
[22]	Organization	Justifies the provision of Internet-only therapy.
[23]	Organization, Community education	Quantifies preferences among young people for online help, face-to-face help, and tele-help.
		Identifies factors that may influence appeal of online help via health promotion.
[24]	Community education	Identifies text-based methods as best means of delivering information about e-mental health.
[26]	Finance/payment	The paper itself does not make the following argument; however, the paper identifies that financial incentives could nudge approximately 20% of participants to engage with e-mental health.
[27]	Organization	Establishes feasibility and acceptability of iCBT for adults 60 years and over with depression.
[28]	Organization, Finance/payment	Establishes feasibility and acceptability of iCBT for adults over 60 years old with depression and anxiety.
		Quantifies economic health costs associated with participating in the programs at around $60 per person.
[29]	Organization, Community education, Information communication technology	Highlights the feasibility and acceptability of service providers in remote Aboriginal and Torres Strait Islander communities using mobile apps to engage with consumers.
		Highlights the need for training and informational materials for service providers.
		Highlights infrastructural and technical barriers to information communication technology use in remote areas.
[31]	Organization, Community education	Showed that young people preferred websites with information or online clinics to websites with question and answer or interactive games.
[32]	Organization, Community education	Suggests tailoring online services (informational and treatment) to different tastes.
[33]	Regulation, Organization, Community education, Information communication technology	Quantifies preferences for Internet treatment compared with face-to-face treatments.
		Identifies concerns with liability as an issue for health professionals recommending Internet-based treatments.
		Identifies health professionals' and lay persons' needs for more information about Internet-based treatments, including information about effectiveness.
		Identifies infrastructure and computer literacy as barriers to use among a minority of health professionals and lay people.
[34]	Organization	Justifies feasibility of Internet-only therapy for young people.
[35]	Community education	Highlights (and quantifies) characteristics of potential user groups for e-mental health. Middle-aged rural females most disposed, older rural males least disposed.
[36]	Organization, Community education	Justifies feasibility of delivering iCBT via not-for-profit organizations' websites.
		Registered clinicians not necessary for delivery, can train other staff.
[41]	Community education	Internet-delivered self-help messages are a low-cost, automated, and easily disseminated prevention option.
[42]	Organization	Justifies school-based delivery of online interventions for depressive and anxiety disorders for adolescents.
[43]	Organization	Justifies delivery of MoodGYM in school settings.
[44]	Organization	Justifies delivery of iCBT for panic disorder with either face-to-face support from general practitioner or email support from psychologist.
[45]	Organization, Regulation, Information communication technology	Privacy and security are important to people using mobile health.
		Not suitable for those who dislike the use of technology.
		Highlights feasibility of mobile mental health.
[46]	Organization	Justifies use of comprehensive eHealth system for management of depression, including adherence to medication (including consultations, monitoring, psychoeducation, and therapy).

Reference	Implications for governing mechanisms	Details related to governing mechanisms[a]
[47]	Organization, Community education	Overall, rural clinicians supported implementation of Internet-assisted therapies, as an adjunct to face-to-face consultations.
		Highlights need for informational materials for rural clinicians and consumers.
[48]	Organization	Justifies iCBT for anxiety and depressive disorders for the wider population.
[49]	Organization, Regulation	Justifies demand for Internet-based treatments for obsessive compulsive disorder.
		Privacy and anonymity important to using face-to-face treatment.
[50]	Organization	Justifies feasibility of iCBT for older adults with anxiety.

[a]iCBT=Internet-based cognitive behavioral therapy.

Ten studies (10/23, 43%) provided insights on community education. One study investigated the usefulness of different modes of delivery of information (eg, by text or by film) about e-mental health care and found that providing text-based information increased likelihood to use e-mental health services in the future [24]. Studies that included information about service providers' views highlighted the need for informational materials and training about e-mental health, including evidence about its efficacy and also the need to distribute information about liability.

Two studies provided some information relevant to financing and payment [26,28]. One study provided an estimate of total health care costs associated with using Internet-based cognitive behavioral therapy (iCBT), showing that iCBT use was associated with marginally higher health care costs [28]. The other identified that participants' willingness to complete iCBT interventions might be enhanced by appropriate financial incentives (ie, nudges) [26].

