Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 18 (2016), Issue 8 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Letter to the Editor

Original Paper

The Effect of Guided Web-Based Cognitive Behavioral Therapy on Patients With Depressive Symptoms and Heart Failure: A Pilot Randomized Controlled Trial

Johan Gustav Lundgren[1*], MSc Nursing; Örjan Dahlström[2*], PhD; Gerhard Andersson[3*], PhD; Tiny Jaarsma[1*], PhD; Anita Kärner Köhler[1*], PhD; Peter Johansson[4*], PhD

[1]Division of Nursing Science, Department of Social and Welfare Studies, Linköping University, Norrköping, Sweden

[2]Disability Research Division, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[3]Psychology, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[4]Cardiovascular medicine, cardiology, Department of Medical and Health Sciences. Department of Social and Welfare Studies, Linköping University, Linköping, Sweden

[*]all authors contributed equally

Corresponding Author:
Johan Gustav Lundgren, MSc Nursing
Division of Nursing Science
Department of Social and Welfare Studies
Linköping University
Department of Social and Welfare Studies
Linköping University
Norrköping, SE 601 74
Sweden
Phone: 46 011363592
Fax: 46 011125448
Email: johan.lundgren@liu.se

Abstract

Background: Depressive symptoms, and the associated coexistence of symptoms of anxiety and decreased quality of life (QoL), are common in patients with heart failure (HF). However, treatment strategies for depressive symptoms in patients with HF still remain to be established. Internet-based cognitive behavioral therapy (ICBT), as guided self-help CBT programs, has shown good effects in the treatment of depression. Until now, ICBT has not been evaluated in patients with HF with depressive symptoms.

Objective: The aims of this study were to (1) evaluate the effect of a 9-week guided ICBT program on depressive symptoms in patients with HF; (2) investigate the effect of the ICBT program on cardiac anxiety and QoL; and (3) assess factors associated with the change in depressive symptoms.

Methods: Fifty participants were randomized into 2 treatment arms: ICBT or a Web-based moderated discussion forum (DF). The Patient Health Questionnaire-9 was used to measure depressive symptoms, the Cardiac Anxiety Questionnaire (CAQ) was used to measure cardiac-related anxiety, and the Minnesota Living with Heart Failure questionnaire was used to measure QoL. Data were collected at baseline and at follow-up at the end of the 9-week intervention. Intention-to-treat analysis was used, and missing data were imputed by the Expectation-Maximization method. Between-group differences were determined by analysis of covariance with control for baseline score and regression to the mean.

Results: No significant difference in depressive symptoms between the ICBT and the DF group at the follow-up was found, [F(1,47)=1.63, *P*=.21] and Cohen´s d=0.26. Secondary within-group analysis of depressive symptoms showed that such symptoms decreased significantly in the ICBT group from baseline to the follow-up (baseline M=10.8, standard deviation [SD]=5.7 vs follow-up M=8.6, SD=4.6, t(24)=2.6, *P*=.02, Cohen´s d=0.43), whereas in the DF group, there was no significant change (baseline M=10.6, SD=5.0, vs follow-up M=9.8, SD=4.3, t(24)=0.93, *P*=.36. Cohen´s d=0.18). With regard to CAQ and QoL no significant differences were found between the groups (CAQ [d(1,47)=0.5, *P*=.48] and QoL [F(1,47)=2.87, *P*=.09]). In the ICBT group in the CAQ subscale of fear, a significant within-group decrease was shown (baseline M=1.55 vs follow-up M=1.35, *P*=.04). In the ICBT group, the number of logins to the Web portal correlated significantly with improvement in depressive symptoms (*P*=.02),

whereas higher age (*P*=.01) and male sex (*P*=.048) were associated with less change in depressive symptoms. This study is underpowered because of difficulties in the recruitment of patients.

Conclusions: Guided ICBT adapted for persons with HF and depressive symptoms was not statistically superior to participation in a Web-based DF. However, within the ICBT group, a statically significant improvement of depressive symptoms was detected.

ClinicalTrial: Clinicaltrials.gov NCT01681771; https://clinicaltrials.gov/ct2/show/NCT01681771 (Archived by WebCite at http://www.webcitation.org/6ikzbcuLN)

(J Med Internet Res 2016;18(8):e194) doi:10.2196/jmir.5556

KEYWORDS

heart failure; depression; Internet-based cognitive behavioral therapy; cognitive behavioral therapy; Internet; eHealth

Introduction

Depressive symptoms are common in patients with heart failure (HF), affecting about 20%-40% of the HF population [1-3]. They lead to higher morbidity and mortality and diminish self-care and health-related quality of life (QoL) [3]. However, treatment strategies for depressive symptoms in patients with HF still remain to be established [3,4].

HF has an unpredictable trajectory with disturbing and limiting symptoms that frequently change, leading to a shift between good and bad days [5] and with a constant risk of hospitalization or death [6,7]. Patients with HF may therefore be prone to developing negative thoughts, rumination, and feelings of hopelessness about loss of health and independence and an uncertain future [8], and this can lead to the development of depression.[9,10]. A vast majority of patients with HF and depressive symptoms also have symptoms of anxiety [11]. Depressive symptoms have a strong negative impact on QoL in patients with HF [3,4]. Because anxiety and depression are closely related, an intervention focusing on decreasing depression may also improve symptoms of anxiety and increase QoL. In general, depressive symptoms can be treated, either by psychotherapy or by pharmacology [12]. However, the impact of pharmacological treatment of depression in HF is not clear [4] and may be complicated due to an already complex pharmacological treatment regime [2]. Furthermore, patients with heart disease seem to prefer talking therapies such as cognitive behavioral therapy (CBT) over pharmacological treatment [8].

In CBT, patients become active participants in their treatment and perform tasks to become aware of and to modify negative thoughts and unhelpful behaviors. By developing skills to cope with these negative thoughts and behaviors, CBT also contributes to a decrease of negative emotions [13]. In HF patients, Freedland et al [14] demonstrated that undergoing CBT for 6 months decreased depression and Gary et al [15] found CBT to be beneficial, especially when combined with physical exercise. In these studies, CBT was provided face-to-face. Due to the lack of CBT therapists to deliver such face-to-face CBT, combined with the large number of HF patients with depression, most HF patients with depressive symptoms might not get access to CBT. Internet-based cognitive behavioral therapy (ICBT) may be an alternative to face-to-face CBT. ICBT has been shown to be a good and time-efficient method for the treatment of depressive symptoms and also effective when delivered by professionals other than psychotherapists. ICBT might, therefore, be considered as an attractive treatment strategy for depression in HF [16], but this is an area still waiting to be explored. Furthermore, since the frequency of participation in CBT treatment [17], level of depressive symptoms pre-intervention [18], age [19], sex, and New York Heart Association Class (NYHA class) [2] may impact changes in depressive symptoms, it is important to investigate these factors in intervention programs.

Recently our group showed that an ICBT program designed for HF patients was feasible [20], but the effect of ICBT on depressive symptoms in patients with HF has, to our knowledge, not been tested in a randomized controlled trial. The primary aim of this study was therefore to evaluate the short-term effect of ICBT on depressive symptoms in patients with HF. A second aim was to investigate the effect of the ICBT program on cardiac anxiety and QoL (secondary outcomes). A third aim was to assess these factors' associations with the change in depressive symptoms.

Methods

Design

An open label, randomized control design was used.

Recruitment Procedure and Inclusion

To recruit participants, an information letter was sent to all patients who had an outpatient appointment or who had been admitted to hospital with the main diagnosis of HF during 2013 and 2014 in 4 hospitals in the southeast of Sweden (Figure 1). Inclusion criteria were at least mild depressive symptoms (the Patient Health Questionnaire-9 (PHQ-9)≥5), regular access to a computer with an Internet connection, access to a cellphone, being motivated to participate in treatment of depressive symptoms, and being aged older than 18 years. Exclusion criteria were suffering from other severe disease or illness that hindered participation in the study, admission to hospital during the last month due to HF, other treatment planned during the intervention that had been assessed as likely to hinder participation (such as surgery or planned inpatient treatment), severe level of depressive symptoms assessed as needing inpatient treatment, and high level of suicide risk or other psychiatric disorder assessed as making the intervention unsuitable. Patients who had perceived themselves as depressed or feeling down during (at least) the last 2 weeks and felt motivated and ready to participate in the study were invited to

register their interest and perform a Web-based screening on the study website. Computer/Internet literacy was not a criterion for inclusion or exclusion. However, recipients interested in participation had to register on the study website and complete a Web-based screening form including an assessment of depressive symptoms by means of the PHQ-9, self-reported use of medication, comorbidities, NYHA class, and demographics.

A total of 64 patients completed the Web-based screening form and 58 were found to be possible candidates for inclusion. Candidates were contacted by telephone to check any uncertainties in the screening forms and to prevent multiple registrations. A structured phone assessment using the Mini International Neuropsychiatric Interview Swedish revised version 5.0.0 [21] was conducted to detect symptoms of other psychiatric health problems or suicidality that could hinder participation in the intervention, as shown in Figure 1. Two candidates were excluded because of suspicion of other mental illness and 3 candidates were excluded because they reported no depressive symptoms during the phone interview despite a screening PHQ-9 >4. During the phone call, the participant received detailed information about the study procedures. The 50 patients remaining after the telephone interview underwent baseline assessment and were randomized to either the ICBT group or the discussion forum group (DF group). Randomization was performed by a person blinded to screening and baseline data using Stata v.13 proc Ralloc with a block size of two. All participants gave written informed consent. No financial compensation was given to the participants. The regional ethical review board in Linköping, Sweden approved the study (dnr 2011/166-31). The study is registered at clinicaltrials.gov (NCT01681771).

Figure 1. Flow diagram of participants. DF, discussion forum; HF, heart failure; ICBT, Internet-based cognitive behavioral therapy; PHQ-9, Patient Health Questionnaire-9.

Intervention Procedure

Each participant received a password and a user name. Login to the Web portal (to access the treatment program, feedback, secure email, and assessment forms) required a 2-factor authentication system (requiring both a user name and password login and a single use code sent to a preregistered mobile phone) to protect sensitive information. If technical problems occurred, both the therapist and the participants could get support from a computer technician.

The ICBT program has been described in detail elsewhere [20]. Compared with the ICBT program tested in the proof-of-concept study and based on the findings that many of the participants in proof-of-concept study reported that the intervention was quite demanding in some parts, a short nonmandatory relaxation exercise was added in module 3. No changes were made to the program during the trial. To summarize, the ICBT program consisted of 7 consecutive modules that were worked with over 9 weeks (Table 1). Each module contained texts to be read and assignments to be completed by the participants. Written

feedback was provided on each assignment. Participants could ask questions about the feedback or the content of the module using the secure email system. A mental health specialist nurse with experience of HF care provided feedback and answers within 24 hours during weekdays. The mental health nurse was supervised by a clinical psychologist and could consult a cardiologist and nurses specializing in HF care if needed. Participants who did not complete modules were reminded by personalized manually written emails, a maximum of 3 reminders were sent during a consecutive period of 2 weeks. Screenshots from different parts of the treatment platform are available (Multimedia Appendix 1), and a more comprehensive tour of the platform is available (Multimedia Appendix 2). Patients who were randomized to the DF group participated in a moderated discussion Web-based forum where new discussion topics were presented each week over a 9-week period. The topic was introduced without any extended background in a presentation consisting of statements and questions (Table 1). The discussion was performed in writing. Participants made their posts in discussion threads for each topic. To minimize waiting time, the participants in the DF group were allocated to 2 groups (n=12 and n=13) based on the dates they were enrolled in the study. All participants in the DF group were offered ICBT after the completion of the study.

Table 1. Overview of the guided Internet-based cognitive behavioral therapy program and the discussion forum.

Module	ICBT[a] (content and CBT[b] component)	DF[c] (topic/question for discussion)	Week
1	Introduction (values and goals)	HF[d]: what do you know about HF?	1
2	Living with heart failure (psychoeducation)	The effect of HF on everyday life: do you have any tips you would like to share about handling HF?	2
3	Depression/depressive symptoms and heart failure (psychoeducation) Nonmandatory relaxation exercise	Self-care: do you have any methods that make self-care easier that you can share with the others in the DF?	3
4	Behavior activation 1: to enable change	Physical activity: have you been recommended physical activity? What is good or bad about physical activity when suffering from HF?	4
		Health care contacts: do you prepare yourself before health care appointments? Do you have any tips you can share with the others?	5
5	Behavior activation 2: to implement change	Health literacy: if you do not get answers from the health care system, do you look for information in other ways? Do you have any tips on where one can find information about health and diseases such as HF and depression?	6
		The effect of HF and depression on significant others: do you think that your health affects your relationships with others? If so, in what ways?	7
6	Problem solving: a tool for dealing with problems	The effect of HF and depression on significant others: how do you handle situations where your health affects other? Do you have any good examples of how to handle this that you can share?	8
7	Consummation	Summarizing: are there questions/topics that have not been discussed that you would like to address? How did you perceive the DF?	9

[a]ICBT: Internet-based cognitive behavioral therapy.

[b]CBT:cognitive behavioral therapy.

[c]DF:discussion forum.

[d]HF:heart failure.

Measurements

Self-assessed data were collected on the Web at baseline (before the start of the intervention) and after the end of week 9 in the intervention. The data collection system for the follow-up was accessible for the participants from the 63rd day after the start of the intervention and could be completed during a 3-week period. All data except activity in the program was self-reported. Participants who did not complete outcome measures were reminded to do so by email up to 3 times.

Depressive Symptoms (Primary Outcome Measurement)

Depressive symptoms were measured with the self-administered PHQ-9 [22]. The PHQ-9 is a 9-item instrument for measurement of depressive symptoms during the previous 2 weeks. Each item is answered on a 4-grade scale where zero means that the item does not affect the person, and scores 1 to 3 indicate that the item affects the person for periods ranging from several days to almost every day. The answers are summed to a total sum score in the range 0-27, with higher numbers representing a

higher level of depressive symptoms. Proposed cutoff values are 0-4 for no depressive symptom, 5-9 for mild depression, 10-14 for moderate depression, 15-19 for moderately severe depression, and 20-27 for severe depression [23]. PHQ-9 has been tested for reliability and validity in patients with HF [24]. The Web-based version of PHQ-9 has demonstrated good interformat reliability [25]. Cronbach's alpha of the PHQ-9 in this study was .81 (baseline) and .82 (follow-up).

Cardiac Anxiety (Secondary Outcome Measurement)

The Cardiac Anxiety Questionnaire (CAQ) [26] was used to measure cardiac-related anxiety. CAQ is an 18-item self-rating instrument. Item scores range from 0 (never) to 4 (always). The total sum and mean total (range from 0 to 4) can be calculated for CAQ. The CAQ consists of 3 subscales: fear, avoidance, and heart-focused attention. CAQ has demonstrated good psychometric properties [26]. Cronbach's alpha of the CAQ in this study was total scale .87 (baseline) and .85 (follow-up); subscale of fear .83 and .80; subscale of avoidance .89 and .88; subscale of heart-focused attention .69 and .70.

QoL (Secondary Outcome Measurement)

QoL was measured with the disease-specific instrument Minnesota Living with Heart Failure questionnaire (MLHF) [27]. MLHF is a 21-item self-rating instrument. Each item is scored on a 6-point Likert scale (no, 0 to very much, 5). The total score is in the range 0-105, and a lower score indicates better QoL. The MLHF can be divided into physical and emotional factors. The reliability of MLHF has been reported as good [28,29]. A change of 5 points has been suggested as clinically important [30], Cronbach's alpha of the MLHF in this study was total score .93 (baseline) and .93 (follow-up); physical .91 and .90; emotional .93 and .92.

Activity in the ICBT Program

Activity in the program was calculated by the number of modules that the participants worked with (ie, the module had been assigned to the participant and the participant had done some activity related to the module, eg, handed in an assignment or posted messages regarding the module to the feedback provider; ICBT group only) as well as the number of logins to the Web portal during the 9-week period (both groups). Data concerning activity was aggregated from the Web portals log.

Statistical Methods and Power Analysis

Analysis of participants' characteristics was performed with descriptive statistics (mean, standard deviation, percent, and frequencies). For continuous variables, assumptions of normality were checked and primary outcome measurements were found suitable for parametric analysis. Analysis of covariance (ANCOVA) adjusting for baseline scores and regression to the mean [31] was used for comparison between groups (ie, ICBT vs DF). Paired samples *t* tests were used for within-group comparisons. Effect size was calculated with Cohen´s *d*. A small effect is considered to be between 0.2 and .0.5, a medium effect is considered between 0.5 and 0.8, and a value above 0.8 is considered to be a large effect. Pearson`s *r* or Kendall´s tau-b were used to analyze associations with change in the level of depressive symptoms. A chi-square test was used for nominal data, and if the expected number of observations was less than 5, Fisher´s exact test was used. Subtracting the baseline sum from the follow-up sum gave a figure indicating the change in level of depressive symptoms. A negative value meant a decrease of depressive symptoms, whereas a positive value meant an increase of depressive symptoms at the follow-up compared with baseline. All analyses were performed according to the intention-to-treat principle, regardless of actual completion of the ICBT program or DF.

A total of 18% (n=9) of the participants had missing data at the follow-up measurement. Missing values analysis was performed and data missing completely at random was assumed because there were no significant differences between background variables for participants with complete data versus incomplete data, and Little´s test for missing completely at random was not significant (χ^2(111, N=50)=82.07, *P*=.98). Missing values were imputed using the Expectation-Maximization (EM) method. Based on observed values, EM imputes missing values based on maximum likelihood estimates in an iterative process [32]. Subgroup analysis was performed on participants with complete data. Power analysis showed that a total of 104 participants were needed (effect size=0.5, alpha=.05 (*Z*=1.96), power 0.80 (*Z* −0.84). Statistical analysis was performed using IBM SPSS, version 23 and Microsoft Excel 2013. *P* value <.05 was considered as significant.

Results

The characteristics of the participants are presented in (Table 2). Most (n=29, 59%) of the participants were men, and the mean age was 63 years (range 23-80). Participants in the DF group reported significantly more prescription of diuretics (χ^2(1, N = 50)=4.67, *P*=.03) and sleep medications (χ^2(1, N=50)=3.95, *P*=.047). Participants who did not complete the follow-up assessment (n=9) did not significantly differ at baseline in level of depression (PHQ-9 *t* (48)=1.89, *P*=.07), cardiac-related anxiety (CAQ *t* (48)=-0.55, *P*=.60), or QoL (MLHF *t* (48)=0.69, *P*=.50) from those who completed the assessment.

Table 2. Participants' characteristics.

		Total (n=50)	ICBT group (n=25)	DF group (n=25)
Demographics				
	Age, M (SD)	62.9 (12.8)	63.6 (13.9)	62.3 (11.7)
	Men, n (%)	29 (59)	15 (60)	14 (58)
	Cohabitation[a], n (%)	37 (76)	19 (76)	18 (75)
	Level of depression at screening PHQ-9, M (SD)	11.5 (4.8)	11.8 (4.4)	11.2 (5.2)
HF[b]symptoms and treatment				
	NYHA[c] class, n (%)			
	I	11 (22)	8 (32)	3 (12)
	II	20 (40)	12 (48)	8 (32)
	III	18 (36)	5 (20)	13 (52)
	IV	1 (2)	0 (0)	1 (4)
	Dyspnea[d], n (%)	48 (96)	24 (96)	23 (92)
	Fatigue[d], n (%)	49 (98)	25 (100)	24 (96)
	Swollen legs or feet[d], n (%)	23 (46)	14 (56)	12 (48)
	Time with HF>6 month/<6 month, n (%)	45/5 (88/10)	22/3 (88/12)	23/2 (92/8)
	Previously hospitalized due to HF, n (%)	36 (72)	17 (68)	19 (76)
	Beta blocker, n (%)	44 (88)	22 (88)	22 (88)
	ACE-I[e]/ARB[f], n (%)	47 (94)	22 (88)	25 (100)
	Diuretics, n (%)	34 (68)	14 (56)	20 (80)[g]
Comorbidities, n (%)				
	Ischemic heart disease	18 (36)	8 (32)	10 (40)
	Hypertension	26 (52)	11 (44)	15 (60)
	Arrhythmia	26 (52)	14 (56)	12 (48)
	Diabetes	7 (14)	2 (8)	5 (20)
	Pulmonary disease	6 (12)	1 (4)	5 (20)
	Stroke or TIA	11 (22)	4 (16)	7 (28)
	Kidney disease	1 (2)	1 (4)	0 (0)
	Cancer	5 (10)	3 (12)	2 (8)
	Other psychiatric disorder[h]	2 (4)	2 (8)	0 (0)
Pharmacological antidepressive, anxiolytic, or sleep medication				
	Antidepressives	9 (18)	9 (12)	6 (24)
	Anxiolytics	2 (4)	1 (4)	1 (4)
	Sleep medication	14 (28)	4 (16)	10 (40)[g]

[a]Cohabitation includes participants that live with someone in a long-term relationship (including married). Not living with partner includes participants who were divorced, with partner deceased or living alone.

[b]HF, heart failure.

[c]NYHA, New York Heart Association.

[d]Symptoms reported to affect the participant very severely to little have been collapsed and reported as presence of symptoms.

[e]ACE-I, angiotensinogen-converting enzyme inhibitor.

[f]ARB, angiotensin receptor blocker.

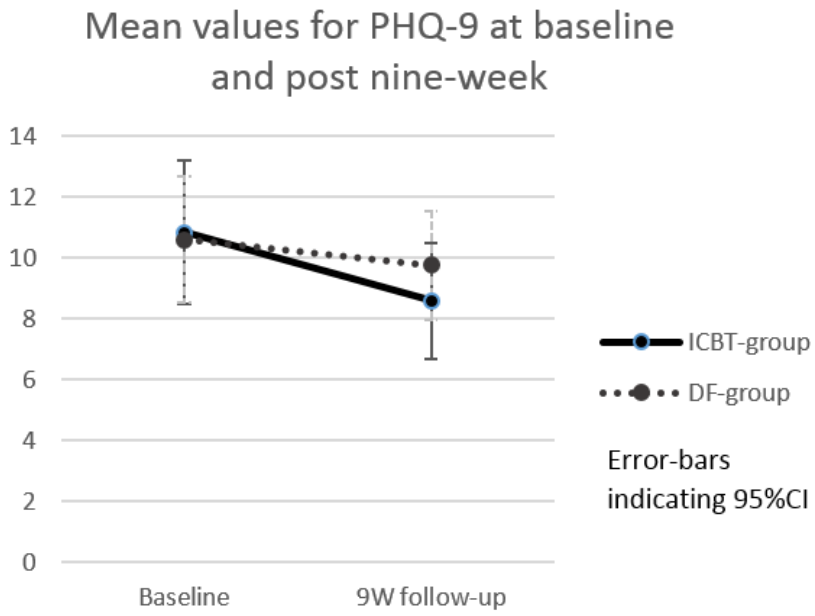
[g]Significant difference between CBT and discussion groups (P<.05).

[h]Self-reported: anxiety disorder (n=1) and drug dependence (n=1).

Primary Outcome: Level of Depressive Symptoms

In the primary ANCOVA analysis, there was no significant difference in depressive symptoms between the ICBT and the DF group at the follow-up [*F* (1,47)=1.63, *P*=.21] Cohen´s *d*= 0.26. Secondary within-group analysis showed that depressive symptoms in patients in the ICBT group decreased significantly from baseline to the follow-up (Figure 2) (baseline M = 10.8, SD=5.7 vs follow-up M=8.6, SD=4.6, *t* (24)=2.6, *P*=.02. Cohen´s *d*=0.43). In the patients in the DF group, a small nonsignificant change in depressive symptoms was found (baseline M = 10.6, SD=5.0, vs follow-up M = 9.8, SD=4.3, *t* (24)=0.93, *P*=.36. Cohen´s *d*= 0.18).

Figure 2. Mean values for PHQ-9 at baseline and follow-up in the 2 groups (n=25 ICBT and n=25 DF). DF, discussion forum; ICBT, Internet-based cognitive behavioral therapy; PHQ-9, Patient Health Questionnaire-9.

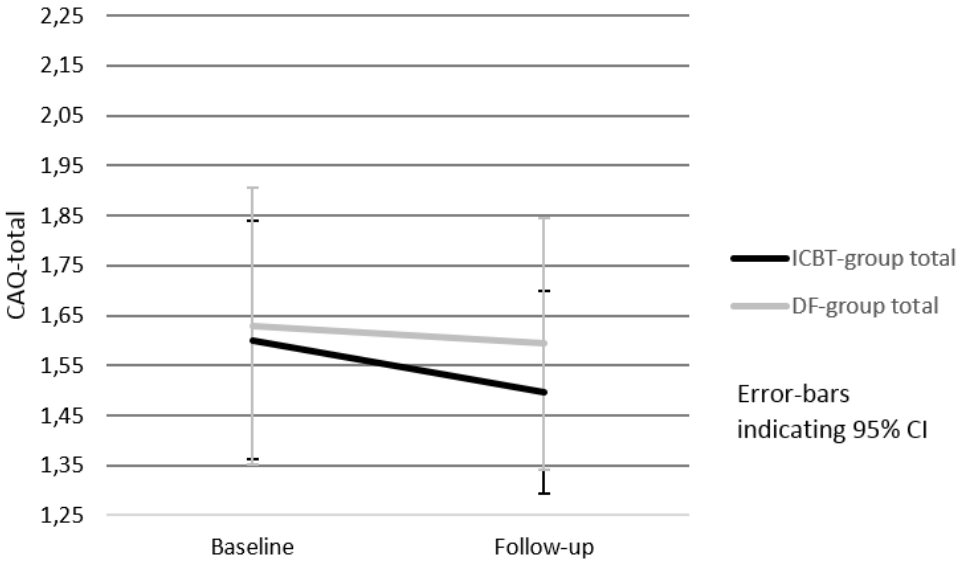


Secondary Outcomes: Cardiac-Related Anxiety and QoL

Between group comparison (ie, ANCOVA, ICBT vs DF) showed no statistically significant difference in the CAQ total score [F(1,47)=0.51, *P*=.48], Cohen´s *d*= 0.18, subscale of fear [*F* (1, 47)=1.57, *P*=.22], Cohen´s *d*=0.43, subscale of avoidance [F(1,47)=0.11, *P*=.74], Cohen´s *d*= 0.17, and subscale of heart-focused attention [F(1,47)=0.39, *P*=.54], Cohen´s *d*= 0.08. In the secondary within-group analysis, the ICBT group showed a decrease in the total CAQ score and in the subscale of fear. The decrease in the subscale of fear was statistically significant (baseline M=1.55, SD=0.73 vs follow-up M=1.35, SD=0.60, *t* (24)=2.18 *P*=.04. Cohen´s *d*=0.30), see Figure 3, but the decrease in the total score was not significant (baseline M=1.60, SD=0.58 vs follow-up M=1.49, SD=0.49, *t* (24)=1.25, *P*=.22. Cohen´s *d*= 0.31). In the subscales of avoidance and heart-focused attention, no significant changes were found (Multimedia Appendix 3). In the DF group, no significant changes in any of the CAQ scales were found (total *P*=.86, fear *P*=.92, avoidance *P*=.82, heart-focused attention *P*=.83).

Between-group analysis (ie, ANCOVA, ICBT vs DF) of MLHF revealed no significant differences for the total score [*F* (1,47)=2.87, *P*=.09], Cohen´s *d*= 0.51, the physical factor [*F* (1,47)=3.35, *P*=.07], Cohen´s *d*= 0.56, and the emotional factor [F(1,47)=0.20, *P*=.66], Cohen´s *d*= 0.37. The change in scores from baseline to the follow-up for the total score and the physical and emotional factors in the MLHF is shown in Multimedia Appendix 4. In the ICBT group, the mean total score decreased by 6.0 points, by 2.4 points in the physical factor and by 0.3 points in the emotional factor. None of the differences were statistically significant; total score baseline M=41.8, SD=20.5 vs follow-up M=35.8, SD=15.3, *t* (24)=1.79, *P*=.09, Cohen´s *d*= 0.33; the physical factor, baseline M=17.5, SD=8.7 vs follow-up M = 15.1, SD=7.5, *t* (24)=1.62, *P*=.12, Cohen´s *d*= 0.28; baseline M=10.8, SD = 7.2 vs follow-up M = 10.5, SD = 6.4, *t* (24)=0.31, *P*=.76, Cohen´s *d*= 0.05. In the DF group, the mean total score decreased by 1.9 points, by 0.2 points in the physical factor, and by 0.8 points in the emotional factor. None of the differences were significant; baseline M=47.1, SD=24.0 vs follow-up M=45.3 SD=21.3, *t* (24)=0.64, *P*=.53, Cohen´s *d*= 0.08; the physical factor baseline M=20.0, SD=10.6 vs follow-up M = 19.8, SD=8.9, *t* (24)=0.14, *P*=.89, Cohen´s *d*= 0.02; the emotional factor baseline M=13.7, SD = 6.3 vs follow-up M = 12.9, SD = 5.6, *t* (24)=0.91, *P*=.37, Cohen´s *d*= 0.14.

Figure 3. Change in cardiac anxiety—mean values for subscale of fear in the 2 groups (n=25 ICBT and n=25 DF). ANCOVA: [F(1,47)=1.57, *P*=.22], total scale range for CAQ (mean values): 0-4. ANCOVA, analysis of covariance; CAQ, Cardiac Anxiety Questionnaire; DF, discussion forum; ICBT, Internet-based cognitive behavioral therapy.



Relationship Between Factors and Changes in Depressive Symptoms

The median number of modules performed in the ICBT group was 4. Six (24%) of the participants in the ICBT group had worked with all 7 modules and 15 (60%) had worked with at least 4 modules (ie, 57% of the program). There was no significant relationship between the number of modules completed and the change in depressive symptoms (τ_b=.13, *P*=.46). In the ICBT group, the number of logins to the Web portal was significantly related to the change in depressive symptoms (ie, improvement in depressive symptoms) (*r*=−.50, *P*=.02) a similar relationship was found in the DF group although not as strong and not significant (*r*=−.32, *P*=.17). Age had a negative correlation with number of logins in the ICBT group (*r*=−.67, *P*<.001). In the DF group, this correlation was not so strong and not significant (*r*= −.24, *P*=.25).