Two studies addressed regulatory issues. These included participants' concerns about privacy and anonymity [45], which has relevance to data collection, storage, and security, and health care professionals' concerns about legal liability [33] for recommending and using Internet-based treatments. Finally, three studies highlighted infrastructure and technical issues [29,33,45] associated with deploying mobile-health technologies, including in remote or Aboriginal and Torres Strait Islander communities [29,33]. Computer literacy was seen as a minor issue [33].

Discussion

Principal Findings

Meta-analyses show that Internet-based and Internet-assisted therapies are an effective means of treating many individuals with depressive and anxiety disorders, and that those who use these therapies tend to be satisfied with them [1]. While these results show that e-mental health has a potentially important role in the Australian health system, the evidence base does not adequately define the population for whom e-mental health care is, and could be, most suitable. It does not accurately benchmark current use or provide indications of likely future levels of e-service use compared to other treatments. It also does not present sufficient information to inform policies that could facilitate its broad-scale adoption. These findings corroborate a recent review and NHMRC Case for Action [7,9], a review that found no policy-focused research has been undertaken on e-mental health [19] and calls for further translational research in this area [51].

Current knowledge on determinants of e-mental health service use presents a program development perspective on e-mental health establishment. The primary focus of proposed translational activities has been on closing the evidence-practice gap, ensuring the viability of e-services through financing, and enhancing the reach of, and adherence to, e-services including through promotion [9]. These are important and necessary translational activities. However, facilitating the establishment of e-mental health care within the Australian health system requires additional translational research to provide, what we term, a "policy-making" perspective. Distinctively from translational research activities focused on consolidating and expanding e-services within the Australian health system [9], a policy-making perspective approaches the question of implementing e-mental health, exogenously, based on two primary considerations: (1) the kinds of mechanisms available to government to facilitate implementation and (2) the imperative to fit e-mental health care within a population-based, stepped-care model that includes a range of treatment types for depressive and anxiety disorders and incorporates contingency planning.

The studies we reviewed were mostly clinical trials conducted with self-selected e-therapy users. Information about culture/ethnicity and socioeconomic status are infrequently reported. Based on the studies we reviewed, there seems to be a sex bias, with females more likely to use e-mental health care than males. These patterns of use probably reflect patterns of utilization in face-to-face treatment seeking [52]. Highlighting these biases does not undermine the value of e-services but is important to ensuring that integrating e-services into the mental health system works to overcome inequalities, rather than exacerbate them. How best to respond to these biases is unclear, as three courses of action are possible: (1) invest in promoting existing e-services to under-using demographics, (2) design new services tailored for these populations, and (3) invest in funding alternative treatment modalities that may be more attractive to groups who underutilize e-services. Further policy-focused research on non-use of e-mental health care is important to informing appropriate future courses of action with respect to these biases.

Different studies investigated and reported different possible facilitators and barriers to use and the concepts investigated

proved to be fairly slippery. Factors that may facilitate or impede use operate at different scales and levels and can be viewed differently from different perspectives. In other words, constructs can be worded as both facilitators and barriers while reflecting a similar process. Additionally, there was evident symmetry as to what is a facilitator or barrier. For example, different beliefs about whether or not online therapies are private as well as whether or not anonymity is an appealing or undesirable quality in a treatment, highlights the importance of different interpretations as well as preferences. While lack of consistency in the definition of constructs across studies likely contributes to a lack of unequivocally identifiable facilitators or barriers, we think that further diverse examination of facilitators and barriers is needed before any calls for standardization of constructs is warranted. More pressing is the need to examine different interpretations of online health interventions to inform the detail (wording) of community education campaigns.

Future Research Directions

Policy-focused research is required to (1) prioritize ongoing research and development of e-services that will ensure adequate coverage of mental health care for prospective e-help-seekers, (2) provide accurate estimates of current e-mental health usage and identify realistic future targets for e-service use, relative to other service and treatment types, (3) elucidate the factors underlying preferences for and against therapies, particularly to inform promotional materials that resonate with different perceptions and values of self-help, privacy, and anonymity, and (4) inform the establishment of appropriate governing mechanisms for e-services, giving highest priority to privacy and data security, liability, and modes of financing and payment [9].