The level of depressive symptoms at screening was not associated with the level of depressive symptoms at the follow-up. A separate analysis of participants with PHQ-9 ≥10 at screening (ICBT n=18, DF n=15) showed no significant difference between groups in the level of depressive symptoms at the follow-up [*F* (1,29)=1.30, *P*=.26] nor of participants with PHQ-9 ≤15 (ICBT n=19, DF n=19) [*F* (1,34)=0.82, *P*=.37]. Higher age correlated significantly with less change in depressive symptoms (*r*=.54, *P*=.01) in the ICBT group and women (n=8, complete cases) (M = −3.4, SD=4.6) had a significantly higher mean change in depressive symptoms compared with men (n=13, complete cases) (M=−0.08, SD=2.6) *t* (19)=2.12, *P*=.048. Cohen´s *d*= 0.89. The severity of HF as assessed by NYHA class was not associated with a change in depressive symptoms.

Discussion

Principal Findings

To our knowledge, this is the first study evaluating an ICBT program aimed at decreasing depressive symptoms in patients with HF. The recruitment of participants was more difficult than expected. Based on a prevalence of depressive symptoms among HF patients at approximately 20% [1], we expected about 571 of the 2852 contacted the patients have a significant level of depressive symptoms. However, only 80 patients registered as interested and 50 were found to be eligible for inclusion. Other studies of CBT in HF also appear to have difficulty in achieving sample sizes corresponding to power analysis [14,15,33]. Due to practical and economic reasons, we chose to end the study without achieving the targeted sample size. Other recruitment strategies may be more effective. However, there may also be a structural problem in reaching out to patients with HF and depressive symptoms. This is because depressive symptoms can reduce patients' decision-making process due to ambivalence and impaired cognitive functioning [34]. Our primary analysis could not detect any significant difference in the reduction of depressive symptoms between the ICBT and DF groups; this may be explained by the slight reduction of depressive symptoms in the DF group. Dekker et al [35] reported similar results where depressive symptoms decreased in HF patients who received a brief CBT intervention or received standard care before discharge from hospital. The within-group analysis of depressive symptoms showed a significant decrease in the ICBT group but not in the DF group. These findings are in line with previous studies reporting that CBT for depression is significantly superior compared with no treatment, but only nominally superior compared with psychological placebo [33,36] such as the DF in our study. Studies specifically on HF patients with depressive symptoms also suggest that an active intervention, such as CBT or placebo with attention control such as DF can

reduce depressive symptoms to a greater extent than standard care [14,15,37]. Designing a new ICBT program is a complex process [38]. Our program is to our knowledge one of the first ICBT programs for patients with HF and depressive symptoms. Furthermore, HF patients are often older compared with other patient groups treated with ICBT. Therefore, there is a need for more research to gain knowledge on how to design or redesign ICBT programs for the HF population. A possible future approach to achieve optimal results in treatment of depressive symptoms in HF could be a stepped care model [39] where patients could start with a type of DF or physical exercise, and if the depressive symptoms did not improve, ICBT could be added. However, such models have to be evaluated in further studies.

The secondary outcomes of CAQ and MLHF did not show any significant difference between the ICBT group and the DF group. However, in the ICBT group, a lower cardiac-related anxiety in the subscale of fear and an increased QoL was found in the within-group analysis. The increase in QoL of 6 points in the total MLHF score was not statistically significant; however, a change of 5 points in MLHF has been proposed as a measure of a clinically important change [30].

It is common for depressive symptoms in patients with HF to coexist with anxiety [11]. Anxiety CBT treatment can have a better effect on depressive symptoms than CBT for depression in patients with HF [40]. Thus, anxiety and worries may need special attention when designing or redesigning ICBT interventions for patients with HF and depressive symptoms.

We also found that the age and sex of the participant may need to be taken into account. Higher age and male sex correlated with less change in depressive symptoms in the ICBT group. Older people to some extent seem to benefit from CBT [20,41]; however, the evidence of benefit for them is greater in problem-solving therapy [42,43]. Our ICBT program relied to a large extent on behavioral activation and to a lesser extent on problem-solving therapy [20]. Most patients with HF are older and vulnerable, which raises an important question of whether problem-solving therapy should be used to a greater extent in future studies, as proposed by Alexopoulos et al [44]. Women showed more positive effects compared with men in the ICBT group. This is in line with the results of other ICBT studies [45,46]. However, sex as a predictor of outcome of CBT delivered by other modalities has shown inconsistent results; men have been reported to have a better response in telephone and face-to-face CBT [47]. More research is needed to determine whether CBT and ICBT should be adapted to the different sexes.

The cornerstone of CBT is to encourage participants and involve them in the treatment [48]. We found that activity in the program, as indicated by the number of logins, correlated significantly with a change in depressive symptoms. This suggests that helping participants in ICBT programs to be active is important for a positive outcome. In our program, participants who did not follow the pace of the program were reminded to do so by email. In contrast, others have used more intensive reminder techniques of both text messages and phone calls, and thereby achieved a higher adherence to ICBT treatment [49,50]. This indicates that more direct reminders can motivate less active participants. The number of modules performed did not correlate with level of depression at follow-up. The reasons for this have not been investigated in this study. However, during the study, we experienced that some patients early in the program reported that they felt better and therefore did not proceed with the next module. This may have affected the result negatively, thus one may speculate that patients might have improved even more if they had completed the program. Our study has shown that ICBT for HF patients with depressive symptoms is feasible. However, further research is needed to develop effective ICBT programs for depressive symptoms in HF patients. Furthermore, this study only evaluated the effect directly after the intervention; thus, the long-term effects of ICBT on depression in patients with HF need to be evaluated in future studies.

The generalizability of the results is limited for several reasons. One major limitation of our study is that it is underpowered. Post hoc power calculation for this study showed a power of 16% and a need for 462 patients to be included to achieve a statistically significant result as regards depressive symptoms. A reason for the need for such a large sample could be floor effects, because patients with mild depression were also included (PHQ-9≥5). A reason for including these patients is that even mild depression has a strong negative impact on QoL in HF patients [51]. The mean age in the study sample was lower compared with the HF population in the community (approximately 78 years) [52]; thus, it is unclear how the program works in older HF patients. Furthermore, the study could only include patients with access to a computer and the Internet, and therefore, the results cannot be generalized to patients without such access. On the other hand, in Sweden, 42% (n=355 793) of the population aged older than 75 years use computers and the Internet at home. This figure is expected to rise to approximately 80% in the coming 5-10 years [53]. There is a potential limitation in relying on self-reported data. However, all patients were identified by diagnostic codes for HF from electronic hospital records and were contacted by telephone to verify their reported medication and comorbidities. There were twice as many participants taking an antidepressant medication in the DF group compared with the ICBT group. The reason for this not being statistically significant is probably due to low power in the study. Subgroup analysis in this study has to be interpreted with caution due to the limited sample size. Missing data were imputed using the EM method. Although Littles´ test indicated that the missing data in our study appear to be missing at random, this can never be certain, and missing data and imputation can carry a risk of bias. Initially, all analysis was performed on both nonimputed data and imputed data, and we found no significant differences between the analysis with the exception of the ANCOVA on the physical factor of QoL, which presented a significant difference between the groups on nonimputed data but not on imputed data. There were more patients reporting use of antidepressive medication (nonsignificant) and sleep medication (significant) in the DF group. We do not think that this had any major impact on our result. For example, pharmacological antidepressive treatment in HF patients has shown a poor effect on depression [37]. The lack of power of this study may also have resulted in differences not being detected between participants completing the

intervention and dropouts. Nevertheless, we think that our study is important since interventions for patients with HF and depressive symptoms are not widely studied. To our knowledge, our study is one of the first to investigate the effect of ICBT in HF and depressive symptoms. Development of new interventions is an iterative process [38] and although clear-cut results may be desirable, the novelty of the research area makes this unlikely. With our study's limited sample size and its recruitment difficulties, the results may best serve as a foundation for further research rather than as clinical recommendations.

Conclusion

Guided ICBT adapted for persons with HF and depressive symptoms was not statistically superior to participation in a Web-based DF. However, within the ICBT group, a statically significant improvement of depressive symptoms was detected.

Acknowledgments

The authors would like to thank the computer technicians George Vleascu and Alexander Alasjö for technical support during the intervention.

This study was funded by grants from the Swedish Heart and Lung Association (grant number E087/13 and E08/14), the Medical Research Council of Southeast Sweden (grant numbers FORSS-374721 and FORSS-470121) and the Region Östergötland (grant numbers LIO-355611, LIO-374831, LIO-443711 and LIO-470271). The funding sources had no influence on the design, procedure, analysis, or interpretation of the results in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the treatment platform.

[PPTX File, 299KB - jmir_v18i8e194_app1.pptx]

Multimedia Appendix 2

Video of treatment platform.

[PPTX File, 47MB - jmir_v18i8e194_app2.pptx]

Multimedia Appendix 3

Change in cardiac anxiety—mean values for total score and subscales of avoidance and heart-focused attention in the 2 groups (n=25 ICBT and n=25 DF).

[PPTX File, 49KB - jmir_v18i8e194_app3.pptx]

Multimedia Appendix 4

Change in Minnesota Living with Heart Failure (MLHF)—total score and factors in the 2 groups (n=25 ICBT and n=25 DF).

[PPTX File, 50KB - jmir_v18i8e194_app4.pptx]

Multimedia Appendix 5

CONSORT-EHEALTH Checklist V1.6.

[PDF File (Adobe PDF File), 1MB - jmir_v18i8e194_app5.pdf]

References

1. Johansson P, Dahlström U, Broström A. The measurement and prevalence of depression in patients with chronic heart failure. Prog Cardiovasc Nurs 2006;21(1):28-36. [Medline: 16522966]
2. Rutledge T, Reis VA, Linke SE, Greenberg BH, Mills PJ. Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. J Am Coll Cardiol 2006 Oct 17;48(8):1527-1537 [FREE Full text] [doi: 10.1016/j.jacc.2006.06.055] [Medline: 17045884]
3. Rustad JK, Stern TA, Hebert KA, Musselman DL. Diagnosis and treatment of depression in patients with congestive heart failure: a review of the literature. Prim Care Companion CNS Disord 2013;15(4):- [FREE Full text] [doi: 10.4088/PCC.13r01511] [Medline: 24392265]

4. Hare DL, Toukhsati SR, Johansson P, Jaarsma T. Depression and cardiovascular disease: a clinical review. Eur Heart J 2014 Jun 1;35(21):1365-1372 [FREE Full text] [doi: 10.1093/eurheartj/eht462] [Medline: 24282187]
5. Pattenden JF, Roberts H, Lewin RJ. Living with heart failure; patient and carer perspectives. Eur J Cardiovasc Nurs 2007 Dec;6(4):273-279. [doi: 10.1016/j.ejcnurse.2007.01.097] [Medline: 17383236]
6. Muzzarelli S, Leibundgut G, Maeder MT, Rickli H, Handschin R, Gutmann M, et al. Predictors of early readmission or death in elderly patients with heart failure. Am Heart J 2010 Aug;160(2):308-314. [doi: 10.1016/j.ahj.2010.05.007] [Medline: 20691837]
7. Kheirbek RE, Alemi F, Citron BA, Afaq MA, Wu H, Fletcher RD. Trajectory of illness for patients with congestive heart failure. J Palliat Med 2013 May;16(5):478-484. [doi: 10.1089/jpm.2012.0510] [Medline: 23545095]
8. Simmonds RL, Tylee A, Walters P, Rose D. Patients' perceptions of depression and coronary heart disease: a qualitative UPBEAT-UK study. BMC Fam Pract 2013;14:38 [FREE Full text] [doi: 10.1186/1471-2296-14-38] [Medline: 23509869]
9. Shimizu Y, Suzuki M, Okumura H, Yamada S. Risk factors for onset of depression after heart failure hospitalization. J Cardiol 2014 Jul;64(1):37-42 [FREE Full text] [doi: 10.1016/j.jjcc.2013.11.003] [Medline: 24331765]
10. Alosco ML, Spitznagel MB, Miller L, Raz N, Cohen R, Sweet LH, et al. Depression is associated with reduced physical activity in persons with heart failure. Health Psychol 2012 Nov;31(6):754-762 [FREE Full text] [doi: 10.1037/a0028711] [Medline: 22924448]
11. Dekker RL, Lennie TA, Doering LV, Chung ML, Wu J, Moser DK. Coexisting anxiety and depressive symptoms in patients with heart failure. Eur J Cardiovasc Nurs 2014 Apr;13(2):168-176 [FREE Full text] [doi: 10.1177/1474515113519520] [Medline: 24408885]
12. Möller H, Bitter I, Bobes J, Fountoulakis K, Höschl C, Kasper S, et al. Position statement of the European Psychiatric Association (EPA) on the value of antidepressants in the treatment of unipolar depression. Eur Psychiatry 2012 Feb;27(2):114-128. [doi: 10.1016/j.eurpsy.2011.08.002] [Medline: 22119161]
13. Beck AT, Dozois DJ. Cognitive therapy: current status and future directions. Annu Rev Med 2011 Feb;62:397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
14. Freedland KE, Carney RM, Rich MW, Steinmeyer BC, Rubin EH. Cognitive Behavior Therapy for Depression and Self-Care in Heart Failure Patients: A Randomized Clinical Trial. JAMA Intern Med 2015 Nov 1;175(11):1773-1782. [doi: 10.1001/jamainternmed.2015.5220] [Medline: 26414759]
15. Gary RA, Dunbar SB, Higgins MK, Musselman DL, Smith AL. Combined exercise and cognitive behavioral therapy improves outcomes in patients with heart failure. J Psychosom Res 2010 Aug;69(2):119-131 [FREE Full text] [doi: 10.1016/j.jpsychores.2010.01.013] [Medline: 20624510]
16. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
17. Flynn H. Setting the Stage for the Integration of Motivational Interviewing With Cognitive Behavioral Therapy in the Treatment of Depression. Cognitive and Behavioral Practice 2011 Feb;18(1):46-54. [doi: 10.1016/j.cbpra.2009.09.006]
18. Driessen E, Hollon S. Cognitive Behavioral Therapy for Mood Disordersfficacy, Moderators and Mediators. The Psychiatric clinics of North America 2010;33(3):537-555.
19. Serfaty MA, Haworth D, Blanchard M, Buszewicz M, Murad S, King M. Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: a randomized controlled trial. Arch Gen Psychiatry 2009 Dec;66(12):1332-1340. [doi: 10.1001/archgenpsychiatry.2009.165] [Medline: 19996038]
20. Lundgren J, Andersson G, Dahlström Ö, Jaarsma T, Köhler A, Johansson P. Internet-based cognitive behavior therapy for patients with heart failure and depressive symptoms: A proof of concept study. Patient Education and Counseling 2015;98(8):935-942.
21. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
22. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [Medline: 10568646]
23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
24. Hammash MH, Hall LA, Lennie TA, Heo S, Chung ML, Lee KS, et al. Psychometrics of the PHQ-9 as a measure of depressive symptoms in patients with heart failure. Eur J Cardiovasc Nurs 2013 Oct;12(5):446-453. [doi: 10.1177/1474515112468068] [Medline: 23263270]
25. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
26. Eifert GH, Thompson RN, Zvolensky MJ, Edwards K, Frazer NL, Haddad JW, et al. The cardiac anxiety questionnaire: development and preliminary validity. Behav Res Ther 2000 Oct;38(10):1039-1053. [Medline: 11004742]

27. Rector T, Kubo S, Cohn J. Patients?self-assessment of their congestive heart failure. Part 2: Content, reliability and validity of a new measure, the Minnesota Living with Heart Failure questionnaire. Heart Failure (Oct/Nov) 1987:198-209.
28. Middel B, Bouma J, de JM, van SE, Niemeijer MG, Crijns H, et al. Psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q). Clin Rehabil 2001 Oct;15(5):489-500. [Medline: 11594639]
29. Garin O, Ferrer M, Pont A, Rué M, Kotzeva A, Wiklund I, et al. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. Qual Life Res 2009 Feb;18(1):71-85. [doi: 10.1007/s11136-008-9416-4] [Medline: 19052916]
30. Rector TS, Tschumperlin LK, Kubo SH, Bank AJ, Francis GS, McDonald KM, et al. Use of the Living With Heart Failure questionnaire to ascertain patients' perspectives on improvement in quality of life versus risk of drug-induced death. J Card Fail 1995 Jun;1(3):201-206. [Medline: 9420652]
31. Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ 2001 Nov 10;323(7321):1123-1124 [FREE Full text] [Medline: 11701584]
32. Blankers M, Koeter MW, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]
33. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5):e1001454 [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
34. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder, 5th edition. Arlington, VA: American Psychiatric Association; 2013.
35. Dekker RL, Moser DK, Peden AR, Lennie TA. Cognitive therapy improves three-month outcomes in hospitalized patients with heart failure. J Card Fail 2012 Jan;18(1):10-20 [FREE Full text] [doi: 10.1016/j.cardfail.2011.09.008] [Medline: 22196836]
36. Honyashiki M, Furukawa T, Noma H, Tanaka S, Chen P, Ichikawa K, et al. Specificity of CBT for Depression: A Contribution from Multiple Treatments Meta-analyses. Cognitive Therapy & Research 2014;38(3):A.
37. O'Connor CM, Jiang W, Kuchibhatla M, Silva SG, Cuffe MS, Callwood DD, et al. Safety and efficacy of sertraline for depression in patients with heart failure: results of the SADHART-CHF (Sertraline Against Depression and Heart Disease in Chronic Heart Failure) trial. J Am Coll Cardiol 2010 Aug 24;56(9):692-699 [FREE Full text] [doi: 10.1016/j.jacc.2010.03.068] [Medline: 20723799]
38. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]
39. Rozanski A, Blumenthal JA, Davidson KW, Saab PG, Kubzansky L. The epidemiology, pathophysiology, and management of psychosocial risk factors in cardiac practice: the emerging field of behavioral cardiology. J Am Coll Cardiol 2005 Mar 1;45(5):637-651 [FREE Full text] [doi: 10.1016/j.jacc.2004.12.005] [Medline: 15734605]
40. Tully PJ, Selkow T, Bengel J, Rafanelli C. A dynamic view of comorbid depression and generalized anxiety disorder symptom change in chronic heart failure: the discrete effects of cognitive behavioral therapy, exercise, and psychotropic medication. Disabil Rehabil 2015;37(7):585-592. [doi: 10.3109/09638288.2014.935493] [Medline: 24981015]
41. Hyer L, Yeager CA, Hilton N, Sacks A. Group, individual, and staff therapy: an efficient and effective cognitive behavioral therapy in long-term care. Am J Alzheimers Dis Other Demen 2008;23(6):528-539. [doi: 10.1177/1533317508323571] [Medline: 19001352]
42. Gellis ZD, Bruce ML. Problem solving therapy for subthreshold depression in home healthcare patients with cardiovascular disease. Am J Geriatr Psychiatry 2010 Jun;18(6):464-474 [FREE Full text] [Medline: 20871804]
43. Gellis ZD, Kenaley BL, Ten HT. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. J Am Geriatr Soc 2014 May;62(5):889-895. [doi: 10.1111/jgs.12776] [Medline: 24655228]
44. Alexopoulos GS, Raue PJ, Sirey JA, Arean PA. Developing an intervention for depressed, chronically medically ill elders: a model from COPD. Int J Geriatr Psychiatry 2008 May;23(5):447-453. [doi: 10.1002/gps.1925] [Medline: 17932995]
45. Spek V, Nyklícek I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. J Affect Disord 2008 Jan;105(1-3):137-145. [doi: 10.1016/j.jad.2007.05.001] [Medline: 17543392]
46. 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]
47. Stiles-Shields C, Corden ME, Kwasny MJ, Schueller SM, Mohr DC. Predictors of outcome for telephone and face-to-face administered cognitive behavioral therapy for depression. Psychol Med 2015 Nov;45(15):3205-3215. [doi: 10.1017/S0033291715001208] [Medline: 26077620]
48. Lundgren J, Andersson G, Johansson P. Can cognitive behaviour therapy be beneficial for heart failure patients? Curr Heart Fail Rep 2015 Apr;12(2):166-172. [doi: 10.1007/s11897-014-0244-2] [Medline: 25475179]

49. Kaldo V, Jernelöv S, Blom K, Ljótsson B, Brodin M, Jörgensen M, et al. Guided internet cognitive behavioral therapy for insomnia compared to a control treatment - A randomized trial. Behav Res Ther 2015 Aug;71:90-100 [FREE Full text] [doi: 10.1016/j.brat.2015.06.001] [Medline: 26091917]
50. Blom K, Tarkian TH, Wiklund T, Danlycke E, Forssén M, Söderström A, et al. Internet-vs. group-delivered cognitive behavior therapy for insomnia: A randomized controlled non-inferiority trial. Behav Res Ther 2015 Jul;70:47-55 [FREE Full text] [doi: 10.1016/j.brat.2015.05.002] [Medline: 25981329]
51. Faller H, Störk S, Schuler M, Schowalter M, Steinbüchel T, Ertl G, et al. Depression and disease severity as predictors of health-related quality of life in patients with chronic heart failure--a structural equation modeling approach. J Card Fail 2009 May;15(4):286-292.e2. [doi: 10.1016/j.cardfail.2008.10.022] [Medline: 19398075]
52. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, ESC Committee for Practice Guidelines. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012 Jul;33(14):1787-1847 [FREE Full text] [doi: 10.1093/eurheartj/ehs104] [Medline: 22611136]
53. Findahl O, Davidsson P. The Swedes and the Internet. Sweden: The Internet Foundation; 2015. Svenskarna och internet 2015 URL: http://www.soi2015.se/ [accessed 2016-07-05] [WebCite Cache ID 6imJRIi3b]

Abbreviations

CAQ: Cardiac Anxiety Questionnaire.
CBT: cognitive behavioral therapy.
DF: discussion forum.
EM: Expectation-Maximization.
HF: heart failure.
ICBT: Internet-based cognitive behavioral therapy.
MLHF: Minnesota Living with Heart Failure questionnaire.
PHQ-9: Patient Health Questionnaire-9.
QoL: quality of life.

Edited by G Eysenbach; submitted 26.01.16; peer-reviewed by C Nahlén Bose, S Alfonsson; comments to author 03.04.16; revised version received 12.06.16; accepted 17.06.16; published 03.08.16.

Please cite as:
Lundgren JG, Dahlström Ö, Andersson G, Jaarsma T, Kärner Köhler A, Johansson P
The Effect of Guided Web-Based Cognitive Behavioral Therapy on Patients With Depressive Symptoms and Heart Failure: A Pilot Randomized Controlled Trial
J Med Internet Res 2016;18(8):e194
URL: http://www.jmir.org/2016/8/e194/
doi:10.2196/jmir.5556
PMID:27489077

Original Paper

Integrating Evidence From Systematic Reviews, Qualitative Research, and Expert Knowledge Using Co-Design Techniques to Develop a Web-Based Intervention for People in the Retirement Transition

Nicola O'Brien[1], PhD; Ben Heaven[1], PhD; Gemma Teal[3], MEng; Elizabeth H Evans[1], PhD; Claire Cleland[3], PhD; Suzanne Moffatt[1], PhD; Falko F Sniehotta[1], PhD; Martin White[4], MD; John C Mathers[5], PhD; Paula Moynihan[6], PhD

[1]Institute of Health and Society, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]Institute of Design Innovation, The Glasgow School of Art, Glasgow, United Kingdom

[3]School of Planning, Architecture and Civil Engineering, Queens University, Belfast, United Kingdom

[4]UKCRC Centre for Diet & Activity Research (CEDAR), MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

[5]Institute of Cellular Medicine, Human Nutrition Research Centre, Newcastle University, Newcastle upon Tyne, United Kingdom

[6]Institute of Health and Society, Centre for Oral Health Research, Newcastle University, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Paula Moynihan, PhD
Institute of Health and Society
Centre for Oral Health Research
Newcastle University
Framlington Place
Newcastle upon Tyne, NE24BW
United Kingdom
Phone: 44 1912088241
Fax: 44 191 2225928
Email: paula.moynihan@ncl.ac.uk

Abstract

Background: Integrating stakeholder involvement in complex health intervention design maximizes acceptability and potential effectiveness. However, there is little methodological guidance about how to integrate evidence systematically from various sources in this process. Scientific evidence derived from different approaches can be difficult to integrate and the problem is compounded when attempting to include diverse, subjective input from stakeholders.

Objective: The intent of the study was to describe and appraise a systematic, sequential approach to integrate scientific evidence, expert knowledge and experience, and stakeholder involvement in the co-design and development of a complex health intervention. The development of a Web-based lifestyle intervention for people in retirement is used as an example.

Methods: Evidence from three systematic reviews, qualitative research findings, and expert knowledge was compiled to produce evidence statements (stage 1). Face validity of these statements was assessed by key stakeholders in a co-design workshop resulting in a set of intervention principles (stage 2). These principles were assessed for face validity in a second workshop, resulting in core intervention concepts and hand-drawn prototypes (stage 3). The outputs from stages 1-3 were translated into a design brief and specification (stage 4), which guided the building of a functioning prototype, Web-based intervention (stage 5). This prototype was de-risked resulting in an optimized functioning prototype (stage 6), which was subject to iterative testing and optimization (stage 7), prior to formal pilot evaluation.

Results: The evidence statements (stage 1) highlighted the effectiveness of physical activity, dietary and social role interventions in retirement; the idiosyncratic nature of retirement and well-being; the value of using specific behavior change techniques including those derived from the Health Action Process Approach; and the need for signposting to local resources. The intervention principles (stage 2) included the need to facilitate self-reflection on available resources, personalization, and promotion of links between key lifestyle behaviors. The core concepts and hand-drawn prototypes (stage 3) had embedded in them the importance of time use and work exit planning, personalized goal setting, and acceptance of a Web-based intervention. The design brief detailed the features and modules required (stage 4), guiding the development of wireframes, module content and functionality,

virtual mentors, and intervention branding (stage 5). Following an iterative process of intervention testing and optimization (stage 6), the final Web-based intervention prototype of LEAP (Living, Eating, Activity, and Planning in retirement) was produced (stage 7). The approach was resource intensive and required a multidisciplinary team. The design expert made an invaluable contribution throughout the process.

Conclusions: Our sequential approach fills an important methodological gap in the literature, describing the stages and techniques useful in developing an evidence-based complex health intervention. The systematic and rigorous integration of scientific evidence, expert knowledge and experience, and stakeholder input has resulted in an intervention likely to be acceptable and feasible.

(J Med Internet Res 2016;18(8):e210) doi:10.2196/jmir.5790

KEYWORDS

intervention studies; health behavior; retirement; Internet

Introduction

Leading international bodies in health and social care research and governance advocate the integration of stakeholder involvement in the design and development of novel health interventions [1-3]. Stakeholder input in intervention development is important to ensure that the intervention is relevant and useful for those people or groups who have or could have an interest in it. This, in turn, has the potential to maximize the acceptability and potential effectiveness of the intervention. Contemporary methods for designing products or services have moved away from using material and supplier-centered processes to more social and user-centered processes [4]; consequently, design-oriented approaches to health care innovation are being more widely recognized [5-7]. Involving relevant stakeholders as co-designers of health interventions allows the stakeholders to help define the health care problem and identify preferred intervention solutions [8-12]. Divergent and convergent thinking may result in the generation of new intervention ideas and selection of the best idea available. Intervention ideas are prototyped and explored hands-on, through sequential processes to rehearse the future [4,7]. Despite the growing use of co-design techniques in health care innovation, there is no explicit, replicable, and accepted description of their application in the development of complex health interventions.

Stakeholder involvement alone is not sufficient for effective intervention development. A range of research methods also needs to be applied, including careful consideration of existing evidence of need, and for effectiveness and cost-effectiveness of interventions to tackle the specific problem. Qualitative research provides depth of understanding to the relevant issues. Evidence-based medicine has been formally recognized as one of modern medicine's most important milestones [13] and is applied increasingly across the fields of public health, behavioral medicine and health, and social care. Systematic and rigorous methods, including systematic reviews, meta-analyses, and meta-syntheses, for identifying and evaluating the evidence base and identifying and developing theory are key elements in the process of developing complex health interventions [14]. Quantitative and/or qualitative data are synthesized to draw conclusions about likely intervention effects and potential effect modifiers. For example, the active ingredients or features of interventions, such as behavior change techniques (BCT) associated with more positive behavior change and theory underpinning effective interventions, can be identified [15-19]. These conclusions can then be used to inform the development of novel interventions including the features that are most likely to be effective [20,21]. This systematic and theoretical approach to intervention development, along with accuracy of reporting the intervention protocol, facilitate the evaluation and replication of the intervention. While this approach to evidence synthesis is desirable, in practice it can be challenging because of the absence of established methods to guide the application of evidence to the specific population or clinical context in building a complex intervention [22,23].

Qualitative research methods help inform the development of complex health interventions [24,25]. Interviews, focus groups, and observational methods can explore the needs, attitudes, behavior, and contextual factors of the specific population and health topic under investigation [26]. The outcomes of such qualitative research can help intervention developers identify potential further intervention effect modifiers, which may inform tailoring of the intervention, thereby increasing the likelihood that the intervention will be accepted and effective.

In practice, intervention developers are likely to use deductive and inductive research methods to generate the evidence base on which the novel intervention is based. Mixed-method studies attempt to bridge the epistemological differences between quantitative and qualitative data acquisition approaches [27]. Such an inclusive approach to evidence synthesis, integrating different forms of scientific evidence, can be challenging, especially when mixed-method findings conflict [28]. Moreover, methodological guidance on how to integrate this evidence with stakeholder needs and preferences is lacking. This paper aims to fill this important methodological gap, describing and appraising a systematic and sequential approach to intervention development, drawing on techniques of co-design. Specifically, we detail the stages and techniques used to integrate quantitative (systematic review) and qualitative (interviews and focus groups) evidence, and expert knowledge and experience to engage stakeholders in a co-design process. The process is illustrated through the development of a Web-based lifestyle intervention (Living, Eating, Activity, and Planning in retirement: LEAP) to promote health and well-being of people in retirement.

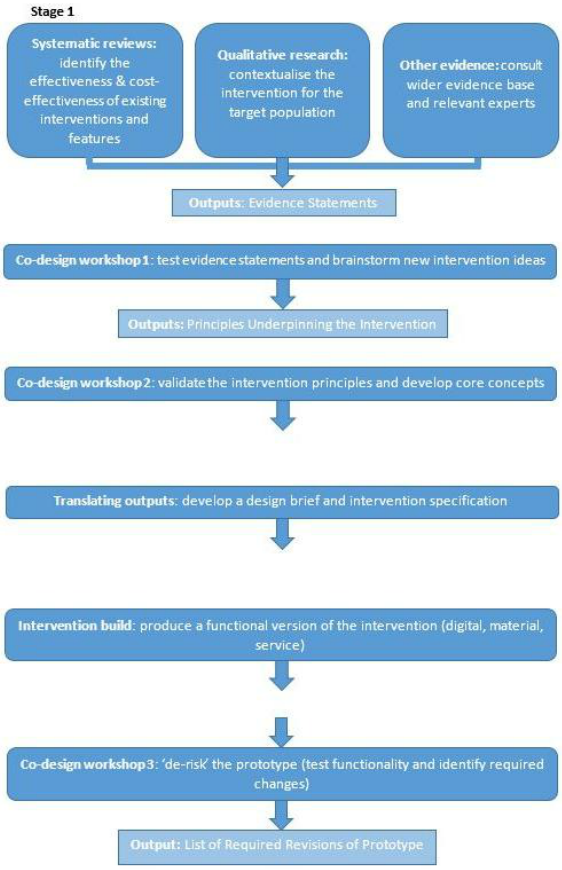
Methods and Results

Overview

An iterative co-design process involving sequential validation of the evidence, generation of intervention ideas, and prototyping, testing, analyzing, and optimizing the intervention was followed. This process is described as a series of stages in which each stage in the process resulted in output(s) to inform the design of the intervention. After each stage, the research team discussed and analyzed the output(s) and critically reflected on the process. Outputs from each stage were used subsequently as inputs to the next stage of development. The methods and results of each stage are therefore presented sequentially. Figure 1 displays an overview of the methods employed and outputs derived at each stage.

The context and underlying rationale behind developing a Web-based lifestyle intervention to promote health and well-being of people in retirement is presented in the following summary.

Figure 1. Overview of the systematic, sequential approach to intervention co-design and development.



Summary of the Intervention Development Context

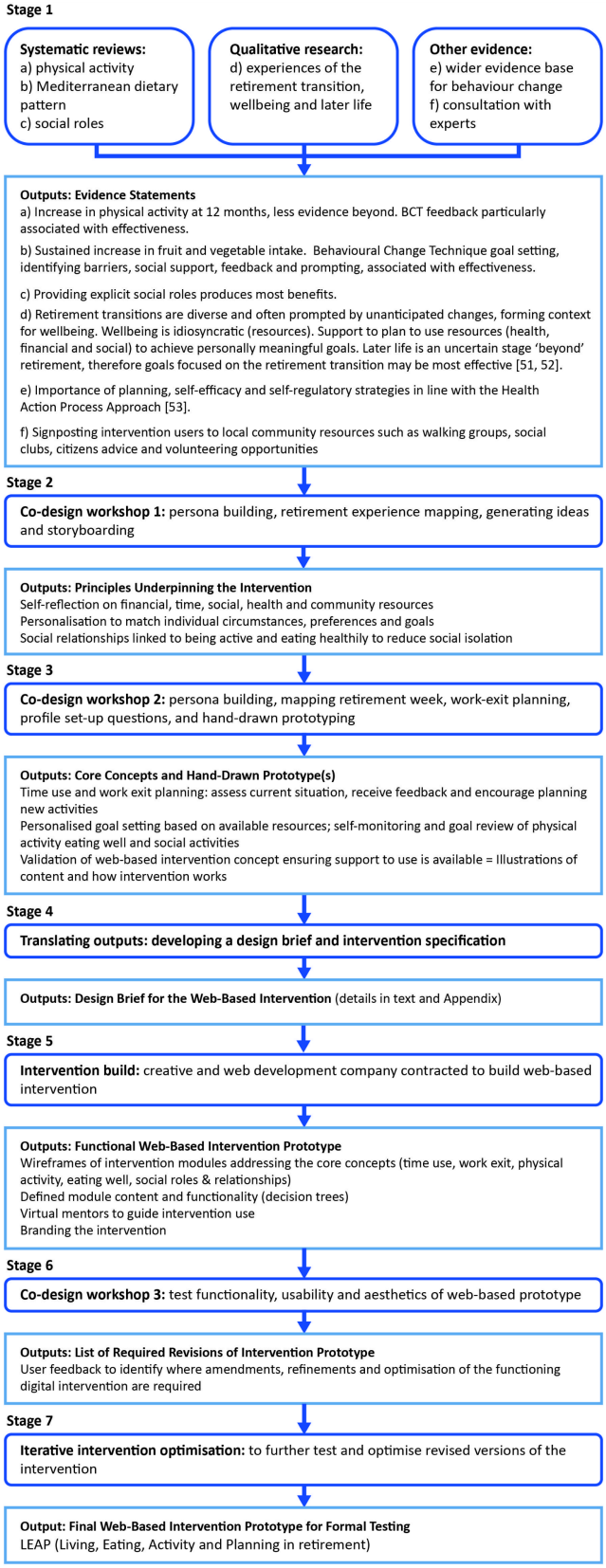
Retirement from full-time work is a life transition that has been shown to be associated with changes in key lifestyle factors. Some cross-sectional and longitudinal studies have shown that people engage in more healthy behaviors with retirement [29-31]. However, the evidence is inconsistent and other studies have shown reduced physical activity (PA) [32,33], less healthy dietary behavior [34], and a loss of perceived status and purpose [35]. The population health and well-being benefits of physical activity, a healthy diet, in particular one based on a Mediterranean dietary pattern, and social engagement are well documented [36-38]. Despite recent evidence that today's older adults are healthier than they were 10-20 years ago [39], with a globally ageing population and the accompanying increase in the prevalence of chronic ill health and morbidity [40], maintaining a healthy lifestyle into later years is vital for individual well-being and to lessen the burden on society.

As engagement with key health and social behaviors may change in retirement, the retirement transition offers a unique window of opportunity to intervene to improve health and well-being of older adults [41]. A small number of studies have delivered lifestyle interventions in the retirement transition [42-45], and systematic reviews synthesizing data from this life stage provide support for their effectiveness [46-49].

A predefined priority for the research team was to develop a personalized, scalable, sustainable, and potentially cost-effective intervention. Web-based interventions can be tailored to the individual user and may be more compatible with modes of accessing information and support (eg, by mobile phone) in future cohorts of older people. They also have greater potential for wide scale use in the target population. As such, the possibility of a Web-based intervention was a planned consideration of the research team.

Figure 2 details the process by which the sequential intervention development approach (detailed in Figure 1) was applied to develop LEAP, a Web-based lifestyle intervention in retirement. The specific outputs of each stage in the development of LEAP are presented.

Figure 2. Applied example: integrating systematic review, qualitative research and other evidence with stakeholder engagement in a co-design process to develop the LEAP intervention.



Stage 1: Compiling the Evidence Base

Stage 1 Procedure

Evidence from systematic reviews, qualitative research, and other sources, including the wider evidence base for behavior change and consultation with relevant experts, was summarized by the research team.

Stage 1 Analysis

The evidence was recorded as a list of "evidence statements" [50] that informed the aims and content of co-design workshop 1 (stage 2).

Stage 1 Outputs

The evidence statements are listed in Figure 2 and descriptions of the evidence are enlarged upon in Multimedia Appendix 1. The systematic reviews, which in part underpinned these statements, have previously been published [18,19,47-49].

Stage 2: Co-design Workshop 1

Stage 2 Participants

A total of 42 stakeholders participated in co-design workshop 1. Participants included 12 members of the research team (6 workshop facilitators and 6 scribes), 22 adults aged 55 years or over (9 males) as potential intervention users, and 8 health and social care professionals (3 males) from the voluntary sector and public health organizations, whose role was related to improving health and well-being of people in retirement. The research team included health researchers from a range of disciplines involved in improving health and well-being in older people and with combined expertise in design, behavior change, public health, physical activity, nutrition and dietetics, and social gerontology.

Older adults from local older people's forums were sampled purposively to represent men and women at different stages in the retirement transition and from diverse socioeconomic backgrounds.

Stage 2 Procedure

The aim of stage 2 (co-design workshop 1) was to determine the face validity of the evidence statements and brainstorm new intervention ideas, which were informed and inspired by the evidence statements. In preparation, discussion among the research team identified that the qualitative work provided the context in which the intervention would be built. The qualitative work emphasized individual experiences through retirement; retirement was commonly experienced as a process rather than a discrete event. The systematic review evidence and the theoretical framework of the Health Action Process Approach provided recommendations regarding intervention content, modalities, and timings. Through these discussions, the research team identified the need for co-design techniques that could combine descriptive, context-rich narratives with more discrete evidence regarding health and social behaviors.

The workshop took place in a university space. Participants were divided into small groups, each comprising older adults, health and social care professionals, one facilitator, and one scribe. Facilitators guided the structure and timing of the workshop, and scribes recorded participants' comments and ideas.

The first technique used in the workshop was the co-design of a persona. As retirement is both a process and idiosyncratic, "persona building" [51] was a useful technique in order to orient each group to real world issues in retirement. Each group of workshop participants was assigned a different persona with a description that was a composite of different accounts and experiences of participants in the qualitative study and from clinical experience of the research team. The personas represented male and female older adults from a range of socioeconomic positions (see Multimedia Appendices 2 and 3). Each workshop group considered one of the behaviors examined in the systematic reviews [47-49]. For example, Jeff was said to be physically inactive but had enjoyed being active when he was younger.

The second technique was "experience mapping" [52] of the retirement transition. To generate intervention ideas including considering when it would be most needed, accepted, and potentially effective, we mapped different retirement pathways identified in the qualitative study. Each group considered the possible key stages in the persona's retirement transition such as "Jeff gets made redundant," "Jeff retires," and "Jeff gets a part-time job." This technique allowed participants to discuss how the persona might feel during the retirement transition, reflecting on their own experiences. Each group generated intervention ideas that would help tackle the specific health, social, or resource challenges of the persona in each scenario. In particular, groups were asked to consider Web-based intervention ideas. Drawing on evidence statements regarding the importance of the local environment and community resources to facilitate behavior change, groups were presented with a map of the persona's local area and asked to think about whether local resources could support the intervention ideas generated.

"Wild cards" representing random events that might disrupt the retirement story were introduced to mimic the unpredictable nature of real life and challenge the participants to consider whether these events would alter the retirement story. The wild cards also provided opportunities to discuss how specific intervention features, such as behavioral change techniques, could be incorporated in the persona's retirement pathway.

The third technique was "storyboarding" [53]. This allowed the group to pull their different ideas together to form a new intervention to support an ideal retirement experience for their persona. The intervention ideas from each group provided an outline of a potential intervention, which was defined to include its name, how it would be signposted or advertised, features to encourage initial and longer-term engagement, and the lifestyle behaviors it would help to promote.

Stage 2 Analysis

Shortly after the workshop, the facilitators prepared detailed notes to capture group discussions and describe how the participants tackled each activity, reflecting on the key insights and ideas discussed within each group. Analyzing the outputs of the workshop activities revealed recurring design ideas (ie,

broad goals for the intervention), which became the "principles underpinning the intervention." The intervention principles were derived from validating the acceptability and importance of the evidence statements with potential users, and therefore the principles reflected the context in which the intervention would be built and ideas for intervention content, modality, and timing. These principles were further explored and developed by stakeholders at subsequent workshops (co-design workshops 2 and 3, described below) to generate more tangible ideas for products and features for the intervention. The method used to analyze the facilitator notes and workshop materials resembles thematic analysis [54], a technique that allows for the identification of repeated patterns of meaning. As used here, it captured design ideas in addition to themes.

Stage 2 Outputs

The outputs of co-design workshop 1 were the potential intervention ideas from each of the six groups. The potential intervention components, resources for the intervention, and the key design priorities were identified by the research team. The most common themes and features in the intervention ideas formed the following intervention principles:

1. *Self-reflection* on financial, time, social, health, and community resources; the intervention should provide practical assistance in planning or structuring activities focusing on key life events rather than age;

2. *Personalization* to individual circumstances, preferences, and goals, providing a flexible intervention with individual feedback and tailored support from a mentor;

3. *Social relationships* linked to being physically active, eating healthily, in order to reduce risks of social isolation, promote a sense of social support, and share experiences in an engaging way.

Stage 3: Co-design Workshop 2

Stage 3 Participants

A total of 20 stakeholders participated in co-design workshop 2: 6 members of the research team (3 facilitators and 3 creative facilitators) and 14 older adults (6 males). Older adults were recruited from local forums as before.

Stage 3 Procedure

Co-design workshop 2 aimed to obtain user feedback on the intervention principles derived from co-design workshop 1. Feedback was used to assess face validity of the principles and to develop the core intervention concepts. The workshop took place in a local community meeting space and lasted 4 hours including refreshment breaks. Participants were divided into three groups, each of which was led by a workshop facilitator. Each group was also supported by a creative facilitator with design expertise, who sketched the intervention ideas as they were being generated and facilitated the development of hand-drawn prototypes of potential interventions using paper Web browser templates. Web browser templates were used to explicitly explore the acceptability of a Web-based intervention.

In preparation for this workshop, the design expert identified further co-design techniques to facilitate the presentation and interaction with specific aspects of the intervention principles using Web browser visual materials and prompts. "Prototyping" was a key technique used to communicate ideas, which enabled the progression of thinking through physical making, a safe space for failure leading to faster learning, and encouragement and permission to explore new behaviors [55].

Validation of the self-reflection intervention principle was conducted using mock-ups of a work transition tool with interactive graphs and texts. This tool, which supported individuals to reflect on possible work exits (eg, retiring fully, reducing hours) or re-entry (eg, returning to employment) as identified in the qualitative work [56], was developed in each workshop group.

To further explore the personalization intervention principle, participants were asked to consider what questions the intervention platform should ask to learn about the persona's attitudes and habits in relation to the target lifestyle behaviors. The answers to these questions would shape how the intervention could be personalized to meet the persona's needs, circumstances, and goals. Participants wrote the questions on cards, placed them in a natural conversational order, and considered options for how they could be presented (eg, written text, video clips, and images).

The co-design technique of persona building was used to further explore and validate the social relationships intervention principle, while also providing further opportunity to explore the act of planning included in the self-reflection principle. Groups mapped a typical week during retirement, focusing on the absolute and relative time the persona engaged in lifestyle behaviors related to being physically active, eating healthily, and spending time with other people.

Stage 3 Analysis

Detailed notes capturing each group's discussion, how the participants tackled each activity, and key feedback on the intervention principles were produced by the facilitators. Using thematic-based analysis as in stage 2, facilitator notes and the hand-drawn prototypes of potential interventions were analyzed to identify recurring design ideas and intervention user requirements, and to define the core intervention concepts, which reflected the intervention principles and the target lifestyle behaviors of physical activity, healthier eating, and social roles.

Stage 3 Outputs

The outputs of co-design workshop 2 were core intervention concepts and the hand-drawn *prototypes* of novel interventions, which served to document how the intervention principles were validated through user feedback. The "work transition" and "mapping a retirement week" tools were evaluated as enabling self-reflection; providing feedback on financial, time, and social resources (including considering how social relationships can be linked with other activities during a week); and facilitating future goal setting and planning. Seeing when new activities could take place was deemed to be extremely valuable, providing insight into potential spare time. It also served to prompt people to set boundaries (eg, ensuring they did not overcommit to looking after grandchildren) and goals.

A core set of personalization questions were defined whose answers would allow the intervention to be tailored to individual needs and goals. The result was a low-fidelity (limited function) prototype of the user registration component of the intervention, with each webpage hand-drawn on a deck of paper templates.

The majority of participants welcomed a Web-based intervention, acknowledging the benefits of having access at home and at convenient times, and the intrinsic ability of a Web-based intervention to be tailored to the individual. However, there were concerns that some individuals may feel unsupported by technology and consequently disengage from the intervention, further reinforcing the need for support from a mentor outlined in the personalization design principle. Potential cost-effectiveness and scalability of the intervention were predefined priorities of the research team. Therefore, providing access to someone to support use of the intervention, such as a health care assistant, was considered unfeasible. Consequently, the role of a virtual mentor within the Web-based intervention, who could help the user explore retirement transition options and lifestyle behaviors, was explored and positively appraised.

The research team identified common themes and features of the hand-drawn prototypes that related to the target lifestyle behaviors to form the following core intervention concepts: (1) time use and work exit planning as an opportunity to assess current financial, time, and social resources, receive feedback, and encourage the planning of new activities, (2) personalized goal setting based on identified available resources, self-monitoring of behavior, and regular reviewing of lifestyle goals, and (3) a Web-based intervention as an acceptable mode of delivery, providing that support to use is available.

Stage 4: Translating Outputs Into a Design Brief and Specification

Stage 4 Procedure

The aim of this stage was for the research team to examine, critically evaluate, and translate the outputs from the previous stages into a detailed design brief and specification document to inform the intervention build.

Stage 4 Analysis

The evidence statements (stage 1), design principles (stage 2), hand-drawn prototypes, and core intervention concepts (stage 3) were examined for recurring design ideas and intervention requirements across all outputs. These ideas and requirements were evaluated critically by the research team for concurrence with the team's predefined priorities, the intervention development context, and their suitability to support the promotion of the target lifestyle behaviors.

Stage 4 Outputs

The output was a design brief and specification document detailing the aim of the intervention and the design features it should include (see Multimedia Appendix 4). The design brief stated the need for an interactive website including a set of intervention tools to support people to have a healthier and more fulfilling retirement. The design specification detailed the following design features that the intervention should include: personalized, scalable, sustainable, interactive, digital, user flow through the intervention, and visually and functionally engaging. The following intervention sections or modules to be included were also detailed: user profile, work-exit and cost of living, time and activity planner, physical activity, eating well, and social relationships.

Stage 5: Intervention Build

Stage 5 Procedure

The aim of this stage was to produce a functional version of the intervention prototype. This involved a tendering process to identify a Web development company that would support the building of a functional Web-based intervention prototype. The design brief and specification were included in the tender. The research team worked closely with the contracted company throughout the process, holding regular face-to-face meetings to discuss emerging ideas for presenting the intervention content, to order and structure the intervention modules, and to maximize user engagement with the intervention.

Stage 5 Outputs

Wireframes for the intervention modules, detailed module content, and decision trees guiding user flow through the intervention were developed. Wireframes are simple images that show how a website and its webpages are structured and how the content is arranged. A set of six "virtual mentors" connected, if desired, to audio files recorded by local actors to provide cultural links were also developed to guide and support users through the intervention. The final output of this stage was a functioning Web-based intervention prototype for testing and optimization with stakeholders (see Multimedia Appendices 5-14).

Stage 6: Co-design Workshop 3

Stage 6 Participants

A total of 37 stakeholders participated in co-design workshop 3: 8 members of the research team as facilitators and 29 older adults (12 males). Older adults were recruited from local forums.

Stage 6 Procedure

The aim of the third and final co-design workshop was to "de-risk" the prototype [57] through testing intervention functionality and identifying necessary modifications using a cognitive walkthrough activity [58]. The final workshop took place in a university space. Participants were divided into small groups, each of which was led by a workshop facilitator. The intervention de-risking techniques focused on exploiting user experience testing. Participants were provided with a tablet and asked to use the intervention with the aim of testing its functionality, usability, and aesthetics. Feedback, queries, technical and functional issues that participants expressed were recorded by the group facilitator on printed screenshots of each page of the intervention. The technique of persona building was used as the vehicle for the group to navigate the intervention from the perspective of the persona.

Stage 6 Analysis

The feedback and issues identified by each group were collated by the facilitators. Technical and functional issues were added

to a list of required revisions to the intervention prototype. Other feedback, such as comments relating to the design, esthetics, or content of the intervention, was considered by the research team to ascertain whether the revisions were feasible and essential.

Stage 6 Outputs

The output of this stage was a comprehensive list of revisions to the intervention prototype required to improve user experience and acceptability of the intervention. Identified revisions to the prototype included (1) refining color contrasts and font size, (2) revising text content, order, and position, (3) including an intervention overview page, a dashboard summarizing the parts of the intervention with which the user has already engaged, and a diary summarizing the user's activities scheduled for the following weeks, (4) adding progress bars for questionnaires, and (5) providing options to hear the mentor's voice and viewing the time planner as a calendar or pie chart. An optimized functioning intervention prototype was produced following the amendments and refinements.

Stage 7: Iterative Intervention Optimization

Stage 7 Participants

A group of 30 representatives of stakeholders (potential intervention users, researchers, and health and social care professionals) provided feedback on the revised intervention prototype.

Stage 7 Procedure

The aim of this final stage was for the revised intervention prototype to be further tested by stakeholders to identify additional ways to improve and refine the intervention. This stage adopted an iterative testing, user feedback, and intervention refinement process whereby optimization occurred in parallel with testing to ensure that new or revised features were also tested. The research team liaised closely with the Web development company to ensure that optimization occurred promptly and efficiently.

Stage 7 Outputs

The output was a final prototype Web-based intervention, ready for formal field testing in a pilot randomized controlled trial (to be reported elsewhere). The intervention was named LEAP (Living, Eating, Activity, and Planning in retirement). Table 1 [16,18,19,38,56,59-64] presents a summary of LEAP intervention modules, tools and interactive features, and the evidence on which each element was based.

Ethical Approval

This work was conducted as part of the LiveWell program. Ethical approval was acquired from Newcastle University Ethics Committee (No 00423). Informed consent was obtained from participants in the qualitative study. The workshops were based on a co-design methodology where all stakeholders (research team, older adults, and health professionals) held shared "power" in the development of the new intervention. No personal data were collected, thus ethical consent to participate in the workshops was not obtained. However, informed consent was obtained for the purpose of photographically recording the activities at each workshop.

Table 1. The LEAP features and modules, and the objective(s), tools, and evidence on which they were based at different stages of the intervention development process.

LEAP section/ module	Objective(s)	LEAP tools	Evidence base
User profile	Register the user and set preferences		Qualitative research found that retirement transitions and available resources are idiosyncratic. The user profile supports the tailoring of LEAP to address the variable nature of retirement transitions [56,59]. Preliminary information about the user's retirement stage, physical activity, diet, and social circumstances is captured and used to tailor the introduction of the related module.
Intervention overview (Multimedia Appendices 5-7)	Provide an overview of the modules and their objectives, and guidance on the general functions and features of LEAP. Set personal preferences (mentor, email bulletin).	Interactive carousel overview of modules.	Co-design workshop 3 identified the need for an overview to provide a guide to the intervention modules and tools, including emphasizing the intended dip-in and dip-out nature and user-determined flow through the intervention.
		Opt-in for LEAP to be personalized and supported by a virtual mentor. One of eight mentors could be chosen with the option of hearing their voice or reading the text.	Co-design workshop 1 identified the need for a mentor to support user journey through the intervention. This idea was appraised positively during co-design workshop 2. Virtual mentors were developed and optimized during co-design workshop 3.
		Opt-in to receive weekly email bulletin summarizing recent usage of LEAP and prompt revision of goals and plans (BCT: follow-up prompts, goal review).	Systematic review of dietary interventions found that the BCT follow-up prompts was associated with greater intervention effectiveness [19]. Self-regulation prompts for action control [60] is an effective behavior change strategy [61].
Time module (Multimedia Appendix 8)	Reflect on current and desired future time use over the retirement transition.	Interactive calendar or pie chart time planner.	Qualitative research indicated that assistance to reflect on current and future time use was important. Considering how time might be spent in retirement (eg, additional care of relatives, unstructured time) might help identification of personalized goals (eg, a need for a structured role or activity) and potential barriers/facilitators to goal achievement [56]. This module provides space and tools for the user to think through the possibilities and opportunities for lifestyle behaviors, goals, and aspirations. Co-design workshop 1 found that a time reflection tool would be useful. Co-design workshop 2 found that providing a choice of how the time planner tool is presented (calendar or pie chart style) is desirable and that this module was valuable to "set the scene" for other modules, where activities could be considered and scheduled.
Changing Work module (Multimedia Appendix 9)	Consider financial and work situation as participant moves through the retirement transition.	Interactive bar charts and graphs to visualize different retirement trajectories and the effect on income and expenditure.	Qualitative research indicated that finances and modes of work transition (eg, full to part-time, fully retire, retired to part-time) are idiosyncratic and lay the foundation for different retirement experiences and the adoption and maintenance of lifestyle choices [59]. This module allows the user to consider their circumstances and opportunities to engage in new activities (eg, continuing working reduces available free time but the continued income could mean can retire earlier). Co-design workshop 1 showed that a work exit tool was appraised positively by potential users. Co-design workshops 2 and 3 further developed and refined this tool.

LEAP section/ module	Objective(s)	LEAP tools	Evidence base
Moving More module (Multimedia Appendix 10)	Awareness of current physical activity level. Opportunity and tools to engage in self-regulation of PA.	Pedometer to facilitate self-monitoring, goal setting including step-count, receive feedback, schedule activities, identify barriers, and revisit and review step and activity goals (BCT self-monitoring, goal setting behavior and outcome, goal review, action planning, barrier identification).	Physical activity was a predefined target behavior. Systematic review of physical activity suggested that the BCT providing feedback was associated with greater long-term effectiveness [18]. Evidence for the effectiveness of other self-regulation BCTs to promote PA, in line with the Health Action Process Approach [16,62-64]. Co-design workshop 1 confirmed that the BCTs of self-monitoring, goal setting, and action planning were acceptable to stakeholders and potentially valuable for most. Co-design workshops 2 and 3 further developed and refined this module.
Being Social module (Multimedia Appendix 11)	Explore potential benefits of having a meaningful occupation/ role or spending time with significant others.	Interactive social relationship mapping, social role case studies and schedule activities (BCT: action planning).	Social connectedness was a predefined target behavior. Systematic review of social roles suggested that interventions providing explicit roles with group support were effective [48]. The social roles tool provides resources to explore explicit roles. Participating in social relationships has been identified in the literature as key to well-being in later life [38]. Qualitative research confirmed the importance of social relationships but did not identify a clear opportunity for intervention [56]. Co-design workshop 2 identified a potential intervention mechanism through a relationship reflection tool, supporting by structured suggestions for maintaining and building social relationships.
Eating Well module (Multimedia Appendix 12)	Awareness of current diet and provision of information to make diet more Mediterranean in style.	Mediterranean diet quiz and feedback, goal setting, recipe book, schedule trying a new recipe, identify barriers, and revisit and review goals (BCTs: information about consequences of behavior, goal setting behavior and outcome, goal review, action planning, barrier identification).	A predefined target behavior. Systematic review of Mediterranean dietary patterns suggested that the BCTs of goal setting, identifying barriers, feedback, and follow-up prompts were associated with greater effectiveness [19]. Co-design workshop 1 identified the need for a self-assessment tool to appraise fit between current diet and Mediterranean eating pattern, with personalized feedback and suggestions to improve. Co-design workshop 2 confirmed acceptability of the module's core functions, including personal goal setting, feedback, and follow-up prompts, in line with BCTs identified in systematic review [19]. A meal planner and recipe guide were also judged acceptable. Co-design workshop 3 confirmed acceptability of the barrier identification and coping planning features. Stakeholders suggested improvements to interface usability and clarity.
Diary (Multimedia Appendix 13)	Schedule PA, trying a new Mediterranean diet recipe or social activity for the current and following week.	Intelligent design remembers previously named significant others and prompts to add them to the scheduled activity (BCT: action planning).	This feature arose in co-design workshop 3 and was developed subsequently as a way to summarize the activities a user had scheduled and link with the weekly email bulletin to encourage revisiting the intervention to update data, get feedback, revise goals and plans, and schedule new activities. Evidence for the effectiveness of self-regulation behavioral change techniques to promote health behaviors, in line with the Health Action Process Approach [16,18,19,62-64].

Discussion

Principal Findings

This paper's key contribution is providing a description of a systematic, sequential approach to integrating scientific evidence from systematic reviews, qualitative research, and expert knowledge and experience with stakeholder involvement to develop an evidence-based complex health intervention. We have detailed the stages employed including the co-design techniques used and the outputs produced, and have demonstrated the application of the approach in the development of LEAP, a Web-based lifestyle intervention for people in the

retirement transition. Here we provide a critical appraisal of this approach.

Strengths and Limitations

The approach presented in this paper follows and complements the Medical Research Council (MRC) guidance for the development of complex health interventions [14]. As advocated in this guidance, our approach applied systematic and rigorous methods to identify and evaluate the evidence base and the theoretical basis for a novel intervention. In addition, we have described the practical stages and methods required to integrate this evidence with stakeholder input. Specifically, we have utilized co-design methodology to facilitate stakeholder engagement and input, which can be modified and refined to suit the specific intervention context and target population. Our approach adds to recent studies using co-design techniques for health care innovation [5-7] providing a concrete example of how to apply these methods in the development of a Web-based lifestyle intervention in retirement.