Conducting this research independently of e-mental health program development will allow for resourcing across research and development and service delivery to be informed by a critical appraisal that includes contingency planning. Research focused on increasing adoption and adherence is focused on engaging with consumer preferences as well [9]. However, from a policy-making perspective, understanding preferences, and how malleable these might be, has a slightly different function insofar as it can inform decisions about how to allocate funds to different activities along the translational spectrum from program development to promotion. Understanding preferences is also important in deciding how to allocate resources to other treatment modalities and institutions that address the downstream impacts for those who, for whatever reason, remain untreated.

Methodologically speaking, in addition to translational research identified elsewhere [9], we recommend:

1. Further reviews of eHealth policy from Australia and internationally to inform policies on privacy, data security, liability, and modes of financing and payment for services. These reviews should draw on academic and gray literature across a range of eHealth and telehealth areas, with the aim of identifying suitable regulatory mechanisms for governing e-mental health. Literature reviews can be enhanced through stakeholder interviews with Australian e-service developers and providers as well as policy makers.

2. Qualitative interview studies of current users and non-users of e-mental health services, including semi-structured interviews and think-aloud exercises, should be conducted to inform the details of promotional materials that will resonate with disparate perceptions of e-mental health services with respect to issues of stigma, privacy, anonymity, and self-help.

3. Surveys using discrete choice experiment methodologies are important for accurately characterizing preferences for e-mental health care, face-to-face therapies, and prescription medications. Prescription medications, in particular, are the "elephant in the room" of e-mental health studies; including this treatment in comparisons is important given the biases evident in e-mental health care use and in understanding the scope and limits of e-mental health care for those who are not fluent in English or have low literacy or comprehension. A course of prescription medications has minimum language or comprehension requirements.

Limitations

We elected to focus on Australian research because policy development is importantly context-specific [14]. Nonetheless, our conceptual framework and methodological approach for this study and the implications drawn for future research all have international relevance. The inferences made under the theme "Target Demographic" must be understood in relation to our search criteria, which focused on factors influencing service use and thus did not include feasibility or effectiveness studies for programs targeting different cultural groups that did not provide data on service use factors [53-55]. Policy studies and economics research relevant to the topic may have been excluded because they are not found in medical databases searched. Our capacity to undertake truly multidisciplinary systematic reviews may have been limited by differences in the meanings of words in medical research versus political and social sciences, the specificities of MeSH terms, and other conventions for identifying search terms, and the different framings and focus of research in different disciplines. However, we attempted to overcome this limitation by searching databases such as ProQuest Social Science and search engines like Google Scholar. In addition, our search was conducted using a combination of headings as well as keywords and synonyms across the different disciplines. Our findings corroborate other reviews that point to a lack of translational research in this area. Therefore, we are reasonably confident about our results.

Our classification scheme for policy-relevant research does not acknowledge the "behind-the-scenes" development of implementation-focused thinking that can inform research design and questions nor policy advocacy work that addresses implementation issues. Finally, our review excluded general e-mental health studies (eg, [56]) that focused on service use types but did not investigate disorder type specifically, as the scope of our review included only depressive and anxiety disorders. We do not think such omissions invalidate our conclusions.

Conclusion

Successfully establishing e-mental health care within the health system will depend on the skillful coordination of activities within clinical, community, research and development, and policy-making realms. This, in turn, will depend on appropriate translational research being conducted that is relevant to each of these domains. This review provides a rationale and framework for undertaking dedicated policy-focused research on e-mental health in the future.

Authors' Contributions

All authors contributed to the conception and design of this review. JL and CM were responsible for data extraction and analysis. CM and JL were responsible for the initial drafting of the manuscript. All authors contributed to revising the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of excluded studies with exclusion reasons.

[PDF File (Adobe PDF File), 37KB - jmir_v18i1e10_app1.pdf]

Multimedia Appendix 2

Target demographic.

[PDF File (Adobe PDF File), 60KB - jmir_v18i1e10_app2.pdf]

Multimedia Appendix 3

Measurements related to facilitating uptake and evidence on e-mental health utilization.