Our intervention development approach follows seven distinct stages, each of which has the following strengths, limitations, and challenges. Stage 1, "compiling the evidence base," is an essential component of intervention development [14], but depending on the size and extent to which the evidence base has been interrogated previously, this stage can be resource intensive, which may be a barrier for projects with scarce resources. In the example of developing LEAP, there was limited existing evidence for the effectiveness of lifestyle interventions in retirement and on experiences of the retirement transition. The wider evidence base for behavior change and knowledge of local resources with which the intervention could link were evidence sources that already existed yet required explicit interrogation in relation to the intervention objectives.

A challenge faced during stage 2 co-design workshop 1 involved the use of personas. Some stakeholders struggled with this concept initially but when the purpose of using a persona was further explained by researchers, stakeholders engaged with the process. In addition, many stakeholders found that the wild cards, used to assess how specific BCTs could be incorporated in the persona's retirement pathway, were too abstract and difficult to grasp, thus limiting the assessment of their potential value and acceptability in the intervention. However, allowing stakeholders to explore how BCTs work in practice using a prototype intervention at a later stage in the intervention development process (stage 4) was found to be more effective.

The possibility of a Web-based intervention was a planned consideration of the research team as the aim of the LiveWell program, set out in the funding agreement, was to develop a personalized, scalable, sustainable, and potentially cost-effective intervention. However, a Web-based intervention was suggested as a potential mode of delivery for the intervention by several of the groups of stakeholders in this workshop. A strength of using co-design techniques is that they can be used to support stakeholders to explore the evidence base for a novel intervention but they can also be used to stimulate the creation of other intervention ideas. Although a Web-based intervention was a planned consideration of the research team, we were cognizant of the potential limitations related to the so-called digital divide and subsequent health inequalities in intervention access and use [65]. This issue was further explored in the formal pilot evaluation of LEAP (reported elsewhere).

The success of stage 2 co-design workshop 1, stage 3 co-design workshop 2, and stage 4 translating outputs into a design brief and specification, relied critically on contributions from our design expert. During stage 2, the design expert identified established co-design techniques, including persona-building and storyboarding, to facilitate the development of persona narratives within the scope of the qualitative evidence, and as a means to punctuate these narratives with opportunities to pursue the specific lifestyle behaviors identified in the systematic reviews. They also ensured that the evidence statements, written in scientific terms, were translated into plain English and presented in a visually engaging way so that they were accessible and interesting for all participants. During stage 3, the design expert guided the creative facilitators to sketch the intervention ideas generated and prototype potential interventions. When testing prototypes, the researchers observed that the level of fidelity of the prototypes was important; sketched ideas were easier for participants to engage with than more detailed mock-ups of one part of the intervention were, which was interpreted as a finished product inviting little useful feedback [66]. These visual aids served both as prompts for discussion in the workshop and as illustrations of design ideas for use in stage 4.

The challenge of stage 4 was to ensure that recurring design ideas contained in the prototypes were translated into specific features for tools, modules, or requirements of the intervention. Here, the design expert supported the production of a design brief and specification that reflected the outputs from previous stages while also detailing the intervention content and function that should be included in the intervention build.

A challenge of stage 6 co-design workshop 3 was that some user feedback suggested revisions to the prototype that were deemed unfeasible or non-essential by the research team and therefore were not addressed in the revisions. For example, idiosyncratic feedback about the value of particular modules or features indicated that not all parts of the intervention would be useful to all users. Rather than trying to anticipate which parts of the intervention would be most valuable to a user on the basis of their user profile, the team decided to emphasize in the intervention overview that LEAP is designed to allow a user to choose which modules or features to engage with, in an order of their choice. This would also allow an individual to revisit other parts of the intervention at a later date when perhaps their needs and priorities had changed.

The challenge of translating requirements from multiple perspectives and evidence sources (i.e., people with experience of retirement, organizations, and subject experts) alongside the scope and stated aims of the research program required that pragmatic compromises were made. Decisions were taken through discussion by the research team. In addition, the contracted Web-development company had many design decisions to make in the functional intervention build that were not directly influenced by co-design stakeholders until the prototype testing stages.

A strength of our approach is that it draws on the diverse skills of a multidisciplinary team with expertise in a range of research methodologies, including systematic reviewing, qualitative enquiry, and intervention co-design and development. Teams also need the subject expertise required to develop the particular intervention, which in our case included PA, nutrition and dietetics, social gerontology, and information technology expertise.

Comparison of our Intervention Development Approach With Other Approaches

Our sequential approach fills an important methodological gap in the complex health intervention development literature providing the description and appraisal of how to integrate systematic review, qualitative research, and other evidence with stakeholder engagement in a co-design process. Other approaches have advocated the integration of user perspectives in intervention development, demonstrating the importance of conducting qualitative research with a wide range of people from the target user populations at every stage of intervention development, from planning to feasibility testing and implementation [25,67]. In addition, the application of co-design techniques in health care intervention development has been demonstrated [8-11]. Our approach values the role of qualitative research and stakeholder input in intervention development but also details *how* to integrate systematic review evidence in the process, which is an important component of the MRC guidance [14]. Moreover, we provide detailed information on the stages and methods required to follow the approach to develop an intervention that is not only evidence-based but also fits the needs of intervention stakeholders, thereby increasing the likelihood of the intervention being acceptable and feasible. A 6-step guide [68] attempts to fill the methodological gap in the literature by providing a guide of how to develop public health interventions from defining the problem and identifying the modifiable causal and contextual factors through to collecting preliminary evidence of effectiveness. Our approach complements this guide describing the specific methods and co-design techniques that can be employed at each step.

As illustrated by our example of developing LEAP, the approach we have tested enables a clearly documented description of the intervention development process including the evidence on which each intervention feature/characteristic was based and the potential causal mechanisms of change in terms of BCTs used (see Table 1). Documenting the process in this way ensures that the intervention can be clearly described and reported, facilitating future replication. Thus our approach supports researchers to conform to the Template for Intervention Description and Replication (TIDieR) intervention reporting guidelines [69] and to develop an intervention logic model or "theory of change," which can direct an evaluation of the effectiveness of the intervention as advocated in the MRC guidance on process evaluation of complex interventions [70].

Implications of our Approach

The approach we present provides a sequential description of the methods needed to pursue evidence and theory-based complex health intervention development. The co-design techniques we employed, namely persona building [56], experience mapping [59], storyboarding [60], and prototyping [43], originate from product and service design adopting a social and user-centered process [4]. Co-design techniques have been used to involve stakeholders as co-designers in health care innovation and intervention development [5-11]. Our approach adds to this growing body of literature providing an explicit and replicable description of how to apply the techniques using the example of the development of LEAP, a complex lifestyle intervention for people in the retirement transition.

As discussed, the approach can be labor and time intensive. In the illustrative example of developing LEAP, a large proportion of the project timeline was attributed to delivering the outputs of stage 1. Conducting high-quality systematic reviews is a lengthy and resource-intensive process, which in our example, included additional work to identify the associations between intervention features and effectiveness. This was a necessary stage in the process as current systematic review evidence of interventions for PA, Mediterranean dietary patterns, and social roles for people in retirement did not exist. Where recently conducted, high-quality systematic reviews exist, these can be used to develop the evidence statements to inform intervention co-design (stage 2), maximizing the time and resources available for the later stages of designing, building, and de-risking the intervention.

Further Work

We have demonstrated that a sequential approach can be applied to the development of a Web-based lifestyle intervention for people in retirement. Further work is needed to apply this approach to other areas of health intervention development. Further application and refinement of this approach would help build evidence about its utility and acceptability. This in turn could support the development of formal guidance on this process.

The final output of our approach to intervention development is a functional prototype (in our case, the Web-based intervention LEAP) ready for formal testing to ascertain the effectiveness of the intervention. Web-based interventions have significant promise to reach the rapidly expanding older adult population who are increasingly becoming routine Internet users [71] and have been shown to have positive effects on lifestyle behaviors, including PA, in older adults [72-74]. The feasibility and acceptability of LEAP has been formally tested in a pilot randomized controlled trial (NCT02136381), which will be reported elsewhere. The pilot data will be used to inform the design of a definitive evaluation of the effectiveness and cost-effectiveness of LEAP.

Conclusions

This paper fills an important methodological gap in the complex health intervention development literature by describing and appraising a systematic, sequential approach to the co-design and development of an evidence-based complex health intervention. Using the example of the development of the LEAP intervention, we have illustrated the application of this approach and detailed the stages and techniques followed, integrating quantitative and qualitative evidence (derived from systematic

reviews and qualitative research), expert knowledge and experience, and stakeholder involvement.

Acknowledgments

We would like to thank the stakeholders who provided invaluable input during the co-design process and the other members of the LiveWell team for their contributions at different stages of the project.

This work was conducted as part of the LiveWell program (Grant Reference Number G0900686), which is funded through the Lifelong Health and Wellbeing cross-councils initiative in partnership with the following: Biotechnology and Biological Sciences Research Council; Engineering and Physical Sciences Research Council; Economic and Social Research Council; Medical Research Council; Chief Scientist Office of the Scottish Government Health Directorates; National Institute for Health Research/The Department of Health; The Health and Social Care Research & Development of the Public Health Agency (Northern Ireland); and Wales Office of Research and Development for Health and Social Care, Welsh Assembly Government.

FFS is funded by Fuse, the UK Clinical Research Collaboration Centre for Translational Research in Public Health, and, at the time of this study, MW was director of Fuse. MW is also a member of the Centre for Diet and Activity Research (CEDAR). Both Fuse and CEDAR are UK Clinical Research Collaboration (UKCRC) Public Health Research Centres of Excellence. Funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the National Institute for Health Research, and the Wellcome Trust under the auspices of the UKCRC, is gratefully acknowledged.

The views expressed in this paper are those of the authors and do not necessarily reflect those of the above named funders.

Authors' Contributions

JM, PM, MW, FS, and SM conceived the project and secured funding for the study. NO, BH, GT, and PM developed the study design. The methods were further developed and fieldwork was undertaken by NO, BH, GT, EE, CC, SM, and PM. Data analyses and interpretation were conducted by all authors. NO wrote the first draft of the paper. PM wrote revisions. All authors commented on drafts and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Further description of outputs of Stage 1 that underpinned evidence statements.

[PDF File (Adobe PDF File), 24KB - jmir_v18i8e210_app1.pdf]

Multimedia Appendix 2

Example of persona used in Stage 2 Meet Helen.

[JPG File, 73KB - jmir_v18i8e210_app2.JPG]

Multimedia Appendix 3

Example of personal used in Stage 2 Meet Jeff.

[JPG File, 70KB - jmir_v18i8e210_app3.JPG]

Multimedia Appendix 4

Summary of design brief and specification.

[PDF File (Adobe PDF File), 47KB - jmir_v18i8e210_app4.pdf]

Multimedia Appendix 5

Screenshot of LEAP welcome page.

[JPG File, 63KB - jmir_v18i8e210_app5.JPG]

Multimedia Appendix 6

Screenshot of mentor selector.

[JPG File, 104KB - jmir_v18i8e210_app6.JPG]

Multimedia Appendix 7

Screenshot of mentor overview.

[JPG File, 71KB - jmir_v18i8e210_app7.JPG]

Multimedia Appendix 8

Screenshot of My Time Module.

[JPG File, 67KB - jmir_v18i8e210_app8.JPG]

Multimedia Appendix 9

Screenshot of Changing Work Module.

[JPG File, 56KB - jmir_v18i8e210_app9.JPG]

Multimedia Appendix 10

Screenshot of Moving More Module.

[JPG File, 62KB - jmir_v18i8e210_app10.JPG]

Multimedia Appendix 11

Screenshot of Being Social Module.

[JPG File, 68KB - jmir_v18i8e210_app11.JPG]

Multimedia Appendix 12

Screenshot of Eating Well Module.

[JPG File, 60KB - jmir_v18i8e210_app12.JPG]

Multimedia Appendix 13

Screenshot of Diary.

[JPG File, 61KB - jmir_v18i8e210_app13.JPG]

Multimedia Appendix 14

Screenshot of dashboard.

[JPG File, 75KB - jmir_v18i8e210_app14.JPG]

References

1. Agency for Healthcare Research and Quality. The Effective Health Care Program Stakeholder Guide. Rockville, MD; 2014. URL: http://www.ahrq.gov/research/findings/evidence-based-reports/stakeholderguide/index.html [accessed 2016-07-13] [WebCite Cache ID 6iy3fNl2E]
2. Canadian Institutes of Health Research. A discussion paper for a 10-year plan to change health care using the levers of research. 2010. Strategy for Patient-Oriented Research URL: http://www.cihr-irsc.gc.ca/e/41232.html [accessed 2016-07-14] [WebCite Cache ID 6iztmKYcm]
3. Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, et al. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. Health Technol Assess 2004 Apr;8(15):1-148, III [FREE Full text] [Medline: 15080866]
4. Bjögvinsson E, Ehn P, Hillgren P. Design Things and Design Thinking: Contemporary Participatory Design Challenges. Design Issues 2012 Jul;28(3):101-116. [doi: 10.1162/DESI_a_00165]
5. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based Co-design and Healthcare Improvement: Realizing Participatory Design in the Public Sector. Design J 2015 Jun 01;18(2):227-248. [doi: 10.2752/175630615X14212498964312]
6. Bate S, Robert G, Maher L. Bringing User Experience to Healthcare Improvement: The Concepts, Methods and Practices of Experience-based Design. Oxford, UK: Radcliffe Publishing; 2007.

7. Robert G, Cornwell J, Locock L, Purushotham A, Sturmey G, Gager M. Patients and staff as codesigners of healthcare services. BMJ 2015 Feb 10;350(feb10 14):g7714-g7714. [doi: 10.1136/bmj.g7714]
8. Macdonald A, Teal G, Bamford C, Moynihan P. Hospitalfoodie: an interprofessional case study of the redesign of the nutritional management and monitoring system for vulnerable older hospital patients. Quality in primary care 2012;20(3):a.
9. Bowen S, McSeveny K, Lockley E, Wolstenholme D, Cobb M, Dearden A. How was it for you? Experiences of participatory design in the UK health service. CoDesign 2013 Dec;9(4):230-246. [doi: 10.1080/15710882.2013.846384]
10. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
11. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
12. Springham N, Robert G. Experience based co-design reduces formal complaints on an acute mental health ward. BMJ Quality Improvement Reports 2015;4(1):-. [doi: 10.1136/bmjquality.u209153.w3970]
13. Ferriman A. BMJ readers choose the “sanitary revolution” as greatest medical advance since 1840. BMJ 2007 Jan 20;334(7585):111-111. [doi: 10.1136/bmj.39097.611806.DB]
14. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29:a1655. [doi: 10.1136/bmj.a1655]
15. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. Soc Sci Med 2010 May;70(10):1618-1625. [doi: 10.1016/j.socscimed.2010.01.039] [Medline: 20207464]
16. 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]
17. Dombrowski S, Sniehotta F, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychology Review 2012 Mar;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
18. O'Brien N, McDonald S, Araújo-Soares V, Lara J, Errington L, Godfrey A, et al. The features of interventions associated with long-term effectiveness of physical activity interventions in adults aged 55-70 years: a systematic review and meta-analysis. Health Psychol Rev 2015;9(4):417-433. [doi: 10.1080/17437199.2015.1012177] [Medline: 25689096]
19. Lara J, Evans E, O'Brien N, Moynihan P, Meyer T, Adamson A, et al. Association of behaviour change techniques with effectiveness of dietary interventions among adults of retirement age: a systematic review and meta-analysis of randomised controlled trials. BMC Medicine 2014;12:a. [doi: 10.1186/s12916-014-0177-3]
20. Sniehotta FF, Dombrowski SU, Avenell A, Johnston M, McDonald S, Murchie P, et al. Randomised controlled feasibility trial of an evidence-informed behavioural intervention for obese adults with additional risk factors. PLoS One 2011;6(8):e23040 [FREE Full text] [doi: 10.1371/journal.pone.0023040] [Medline: 21897841]
21. Kok G, Gottlieb N, Peters G, Mullen P, Parcel G, Ruiter RAC, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2015 Oct 15:1-16. [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
22. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ 2014 Jun 13;348(Jun13 4):g3725-g3725. [doi: 10.1136/bmj.g3725]
23. Lemmer B, Grellier R, Steven J. Systematic Review of Nonrandom and Qualitative Research Literature: Exploring and Uncovering an Evidence Base for Health Visiting and Decision Making. Qualitative Health Research 1999 May 01;9(3):315-328. [doi: 10.1177/104973299129121884]
24. Lewin S, Glenton C, Oxman A. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ 2009;339:b3496 [FREE Full text] [Medline: 19744976]
25. Yardley L, Williams S, Bradbury K, Garip G, Renouf S, Ware L, et al. Integrating user perspectives into the development of a web-based weight management intervention. Clin Obes 2012 Oct;2(5-6):132-141. [doi: 10.1111/cob.12001] [Medline: 25586248]
26. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995 Jul 1;311(6996):42-45 [FREE Full text] [Medline: 7613329]
27. O?Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ 2010;341:-. [Medline: 20851841]
28. Moffatt S, White M, Mackintosh J, Howel D. Using quantitative and qualitative data in health services research - what happens when mixed method findings conflict? [ISRCTN61522618]. BMC Health Serv Res 2006;6:28 [FREE Full text] [doi: 10.1186/1472-6963-6-28] [Medline: 16524479]
29. Godfrey A, Lord S, Galna B, Mathers JC, Burn DJ, Rochester L. The association between retirement and age on physical activity in older adults. Age Ageing 2014 May;43(3):386-393 [FREE Full text] [doi: 10.1093/ageing/aft168] [Medline: 24171945]

30. Touvier M, Bertrais S, Charreire H, Vergnaud A, Hercberg S, Oppert JM. Changes in leisure-time physical activity and sedentary behaviour at retirement: a prospective study in middle-aged French subjects. International Journal of Behavioral Nutrition and Physical Activity 2010;7(1):a. [Medline: 20181088]
31. Zantinge EM, van den Berg M, Smit HA, Picavet S. Retirement and a healthy lifestyle: opportunity or pitfall? A narrative review of the literature. Eur J Public Health 2014 Jun;24(3):433-439 [FREE Full text] [doi: 10.1093/eurpub/ckt157] [Medline: 24148557]
32. Slingerland AS, van Lenthe FJ, Jukema JW, Kamphuis CBM, Looman C, Giskes K, et al. Aging, retirement, and changes in physical activity: prospective cohort findings from the GLOBE study. Am J Epidemiol 2007 Jun 15;165(12):1356-1363 [FREE Full text] [doi: 10.1093/aje/kwm053] [Medline: 17420180]
33. Berger U, Der G, Mutrie N, Hannah M. The impact of retirement on physical activity. Ageing and Society 1999;25(2):181-195. [doi: 10.1017/S0144686X04002739]
34. Nooyens ACJ, Visscher TLS, Schuit A, van Rossum CTM, Verschuren WMM, van Mechelen W, et al. Effects of retirement on lifestyle in relation to changes in weight and waist circumference in Dutch men: a prospective study. Public Health Nutr 2005 Dec;8(8):1266-1274. [Medline: 16372922]
35. Hobbis S, Thirlaway K, Sander L, Hendry L. Retirement and lifestyle behaviours: a thematic analysis of a pilot study. Health Psychology Update 2011;20(2):a.
36. Lee I, Shiroma E, Lobelo F, Puska P, Blair S, Katzmarzyk P. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet 2012 Jul;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
37. Sofi F, Abbate R, Gensini G, Casini A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr 2010 Nov;92(5):1189-1196 [FREE Full text] [doi: 10.3945/ajcn.2010.29673] [Medline: 20810976]
38. Holt-Lunstad J, Smith T, Layton J. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
39. Matthews F, Arthur A, Barnes L, Bond J, Jagger C, Robinson L, Medical Research Council Cognitive Function and Ageing Collaboration. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. Lancet 2013 Oct 26;382(9902):1405-1412 [FREE Full text] [doi: 10.1016/S0140-6736(13)61570-6] [Medline: 23871492]
40. World Health Organization. World Health Statistics. Geneva; 2013. URL: http://www.who.int/gho/publications/world_health_statistics/EN_WHS2013_Full.pdf [accessed 2016-07-14] [WebCite Cache ID 6izyezOVk]
41. Schäfer M, Jaeger-Erben M, Bamberg S. Life Events as Windows of Opportunity for Changing Towards Sustainable Consumption Patterns? J Consum Policy 2012;35(1):65-84.
42. Fujiwara Y, Sakuma N, Ohba H, Nishi M, Lee S, Watanabe N, et al. REPRINTS: Effects of an Intergenerational Health Promotion Program for Older Adults in Japan. Journal of Intergenerational Relationships 2009 Mar 12;7(1):17-39. [doi: 10.1080/15350770802628901]
43. Glass T, Freedman M, Carlson M, Hill J, Frick K, Ialongo N, et al. Experience Corps: design of an intergenerational program to boost social capital and promote the health of an aging society. J Urban Health 2004 Mar;81(1):94-105 [FREE Full text] [doi: 10.1093/jurban/jth096] [Medline: 15047788]
44. Saltz R. Aging Persons as Child-Care Workers in a Foster-Grandparent Program: Psychosocial Effects and Work Performance. The International Journal of Aging and Human Development 2010 Sep 21;2(4):314-340. [doi: 10.2190/AG.2.4.i]
45. Werkman A, Hulshof P, Stafleu A, Kremers S, Kok F, Schouten E. Effect of an individually tailored one-year energy balance programme on body weight, body composition and lifestyle in recent retirees: A cluster randomised controlled trial. BMC Public Health 2010;10:A. [doi: 10.1186/1471-2458-10-110]
46. Barnett I, van Sluijs EMF, Ogilvie D. Physical activity and transitioning to retirement: a systematic review. Am J Prev Med 2012 Sep;43(3):329-336 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.026] [Medline: 22898127]
47. Lara J, Hobbs N, Moynihan P, Meyer T, Adamson A, Errington L. Effectiveness of dietary interventions among adults of retirement age: a systematic review and meta-analysis of randomized controlled trials. BMC Medicine 2014;12(1):a. [doi: 10.1186/1741-7015-12-60]
48. Heaven B, Brown LJE, White M, Errington L, Mathers JC, Moffatt S. Supporting well-being in retirement through meaningful social roles: systematic review of intervention studies. Milbank Q 2013 Jun;91(2):222-287 [FREE Full text] [doi: 10.1111/milq.12013] [Medline: 23758511]
49. Hobbs N, Godfrey A, Lara J, Errington L, Meyer TD, Rochester L, et al. Are behavioral interventions effective in increasing physical activity at 12 to 36 months in adults aged 55 to 70 years? A systematic review and meta-analysis. BMC Med 2013;11:75 [FREE Full text] [doi: 10.1186/1741-7015-11-75] [Medline: 23506544]
50. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance (third edition). 2012. URL: https://www.nice.org.uk/article/pmg4/resources/non-guidance-methods-for-the-development-of-nice-public-health-guidance-third-edition-pdf [accessed 2016-07-14] [WebCite Cache ID 6izzXfCC9]
51. Pruitt J, Adlin T. The persona lifecycle: keeping people in mind throughout product design. Amsterdam: Elsevier; 2006.

52. Bate P, Robert G. Toward More User-Centric OD: Lessons From the Field of Experience-Based Design and a Case Study. The Journal of Applied Behavioral Science 2007 Mar 01;43(1):41-66. [doi: 10.1177/0021886306297014]
53. Buchenau M, Fulton S. Experience prototyping. : ACM; 2000 Presented at: Third Conference on designing interactive systems: Processes, practices, methods, and techniques; Aug. 17-19, 2000; Brooklyn, NY.
54. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Coughlan P, Suri J, Canales K. Prototypes as (Design) Tools for Behavioral and Organizational Change: A Design-Based Approach to Help Organizations Change Work Behaviors. The Journal of Applied Behavioral Science 2007 Mar 01;43(1):122-134. [doi: 10.1177/0021886306297722]
56. Heaven B, O?Brien N, Evans E, White M, Meyer T, Mathers J. Mobilizing Resources for Well-being: Implications for Developing Interventions in the Retirement Transition. The Gerontologist 2015:-. [doi: 10.1093/geront/gnu159]
57. Keeley L, Pikkel R, Quinn B, Walters H. Ten types of innovationthe discipline of building breakthroughs. New Jersey: John Wiley & Sons Inc; 2013.
58. Polson P, Lewis C, Rieman J, Wharton C. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. International Journal of Man-Machine Studies 1992 May;36(5):741-773. [doi: 10.1016/0020-7373(92)90039-N]
59. Moffatt S, Heaven B. 'Planning for uncertainty': narratives on retirement transition experiences. Ageing and Society 2016 Jan 28:1-20. [doi: 10.1017/S0144686X15001476]
60. Schwarzer R. Self-efficacy in the adoption and maintenance of health behaviors: theoretical approaches to a new model. In: Schwarzer R, editor. Self-efficacy: Thought control of action. Washington, DC: Hemisphere; 1992:217-243.
61. Lange D, Richert J, Koring M, Knoll N, Schwarzer R, Lippke S. Self-regulation prompts can increase fruit consumption: a one-hour randomised controlled online trial. Psychol Health 2013;28(5):533-545. [doi: 10.1080/08870446.2012.751107] [Medline: 23282217]
62. Sniehotta F, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
63. Sniehotta F, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul;35(4):565-576. [doi: 10.1002/ejsp.258]
64. Kwasnicka D, Presseau J, White M, Sniehotta F. Does planning how to cope with anticipated barriers facilitate health-related behaviour change? A systematic review. Health Psychology Review 2013 Sep;7(2):129-145. [doi: 10.1080/17437199.2013.766832]
65. 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]
66. Rudd J, Stern K, Isensee S. Low vs. high-fidelity prototyping debate. interactions 1996;3(1):76-85. [doi: 10.1145/223500.223514]
67. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
68. Wight D, Wimbush E, Jepson R, Doi L. Six steps in quality intervention development (6SQuID). Journal of Epidemiology and Community Health 2015:-. [doi: 10.1136/jech-2015-205952]
69. Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
70. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258-h1258. [doi: 10.1136/bmj.h1258]
71. Harte RP, Glynn LG, Broderick BJ, Rodriguez-Molinero A, Baker P, McGuiness B, et al. Human centred design considerations for connected health devices for the older adult. J Pers Med 2014;4(2):245-281 [FREE Full text] [doi: 10.3390/jpm4020245] [Medline: 25563225]
72. Aalbers T, Baars M, Rikkert M. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
73. Mouton A, Cloes M. Web-based interventions to promote physical activity by older adults: promising perspectives for a public health challenge. Arch Public Health 2013;71(1):16 [FREE Full text] [doi: 10.1186/0778-7367-71-16] [Medline: 23819885]
74. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]

Abbreviations

BCT: behavior change technique
LEAP: Living, Eating, Activity, and Planning in retirement
MRC: Medical Research Council
PA: physical activity

Edited by G Eysenbach; submitted 11.04.16; peer-reviewed by A Bub, L Quinlan; comments to author 04.05.16; accepted 31.05.16; published 03.08.16.

Please cite as:
O'Brien N, Heaven B, Teal G, Evans EH, Cleland C, Moffatt S, Sniehotta FF, White M, Mathers JC, Moynihan P
Integrating Evidence From Systematic Reviews, Qualitative Research, and Expert Knowledge Using Co-Design Techniques to Develop a Web-Based Intervention for People in the Retirement Transition
J Med Internet Res 2016;18(8):e210
URL: http://www.jmir.org/2016/8/e210/
doi:10.2196/jmir.5790
PMID:27489143

Original Paper

Long-Term Effects of an Internet-Mediated Pedometer-Based Walking Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial

Marilyn L Moy[1,2], MD; Carlos H Martinez[3], MD; Reema Kadri[4,5], MLIS; Pia Roman[5], MA; Robert G Holleman[5], MPH; Hyungjin Myra Kim[5,6], ScD; Huong Q Nguyen[7], PhD; Miriam D Cohen[8], MSN; David E Goodrich[5], EdD; Nicholas D Giardino[9], PhD; Caroline R Richardson[4,5], MD

[1]Pulmonary and Critical Care Medicine Section, VA Boston Healthcare System, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]Pulmonary & Critical Care Division, University of Michigan Health System, Ann Arbor, MI, United States
[4]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States
[5]Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States
[6]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States
[7]Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States
[8]Pulmonary Medicine Section, VA New York Harbor, Brooklyn, NY, United States
[9]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Caroline R Richardson, MD
Department of Family Medicine
University of Michigan
1018 Fuller St.
Ann Arbor, MI, 48104
United States
Phone: 1 734 998 7120 ext 316
Fax: 1 734 998 7335
Email: caroli@umich.edu

Abstract

Background: Regular physical activity (PA) is recommended for persons with chronic obstructive pulmonary disease (COPD). Interventions that promote PA and sustain long-term adherence to PA are needed.

Objective: We examined the effects of an Internet-mediated, pedometer-based walking intervention, called Taking Healthy Steps, at 12 months.

Methods: Veterans with COPD (N=239) were randomized in a 2:1 ratio to the intervention or wait-list control. During the first 4 months, participants in the intervention group were instructed to wear the pedometer every day, upload daily step counts at least once a week, and were provided access to a website with four key components: individualized goal setting, iterative feedback, educational and motivational content, and an online community forum. The subsequent 8-month maintenance phase was the same except that participants no longer received new educational content. Participants randomized to the wait-list control group were instructed to wear the pedometer, but they did not receive step-count goals or instructions to increase PA. The primary outcome was health-related quality of life (HRQL) assessed by the St George's Respiratory Questionnaire Total Score (SGRQ-TS); the secondary outcome was daily step count. Linear mixed-effect models assessed the effect of intervention over time. One participant was excluded from the analysis because he was an outlier. Within the intervention group, we assessed pedometer adherence and website engagement by examining percent of days with valid step-count data, number of log-ins to the website each month, use of the online community forum, and responses to a structured survey.

Results: Participants were 93.7% male (223/238) with a mean age of 67 (SD 9) years. At 12 months, there were no significant between-group differences in SGRQ-TS or daily step count. Between-group difference in daily step count was maximal and statistically significant at month 4 (P<.001), but approached zero in months 8-12. Within the intervention group, mean 76.7% (SD 29.5) of 366 days had valid step-count data, which decreased over the months of study (P<.001). Mean number of log-ins

to the website each month also significantly decreased over the months of study (P<.001). The online community forum was used at least once during the study by 83.8% (129/154) of participants. Responses to questions assessing participants' goal commitment and intervention engagement were not significantly different at 12 months compared to 4 months.

Conclusions: An Internet-mediated, pedometer-based PA intervention, although efficacious at 4 months, does not maintain improvements in HRQL and daily step counts at 12 months. Waning pedometer adherence and website engagement by the intervention group were observed. Future efforts should focus on improving features of PA interventions to promote long-term behavior change and sustain engagement in PA.

ClinicalTrial: Clinicaltrials.gov NCT01102777; https://clinicaltrials.gov/ct2/show/NCT01102777 (Archived by WebCite at http://www.webcitation.org/6iyNP9KUC)

(J Med Internet Res 2016;18(8):e215) doi:10.2196/jmir.5622

KEYWORDS

bronchitis, chronic; emphysema; pulmonary disease, chronic obstructive; quality of life; exercise; motor activity; Internet

Introduction

Physical activity (PA) is significantly reduced in persons with chronic obstructive pulmonary disease (COPD), even at the earliest stages of disease [1-3]. Its clinical course is punctuated with acute exacerbations, during and following which persons suffer further reductions in PA [4,5]. As a disease with systemic consequences, COPD increases vulnerability to frailty, immobility, and loss of functional independence. Despite optimal pharmacological therapy, persons with COPD suffer from a downward spiral of breathlessness, deconditioning, and physical inactivity [6]. Comorbidities of cardiovascular disease, diabetes mellitus, and osteoporosis contribute to further reductions in PA [7,8].

Physical activity is a modifiable health behavior that affects COPD-specific outcomes [9-14]. It has been shown that a greater quantity of low-intensity PA reduces risk of COPD hospitalizations, whereas high-intensity PA does not result in risk reduction [15]. In a cohort of persons with COPD, those who walk the least have risks that are 2 and 6 times higher for acute exacerbations and COPD-related hospitalizations, respectively, compared to those who walk the most [12]. In addition, persons with COPD with higher PA levels have a significantly lower risk of dying, independent of forced expiratory volume in 1 second (FEV_1) [14]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines recommend regular PA for all persons with stable COPD as part of standard nonpharmacological treatment [6].

Despite the evidence and recommendations, effective long-term PA interventions are lacking in the clinical care of patients with COPD. Most studies of long-term exercise interventions have examined methods to maintain exercise in the subset of persons with COPD who have completed a conventional pulmonary rehabilitation program [16-21]. These interventions have combined weekly- or monthly-supervised exercise classes with unsupervised home exercise, support groups, and/or telephone contact with a health care professional, showing mixed results over the long term [16-21]. Strategies that promote behavior change and long-term adherence to effectively sustain PA in all persons with COPD are needed.

We developed an automated, Internet-mediated, pedometer-based walking program called Taking Healthy Steps to promote PA in persons with COPD. Taking Healthy Steps combines the Omron HJ-720 ITC pedometer (Omron Healthcare, Inc, Bannockburn, IL, USA) with a disease-specific website accessed via a URL. Taking Healthy Steps provides iterative step-count feedback, individualized step-count goals, education on disease self-management, motivational support, and an online community of social support [22-27]. We studied the efficacy of Taking Healthy Steps in a randomized controlled trial (trial registration: Clinicaltrials.gov NCT01102777) [27]. The conceptual framework, study design, and results at 4 months have been described previously [26,27]. We have shown that Taking Healthy Steps is safe and engaging, and improves health-related quality of life (HRQL) and increases daily step count at 4 months [25-27]. In this study, our primary aim was to assess the long-term efficacy of Taking Healthy Steps on HRQL and daily step counts, a marker for walking behavior change, at 12 months. Our secondary aim was to assess long-term engagement with the PA intervention.

Methods

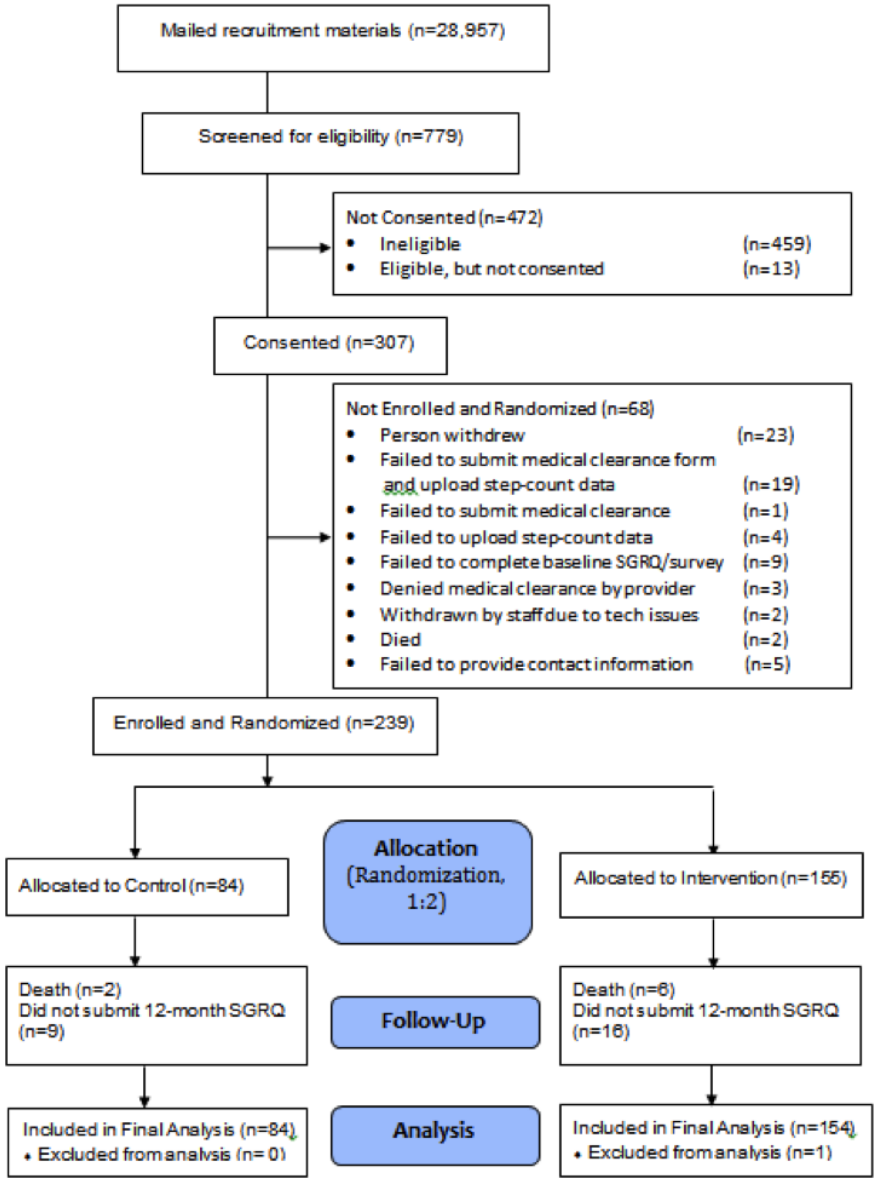
Recruitment

The study design and methods have been reported previously [26,27]. Participants were enrolled from national patient care databases of US Veterans, between December 2011 and January 2013, who had received any treatment services in the previous year and had a COPD diagnosis. Zip codes were matched with the Rural Urban Commuting Area Codes to determine whether one's residence was urban or rural [28]. Of the 21 regional Veteran Integrated Service Networks (VISN) across the 50 United States and Puerto Rico, we excluded Veterans from one VISN (VISN-1) where another COPD research study using the Taking Healthy Steps platform was recruiting participants. The coordinating center was located at the Ann Arbor VA Healthcare System, Ann Arbor, MI, USA. Ethical approval for this study was granted by the VA Ann Arbor Healthcare System Human Studies Subcommittee.

A random sample of 28,957 Veterans (half rural, half urban) with a COPD diagnosis was sent a recruitment letter. Inclusion criteria included having access to a computer with an Internet connection, a USB port, and Windows XP, Vista, Windows 7, or Windows 8. Our a priori exclusion criteria excluded those who did not upload baseline step-count data or who did not

complete the baseline survey to assess HRQL. Per study protocol, participants had to have baseline values for the primary (HRQL) and secondary outcome (daily step count) to be enrolled and randomized. Ultimately, 239 participants were enrolled and randomized in a 2:1 ratio to either Taking Healthy Steps (Internet-mediated, pedometer-based walking program) or wait-list control (pedometer alone), stratified by Modified Medical Research Council (MMRC) dyspnea score and urban versus rural status (Figure 1). All participants were prompted monthly to report new or worsening medical problems; all self-reported adverse events were recorded. There were no face-to-face encounters with staff; all features were automatically delivered via the website.

Figure 1. CONSORT diagram at 12 months.



Outcomes

Primary Outcome

The St. George's Respiratory Questionnaire (SGRQ), a disease-specific instrument with 50 items that has been well validated in COPD [29,30] was used to assess HRQL. It has a summary total score (SGRQ-TS) composed of three domain scores: symptoms (frequency and severity), activities (that cause or are limited by breathlessness), and impact (social functioning and psychological disturbances resulting from airways disease). Scores range from 0 to 100 with lower scores indicating better HRQL. A change of four units is the minimum clinically important difference for the SGRQ-TS [31]. Study participants completed the SGRQ online at study entry, 4 months, and 12 months.

Secondary Outcome

Daily step count was assessed by the Omron HJ-720 ITC pedometer. Once participants completed the baseline survey, study staff mailed them a pedometer that had an embedded USB port, an upload cable, and detailed written instructions on how to install the Java software and upload pedometer data. For users who did not have Java already installed on their computers, the software installation was a one-time event. Thereafter, participants uploaded step-count data using the cable that

connected the pedometer to their home computer. Research staff were available by telephone to assist with software installation and upload of step counts.

A wear day with valid step-count data was defined as one having at least 100 steps and 8 hours of step counts recorded [32]. At baseline, participants wore the pedometer covered with a sticker to blind the participant to device feedback. Baseline daily step count was the mean daily step count calculated using at least 5 days of valid data within a period of seven consecutive days. Follow-up daily step counts were calculated within a window of +/-14 days around day 121 for 4-month values, and +/-14 days around day 366 for 12-month values. Follow-up daily step counts were the means of at least 5 days of valid data within a period of seven consecutive days. We also calculated the mean daily step count each month by examining the data in 30-day increments. We used values from the last valid week (at least 5 days of valid data within a period of seven consecutive days) in each of those months.

Intervention Group

Participants randomized to Taking Healthy Steps completed an intensive 4-month intervention period, followed by a distinct 8-month maintenance phase (Table 1). During the first 4 months, participants were instructed to wear the pedometer every day, reminded to upload at least weekly, and were provided access to the website. The website has four key components [26,27]: individualized goal setting was based on uploaded step counts, iterative feedback allowed self-monitoring of step counts, motivational content provided a new educational tip every other day and a new motivational message each week, and an online community forum enhanced social support [22-27]. During the 8-month maintenance phase, participants continued to wear the pedometer, upload daily step counts, receive weekly step-count goals and feedback, and had access to the online community forum. They could view the initial 4 months of educational content and motivational messages, but no longer received new content. Topics on the online community forum included walking in a variety of weather/seasons, health topics (weight management, COPD disease management), injury prevention, barriers to walking, and technical issues with the pedometer and website.

Table 1. Features available to the Taking Healthy Steps and control groups during the first 4 months versus last 8 months of the study.

Features	0-4 Months		5-12 Months	
	Taking Healthy Steps	Control	Taking Healthy Steps	Control
Wear pedometer	Yes	Yes	Yes	Yes
Upload step-count data	At least weekly	At least monthly	At least weekly	At least monthly
Goal setting	Yes	No	Yes	No
Feedback	Yes	No	Yes	No
New educational and motivational content	Yes	No	No	No
Online community forum	Yes	No	Yes	No

Wait-List Control Group

Participants randomized to the wait-list control group were instructed to wear the pedometer every day, reminded monthly to log in to the website to upload step-count data, and asked to report all adverse events. Veterans in the wait-list control group received neither instruction to increase PA nor step-count goals. They had access to a webpage that showed only a checklist of surveys completed and a count of what week they were in the study. After 12 months, they were given the option to use the Internet-mediated intervention.

Participant Characteristics, Device Adherence, and Website Engagement

At baseline, participants answered questions online that assessed comorbidities, oxygen use, smoking status, and demographics. At study entry, 4 months, and 12 months, dyspnea was assessed using the MMRC scale (range 0-4 with 4 indicating the most severe level of dyspnea) [33]. Events self-reported during the study were defined a priori as COPD-related if persons experienced a combination of symptoms and/or required treatment with antibiotics and/or systemic corticosteroids. The COPD-related events included acute exacerbations or pneumonia, ascertained by self-reported events and/or review of health care utilization (hospitalizations and emergency room visits) and pharmacy data. To assure independence of individual acute exacerbations, participants were considered to have experienced a new acute exacerbation only if it were reported 21 or more days after the previous acute exacerbation [34].

We examined device adherence, overall and by group, by calculating the percentage of days (of 366 days) that were wear days with valid step-count data. For the participants who uploaded valid step-count data at 12 months and completed the study, we also examined percentage of days (of 42 days) that were wear days during the last 6 weeks of the study.

In the intervention group, we objectively examined website engagement by recording the number of log-ins to the website by month of study and assessing the frequency of use of the online community forum. In addition, at 4 and 12 months, participants in the Taking Healthy Steps group answered a structured survey eliciting feedback about their commitment to their step-count goals and various aspects of engagement with the intervention, including participants' ease of finding time to log in to the website, knowledge of step-count goals, and use of the different components of the website.

Statistical Analysis

Proportions, means, and standard deviations described baseline participant characteristics. Two-sample *t* tests and chi-square tests compared baseline characteristics between groups. The occurrence of COPD-related events (acute exacerbations or pneumonia), hospitalizations, emergency room visits, deaths, and adverse events during the study were each compared between groups using a logistic regression model. For the count of hospitalizations, a zero-inflated Poisson regression model was also used to assess the difference in the rate of hospitalizations between groups. These models adjusted for age, gender, treatment group, and oxygen use.

The primary analysis used the intention-to-treat approach, and used a linear mixed-effects model with baseline, 4-month, and 12-month outcome values (eg, SGRQ-TS or daily step count) as dependent variables. No baseline variable was predictive of missingness in models adjusting for stratification variables and treatment group. Thus, the longitudinal data model included participants who had the dependent variable for at least one time point and was expected to give unbiased estimates of the intervention effect assuming missingness at random. The model included participants as random intercepts to adjust for within-participant correlations of repeated measures, fixed predictors of treatment group, 4- and 12-month time indicators, and treatment group by time indicator interactions, MMRC dyspnea score (dichotomized to 0-1 vs 2-4), and urban versus rural residence. We also analyzed the data excluding those who died. The proportion of participants who had at least a 4-unit improvement in SGRQ-TS at 12 months was compared between groups using a chi-square test [31]. For the analysis of mean daily step count by month of study, we used a linear mixed-effect model similar to that for the primary outcome except data were assessed in 30-day increments over the 12-month study period. Predictors were treatment group, month of study as indicator variables (coded as 1-12), group-by-month indicator variables interactions, dichotomized MMRC dyspnea score, and urban versus rural residence.

We assessed website engagement in the intervention group by characterizing the number of log-ins to the website using the mean, median, and interquartile range, and assessed trends over months of study using a linear mixed-effects regression analysis with monthly number of log-ins for each participant as the outcome and time (month since randomization) as the predictor. Trends for device adherence over month of study were examined with percent of days with valid step-count data using a linear mixed-effects model and for use of the online community forum using a generalized mixed-effects model with logit link. The effect of time on participant responses to the online survey about goal commitment and intervention engagement at 4 and 12 months was estimated for each response variable using a mixed-effects model with 4- and 12-month survey data as the dependent variable and predictors including 12-month indicator, baseline dichotomized MMRC dyspnea score, and urban versus rural status. All models, including the model for the number of log-ins, were checked for model assumptions using residuals.

One participant in the Taking Healthy Steps group was considered an outlier given that his change in SGRQ-TS was 4.0 standard deviations greater than the mean for change in SGRQ-TS and his change in daily step count was 8.1 standard deviations greater than the mean for change in daily step count. The extremely high step counts more likely reflected his occupational PA rather than any effects of our intervention. Our main analyses excluded the outlying individual, but we also repeated primary and secondary outcome analyses with this participant included. All analyses were performed with Stata 14.0 (StataCorp LP, College Station, TX, USA).

Results

Participant Characteristics

No information is available on the persons to whom we mailed recruitment materials but who were not screened because they did not go to our website and did not call us (Figure 1). The top three reasons for ineligibility of 459 participants were not sedentary (n=202), could not walk a block (n=120), or no compatible computer access (n=161), with some participants having more than one reason (Figure 1). In all, 68 persons consented but were not enrolled and randomized, including 19 who failed to submit a medical clearance form and did not upload step-count data, one who failed to submit a medical clearance form, four who failed to upload step-count data, and nine who failed to complete the baseline SGRQ (Figure 1).

Participants' (N=238) characteristics include: mean age 67 (SD 9) years, male (93.7%, 223/238), rural residence (45.4%, 108/238), MMRC dyspnea score ≥2 (30.7%, 73/238), current smokers (24.8%, 59/238), and supplemental oxygen use (23.5%, 56/238) (Table 2). There were no significant differences in baseline characteristics between study groups, including current smoking history. Overall, 87.8% (209/238) of participants completed the 12-month online HRQL assessment, and 74.4% (177/238) uploaded 12-month valid step-count data. In the intervention group, 87.7% (135/154) of participants completed the HRQL assessment and 76.6% (118/154) uploaded valid step-count data, compared to 88% (74/84) and 70% (59/84), respectively, in the control group.

Table 2. Baseline participant characteristics (N=238).

Characteristic	Intervention (n=154)	Control (n=84)	Total (N=238)
Age (years), mean (SD)	67 (8.6)	66.4 (9.2)	66.8 (8.8)
Gender (male), n(%)	146 (94.8)	77 (92)	223 (93.7)
Residence, n(%)			
Urban	83 (53.9)	47 (56)	130 (54.6)
Rural	71 (46.1)	37 (44)	108 (45.4)
Hispanic (n=235), n(%)	5 (3.3)	1 (1)	6 (2.6)
Race, n(%)			
Black	7 (4.6)	3 (4)	10 (4.2)
White	142 (92.2)	79 (94)	221 (92.9)
Other	5 (3.3)	2 (2)	7 (2.9)
Current smoker, n(%)	41 (26.6)	18 (21)	59 (24.8)
Oxygen use, n(%)	35 (22.7)	21 (25)	56 (23.5)
SGRQ,[a] mean (SD)			
Symptoms	57.2 (19.1)	56 (19.9)	56.8 (19.3)
Activities	62.3 (20.2)	64.2 (18)	62.9 (19.5)
Impact	32.2 (16.5)	34.1 (17.9)	32.9 (17)
Total	45.6 (15.4)	46.8 (15.6)	46 (15.4)
Baseline daily step count, mean (SD)	3488 (2316)	3521 (2058)	3499 (2224)
MMRC dyspnea score,[b] n (%)			
0-1	108 (70.1)	57 (68)	165 (69.3)
2-4	46 (29.9)	27 (32)	73 (30.7)

[a]SGRQ: St. George's Respiratory Questionnaire. Data for symptoms, activities, and impact were available from 236 participants; total from 233 participants.

[b]MMRC: Modified Medical Research Council.

At 12 months, 29 of 238 (12.2%) participants did not have sufficient data to calculate the SGRQ-TS: 19 Taking Healthy Steps participants and 10 controls. There was no significant difference in baseline SGRQ-TS (mean 49.8, SD 16.1 vs mean 45.6, SD 15.3; *P*=.18) or baseline daily step count (mean 3410, SD 2667 vs mean 3512, SD 2163; *P*=.82) between those for whom SGRQ-TS could not be calculated (n=29) versus those for whom SGRQ-TS was calculated at 12 months (n=209).

The percent of participants with COPD-related events (acute exacerbations or pneumonia) during the study did not differ between groups (control: 18%, 15/84; intervention: 22.7%, 35/154; logistic regression OR 1.4, 95% CI 0.7-2.8; *P*=.33). No between-group difference was found in the percent of participants with hospitalizations (control: 17%, 14/84; intervention: 23.4%, 36/154; logistic regression OR 1.6, 95% CI 0.8-3.2; *P*=.19) or emergency room visits (control: 24%, 20/84; intervention: 29.9%, 46/154; logistic regression OR 1.4, 95% CI 0.8-2.6; *P*=.27) during the 12-month study. For the count of hospitalizations, a zero-inflated Poisson regression model also found no between-group difference. The percent of participants who died during the study did not differ between groups (control: 2%, 2/84; intervention: 3.9%, 6/154; *P*=.53). Finding no between-group differences in the percentage of participants who were hospitalized or died provided assurance that the censoring of the outcome variables (SGRQ-TS or daily step counts) due to these events was not likely to confound the assessment of the between-group outcome differences. However, we repeated the analyses with deaths excluded as well.

Health-Related Quality of Life

There was no significant between-group difference in the primary outcome of SGRQ-TS (mean 1.1 units, 95% CI –2.2 to 4.5; *P*=.50) at 12 months (Table 3). The proportion of participants who achieved at least a 4-unit improvement in SGRQ-TS at 12 months was 45.2% (61/135) in the intervention versus 32% (23/71) in the control group (*P*=.08). There was no significant between-group difference in the SGRQ domain scores of symptoms (mean 0.5 unit, 95% CI –4.2 to 5.2; *P*=.84), activities (mean 0.04 unit, 95% CI –4.2 to 4.2; *P*=.99), and impact (mean 2.3 units, 95% CI –1.6 to 6.1; *P*=.25) at 12 months.

Table 3. Within-group changes and between-group differences in SGRQ scores and daily step counts at 12 months.

Outcome and arm	N	Difference from baseline to 12 months, mean (95% CI)	*P*	Between-group difference, mean (95% CI)	*P* [a]
SGRQ					
Total				1.1 (–2.2, 4.5)	.50
Taking Healthy Steps	154	–2.5 (–4.5, –0.6)	.01		
Control	84	–1.4 (–4.1, 1.3)	.31		
Symptoms				0.5 (–4.2, 5.2)	.84
Taking Healthy Steps	154	–3.2 (–6.0, –0.4)	.02		
Control	84	–2.7 (–6.5, 1.1)	.16		
Activities				0.04 (–4.2, 4.2)	.99
Taking Healthy Steps	154	–1.2 (–3.7, 1.3)	.36		
Control	84	–1.1 (–4.5, 2.3)	.51		
Impact				2.3 (–1.6, 6.1)	.25
Taking Healthy Steps	154	–3.4 (–5.6, –1.1)	.004		
Control	84	–1.1 (–4.2, 2.0)	.48		
Daily step count				–108 (–720, 505)	.73
Taking Healthy Steps	154	270 (–86, 626)	.14		
Control	84	163 (–336, 661)	.52		

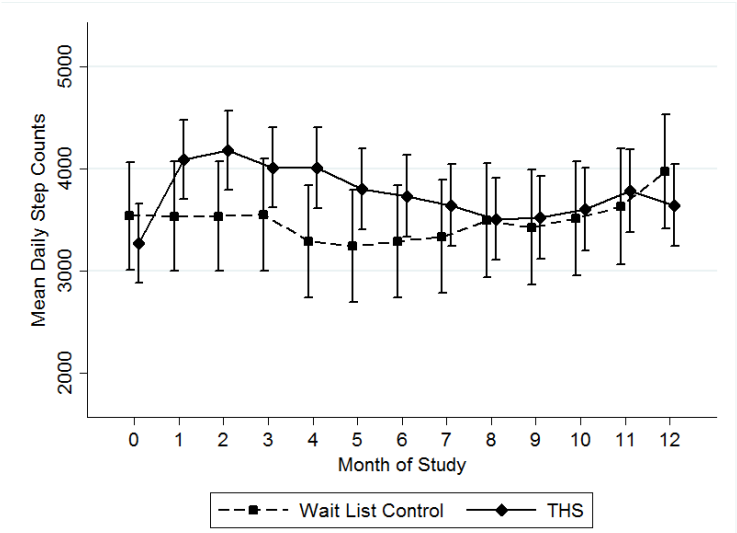
[a] Based on linear mixed-effect models, adjusting for group, 4- and 12-month indicators, group×time indicator interactions, baseline MMRC dyspnea score (dichotomized to 0-1 vs 2-4), and urban versus rural status.

Intervention participants showed an improvement in SGRQ-TS of a mean 2.5 units (95% CI –4.5 to –0.6) at 12 months, compared to baseline (*P*=.01) (Table 3). For domain scores in the intervention group, symptoms improved by a mean 3.2 units (95% CI –6.0 to –0.4, *P*=.02), and impact improved by a mean 3.4 units (95% CI –5.6 to –1.1, *P*=.004) at 12 months. The control group showed no significant changes in the SGRQ-TS and domain scores at 12 months compared to baseline (Table 3). When the analysis was repeated with the outlying individual included, no substantive difference was seen in results, except improvement in symptoms within the Taking Healthy Steps group was marginally significant (*P*=.05). When the analysis excluded the eight deaths, results remained nearly identical.

Daily Step Count

There was no significant difference between groups with respect to the secondary outcome of daily step count at 12 months (*P*=.73) (Table 3). There was no significant change in daily step count in the intervention participants (*P*=.14) or in the control group (*P*=.52) at 12 months, compared to baseline (Table 3). Examination of daily step count by month of intervention showed that differences in daily step counts in the intervention group compared to controls were maximal and statistically significant at month 4, but approached zero in months 8 to 12 (Figure 2). Between-group *P* values were <.001 at 4 months, .28 at 8 months, and .82 at 12 months. Within the intervention group, although daily step counts peaked at 2 months and then declined over the course of the study, daily step counts continued to be higher than baseline values in all months of the study (Figure 2). Analysis including the outlying individual showed improvement in daily step counts at 12 months to be significant in the Taking Healthy Steps group (*P*=.048). Analysis excluding the eight deaths did not change results.

Figure 2. Mean daily step count and 95% confidence intervals by month of study. Note: the Taking Healthy Steps (THS) curve is shifted to the right of the control curve on the x-axis for ease of visual display. Baseline data are included at month zero.



Device Adherence and Website Engagement

Device adherence during the 12-month study period was significantly higher in the intervention group than the control group, with mean 76.7% (SD 29.5) of the 366 days having valid step-count data in the intervention group versus mean 63.7% (SD 32.9) of the 366 days having valid step-count data in the control group (P=.002). For the 177 participants in both groups who uploaded valid step-count data at 12 months and completed the study, mean 83.1% (SD 21.9) of days in the last 6 weeks of the study had valid step-count data. In these last 6 weeks, mean 87.5% (SD 16.5) of days had valid step-count data in the intervention group, which was significantly higher than the mean 74.1% (SD 28.1) of days observed in the control group (P<.001). In the intervention group, device adherence decreased significantly over time (P<.001), with mean 92.1% (95% CI 86.6-97.6) of days having valid step-count data at month 1 versus 70.3% (95% CI 64.9-75.8) of days at month 12 (Figure 3).

In the intervention group, mean number of log-ins to the website decreased significantly over the months of study (P<.001; Figure 4). The number of monthly log-ins was mean 6.8 (SD 3.7; median 6, IQR 3) at month 1, which declined to mean 4.2 (SD 3.5; median 4, IQR 3) by month 9 and mean 3.0 (SD 3.0; median 3, IQR 5) by month 12 (Figure 4). In the intervention group, 83.8% (129/154) of the participants used the online community forum at some point during the 12-month study; 66.2% (102/154) of participants directly viewed an online community forum thread or entry, and an additional 17.5% (27/154) of participants posted a new topic or a reply at least once. More than half of the participants responded “definitely true” (22/121, 18.2%) or “mostly true” (45/121, 37.2%) to the statement: “I learned helpful information when I used the online community forum.” There was a significant trend for decreasing use of the online community forum by month of study (P<.001).

Responses to questions regarding participant’s goal commitment were not significantly different at 12 months compared to 4 months (Table 4). When asked, “Overall, how motivated are you to walk each day?” with responses from 1=not motivated and 10=extremely motivated, the mean response was 6.8 (SD 2.3) at 4 months compared to mean 6.5 (SD 2.5) at 12 months (P=.06). Responses to questions about engagement with the use of Taking Healthy Steps were not significantly different at 12 months compared to 4 months (Table 4).

Table 4. Goal commitment and engagement with Taking Healthy Steps intervention.