[PDF File (Adobe PDF File), 37KB - jmir_v18i1e10_app3.pdf]

References

1. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. Andrews G, Titov N. Is internet treatment for depressive and anxiety disorders ready for prime time? Med J Aust 2010 Jun 7;192(11 Suppl):S45-S47. [Medline: 20528709]
3. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. Med J Aust 2010 Jun 7;192(11 Suppl):S53-S56. [Medline: 20528711]
4. Griffiths KM. A virtual mental health community - a future scenario. Aust N Z J Psychiatry 2013 Feb;47(2):109-110. [doi: 10.1177/0004867412471440] [Medline: 23382506]
5. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
6. Jorm AF, Morgan AJ, Malhi GS. The future of e-mental health. Aust N Z J Psychiatry 2013 Feb;47(2):104-106. [doi: 10.1177/0004867412474076] [Medline: 23382504]
7. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
8. Australian Government Department of Health and Ageing. E-Mental Health Strategy for Australia. Canberra: Australian Government Department of Health and Ageing 2012:1-22 [FREE Full text]
9. Batterham P, Sunderland M, Calear A, Christensen H, Teesson M, Kay-Lambkin F. Case for Action proposal: Translation of e-mental health services for depression. National Health and Medical Research Council 2015:1-21 [FREE Full text]

XSL•FO

RenderX

10. Farah MJ, Illes J, Cook-Deegan R, Gardner H, Kandel E, King P, et al. Neurocognitive enhancement: what can we do and what should we do? Nat Rev Neurosci 2004 May;5(5):421-425. [doi: 10.1038/nrn1390] [Medline: 15100724]

11. Davis A. Ethical issues for psychologists using communication technology: An Australian perspective on service provision flexibility. Professional Psychology: Research and Practice 2014;45(5):303-308.

12. Carter A, Hall W. Addiction neuroethics: The promises and perils of neuroscience research on addiction. UK: Cambridge University Press; 2011:1139504673.

13. Whiteford H, Harris M, Diminic S. Mental health service system improvement: translating evidence into policy. Aust N Z J Psychiatry 2013 Aug;47(8):703-706. [doi: 10.1177/0004867413494867] [Medline: 23814069]

14. Cartwright N, Hardie J. Evidence-Based Policy: A Practical Guide to Doing It Better. UK: Oxford University Press; 2012:-.

15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269, W64. [Medline: 19622511]

17. Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. BMC Med Res Methodol 2009;9:34 [FREE Full text] [doi: 10.1186/1471-2288-9-34] [Medline: 19519887]

18. Calear AL, Christensen H. Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. Med J Aust 2010 Jun 7;192(11 Suppl):S12-S14. [Medline: 20528700]

19. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]

20. Roberts M, Hsiao W, Berman P, Reich M. Getting health reform right: a guide to improving performance and equity. New York: Oxford University Press; 2008:0199888167.

21. Anderson RE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: comparison with clinic delivery and its role in predicting outcome. J Med Internet Res 2012;14(3):e88 [FREE Full text] [doi: 10.2196/jmir.1848] [Medline: 22789657]

22. Bennett K, Reynolds J, Christensen H, Griffiths KM. e-hub: an online self-help mental health service in the community. Med J Aust 2010 Jun 7;192(11 Suppl):S48-S52. [Medline: 20528710]

23. Bradford S, Rickwood D. Adolescent's preferred modes of delivery for mental health services. Child Adolesc Ment Health 2014 Feb 11;19(1):39-45. [doi: 10.1111/camh.12002]

24. Casey LM, Joy A, Clough BA. The impact of information on attitudes toward e-mental health services. Cyberpsychol Behav Soc Netw 2013 Aug;16(8):593-598. [doi: 10.1089/cyber.2012.0515] [Medline: 23679567]

25. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an Internet-based cognitive behaviour therapy program. Aust N Z J Psychiatry 2006 Jan;40(1):59-62. [doi: 10.1111/j.1440-1614.2006.01743.x] [Medline: 16403040]

26. Crisp DA, Griffiths KM. Participating in online mental health interventions: who is most likely to sign up and why? Depress Res Treat 2014;2014:11 [FREE Full text] [doi: 10.1155/2014/790457] [Medline: 24804089]

27. Dear BF, Zou J, Titov N, Lorian C, Johnston L, Spence J, et al. Internet-delivered cognitive behavioural therapy for depression: a feasibility open trial for older adults. Aust N Z J Psychiatry 2013 Feb;47(2):169-176. [doi: 10.1177/0004867412466154] [Medline: 23152358]

28. Dear B, Zou J, Ali S, Lorian C, Johnston L, Terides M, et al. Examining self-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety and depression: Two feasibility open trials. Internet Interventions. Internet Interventions 2015;2(1):17-23.