Goal commitment and engagement	N[a]	4 months mean (95% CI)	12 months mean (95% CI)	*P* [b]
Goal commitment[c]				
It's hard to take my step-count goal seriously.	147	2.1 (1.9-2.2)	2.0 (1.9-2.2)	.69
Quite frankly, I don't care if I reach my step goal or not.	147	1.7 (1.6-1.8)	1.7 (1.6-1.8)	.46
I am strongly committed to pursuing my step-count goal.	146	3.8 (3.6-4.0)	3.7 (3.5-3.9)	.52
It wouldn't take much to make me abandon my step-count goal.	147	1.9 (1.7-2.1)	2.0 (1.8-2.2)	.27
I think my step-count goal is a good goal to shoot for.	146	4.0 (3.8-4.2)	3.9 (3.8-4.1)	.77
Engagement in Taking Healthy Steps[d]				
I would recommend the Taking Healthy Steps walking program to another person with COPD.	146	1.3 (1.2-1.4)	1.2 (1.1-1.3)	.01
It was easy for me to find the time to log in to the website once a week.	146	1.8 (1.6-2.0)	1.8 (1.6-2.0)	.92
I had technical difficulty uploading step-count data from the pedometer to my computer.	146	4.0 (3.7-4.2)	3.9 (3.7-4.1)	.75
I knew what my step goal should be every day.	147	1.5 (1.4-1.6)	1.5 (1.4-1.6)	.48
I was able to comfortably increase my daily step count every week.	147	2.6 (2.5-2.8)	2.8 (2.6-3.0)	.10
I looked at the graphs of the step counts that I walked.	147	1.6 (1.4-1.7)	1.6 (1.4-1.7)	.76
The motivational messages and educational tips were easy to understand.	143	1.9 (1.8-2.0)	1.8 (1.7-1.9)	.21
I learned helpful information when I used the online community forum.	137	2.5 (2.3-2.7)	2.4 (2.3-2.6)	.52
The daily step-count goals were too high for me to walk each day.	147	3.4 (3.2-3.5)	3.4 (3.2-3.5)	.98

[a] Participants with responses at 4 and/or 12 months were included in the models.

[b] Based on linear mixed-effect models with 4 and 12 months as the dependent variable and predictors of 12-month indicator, intervention group indicator and baseline MMRC dyspnea score (dichotomized to 0-1 vs 2-4) and urban versus rural status.

[c] Response scale 1-5 with 1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, 5=strongly agree.

[d] Response scale 1-5 with 1=definitely true, 2=mostly true, 3=not sure, 4=mostly false, 5=definitely false.

Figure 3. Percentage and 95% confidence intervals of days with valid step-count data in the intervention group by month of study.

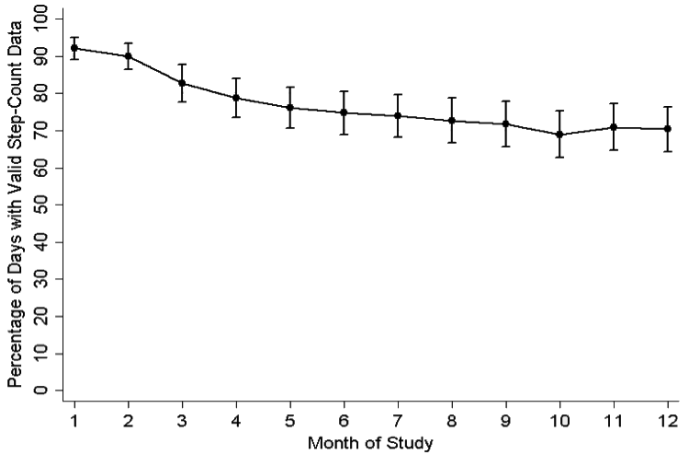
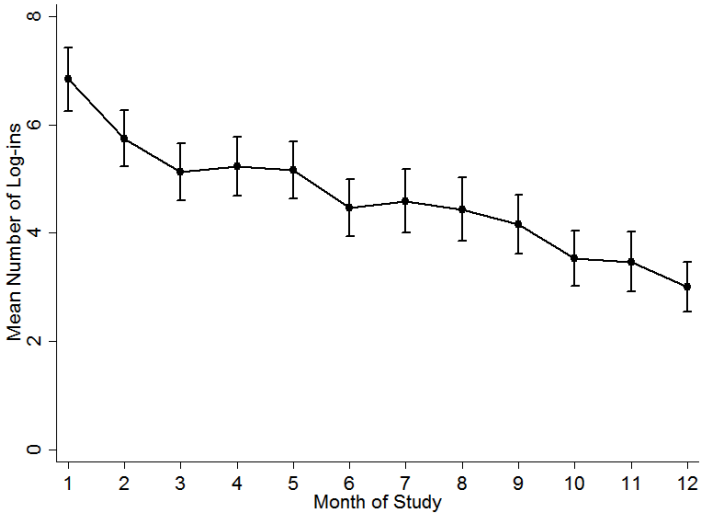


Figure 4. Mean log-ins and 95% confidence intervals in the intervention group by month of study.



Safety

Adverse events were categorized as pulmonary, cardiac, musculoskeletal, or other. A significantly greater percent of participants in the intervention group (27.9%, 43/154) had minor musculoskeletal adverse events than in the control group (10%, 8/84; $P<.001$). There were no differences between groups with respect to pulmonary, cardiac, or other adverse events during the 12 months.

Discussion

We show that our Internet-mediated, pedometer-based walking intervention does not maintain benefits in HRQL and daily step counts at 12 months, despite demonstrated improvements at 4 months [27]. Although we report negative findings for the study overall, lessons learned about device adherence and website engagement are highly informative for guiding the development of future PA interventions that can effectively promote long-term behavior change and sustain PA.

Overall, we found that a COPD population found the study feasible and were engaged. The percentage of participants who completed the 12-month study, providing HRQL and step-count data, was high. In addition, our objective results assessing device adherence and showing that 83% of days for participants in both groups within the last 6 weeks of the study had valid step-count data support that people were not lost to the study and then showing up for the last evaluation period. Importantly, persons in the intervention group had significantly higher device adherence compared to the control group for the study overall and at the end of the study. This finding supports that the goal setting, feedback, educational and motivational content, and online community forum provided on the website significantly increased intervention engagement beyond that observed with the use of a pedometer alone.

For the first time, we rigorously elicited participant survey responses about goal commitment and intervention engagement, and objectively assessed device adherence and website engagement during the 12-month study. In the intervention group, responses to questions about engagement at 12 months were the same compared to those at 4 months, with participants finding time to log in to the intervention, knowing their step-count goal, and using the graphs, tips and messages, and forums. They also reported that they were as committed to their step-count goal at 12 months as they were at 4 months. Although participants reported the same levels of goal commitment and intervention engagement at the end of the study compared to the beginning of the study, sustained behavior change was not observed because there were significant decreases in number of days with valid step-count data, number of log-ins to the website, and use of the online community forum over time. Although we can only speculate as to cause and effect, the decrease in daily step count (a marker of intervention efficacy and walking behavior change) over time mirrors the declines in device adherence and website engagement over time.

The reasons for the observed decline in daily step counts over time require further exploration. Participants may not have continued to wear the pedometer, log in to the website, and walk over the 12 months for a variety of possible reasons that we did not assess, such as waning interest with the intervention, progression of underlying COPD, flare-up of comorbidities, or occurrence of intercurrent life events (eg, spouse illness). The effect of the intervention on daily step counts could potentially have been greater if the control group had not received a pedometer and monthly reminders to upload step counts. We are confident that battery life did not affect the results because we mailed a new battery with replacement instructions to each participant every 4 months. We replaced lost or broken devices reported to us.

These results are similar to published data examining maintenance exercise programs after conventional pulmonary rehabilitation [16-21]. Typically, the unstable clinical course of a chronic lung disease such as COPD makes it difficult for patients to resume or maintain an exercise program [20]. Although we observed no difference in the number of COPD-related events, such as acute exacerbations, between groups, the occurrence of acute exacerbations and flare-up of comorbidities over a period of 12 months may have modified

the response to Taking Healthy Steps within the intervention group.

The failure to obtain long-term benefits with our PA interventions parallels the literature studying other behavior changes, such as smoking cessation [35] and weight loss [36]. Our 8-month maintenance phase retained the key components of goal setting, feedback, and social support. The main feature omitted beginning at month 5 was new educational and motivational content. These findings support that ongoing behavioral modification techniques are critical to sustain PA [37,38]. We speculate that additional intervention components, such as face-to-face contact with peers and/or health care providers, would enhance the social support and motivation needed to sustain PA as a routine behavior. Use of evolving technology, such as wireless transmission and mobile connectivity with cell phones, smartphones/mobile phones, or tablets, could potentially provide anytime/anywhere access to the PA intervention and enhance its long-term efficacy [37,39,40]. Intensive counseling and support at the time of acute exacerbations and flare-up of comorbidities would address medical barriers to PA and motivate patients to continue to walk after an illness. Finally, incorporating the health care provider, health care institutions, communities, and society at large into PA interventions could enhance long-term behavior change and adherence to effectively sustain PA in persons with COPD [41,42].

The exact role of digital walking programs in starting and maintaining exercise in persons with COPD remains to be determined. Both acute and chronic models of digital walking programs are potentially useful. Acute intervention models are needed to initiate and promote PA in the vast majority of patients with COPD who cannot access a conventional pulmonary rehabilitation program [43]. In addition, maintenance models are appropriate and much needed because long-term maintenance of behavior change is challenging. In addition, digital walking programs can potentially be useful adjuncts after conventional pulmonary rehabilitation to maintain benefits, which start to wane as early as 3 to 6 months after program completion [20,21]. They can also be an important component of COPD self-management programs [44]. An interesting future question to address is whether restarting our intervention every 4 to 8 months would be an efficacious long-term strategy.

The potential full impact of our intervention can only be appreciated by performing a future cost-effectiveness analysis. Results from cross-sectional data from our group and others have shown that every step counts. We have not found a “threshold” or “optimal” daily step count to obtain clinical benefits. The benefits appear to be linear such that those with higher step counts have lower risks for acute exacerbations, hospitalizations, hospital admissions and readmissions, and death compared to those with lower step counts [9-14]. Future work is needed to examine whether PA interventions such as ours can decrease health care resource utilization and result in cost savings to our health care system.

Major strengths of our study include the randomized controlled trial design with balanced groups at baseline, objective data on device adherence and website engagement, and the long-term follow-up of 12 months. Our intervention is based on a theoretical model, and informed by previous work eliciting patient feedback to optimize user acceptability and develop the motivational and educational content [25,45]. Our Internet-mediated, pedometer-based intervention focuses specifically on walking, a low-intensity PA that most patients can do. It has already been shown that a greater quantity of low-intensity PA reduces risk of COPD hospitalizations, whereas high-intensity PA does not result in risk reduction [15].

Our study has several limitations. We studied primarily white male Veterans limiting the generalizability of our results. Spirometric confirmation of the COPD diagnosis was not made at study entry. However, any potential misclassification of asthma as COPD was most likely balanced between groups and would not bias the primary results. The vast majority of the patients had MMRC <2. It is justified to include patients with MMRC <2 because patients with newly diagnosed COPD have reduced PA even at the earliest stages of the disease [3]. It is important to promote PA even when patients are not significantly symptomatic, as recommended by the GOLD guidelines for COPD [6]. We found no difference in benefit of the PA intervention in those with MMRC <2 versus MMRC ≥2. We acknowledge the final response rate was likely biased toward responders who had a particular interest in this type of intervention, and the results may not be generalizable to a wider COPD population. Finally, seasonal variation can influence our secondary outcome of daily step counts. We minimized the impact of season by having a 12-month intervention and enrolling participants over all four seasons.

An Internet-mediated, pedometer-based PA intervention for persons with COPD does not maintain improvements in HRQL or daily step count at 12 months, despite demonstrated improvements at 4 months. In addition, waning engagement with the PA intervention support that future efforts should focus on improving features of PA interventions to enhance long-term behavior change and sustain engagement with PA. These findings need to be considered when designing future Internet-mediated PA interventions.

Acknowledgments

We thank the Veterans for their participation in this research study. The study was funded by Department of Veterans Affairs, Health Services Research and Development Service (Grant IIR 09-366, Richardson); Department of Veterans Affairs, Rehabilitation Research and Development Service (Career Development Award, F6847W, Moy); and NIH National Heart, Lung and Blood Institute (Grant T32 HL007749-20, Martinez). None of the funding bodies had any role in the design, collection, analysis or interpretation of the data, in writing the manuscript, or in the decision to submit the manuscript for publication.

Authors' Contributions

MLM, RK, HQN, MDC, DEG, NDG, and CRR were involved in the conception and design of all stages of the study. MLM, CHM, RK, PR, HQN, MDC, and CRR were involved in study data collection. CHM, HMK, RK, PR, RGH, and NDG conducted study analyses. All authors read and approved the final manuscript. CRR, the study PI, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared. This study was initiated by the investigators, who do not receive any financial support from Omron Healthcare. The results of this study do not constitute endorsement of the Omron pedometer by the authors.

Multimedia Appendix 1

CONSORT-EHEALTH checklist.

[PDF File (Adobe PDF File), 786KB - jmir_v18i8e215_app1.pdf]

References

1. Moy ML, Danilack VA, Weston NA, Garshick E. Daily step counts in a US cohort with COPD. Respir Med 2012 Jul;106(7):962-969 [FREE Full text] [doi: 10.1016/j.rmed.2012.03.016] [Medline: 22521225]
2. Waschki B, Kirsten AM, Holz O, Mueller K, Schaper M, Sack A, et al. Disease progression and changes in physical activity in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2015 Aug 1;192(3):295-306. [doi: 10.1164/rccm.201501-0081OC] [Medline: 26020495]
3. Van Remoortel H, Hornikx M, Demeyer H, Langer D, Burtin C, Decramer M, et al. Daily physical activity in subjects with newly diagnosed COPD. Thorax 2013 Oct;68(10):962-963 [FREE Full text] [doi: 10.1136/thoraxjnl-2013-203534] [Medline: 23604460]
4. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med 2010 Sep 16;363(12):1128-1138. [doi: 10.1056/NEJMoa0909883] [Medline: 20843247]
5. Pitta F, Troosters T, Probst VS, Spruit MA, Decramer M, Gosselink R. Physical activity and hospitalization for exacerbation of COPD. Chest 2006 Mar;129(3):536-544. [doi: 10.1378/chest.129.3.536] [Medline: 16537849]
6. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2013 Feb 15;187(4):347-365. [doi: 10.1164/rccm.201204-0596PP] [Medline: 22878278]
7. McNamara RJ, McKeough ZJ, McKenzie DK, Alison JA. Physical comorbidities affect physical activity in chronic obstructive pulmonary disease: a prospective cohort study. Respirology 2014 Aug;19(6):866-872 [FREE Full text] [doi: 10.1111/resp.12325] [Medline: 24909412]
8. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2012 Jul 15;186(2):155-161. [doi: 10.1164/rccm.201201-0034OC] [Medline: 22561964]
9. Moy M, Gould M, Liu I, Lee J, Nguyen H. Physical activity assessed in routine care predicts mortality after a COPD hospitalization. ERJ Open Res [in press] 2016 Sep 01 (forthcoming).
10. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2014 Feb;11(2):149-157. [doi: 10.1513/AnnalsATS.201307-243OC] [Medline: 24308588]
11. Moy ML, Teylan M, Weston NA, Gagnon DR, Danilack VA, Garshick E. Daily step count is associated with plasma C-reactive protein and IL-6 in a US cohort with COPD. Chest 2014 Mar 1;145(3):542-550. [doi: 10.1378/chest.13-1052] [Medline: 24091482]
12. Moy ML, Teylan M, Weston NA, Gagnon DR, Garshick E. Daily step count predicts acute exacerbations in a US cohort with COPD. PLoS One 2013;8(4):e60400 [FREE Full text] [doi: 10.1371/journal.pone.0060400] [Medline: 23593211]
13. Nguyen HQ, Chu L, Amy LI, Lee JS, Suh D, Korotzer B, et al. Associations between physical activity and 30-day readmission risk in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2014 Jun;11(5):695-705. [doi: 10.1513/AnnalsATS.201401-017OC] [Medline: 24713094]
14. Waschki B, Kirsten A, Holz O, Müller K, Meyer T, Watz H, et al. Physical activity is the strongest predictor of all-cause mortality in patients with COPD: a prospective cohort study. Chest 2011 Aug;140(2):331-342. [doi: 10.1378/chest.10-2521] [Medline: 21273294]
15. Donaire-Gonzalez D, Gimeno-Santos E, Balcells E, de Batlle J, Ramon MA, Rodriguez E, PAC-COPD Study Group. Benefits of physical activity on COPD hospitalisation depend on intensity. Eur Respir J 2015 Nov;46(5):1281-1289. [doi: 10.1183/13993003.01699-2014] [Medline: 26206873]

16. Brooks D, Krip B, Mangovski-Alzamora S, Goldstein RS. The effect of postrehabilitation programmes among individuals with chronic obstructive pulmonary disease. Eur Respir J 2002 Jul;20(1):20-29 [FREE Full text] [Medline: 12166571]
17. Griffiths TL, Burr ML, Campbell IA, Lewis-Jenkins V, Mullins J, Shiels K, et al. Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. Lancet 2000 Jan 29;355(9201):362-368. [Medline: 10665556]
18. Ringbaek T, Brondum E, Martinez G, Thogersen J, Lange P. Long-term effects of 1-year maintenance training on physical functioning and health status in patients with COPD: A randomized controlled study. J Cardiopulm Rehabil Prev 2010;30(1):47-52. [doi: 10.1097/HCR.0b013e3181c9c985] [Medline: 20068423]
19. Beauchamp MK, Evans R, Janaudis-Ferreira T, Goldstein RS, Brooks D. Systematic review of supervised exercise programs after pulmonary rehabilitation in individuals with COPD. Chest 2013 Oct;144(4):1124-1133. [doi: 10.1378/chest.12-2421] [Medline: 23429931]
20. Ries AL, Kaplan RM, Myers R, Prewitt LM. Maintenance after pulmonary rehabilitation in chronic lung disease: a randomized trial. Am J Respir Crit Care Med 2003 Mar 15;167(6):880-888. [doi: 10.1164/rccm.200204-318OC] [Medline: 12505859]
21. Spencer LM, Alison JA, McKeough ZJ. Maintaining benefits following pulmonary rehabilitation: a randomised controlled trial. Eur Respir J 2010 Mar;35(3):571-577 [FREE Full text] [doi: 10.1183/09031936.00073609] [Medline: 19643944]
22. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
23. Richardson CR, Mehari KS, McIntyre LG, Janney AW, Fortlage LA, Sen A, et al. A randomized trial comparing structured and lifestyle goals in an internet-mediated walking program for people with type 2 diabetes. Int J Behav Nutr Phys Act 2007;4:59 [FREE Full text] [doi: 10.1186/1479-5868-4-59] [Medline: 18021411]
24. Moy ML, Janney AW, Nguyen HQ, Matthess KR, Cohen M, Garshick E, et al. Use of pedometer and Internet-mediated walking program in patients with chronic obstructive pulmonary disease. J Rehabil Res Dev 2010;47(5):485-496 [FREE Full text] [Medline: 20803392]
25. Moy ML, Weston NA, Wilson EJ, Hess ML, Richardson CR. A pilot study of an Internet walking program and pedometer in COPD. Respir Med 2012 Sep;106(9):1342-1350 [FREE Full text] [doi: 10.1016/j.rmed.2012.06.013] [Medline: 22795984]
26. Martinez CH, Moy ML, Nguyen HQ, Cohen M, Kadri R, Roman P, et al. Taking Healthy Steps: rationale, design and baseline characteristics of a randomized trial of a pedometer-based Internet-mediated walking program in veterans with chronic obstructive pulmonary disease. BMC Pulm Med 2014;14:12 [FREE Full text] [doi: 10.1186/1471-2466-14-12] [Medline: 24491137]
27. Moy ML, Collins RJ, Martinez CH, Kadri R, Roman P, Holleman RG, et al. An Internet-mediated pedometer-based program improves health-related quality-of-life domains and daily step counts in COPD: a randomized controlled trial. Chest 2015 Jul;148(1):128-137. [doi: 10.1378/chest.14-1466] [Medline: 25811395]
28. Weeks WB, Kazis LE, Shen Y, Cong Z, Ren XS, Miller D, et al. Differences in health-related quality of life in rural and urban veterans. Am J Public Health 2004 Oct;94(10):1762-1767. [Medline: 15451747]
29. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis 1992 Jun;145(6):1321-1327. [doi: 10.1164/ajrccm/145.6.1321] [Medline: 1595997]
30. Kaplan RM, Ries AL, Reilly J, Mohsenifar Z, National Emphysema Treatment Trial Research Group. Measurement of health-related quality of life in the national emphysema treatment trial. Chest 2004 Sep;126(3):781-789. [doi: 10.1378/chest.126.3.781] [Medline: 15364757]
31. Jones PW. St. George's Respiratory Questionnaire: MCID. COPD 2005 Mar;2(1):75-79. [Medline: 17136966]
32. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
33. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988 Mar;93(3):580-586. [Medline: 3342669]
34. Aaron SD, Fergusson D, Marks GB, Suissa S, Vandemheen KL, Doucette S, Canadian Thoracic Society/Canadian Respiratory Clinical Research Consortium. Counting, analysing and reporting exacerbations of COPD in randomised controlled trials. Thorax 2008 Feb;63(2):122-128. [doi: 10.1136/thx.2007.082636] [Medline: 17702790]
35. Jiménez-Ruiz CA, Andreas S, Lewis KE, Tonnesen P, van Schayck CP, Hajek P, et al. Statement on smoking cessation in COPD and other pulmonary diseases and in smokers with comorbidities who find it difficult to quit. Eur Respir J 2015 Jul;46(1):61-79. [doi: 10.1183/09031936.00092614] [Medline: 25882805]
36. Appel LJ, Clark JM, Yeh H, Wang N, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968 [FREE Full text] [doi: 10.1056/NEJMoa1108660] [Medline: 22085317]
37. Sallis R, Franklin B, Joy L, Ross R, Sabgir D, Stone J. Strategies for promoting physical activity in clinical practice. Prog Cardiovasc Dis 2015;57(4):375-386. [doi: 10.1016/j.pcad.2014.10.003] [Medline: 25459975]

38. Wilson JJ, O'Neill B, Collins EG, Bradley J. Interventions to increase physical activity in patients with COPD: a comprehensive review. COPD 2015 Jun;12(3):332-343. [doi: 10.3109/15412555.2014.948992] [Medline: 25221907]
39. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. A pilot study of a tool to stimulate physical activity in patients with COPD or type 2 diabetes in primary care. J Telemed Telecare 2014 Jan;20(1):29-34. [doi: 10.1177/1357633X13519057] [Medline: 24414397]
40. Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2014;9:935-944 [FREE Full text] [doi: 10.2147/COPD.S60179] [Medline: 25246781]
41. Heath GW, Parra DC, Sarmiento OL, Andersen LB, Owen N, Goenka S, Lancet Physical Activity Series Working Group. Evidence-based intervention in physical activity: lessons from around the world. Lancet 2012 Jul 21;380(9838):272-281. [doi: 10.1016/S0140-6736(12)60816-2] [Medline: 22818939]
42. Spruit MA, Pitta F, McAuley E, ZuWallack RL, Nici L. Pulmonary rehabilitation and physical activity in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2015 Oct 15;192(8):924-933. [doi: 10.1164/rccm.201505-0929CI] [Medline: 26161676]
43. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, ATS/ERS Task Force on Policy in Pulmonary Rehabilitation. An official American Thoracic Society/European Respiratory Society policy statement: enhancing implementation, use, and delivery of pulmonary rehabilitation. Am J Respir Crit Care Med 2015 Dec 1;192(11):1373-1386. [doi: 10.1164/rccm.201510-1966ST] [Medline: 26623686]
44. Effing TW, Vercoulen JH, Bourbeau J, Trappenburg J, Lenferink A, Cafarella P, et al. Definition of a COPD self-management intervention: International Expert Group consensus. Eur Respir J 2016 Jul;48(1):46-54. [doi: 10.1183/13993003.00025-2016] [Medline: 27076595]
45. Danilack VA, Weston NA, Richardson CR, Mori DL, Moy ML. Reasons persons with COPD do not walk and relationship with daily step count. COPD 2014 Jun;11(3):290-299. [doi: 10.3109/15412555.2013.841670] [Medline: 24152213]

Abbreviations

AE: acute exacerbations
COPD: chronic obstructive pulmonary disease
FEV^1: forced expiratory volume in 1 second
GOLD: Global Initiative for Chronic Obstructive Lung Disease
HRQL: health-related quality of life
MMRC: Modified Medical Research Council
PA: physical activity
SGRQ: St George's Respiratory Questionnaire
SGRQ-TS: St George's Respiratory Questionnaire Total Score
VISN: Veteran Integrated Service Network

Edited by G Eysenbach; submitted 09.02.16; peer-reviewed by N Cobb, I Muller, G Jerome, D Ryan, R Grainger, T Poder; comments to author 10.03.16; revised version received 21.04.16; accepted 24.05.16; published 08.08.16.

Please cite as:
Moy ML, Martinez CH, Kadri R, Roman P, Holleman RG, Kim HM, Nguyen HQ, Cohen MD, Goodrich DE, Giardino ND, Richardson CR
Long-Term Effects of an Internet-Mediated Pedometer-Based Walking Program for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial
J Med Internet Res 2016;18(8):e215
URL: http://www.jmir.org/2016/8/e215/
doi:10.2196/jmir.5622
PMID:27502583

Original Paper

The Effect of Tailored Web-Based Feedback and Optional Telephone Coaching on Health Improvements: A Randomized Intervention Among Employees in the Transport Service Industry

Madeleine Solenhill[1,2], PhD; Alessandra Grotta[3], PhD; Elena Pasquali[4], MSc; Linda Bakkman[1], PhD; Rino Bellocco[3,5], ScD; Ylva Trolle Lagerros[1,6], PhD

[1]Karolinska Institutet, Department of Medicine, Clinical Epidemiology Unit, Stockholm, Sweden

[2]Halmstad University, School of Health and Welfare, Halmstad, Sweden

[3]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden

[4]Karolinska Institutet, Department of Clinical Science, Division of Baxter Novum, Intervention and Technology, Huddinge, Sweden

[5]University of Milano-Bicocca, Department of Statistics and Quantitative Methods, Milan, Italy

[6]Karolinska University Hospital Huddinge, Department of Medicine, Clinic of Endocrinology, Metabolism and Diabetes, Stockholm, Sweden

Corresponding Author:
Madeleine Solenhill, PhD
Halmstad University
School of Health and Welfare
Box 823
Halmstad, S-301 18
Sweden
Phone: 46 +46 35 16 71 00
Fax: 46 +46 35 18 61 92
Email: Madeleine.Solenhill@hh.se

Abstract

Background: Lifestyle-related health problems are an important health concern in the transport service industry. Web- and telephone-based interventions could be suitable for this target group requiring tailored approaches.

Objective: To evaluate the effect of tailored Web-based health feedback and optional telephone coaching to improve lifestyle factors (body mass index—BMI, dietary intake, physical activity, stress, sleep, tobacco and alcohol consumption, disease history, self-perceived health, and motivation to change health habits), in comparison to no health feedback or telephone coaching.

Methods: Overall, 3,876 employees in the Swedish transport services were emailed a Web-based questionnaire. They were randomized into: control group (group A, 498 of 1238 answered, 40.23%), or intervention Web (group B, 482 of 1305 answered, 36.93%), or intervention Web + telephone (group C, 493 of 1333 answered, 36.98%). All groups received an identical questionnaire, only the interventions differed. Group B received tailored Web-based health feedback, and group C received tailored Web-based health feedback + optional telephone coaching if the participants' reported health habits did not meet the national guidelines, or if they expressed motivation to change health habits. The Web-based feedback was fully automated. Telephone coaching was performed by trained health counselors. Nine months later, all participants received a follow-up questionnaire and intervention Web + telephone. Descriptive statistics, the chi-square test, analysis of variance, and generalized estimating equation (GEE) models were used.

Results: Overall, 981 of 1473 (66.60%) employees participated at baseline (men: 66.7%, mean age: 44 years, mean BMI: 26.4 kg/m^2) and follow-up. No significant differences were found in reported health habits between the 3 groups over time. However, significant changes were found in motivation to change. The intervention groups reported higher motivation to improve dietary habits (144 of 301 participants, 47.8%, and 165 of 324 participants, 50.9%, for groups B and C, respectively) and physical activity habits (181 of 301 participants, 60.1%, and 207 of 324 participants, 63.9%, for B and C, respectively) compared with the control group A (122 of 356 participants, 34.3%, for diet and 177 of 356 participants, 49.7%, for physical activity). At follow-up, the intervention groups had significantly decreased motivation (group B: P<.001 for change in diet; P<.001 for change in physical activity; group C: P=.007 for change in diet; P<.001 for change in physical activity), whereas the control group reported significantly increased motivation to change diet and physical activity (P<.001 for change in diet; P<.001 for change in physical activity).

Conclusion: Tailored Web-based health feedback and the offering of optional telephone coaching did not have a positive health effect on employees in the transport services. However, our findings suggest an increased short-term motivation to change health behaviors related to diet and physical activity among those receiving tailored Web-based health feedback.

(J Med Internet Res 2016;18(8):e158) doi:10.2196/jmir.4005

KEYWORDS

diet; exercise; Internet; intervention studies; lifestyle; motivation; occupational health; questionnaires; randomized

Introduction

Behaviors related to diet, physical activity, sleep, stress, and use of tobacco and alcohol play an important role in the prevention of lifestyle-related health problems such as obesity, diabetes, and cardiovascular diseases [1]. According to a review by *Lancet*, lifestyle interventions with a holistic perspective on health, such as an individual's overall lifestyle behaviors, are more successful than those only focusing on one specific behavior [2]. Employees in the transport service industry represent one of many target groups that would benefit from tailored interventions focusing on healthy lifestyle behaviors. A needs assessment conducted by our research team among a sample from the Swedish transport service industry indicated that more than 60% of the track technicians were overweight, 70% had an unhealthy diet, and almost 50% had high cholesterol levels. Furthermore, a report by Kecklund et al [3] highlights health risks associated with irregular workdays due to shift work, long working hours, and the lack of recovery—common characteristics of Swedish transport employees.

We had previously reported findings on employees in the Swedish transport service industry. In detail, different methods to encourage study participation of this target group in a large Web-based lifestyle intervention study (described herein) were examined. The effects of email reminders and additional, more practical, reminders on overall study participation were specifically studied. Sending email reminders was an effective approach to encourage study participation in our intervention study [4]. Until now, we have not yet reported the health effects from our Web-based lifestyle intervention study.