29. Dingwall K, Puszka S, Sweet M, Nagel T. "Like Drawing Into Sand": Acceptability, Feasibility, and Appropriateness of a New e‐Mental Health Resource for Service Providers Working With Aboriginal and Torres Strait Islander People. Australian Psychologist 2015;50(1):60-69.

30. Donker T, Batterham PJ, Warmerdam L, Bennett K, Bennett A, Cuijpers P, et al. Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression. J Affect Disord 2013 Oct;151(1):343-351. [doi: 10.1016/j.jad.2013.06.020] [Medline: 23953024]

31. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]

32. Ellis LA, Collin P, Hurley PJ, Davenport TA, Burns JM, Hickie IB. Young men's attitudes and behaviour in relation to mental health and technology: implications for the development of online mental health services. BMC Psychiatry 2013;13:119 [FREE Full text] [doi: 10.1186/1471-244X-13-119] [Medline: 23601273]

33. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]

34. Johnston L, Dear BF, Gandy M, Fogliati VJ, Kayrouz R, Sheehan J, et al. Exploring the efficacy and acceptability of Internet-delivered cognitive behavioural therapy for young adults with anxiety and depression: an open trial. Aust N Z J Psychiatry 2014 Sep;48(9):819-827. [doi: 10.1177/0004867414527524] [Medline: 24622977]

35. Keane MC, Roeger LS, Allison S, Reed RL. e-Mental health in South Australia: impact of age, gender and region of residence. Aust J Prim Health 2013;19(4):331-335. [doi: 10.1071/PY13027] [Medline: 23945369]

36. Kirkpatrick T, Manoukian L, Dear BF, Johnston L, Titov N. A feasibility open trial of internet-delivered cognitive-behavioural therapy (iCBT) among consumers of a non-governmental mental health organisation with anxiety. PeerJ 2013;1:e210 [FREE Full text] [doi: 10.7717/peerj.210] [Medline: 24349897]

37. Kiropoulos LA, Klein B, Austin DW, Gilson K, Pier C, Mitchell J, et al. Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? J Anxiety Disord 2008 Dec;22(8):1273-1284. [doi: 10.1016/j.janxdis.2008.01.008] [Medline: 18289829]

38. Klein B, Meyer D, Austin DW, Kyrios M. Anxiety online: a virtual clinic: preliminary outcomes following completion of five fully automated treatment programs for anxiety disorders and symptoms. J Med Internet Res 2011;13(4):e89 [FREE Full text] [doi: 10.2196/jmir.1918] [Medline: 22057287]

39. Klein B, Mitchell J, Abbott J, Shandley K, Austin D, Gilson K, et al. A therapist-assisted cognitive behavior therapy internet intervention for posttraumatic stress disorder: pre-, post- and 3-month follow-up results from an open trial. J Anxiety Disord 2010 Aug;24(6):635-644. [doi: 10.1016/j.janxdis.2010.04.005] [Medline: 20447802]

40. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]

41. Morgan AJ, Jorm AF, Mackinnon AJ. Email-based promotion of self-help for subthreshold depression: Mood Memos randomised controlled trial. Br J Psychiatry 2012 May;200(5):412-418 [FREE Full text] [doi: 10.1192/bjp.bp.111.101394] [Medline: 22442102]

42. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]

43. 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]

44. Pier C, Austin DW, Klein B, Mitchell J, Schattner P, Ciechomski L, et al. A controlled trial of internet-based cognitive-behavioural therapy for panic disorder with face-to-face support from a general practitioner or email support from a psychologist. Ment Health Fam Med 2008 Mar;5(1):29-39 [FREE Full text] [Medline: 22477844]

45. Proudfoot J, Parker G, Hadzi PD, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]

46. Robertson L, Smith M, Castle D, Tannenbaum D. Using the Internet to enhance the treatment of depression. Australas Psychiatry 2006 Dec;14(4):413-417. [doi: 10.1111/j.1440-1665.2006.02315.x] [Medline: 17116083]

47. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]

48. Titov N, Andrews G, Kemp A, Robinson E. Characteristics of adults with anxiety or depression treated at an internet clinic: comparison with a national survey and an outpatient clinic. PLoS One 2010;5(5):e10885 [FREE Full text] [doi: 10.1371/journal.pone.0010885] [Medline: 20526371]