To intervene at the workplace can be an effective health promotion strategy, which has also been endorsed by the World Health Organization which stated, "the workplace directly influences the physical, mental, economic and social well-being of workers and in turn the health of their families, communities and society. It offers an ideal setting and infrastructure to support the promotion of health of a large audience" [5]. However, it may be challenging to intervene at the workplace in the transport service industry because of different types of professions among the employees, resulting in different work tasks, schedules, and work settings. Tailored lifestyle interventions would therefore be ideal.

The advancements in technology allow for tailored lifestyle interventions. Prior research suggests promising results for Web-based physical activity interventions on improved physical activity habits [6-9]. The effect is even more valuable as Web-based interventions bridge the gap between the individual's need for health interventions and the primary care's lack of capability to support these individuals [9]. Mateo et al and Muntaner et al further support technologies to increase physical activity by using mobile phones, apps, and digital assistants [6,7]. Similar effects and types of interventions have been found for diet [10-12] and weight loss [13].

For instance, we had previously reported findings from a public Web-based weight loss club targeting behavioral change, using interactive progress charts, graphs, recipes, and chat forums [13]. We found that men and foremost participants older than 45 years were particularly successful in changing their eating behaviors, which was positively associated with greater weight loss [12]. In addition, our male participants logged in to the website most frequently, compared with our female participants. Related research carried out by our research team [14] builds on these findings, suggesting that an interactive food plate would serve as an appropriate tool to assess food intake among a male cohort via the Internet.

According to a review conducted by Webb et al [15], Web-based health interventions that are based on health behavior theories are more successful in promoting health behavior change, compared with nontheory-based health interventions. Also, Web-based interventions that incorporate more behavior change techniques and modes of delivery in addition to the Web further enhance the effects of a health intervention as opposed to fewer health behaviors techniques and 1 mode of delivery [15]. Yet, the actual effect of an intervention may also be related to the type of outcome variable measured [16].

A recent review [17] compared the effects from interventions using Web-based, printed materials, or the telephone to support healthier behaviors with respect to physical activity, diet, and weight control. Telephone intervention was particularly effective to promote healthier habits. Further research supports the telephone as an intervention tool in smoking and alcohol cessation programs [18-20]. In Sweden, there are mainly 2 helplines to encourage healthier behaviors among the public with respect to tobacco and alcohol. These are free of charge and have shown promising effects in reducing alcohol use and alcohol problems [19,21].

Consequently, the aim of this study was to evaluate the effect of tailored Web-based health feedback and the offering of optional telephone coaching with respect to improved health, in comparison to no health feedback or telephone coaching among employees in the transport service industry.

Methods

Participants

We asked 3,876 employees (18-65 years) at 4 transportation companies in the Swedish transport service industry to participate in a lifestyle health intervention. The invited companies represented (1) The Swedish Transport Administration (n=453), (2) The Swedish national train operator (n=2,391), (3) The local train operator in Stockholm (n=573), and (4) The subway transportation operator in Stockholm (n=459). Because the intervention study had a Web-based approach, only those employees with a work email were asked to participate.

Study Design

The employees were emailed a link to the lifestyle health intervention with personal login details (username and password), information about the study, and instructions on how to participate in the study.

When the link was emailed, the participants were randomized to 1 of the 3 groups: (A) control group or (B) intervention Web or (C) intervention Web + telephone. All the 3 groups received an identical questionnaire on health and lifestyle behaviors, only the interventions differed. The intervention group B received tailored Web-based health feedback, and intervention group C received tailored Web-based health feedback + additional optional telephone health coaching for those participants who were motivated to change health behaviors. The Web-based feedback and the optional telephone coaching were based on the participants' responses to the questions in the questionnaire. The CONSORT-EHEALTH has been used to report this study (Multimedia Appendix 1).

Questionnaire

When the participants opened the link, they entered a welcome page with participant information and an informed consent form. By giving informed consent, they entered the questionnaire. Both the baseline and the follow-up questionnaires assessed the following 9 health areas divided into sections: body mass index (BMI), dietary intake, physical activity, stress, sleep, tobacco consumption, alcohol consumption, history of disease, and self-perceived health.

The questionnaire was originally developed to be filled out by using a computer, but it was possible to access the questionnaire using mobile phones.

Outcome Measures

In details, our primary outcome was BMI (kg/m^2), computed by dividing the body weight (in kilograms) by the height (in meters) squared. Secondary outcomes were: eating breakfast (times/week), carbohydrates intake (times/week), sugar intake (times/week), saturated fats intake (times/week), unsaturated fats intake (times/week), leisure-time physical activity (days/week), minutes of leisure-time activity/week ("0-10," "11-20," "21-30", "31-40," "41-50," "51-60," ">60" minutes/week), and total physical activity/day in metabolic equivalent of task (MET) hours, feeling stressed ("no," "little," "somewhat," "quite a lot," "very much"), sleeping ≤5 hours/night ("never," "seldom," "sometimes," "frequently," "mostly"), sleeping ≥9 hours/night ("never," "seldom," "sometimes," "frequently," "mostly"), feeling well-rested after sleeping, ("never," "seldom," "sometimes," "often," "mostly," "always," "do not know"), number of occasions of alcohol consumption ("never," "1 time/month or less," "2-3 times/month," "2-3 times/week," "≥4-6 times/week"), number of glasses during a typical alcohol occasion ("1-2 glasses," "3-4 glasses," " ≥ 5 glasses"), smoke ("daily," "sometimes," "no"), number of cigarettes per week, use of moist oral snuff, "snus" ("daily," "sometimes," "no"), and number of packages of snus per week.

All questionnaire sections (besides self-perceived health and BMI) ended with a follow-up question on the participants' "motivation to change" their health habits in that specific health area. The response options for these question was in accordance with Prochaskas' transtheoretical model (TTM) of behavior change using the 5 phases—precontemplation, contemplation, preparation, action, and maintenance—to assess participants' readiness to change now, within a month, within 6 months, not now, or not at all [22]. If the participants were motivated to change their health habits, they were recommended to contact any of the telephone helplines to receive coaching for healthier habits, presumed that they were assigned to group C intervention Web + telephone (see more details in the following section).

Development of Questionnaire and Randomization

The researchers developed the questionnaire and the content of the feedback and were involved in the development of content of the telephone coaching. We asked experts in respective health field to approve the questions asked, and its feedback, for each part of the questionnaire. The questionnaire was pilot tested by truck drivers, the target group, and the research unit at Karolinska Institutet. Most questionnaire sections used validated items from previous research. We used a questionnaire to assess food frequency including the intake of carbohydrates or fibers; sugar; saturated fats or high fat intake; and unsaturated fats or low fat intake. To complement the food frequency questionnaire, we used an interactive virtual food plate developed by the research team to assess lunch intake, by allowing the participants to complete a typical lunch meal by dragging pictures of food items to a virtual food plate. This interactive food plate has been validated and described previously [14]. Alcohol consumption was studied using the Alcohol Use Disorders Identification Test, AUDIT-C [23], and tobacco use was assessed using validated questionnaires provided by the Swedish National Tobacco Quitline. Sleep was assessed using the validated Karolinska Sleep Questionnaire [23]. Physical activity was measured using a validated questionnaire for total energy expenditure [24]. The instrument has 9 ordered intensity levels, each assigned a value expressed as a multiple of metabolic energy turnover [25] and exemplified by common activities. The participants reported the time spent on each intensity level, in total 24 hours, which allowed for an estimate of MET hours per day. In addition, we asked the participants to indicate the average number of times per week spent on leisure-time physical activities such as brisk walking, running, aerobics or weight lifting and skiing, and number of minutes per session. These self-reported questions about leisure-time physical activity are often used in

epidemiological research, for example, the question about leisure-time activity is similar to the questions used in studies by McTiernan et al [26] and Godin et al [27] and extensively validated against other measures such as maximal oxygen uptake and accelerometers.

A Web company [28] was responsible for developing a Web-based version of the research material in collaboration with the research group. They were also responsible for administrating the website enclosing the feedback to ensure overall feasibility and usability. A helpdesk was administered by the research group and the Web company if participants needed any help to access or use the website or had any questions regarding the content of the feedback.

The research group collected lists of potential participants (employees) from the company manager at the 4 companies. All employees with an email address were eligible for the study. The participant lists were given to the Web company, which was responsible for distributing the questionnaires. Randomization lists were automatically created by the Web company before distributing the questionnaire link.

Reminders to take part in the study were sent to all nonresponders to encourage study participation at baseline and follow-up, which has been described in detail previously [4]. In brief, a total of 4 to 5 and 11 email reminders were sent at baseline and follow-up, respectively. The difference of reminders (4 or 5) depended on when the participants entered the study. The email reminders increased the total response rate by 15% and 21% at baseline and follow-up, respectively. Additional reminders such as informational texts about the study, fliers, oral presentations, and visits by the research group as well as texting were also distributed in addition to the email reminders. Thus, we had several strategies to encourage our target group to participate in this intervention study.

Only those who submitted the questionnaire(s) were considered participants in this study.

Intervention Web

The participants received icons such as colored smiley faces and icons symbolizing “information” and “advice” to guide them throughout the completion process of the questionnaire and with respect to health changes. The smiley faces provided the participants with tailored feedback on level of satisfaction regarding the reported health habit according to national health guidelines [29]. A green and happy smiley indicated “your reported health habits are good”; a yellow smiley indicated “your reported health habits need to be improved”; a red, sad smiley indicated “your reported health habits mean health risks.” The icons for “information” and “advice” allowed the participant to click forward to get more information about the health habit and to receive specific advice and recommendations to improve that specific health habit (see Figure 1).

In addition to the icons shown in Figure 1, tailored feedback appeared instantly and automatically on the computer screen after the participants had responded to each of the sections with questions (Figure 2). For example, if the participants reported a low physical activity level, that is, not meeting the national health guidelines [29], information about the significance of being physically active and advice on how to increase physical activity as a natural part of life were provided via the computer screen. The participants were able to read the feedback instantly. All feedback was summarized and saved at the end of the questionnaire and on a tailored website, which the participants could visit at any time using their personal login details.

Figure 1. Screenshot of the icons used to guide the participants throughout the completion process of the questionnaire.

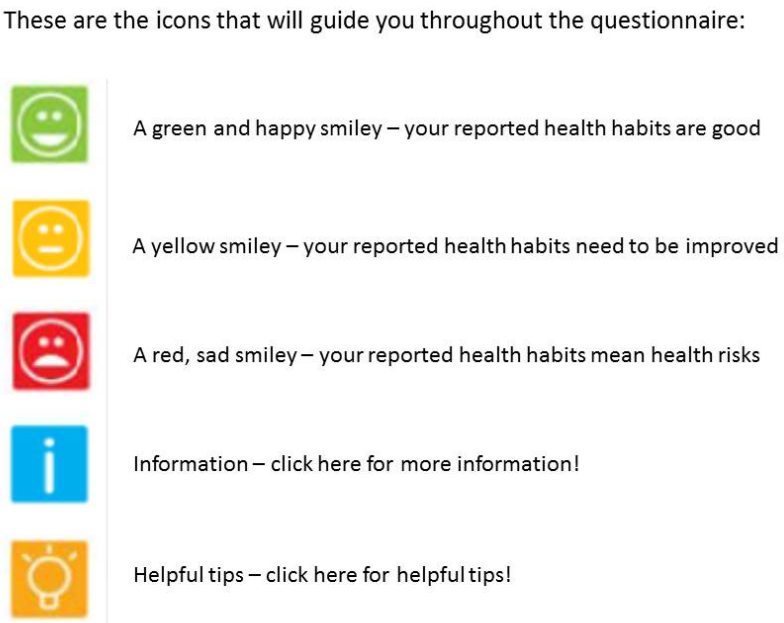
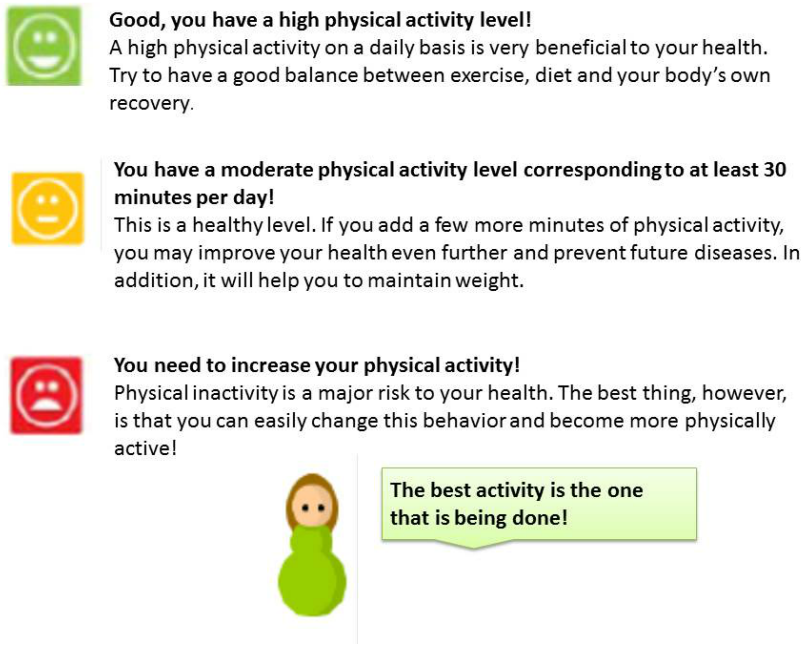


Figure 2. Screenshot of examples of the tailored web-based health feedback given for physical activity.



Intervention Web + Telephone

The participants were offered the opportunity for telephone coaching in addition to Web-based tailored health feedback, if their reported health habits did not meet the national guidelines or if they expressed motivation to change health habits. If the participants were referred to telephone coaching, they could either leave their mobile phone number to be dialed (reactive help) by the telephone helplines or dial themselves (proactive help) (Figure 3). The 3 helplines that were offered were: (1) The Swedish National Tobacco Quitline, (2) The Swedish National Alcohol Helpline, and (3) The Diet and Exercise Helpline. The helplines for tobacco and alcohol are well-established public helplines in Sweden [30]. The Diet and Exercise Helpline was developed for this study. All helplines used counselors who were specifically trained to work effectively with behavior changes using traditional counseling approaches such as the motivational interviewing technique [31] and TTM [22].

The tailored Web-based health feedback and the telephone coaching services were available during 9 months. After 9 months, all participants received a follow-up questionnaire including intervention Web + telephone coaching. The following differences between baseline and follow-up applied: at follow-up, the participants had access to the website encompassing their personalized feedback. This website was however only available during 2 weeks after the study had ended. All participants were offered telephone coaching by The Swedish National Tobacco Quitline and The Swedish National Alcohol Helpline but not by the Diet-and Exercise Helpline. The reasons for this were that the helplines for tobacco and alcohol are national public health services, whereas the helpline for diet and exercise habits was part of the intervention itself and led by the research group.

Data were collected by the Web company [28] and sent to the researchers after the intervention had ended.

The Ethics Committee of the Karolinska Institutet, Stockholm, Sweden, approved this study. See Figure 4 for a flowchart of the study design.

Figure 3. Screenshot of referral to The Swedish National Tobacco Quitline.

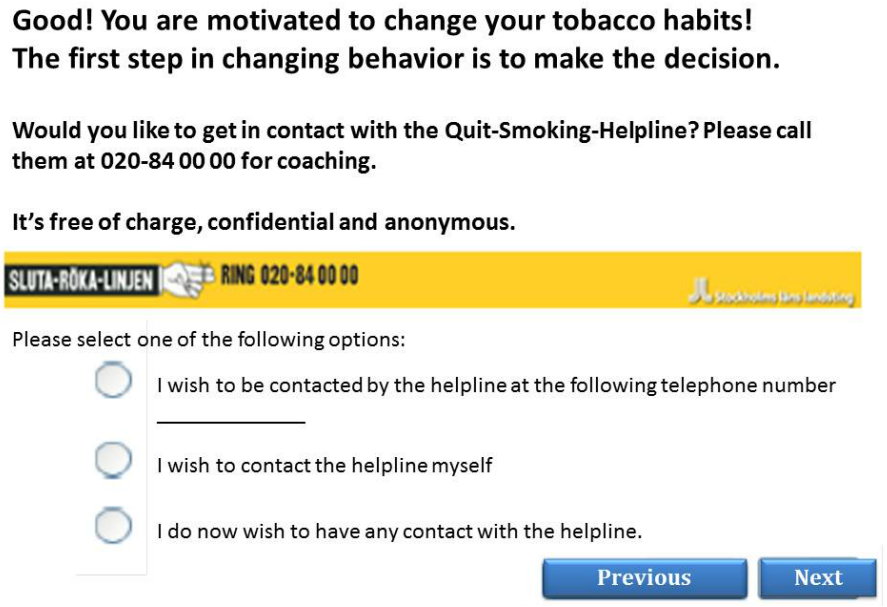
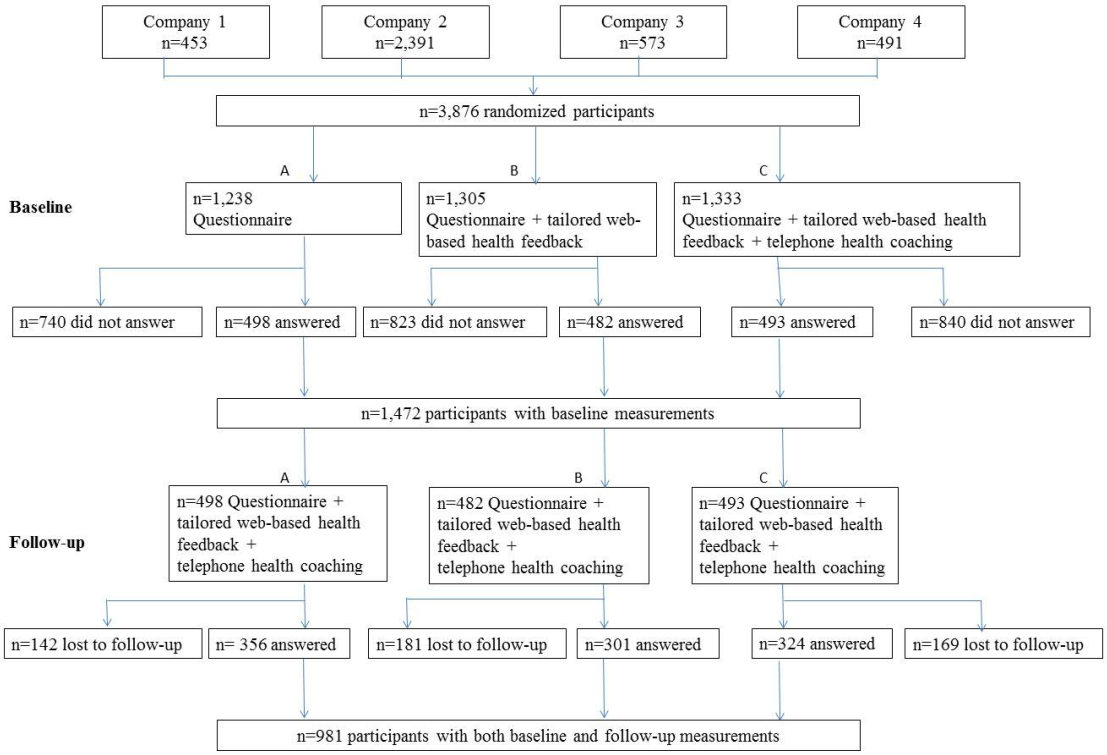


Figure 4. Flowchart of the study design. A total of 3876 employees were asked to participate in the study and were automatically randomized to 1 of the 3 groups: (1) control group or (2) intervention Web, or (3) intervention Web + telephone. All the 3 groups received an identical questionnaire on health and lifestyle behaviors, only the interventions differed. Nine months later, all participants received a follow-up questionnaire and interventions web + telephone.



Statistics

Descriptive statistics (proportions, medians, and means with standard deviations) of the participants' characteristics (sex, age, BMI, physical activity level, use of tobacco—snus and smoking—and perceived health) at baseline (all sample, completers, noncompleters) and follow-up were computed. We also categorized the participants into office workers or field workers (if mostly office based=office worker; if mostly non–office-based with limited access to a computer=field worker). The participants' referral and usage of the 3 telephone helplines were also examined.

We studied whether differences in participants' reported health habits at baseline and follow-up varied across the 3 groups (A, B, and C). Chi-square tests were performed to study possible differences in categorical variables and analysis of variance (or Kruskal–Wallis) for continuous variables.

The participants' levels of "motivation to change" were specifically studied. Both at baseline and follow-up, the participants answered the question "are you motivated to change your X behaviors," in which X refers to the specific health behavior that was examined (ie, physical activity, diet, tobacco, or alcohol). The response options to these questions were: "yes, within a month," "yes, within six months," "yes, but in the future," "do not know," and "no." We summarized the proportions of responses within the yes-options and the proportions within "do not know" and "no," separately.

We used generalized estimating equation (GEE) to study specific effects of the interventions over time in relation to changes in reported health habits and motivation to change. GEE models are an extension of the simple regression model to contexts in which the outcome can be noncontinuous and measured more than once on the same subject. The GEE models were adjusted for age and gender.

Only those who participated at both baseline and follow-up were considered for analyses. STATA version 13.1 (Statacorp LP, College Station, TX) for Windows was used for all statistical calculations and analyses. All reported *P* values were 2-sided. $P<.05$ was considered statistically significant.

Results

Basic characteristics of the study participants

In total, 1473 of 3876 (38.00%) employees completed the Web-based lifestyle questionnaire, and 981 of 1473 (66.60%) employees participated both at baseline and follow-up. Table 1 summarizes baseline characteristics of subjects who started the study, stratified by being, or not being, lost to follow-up. At baseline, 984 of 1473 (66.80%) participants were men. They had a mean age of 44 years (SD 10.2), a BMI of 26.4 kg/m^2(SD 4.2), and perceived their personal health as "rather good" or better. In total, 500 of 1473 (33.94%) participants were categorized as office workers, the remaining as field workers. There were no substantial differences between completers and noncompleters, even if subjects who were lost to follow-up were slightly younger and more inclined to smoke and use snus than subjects who proceeded in the study.

Table 1. The participants' basic characteristics at baseline (all sample, completers, and noncompleters) and follow-up.

Characteristic	Baseline (all), N=1473[a]	Baseline (completers), N=981[b]	Baseline (non-completers), N=492[c]	Follow-up, N=981[b]
Sex, male (%)	984 (66.80)	655 (66.7)	329 (66.9)	656 (66.8)
Age, years (SD[f])	43 (10.68)	44 (10.2)	42 (11.5)	44 (10.2)
Physical activity level/day, MET[e]hours (SD)	44.1 (13.66)	43. 8 (12.8)	44.8 (15.2)	44.0 (12.1)
Smoking, yes (%)	261 (17.72)	158 (16.1)	103 (20.9)	154 (15.7)
Snus, yes (%)	255 (17.31)	160 (16.3)	95 (19.3)	167 (17.0)
BMI[d], kg/m^2(SD)	26.3 (4.25)	26.4 (4.3)	26.1 (4.1)	26.5 (4.2)
Perceived health habits as "rather good" or better, yes (%)	1149 (78.00)	769 (78.4)	380 (77.2)	803 (81.8)

[a]Number of subjects who participated at baseline.

[b]Number of subjects who participated at both baseline and follow-up.

[c]Number of subjects who participated at baseline but were lost to follow-up.

[d]BMI: body mass index.

[e]MET: metabolic equivalent of task.

[f]SD: standard deviation.

Interventions

Among the 981 employees who participated at both baseline and follow-up, 357 of 981 (36.4%) represented group A (control group), 301 of 981 (30.7%) group B (intervention Web), and 324 of 981 (33.0%) group C (intervention Web + telephone). See Figure 4.

In group C, at baseline, 173 (53.4%), 23 (7.1%), and 19 (5.8%) of 324 participants reported a wish to be dialed by or to dial themselves to the Diet and Exercise Helpline, the Swedish National Tobacco Quitline, and the Swedish National Alcohol Helpline, respectively. In details, 77 of 173 (44.5%) participants asked to be dialed by the Diet and Exercise Helpline, 9 of 23 (39.1%) participants by The National Tobacco Quitline, and 5 of 19 (26.3%) participants by The Swedish National Alcohol Helpline. All participants were contacted accordingly. The remaining participants reported that they would contact the helplines themselves. No participant contacted the Diet and Exercise Helpline voluntarily.

At baseline, 550 of 981 participants (56.1%) reported "do not know" or "no" motivation to improve dietary habits. Intervention groups B and C reported higher motivation to improve dietary habits (144 of 301 participants, 47.8%, and 165 of 324 participants, 50.9%, for groups B and C, respectively) and physical activity habits (181 of 301 participants, 60.1%, and 207 of 324 participants, 63.9%, for groups B and C, respectively) compared with the control group A receiving no health feedback (122 of 356 participants, 34.3%, for diet and

177 of 356 participants, 49.7%, for physical activity). At follow-up, the intervention groups had significantly decreased motivation (group B: $P<.001$ for change in diet; $P<.001$ for change in physical activity; and group C: $P=.007$ for change in diet; $P<.001$ for change in physical activity), whereas control group A reported a significant increase in motivation to change over time, at time of receiving health feedback in the follow-up questionnaire ($P<.001$ for change in diet; $P<.001$ for change in physical activity). See Table 2.

Table 2. Motivation to change diet and physical activity at baseline and follow-up across groups.

Motivation		Group A (n=356), Control		Group B (n=301), Intervention Web		Group C (n=324), Intervention Web + telephone	
		Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Plan to change diet, n (%)							
	Yes, within 1 month	58 (16.3)	85 (23.9)	78 (25.9)	66 (21.9)	94 (29.0)	65 (20.1)
	Yes, within 6 months	31 (8.7)	32 (8.9)	31 (10.2)	32 (10.6)	38 (11.7)	40 (12.3)
	Yes, but not now	33 (9.2)	38 (10.7)	35 (11.6)	25 (8.3)	33 (10.2)	34 (10.5)
	Do not know	119 (33.4)	100 (28.1)	95 (31.6)	98 (32.5)	85 (26.2)	99 (30.5)
	No	115 (32.3)	101 (28.3)	62 (20.6)	80 (26.6)	74 (22.8)	86 (26.5)
P for change[a]		<.001		<.001		.007	
Plan to change physical activity, n (%)							
	Yes, within 1 month	96 (26.9)	113 (31.7)	106 (35.2)	91 (30.2)	114 (35.2)	91 (28.1)
	Yes, within 6 months	38 (10.7)	36 (10.1)	41 (13.6)	31 (10.2)	52 (16.1)	43 (13.3)
	Yes, but not now	43 (12.1)	40 (11.2)	34 (11.3)	43 (14.3)	41 (12.6)	36 (11.1)
	Do not know	79 (22.2)	61 (17.1)	53 (17.6)	65 (21.6)	39 (12.0)	56 (17.3)
	No	100 (28.1)	106 (29.8)	67 (22.2)	71 (23.6)	78 (24.1)	98 (30.2)
P for change[a]		<.001		<.001		<.001	

[a]p-value obtained from GEE models.

Participants' health habits at baseline and follow-up are reported in Multimedia Appendix 2. When analyzing possible changes between baseline and follow-up measurers, no significant differences between groups were found. Our results from the GEE models support no significant changes over time for the various reported health aspects at baseline and follow-up or between the 3 groups. However, our results suggest significant increases in days of eating breakfast ($P<.001$), days of physical activity per week ($P=.002$), and decreases in sugar intake ($P<.001$) at follow-up, with no statistically significant differences regarding type of interventions (group B or C).

Discussion

The results from this randomized Web-based intervention suggest no significant health improvements from tailored Web-based health feedback and the offering of optional telephone coaching, in comparison to no health coaching among employees in the Swedish transport industry.

However, our findings point toward an impact of automatic tailored health feedback on increased motivation to change behaviors. More specifically, at baseline, the intervention groups receiving feedback reported higher motivation to improve dietary and physical activity habits compared with the control group A. At follow-up, all the groups received feedback, but only the control group, now receiving feedback, reported an increased motivation to change dietary and physical activity habits. It may thus be proposed that tailored health feedback positively alters participants' motivation to engage in healthier lifestyle habits, but, the effect does not last over time, not even if the feedback is offered again.

The results from our study are in line with the findings from a randomized Web-based intervention study examining participants' motivation to improve cardiovascular health conducted by Ayres et al [32]. In short, they randomized their participants to receive either motivational feedback in the form of a personalized risk assessment of their current cardiovascular health or to respond to questions assessing intentions, attitudes, and anticipated regret about managing cholesterol levels through

diet. They also randomized their participants to a group combining these 2 interventions or to a control group receiving no intervention. Their results indicate an enhanced motivation to improve cardiovascular health from the combined intervention (personalized health feedback and questions on intentions to change health behavior). The results of our study may therefore be supported by this study; in this study, an amplified effect of tailored feedback was found, even if it was for a short term [32]. However, it is difficult to draw conclusions on the actual cause and effect relationship. Prior research suggests a question–behavior effect, meaning that simply asking a person to answer questions about the intention to change behavior may positively influence a person's process to change that behavior [33]. For instance, participants may then reflect on the registered health aspects, which in turn may lead to an increased motivation to change them. Because many Web-based intervention studies use both interventions and questions as part of their design, it is difficult to study whether it is the act of responding to questions or the intervention itself—or a combination of both—that may cause the effect [34].

Prior Web-based lifestyle interventions using tailored health feedback report modest effects on health improvements [35], and previously published systematic reviews of non–Web-based health promotions conducted among workplace personnel show inconclusive results [36]. Also, few Web-based lifestyle interventions have been conducted among a target group such as ours with predominantly middle-aged overweight men in the transport industry, making it difficult to find support from previous research. Pressler et al report no increased effects from Web-based interactive feedback to promote physical activity among sedentary employees at an automobile company [37], compared with ordinary personal advice to encourage physical activity. Hansen et al, however, found an improved level of leisure-time physical activity from their Web-based population-based intervention study after 3 and 6 months but report no other health improvements such as reductions in BMI, waist circumference, body fat percentage, or blood pressure [8].