49. Wootton BM, Titov N, Dear BF, Spence J, Kemp A. The acceptability of Internet-based treatment and characteristics of an adult sample with obsessive compulsive disorder: an Internet survey. PLoS One 2011;6(6):e20548 [FREE Full text] [doi: 10.1371/journal.pone.0020548] [Medline: 21673987]

50. Zou JB, Dear BF, Titov N, Lorian CN, Johnston L, Spence J, et al. Brief internet-delivered cognitive behavioral therapy for anxiety in older adults: a feasibility trial. J Anxiety Disord 2012 Aug;26(6):650-655. [doi: 10.1016/j.janxdis.2012.04.002] [Medline: 22659078]

51. Burns J, Liacos G, Green F. Advice on Innovative Technologies in e-Mental Health - Briefing Paper for the National Mental Health Commission. Abbotsford, Victoria: Young and Well Cooperative Research Centre; 2014.

52. Harris MG, Diminic S, Reavley N, Baxter A, Pirkis J, Whiteford HA. Males' mental health disadvantage: An estimation of gender-specific changes in service utilisation for mental and substance use disorders in Australia. Aust N Z J Psychiatry 2015 Sep;49(9):821-832. [doi: 10.1177/0004867415577434] [Medline: 25818334]

53. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]

54. Kayrouz R, Dear BF, Johnston L, Gandy M, Fogliati VJ, Sheehan J, et al. A feasibility open trial of guided Internet-delivered cognitive behavioural therapy for anxiety and depression amongst Arab Australians. Internet Interventions 2015 Mar;2(1):32-38. [doi: 10.1016/j.invent.2014.12.001]

55. Shand FL, Ridani R, Tighe J, Christensen H. The effectiveness of a suicide prevention app for indigenous Australian youths: study protocol for a randomized controlled trial. Trials 2013;14:396 [FREE Full text] [doi: 10.1186/1745-6215-14-396] [Medline: 24257410]

56. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample? Electron J Appl Psych 2010;6:27-38 [FREE Full text]

Abbreviations

CBT: cognitive behavioral therapy
EU: study of existing e-mental health service users or self-selected sample
F: facilitating uptake
G: governing mechanisms
iCBT: Internet-based cognitive behavioral therapy
iIPT: Internet-based interpersonal therapy
Mod: Moderate
NHMRC: National Health and Medical Research Council
PU: study was on prospective e-mental health users
SP: study of service providers
T: target demographic

XSL•FO
RenderX

Corrigenda and Addenda

Addition to Acknowledgments: Internet Search and Krokodil in the Russian Federation: An Infoveillance Study

Andrey Zheluk[1], BAppSc, GDip(Hist), GDip(IT), MBA; Casey Quinn[2*], B Comm, MPhil, PhD; Peter Meylakhs[3*], BA(psyc), BA(soc), PhD

[1]Menzies Centre for Health Policy, University of Sydney, Sydney, Australia

[2]PRMA Consulting Ltd, Fleet, Hampshire, United Kingdom

[3]Laboratory for Internet Studies, National Research University Higher School of Economics, St Petersburg, Russian Federation

[*]these authors contributed equally

Corresponding Author:
Andrey Zheluk, BAppSc, GDip(Hist), GDip(IT), MBA
Menzies Centre for Health Policy
University of Sydney
D02 Victor Coppleson Building
Sydney, 2006
Australia
Phone: 61 29036522
Fax: 61 293512818
Email: zheloukandr@gmail.com

Related Article:

Correction of: http://www.jmir.org/2014/9/e212/

The authors of the published article "Internet Search and Krokodil in the Russian Federation: An Infoveillance Study" (J Med Internet Res 2014;16[9]:e212) wish to add additional information to the "Acknowledgments" section, so it now reads as follows:

> *The work of Peter Meylakhs was funded by the Basic Research Program of the National Research University Higher School of Economics, Russia. The authors would also like to thank Ms Svetlana Chernova in Ukraine for her assistance with data collection and coding, Associate Professor James Gillespie at the Menzies Centre for Health Policy, University of Sydney for his advice and support, and Ms Anya Sarang from the Andrey Rylkov Foundation in Moscow for facilitating contacts within Russia.*

The first sentence was unintentionally omitted in the original submission. The publication on the JMIR site was amended accordingly, and a new version was submitted to PubMed Central.

XSL•FO

RenderX