Schulz et al indicate small health behavior improvements for physical activity, vegetable and fruit consumption, alcohol intake, and tobacco consumption. They specifically studied differences in health effects from encouraging their participants with either sequential health behavior changes (one at the time) or simultaneous health behavior changes (all at the same time), suggesting that sequential health changes produce greater effects the first year, but that simultaneous changes generate a more long-term effect found after 2 years [38]. These findings are in line with those of research by Krebs et al, suggesting that tailored interventions using iterative feedback enhance the effects when compared with tailored interventions based on "one assessment feedback only" [35]. In our study, we asked the participants to change all health behaviors at once, meaning that the potential long-term effects on health behavior improvements might be too early to detect from our 9-month intervention. Research studies indicate higher dropout rates for interventions promoting simultaneous behavior changes due to prolonged time for the participants to see individual health improvements [39], which could support the response rate in our study [4].

Our participants who received Web-based health feedback received it instantly when filling out the questionnaire and, if they wanted to, had the opportunity to read it again on the website after submitting their questionnaire. It is possible that once they had read the feedback on the screen (and potentially saved it or printed it), there might not have been a reason to revisit the website because it offered no new tips, updated information, or interactive features such as chat rooms or blogs. The fact that 66% of the study participants were field workers may have influenced the actual effect on the intervention. Even if all study participants had access to computers and mobile phones during their workday, it is possible that field workers were less likely to enter the website enclosing the tailored feedback, due to inconveniences of their work setting from being out in the field. It would be interesting to study how participants would use a questionnaire such as ours with interactive feedback as a continuous mean to assess their own health and receive instant feedback on new health behaviors. However, it is complicated to draw conclusions from interactive data collection, including our results.

Even if the questionnaire was exactly the same in the 3 groups, we cannot rule out that the intervention groups were affected by the fact that they received instant feedback. Thus, although randomization was conducted to produce comparability between the groups with respect to unmeasured confounding, the feedback in itself may have introduced a difference. Results such as comparisons at baseline and completers' analyses should therefore be interpreted with caution.

Moreover, the follow-through on the "opt in" and "I will call you" feature for contact with the telephone helplines in this study is interesting and novel. It appears that 55% to 74% of participants interested in telephone coaching indicated that they would contact the helplines on their own initiative; however, no one voluntarily contacted the helplines. Perhaps it requires a stronger courage by the individual to dial a helpline to ask for help, as opposed to putting the responsibility in the hands of the health counselors. Future research is recommended to study this further, particularly as nonfacial health interventions are rapidly increasing in popularity due to the advancement of technology. Many interventions are indeed based on the individuals' own capacity to take the first step in his or her behavior change.

The contact information to the 3 helplines was only provided in the questionnaire and thus saved at the personalized website. The helplines for tobacco and alcohol were, on the other hand, also advertised at general health care settings as they were open to the public, but the information to the Diet and Exercise helpline was only provided in the questionnaire. Frequent reminders and information about the feedback and telephone helplines could have increased the use of these features and the number of callers to the 3 helplines.

Even if our telephone coaching used well-established coaching techniques such as the motivational interviewing technique, the benefits of the technique may not have been fully realized because telephone coaching was only offered to those with "higher" levels of motivation to change health behaviors. In fact, the strength of motivational interviewing is the ability to

explore ambiguity and to help individuals develop stronger motivation and an interest in behavior change. Thus, our study design might have restricted us to fully make use of this coaching technique, which in turn may have influenced the actual intervention effect of telephone coaching.

We did not track the participants' logins to the tailored website enclosing the personal health feedback, which would have allowed us to study the participants' use of the website. We have previously reported findings on the association between number of logins to a Web-based weight club and greater weight loss among men [13]. Furthermore, van Genugthen et al suggest that participants with adequate skills to self-regulate their health behaviors and who possess action planning skills are more likely to visit a tailored Web-based intervention compared with participants with less self-regulation and action planning skills [40]. However, Robroek et al state that even if web-based messages may increase website use, the actual use of the website's tools might be low [41]. Thus, future studies are strongly recommended to invest time in promoting and documenting the participants' use of interventions. Careful tracking may not only improve the quality of the study but also ensure participants' partaking of proposed interventions.

We have previously presented results on response rate in this study, suggesting that the overall response rate may be vulnerable to specific occupational groups with less screen time, season of study, and the possibility to participate during work hours [4]. The method of using the Internet to deliver health interventions seems however appropriate for our study population. All participants in our study had email addresses and access to computers or mobile phones during time of study. About 90% of the Swedish population has access to the Internet, and 85% has broadband at home. A high computer literacy and experience of use among our study population is therefore likely.

There is no reason to believe that selection bias of highly motivated participants affected this study. Most of the participants were not motivated to change their dietary and physical activity habits at baseline. However, all the measures in this study rely on self-reports, which is a limitation in this study. Future studies would benefit from adding device-based measures, for example, for assessing physical activity. For instance, accelerometers can be used for large surveys at acceptable costs and to provide rather reliable measures [42]. Yet, the advancement in device-based measurers has escalated the past years and at the time of planning this study, little research using these devices had been done, thus explaining the choice of method of this Web- and telephone-based intervention study.

It cannot go unnoticed that our study sample represents a challenging subgroup in health research. Our participants represented mostly middle-aged men working in the transport industry. Extending our findings to other populations is therefore difficult. According to prior literature, overweight, middle-aged men are generally underrepresented in epidemiological lifestyle studies, supporting our results on total response rate [43]. In addition, previous research indicates that men have a tendency to procrastinate until they seek health care [44]. Therefore, any health improvement on our study group, such as increased motivation found in this study is of significance. The optimal intervention still depends on the individual's motivation, personal preference, and level of implementation [38]. Consequently, even if we did not find any health effects from using tailored Web-based health feedback and optional telephone coaching as intervention tools to promote healthier lifestyle behaviors in our target group, it may have increased our participants' motivation to adopt healthier dietary and physical activity habits. Future research is recommended to further study the effects on motivation with respect to short-term and long-term effects.

Conclusion

Tailored Web-based health feedback and the offering of optional telephone coaching did not have a positive health effect on employees in the transport service industry. However, our findings suggest an increased short-term motivation to change health behaviors related to diet and physical activity among those receiving tailored Web-based health feedback.

Acknowledgments

This study was funded by the Swedish Transport Administration, the Erik and Edith Fernstrom's foundation, Sven and Dagmar Saléns foundation, and the regional agreement on medical training and clinical research between Stockholm County Council and Karolinska Institutet, Sweden.The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors would like to acknowledge the contributions from all the collaborating units and assistantship during the conduct of this study: the Swedish Transport Administration's offices in Sweden, (mainly Nils Edström, Ingrid Bråborg, and Berit Gille), SJ AB, Stockholmståg, Veolia Transport, the Quit-Smoking-Helpline, the Alcohol-Helpline, and Twitch Capital AB handling the Diet and Exercise Helpline (particularly Fredrik Karlsson and Nina Svensson). Finally, the authors agree that their students Judith van Broek and Thomas Cammelbeeck deserve a special note for their valuable efforts during the data management of preliminary results.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT eHealth checklist V 1.6.1.

[PDF File (Adobe PDF File), 928KB - jmir_v18i8e158_app1.pdf]

Multimedia Appendix 2

The participants' reported health at baseline and follow-up (n=981).

[PDF File (Adobe PDF File), 43KB - jmir_v18i8e158_app2.pdf]

References

1. World Health Organization. Global status report on noncommunicable diseases 2014 URL: http://www.who.int/nmh/publications/ncd-status-report-2014/en/ [accessed 2016-03-10] [WebCite Cache ID 6fuNG2E9S]
2. Das P, Horton R. Rethinking our approach to physical activity. Lancet 2012 Jul 21;380(9838):189-190. [doi: 10.1016/S0140-6736(12)61024-1] [Medline: 22818931]
3. Kecklund G, Ingre M, Åkerstedt T. Working hours, health and safety - an update of current research (Original title in Swedish). In: Stockholm The Stress Research Institute. Stockholm: Universitetsservice Frescati; 2010.
4. Svensson M, Svensson T, Hansen AW, Trolle LY. The effect of reminders in a web-based intervention study. Eur J Epidemiol 2012 May;27(5):333-340. [doi: 10.1007/s10654-012-9687-5] [Medline: 22531973]
5. World Health Organization. 2015. The workplace: a priority setting for health promotion Internet URL: http://www.who.int/occupational_health/topics/workplace/en/ [accessed 2015-06-15] [WebCite Cache ID 6ZJ6KGb0p]
6. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: A systematic review. Health Informatics J 2015 Feb 3:NA Epub ahead of print. [doi: 10.1177/1460458214567004] [Medline: 25649783]
7. Flores M, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
8. Petersen CB, Severin M, Hansen AW, Curtis T, Grønbæk M, Tolstrup JS. A population-based randomized controlled trial of the effect of combining a pedometer with an intervention toolkit on physical activity among individuals with low levels of physical activity or fitness. Prev Med 2012 Feb;54(2):125-130. [doi: 10.1016/j.ypmed.2011.12.012] [Medline: 22200586]
9. van den Berg M, Schoones JW, Vliet Vlieland T. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
10. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: A systematic literature review. J Telemed Telecare 2015 Feb 12:NA Epub ahead of print. [doi: 10.1177/1357633X15572203] [Medline: 25680388]
11. Kass AE, Trockel M, Safer DL, Sinton MM, Cunning D, Rizk MT, et al. Internet-based preventive intervention for reducing eating disorder risk: A randomized controlled trial comparing guided with unguided self-help. Behav Res Ther 2014 Oct 2;63C:90-98. [doi: 10.1016/j.brat.2014.09.010] [Medline: 25461783]
12. Svensson M, Hult M, van der Mark M, Grotta A, Jonasson J, von HY, et al. The change in eating behaviors in a Web-based weight loss program: a longitudinal analysis of study completers. J Med Internet Res 2014;16(11):e234 [FREE Full text] [doi: 10.2196/jmir.3131] [Medline: 25367316]
13. van der Mark M, Jonasson J, Svensson M, Linné Y, Rossner S, Lagerros YT. Older members perform better in an internet-based behavioral weight loss program compared to younger members. Obes Facts 2009;2(2):74-79. [doi: 10.1159/000209383] [Medline: 20054209]
14. Svensson M, Bellocco R, Bakkman L, Trolle LY. An interactive internet-based plate for assessing lunchtime food intake: a validation study on male employees. J Med Internet Res 2013;15(1):e13 [FREE Full text] [doi: 10.2196/jmir.2217] [Medline: 23335728]
15. 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]
16. Cameron D, Epton T, Norman P, Sheeran P, Harris PR, Webb TL, et al. A theory-based online health behaviour intervention for new university students (U@Uni:LifeGuide): results from a repeat randomized controlled trial. Trials 2015;16:555 [FREE Full text] [doi: 10.1186/s13063-015-1092-4] [Medline: 26643917]
17. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
18. Neri AJ, Momin BR, Thompson TD, Kahende J, Zhang L, Puckett MC, et al. Use and effectiveness of quitlines versus Web-based tobacco cessation interventions among 4 state tobacco control programs. Cancer 2016 Apr 1;122(7):1126-1133. [doi: 10.1002/cncr.29739] [Medline: 26854479]

19. Heinemans N, Toftgård M, Damström-Thakker K, Galanti MR. An evaluation of long-term changes in alcohol use and alcohol problems among clients of the Swedish National Alcohol Helpline. Subst Abuse Treat Prev Policy 2014;9:22 [FREE Full text] [doi: 10.1186/1747-597X-9-22] [Medline: 24893718]
20. Helgason AR, Tomson T, Lund KE, Galanti MR, Ahnve S, Gilljam H. Factors related to abstinence in a telephone helpline for smoking cessation. Eur J Public Health 2004 Sep;14(3):306-310 [FREE Full text] [Medline: 15369039]
21. Nohlert E, Öhrvik J, Helgason AR. Non-responders in a quitline evaluation are more likely to be smokers - a drop-out and long-term follow-up study of the Swedish National Tobacco Quitline. Tob Induc Dis 2016;14:5 [FREE Full text] [doi: 10.1186/s12971-016-0070-2] [Medline: 26843854]
22. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
23. Westerlund A, Brandt L, Harlid R, Åkerstedt T, Lagerros YT. Using the Karolinska Sleep Questionnaire to identify obstructive sleep apnea syndrome in a sleep clinic population. Clin Respir J 2014 Oct;8(4):444-454. [Medline: 25396257]
24. Lagerros YT, Mucci LA, Bellocco R, Nyrén O, Bälter O, Bälter KA. Validity and reliability of self-reported total energy expenditure using a novel instrument. Eur J Epidemiol 2006;21(3):227-236. [doi: 10.1007/s10654-006-0013-y] [Medline: 16547838]
25. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [Medline: 10993420]
26. McTiernan A, Kooperberg C, White E, Wilcox S, Coates R, Adams-Campbell LL, Women's Health Initiative Cohort Study. Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women's Health Initiative Cohort Study. JAMA 2003 Sep 10;290(10):1331-1336. [doi: 10.1001/jama.290.10.1331] [Medline: 12966124]
27. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
28. Netsurvey. About us - From questions to results 2012 URL: http://www.netsurvey.se/en/ [accessed 2015-06-15] [WebCite Cache ID 6ZJ5nQ0t4]
29. Swedish National Institute for Public Health. Physical Actvity in the Prevention and Treatment of Disease 2015 URL: http://www.fyss.se/om-fyss-2/fyss-2015/ [accessed 2015-06-15] [WebCite Cache ID 6ZJ63MSrS]
30. Stockholm Health Care Services, Stockholm County Counsil. 2015. Sluta-Röka-Linjen (The Swedish National Tobacco Quitline) URL: http://slutarokalinjen.org/ [accessed 2015-06-15] [WebCite Cache ID 6ZJ69aFjW]
31. Miller WR. Motivational interviewing: research, practice, and puzzles. Addict Behav 1996;21(6):835-842. [Medline: 8904947]
32. Ayres K, Conner M, Prestwich A, Hurling R, Cobain M, Lawton R, et al. Exploring the question-behaviour effect: randomized controlled trial of motivational and question-behaviour interventions. Br J Health Psychol 2013 Feb;18(1):31-44. [doi: 10.1111/j.2044-8287.2012.02075.x] [Medline: 22519696]
33. Dholakia UM. A Critical Review of Question-Behavior Effect Research. In: Malhotra NK, editor. Review of marketing research. Bingley, UK: Emerald Group Publishing Limited; 2010:145-197.
34. Rodrigues AM, O'Brien N, French DP, Glidewell L, Sniehotta FF. The question-behavior effect: genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses. Health Psychol 2015 Jan;34(1):61-78. [doi: 10.1037/hea0000104] [Medline: 25133835]
35. 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]
36. Malik SH, Blake H, Suggs LS. A systematic review of workplace health promotion interventions for increasing physical activity. Br J Health Psychol 2014 Feb;19(1):149-180. [doi: 10.1111/bjhp.12052] [Medline: 23827053]
37. Pressler A, Knebel U, Esch S, Kölbl D, Esefeld K, Scherr J, et al. An internet-delivered exercise intervention for workplace health promotion in overweight sedentary employees: a randomized trial. Prev Med 2010;51(3-4):234-239. [doi: 10.1016/j.ypmed.2010.07.008] [Medline: 20638409]
38. Schulz DN, Kremers S, Vandelanotte C, van Adrichem MJ, Schneider F, Candel MJ, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
39. Schulz DN, Schneider F, de VH, van Osch L, van Nierop P, Kremers S. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
40. van Genugten L, van Empelen P, Oenema A. Intervention use and action planning in a web-based computer-tailored weight management program for overweight adults: randomized controlled trial. JMIR Res Protoc 2014;3(3):e31 [FREE Full text] [doi: 10.2196/resprot.2599] [Medline: 25057122]
41. Robroek SJW, Lindeboom DEM, Burdorf A. Initial and sustained participation in an internet-delivered long-term worksite health promotion program on physical activity and nutrition. J Med Internet Res 2012;14(2):e43 [FREE Full text] [doi: 10.2196/jmir.1788] [Medline: 22390886]

42. Lee I, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. Br J Sports Med 2014 Feb;48(3):197-201 [FREE Full text] [doi: 10.1136/bjsports-2013-093154] [Medline: 24297837]
43. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
44. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003 Jan;58(1):5-14. [Medline: 12674814]

Abbreviations

BMI: body mass index
GEE model: generalized estimating equation regression models
MET: metabolic equivalent of task
TTM: transtheoretical model

Edited by G Eysenbach; submitted 20.11.15; peer-reviewed by G Jerome, C Marsaux, M Conner; comments to author 31.01.16; revised version received 31.03.16; accepted 30.04.16; published 11.08.16.

Please cite as:
Solenhill M, Grotta A, Pasquali E, Bakkman L, Bellocco R, Trolle Lagerros Y
The Effect of Tailored Web-Based Feedback and Optional Telephone Coaching on Health Improvements: A Randomized Intervention Among Employees in the Transport Service Industry
J Med Internet Res 2016;18(8):e158
URL: http://www.jmir.org/2016/8/e158/
doi:10.2196/jmir.4005
PMID:27514859

Original Paper

Web-Based Video-Coaching to Assist an Automated Computer-Tailored Physical Activity Intervention for Inactive Adults: A Randomized Controlled Trial

Stephanie Alley[1], PhD; Cally Jennings[2], PhD; Ronald C Plotnikoff[3], PhD; Corneel Vandelanotte[1], PhD

[1]Physical Activity Research Group, School of Human, Health and Social Sciences, Central Queensland University, Rockhampton, Australia

[2]Alberta Center for Active Living, Faculty of Physical Education and Recreation, University of Alberta, Edmonton, AB, Canada

[3]Priority Research Centre for Physical Activity and Nutrition, Faculty of Health and Medicine, University of Newcastle, Callaghan, Australia

Corresponding Author:
Stephanie Alley, PhD
Physical Activity Research Group
School of Human, Health and Social Sciences
Central Queensland University
Building 77
Bruce Highway
Rockhampton, 4701
Australia
Phone: 61 749232263
Fax: 61 749232265
Email: s.alley@cqu.edu.au

Abstract

Background: Web-based physical activity interventions that apply computer tailoring have shown to improve engagement and behavioral outcomes but provide limited accountability and social support for participants. It is unknown how video calls with a behavioral expert in a Web-based intervention will be received and whether they improve the effectiveness of computer-tailored advice.

Objective: The purpose of this study was to determine the feasibility and effectiveness of brief video-based coaching in addition to fully automated computer-tailored advice in a Web-based physical activity intervention for inactive adults.

Methods: Participants were assigned to one of the three groups: (1) tailoring + video-coaching where participants received an 8-week computer-tailored Web-based physical activity intervention ("My Activity Coach") including 4 10-minute coaching sessions with a behavioral expert using a Web-based video-calling program (eg, Skype; n=52); (2) tailoring-only where participants received the same intervention without the coaching sessions (n=54); and (3) a waitlist control group (n=45). Demographics were measured at baseline, intervention satisfaction at week 9, and physical activity at baseline, week 9, and 6 months by Web-based self-report surveys. Feasibility was analyzed by comparing intervention groups on retention, adherence, engagement, and satisfaction using *t* tests and chi-square tests. Effectiveness was assessed using linear mixed models to compare physical activity changes between groups.

Results: A total of 23 tailoring + video-coaching participants, 30 tailoring-only participants, and 30 control participants completed the postintervention survey (83/151, 55.0% retention). A low percentage of tailoring + video-coaching completers participated in the coaching calls (11/23, 48%). However, the majority of those who participated in the video calls were satisfied with them (5/8, 71%) and had improved intervention adherence (9/11, 82% completed 3 or 4 modules vs 18/42, 43%, *P*=.01) and engagement (110 minutes spent on the website vs 78 minutes, *P*=.02) compared with other participants. There were no overall retention, adherence, engagement, and satisfaction differences between tailoring + video-coaching and tailoring-only participants. At 9 weeks, physical activity increased from baseline to postintervention in all groups (tailoring + video-coaching: +150 minutes/week; tailoring only: +123 minutes/week; waitlist control: +34 minutes/week). The increase was significantly higher in the tailoring + video-coaching group compared with the control group (*P*=.01). No significant difference was found between intervention groups and no significant between-group differences were found for physical activity change at 6 months.

Conclusions: Only small improvements were observed when video-coaching was added to computer-tailored advice in a Web-based physical activity intervention. However, combined Web-based video-coaching and computer-tailored advice was

effective in comparison with a control group. More research is needed to determine whether Web-based coaching is more effective than stand-alone computer-tailored advice.

Trial Registration: Australian New Zealand Clinical Trials Registry (ACTRN): 12614000339651; http://www.anzctr.org.au/TrialSearch.aspx?searchTxt=ACTRN12614000339651+&isBasic=True (Archived by WebCite at http://www.webcitation.org/6jTnOv0Ld)

(J Med Internet Res 2016;18(8):e223) doi:10.2196/jmir.5664

KEYWORDS

motor activity; health promotion; chronic disease; e-counseling; Internet

Introduction

Physical activity improves physical and mental health and significantly lowers the risk of noncommunicable disease, including cardiovascular disease, diabetes, and cancer [1]. Australian guidelines recommend 150-300 minutes of moderate-intensity activity each week, over 5 days, to receive health benefits and reduce the risk of noncommunicable disease [2]. Despite this, less than 50% of Australians meet these recommendations [3]. As such, there is a need for effective and affordable physical activity interventions with a broad reach.

Health behavior change interventions delivered via the Web have the potential to reach a large audience at low cost, due to the majority of Australians (92%) having access to the Web [4]. Furthermore, they are convenient for participants and enable the content to be delivered in a nonconfrontational manner [5-7]. Many studies have found Web-based physical activity interventions to be effective in the short term. For example, reviews of Web-based physical activity interventions have found more than half of controlled trials to report positive physical activity outcomes [8,9]. However, problems with low engagement and high dropout rates lead to small and short-term behavior changes [8,10]. Furthermore, many interventions struggle to reach those most in need of increasing their physical activity [8,10-12]. Personalized health advice through coaching sessions or computer-tailored feedback engages participants and improves the effectiveness of Web-based health behavior interventions [13-15]. Both Web-based coaching sessions and computer-tailored advice provide support similar to traditional face-to-face coaching sessions, at a lower cost with fewer geographical restrictions [13,16].

Coaching is defined as facilitating health behavior change through interactions between a health professional (coach) and a client [17]. Web-based coaching sessions are most similar to traditional face-to-face coaching as they provide personal interaction. Coaching in Web-based physical activity interventions improves participants' perceptions of their social support, which is associated with greater levels of behavior change [18,19]. The Social Cognitive Theory stipulates that the acquisition of a new behavior is influenced by the individuals' cognitive factors (eg, attitude), behavioral factors (eg, skills), and contextual factors including reinforcement, instructions, and social norms [20]. Including coaching in a Web-based intervention addresses the otherwise overlooked social contextual factors that play an important role in behavior change [20].

Advances in Internet technology and broadband capacity now allow the option of delivering coaching sessions via free Web-based video-calling programs, which allow participants to view the coach and engage in a verbal discussion. The popular video-calling programs Skype, Google Hangouts, and FaceTime are encrypted, which ensures privacy of participant information [21]. Video-coaching facilitates higher engagement, feelings of accountability, and social support, and reduces the risk of misunderstandings compared with emails and instant messaging [22,23]. Video-coaching has been found to be effective in producing changes in parenting behavior [24], to have a high feasibility for smoking cessation [25], and a high feasibility for supporting in-home rehabilitation in the elderly [26]. However, no studies have used video-coaching as part of a Web-based physical activity program. Despite the potential of Web-based video-coaching, its reliance on the time of a behavior change expert leads to higher implementation costs compared with fully automated computer-tailored advice.

Computer tailoring can deliver personalized advice at a low cost by using a computer-based expert system to automatically deliver feedback to participants' responses to a Web-based questionnaire [13]. Computer-tailored physical activity advice is preferred by participants, leads to greater attention [27], and improved health behavior outcomes compared with generic health advice [14]. Although the effectiveness of computer tailoring is well established and it has the benefit of providing personalized advice to large numbers at low cost, it is unknown whether it could be more effective with an element of human support.

To our knowledge, no health behavior interventions have combined computer-tailored advice with Web-based video-coaching. This approach may improve intervention outcomes by utilizing the benefits of both methods. Providing computer-tailored advice can limit the time required from a video-coach, therefore limiting costs, as well as reducing reliance on the knowledge and expertise of the coach. A brief coaching session can add an element of social support, as well as further explanation, personalization, and interpretation of theory-based computer-tailored advice received by participants at an earlier time. It is unknown whether brief video-coaching sessions to reiterate computer-tailored physical activity advice are feasible in terms of retention, adherence, engagement, and satisfaction. It is also unknown whether they lead to improved physical activity and quality of life compared with stand-alone computer-tailored advice. Therefore, this study explores the feasibility and effectiveness of a brief Web-based coaching

session in addition to computer-tailored advice for inactive adults.

The first aim of the study was to determine the feasibility of brief video-coaching, when used to discuss previously received computer-tailored physical activity advice, in a stand-alone Web-based intervention for inactive adults. Feasibility was determined by adherence and satisfaction of the coaching sessions and comparing intervention retention, adherence, website engagement, and satisfaction of the tailoring + video-coaching and tailoring-only groups. The second aim was to test the effectiveness of the video-coaching sessions in terms of physical activity and quality of life outcomes. It was hypothesized that computer tailoring in combination with video-coaching would result in greater retention, adherence, engagement, and satisfaction with the intervention, compared with a computer-tailored–only group and a waitlist control group. It was also hypothesized that computer tailoring and video-coaching would result in greater improvements in quality of life and physical activity compared with a computer-tailored–only group and a waitlist control group.

Methods

Research Procedure

A detailed account of the methods can be found in the protocol paper [28] and a consort eHealth checklist for the paper can be found here (Multimedia Appendix 1) [29]. The recruitment methods, participant eligibility, protocol, intervention description, measures, and data analysis are summarized below.

Recruitment

Print advertising and Web advertising were used to recruit participants from a number of Australian metropolitan and regional cities (Sydney, Melbourne, Perth, Brisbane, Rockhampton, Bundaberg, Mackay, and Townsville). Print advertising included newspaper advertisements and articles, posters and leaflets displayed in health clinics, and leaflets distributed to peoples' homes. The Web advertising included links displayed on community websites and paid advertisements on Google and Facebook. Ethics approval was received from the Central Queensland University Human Research Ethics Committee (H13/04-044), before recruitment took place from March 2014 to January 2015. This study is registered with the Australian New Zealand Clinical Trials Registry (ACTRN12614000339651).

Participants

People were eligible to participate if they were English-speaking Australian adults (older than 18 years). Participants were excluded if they were pregnant, at risk of injury or ill health from their increasing physical activity (as assessed by the Physical Activity Readiness Questionnaire), or if they were already meeting the physical activity recommendations (as assessed via a single item asking if participants participated in 30 minutes of physical activity on most days). It is likely that the intervention attracted participants with a high Internet literacy.

Protocol

Information about the study, including the affiliation with Central Queensland University, was available on the landing page of the intervention website (Multimedia Appendices 2 and 3). To assess individuals' eligibility, how they heard about the program, and collect contact details, a screening questionnaire was delivered through the intervention website. Eligible participants were randomly assigned based on a sequence (not concealed) of random numbers between 1 and 3 to one of three study arms: tailoring + video-coaching, tailoring-only, or waitlist control. This was done in blocks of 15 participants. SA generated the random allocation sequence and assigned participants to groups. Participants remained blinded to their condition until after completing all baseline measures. Participants began the intervention on the Monday following their recruitment. The consent form and then baseline questionnaire were administered through the intervention website for all groups. Upon completing the baseline questionnaire, the intervention groups received module 1 of their personalized advice, whereas the control group received nothing. The intervention "My Activity Coach" delivered 1 module of computer-tailored advice every 2 weeks over 8 weeks (4 modules in total). During the weeks where no new modules were received, participants in the tailoring + video-coaching group received a brief coaching session through a Web-based video-calling program (eg, Skype) to reiterate the advice received in their previous module. Participants in the tailoring-only group received an email reminding them of the tailored advice they received in the previous module to ensure both intervention groups received the same number of contacts. Participants in the waitlist control group were given the opportunity to participate in the intervention without coaching after they completed the final questionnaire. Questionnaires were administered through the intervention website immediately after the end of the intervention (week 9) and 6 months after the end of the intervention. Participants had to log in to complete each survey, which ensured they were only completed once. The week 9 questionnaires were collected from June 2014 to March 2015. It was not possible to blind researchers to participants' group assignment after they had completed the baseline questionnaire. Participants were blinded to their group assignment only when completing their baseline questionnaire. The consent sheet explained the 2 interventions and therefore it is possible that participants worked out whether they were in the intervention of interest or comparator. Participants who completed all surveys went in the draw to win 1 of 30 pedometers, 6 Fitbits, and 3 heart rate monitors. Because many individuals began but failed to complete the screening questionnaire, we tested conducting the screening questionnaire by phone (after receiving an ethics amendment from the Central Queensland University's Human Research Ethics Committee). This was done for 15 prospective participants and discontinued because of failure to increase screening completions. The Web-based booking system for coaching participants was changed halfway through the trial as the first booking system was discontinued. No other changes to the protocol were carried out during the trial.

Intervention

The 8-week "My Activity Coach" Web-based intervention delivered a new module of tailored advice to participants every 2 weeks [30]. Each module required participants to complete a brief Web-based questionnaire about their physical activity and psychosocial correlates of physical activity. Feedback was then provided based on their responses to the questionnaire (Multimedia Appendices 4 and 5). Participants received up to 4 reminder emails and a reminder phone call when they did not complete the survey required for each module. The tailored advice was based on behavior change theory (Theory of Planned Behavior [31]) and communication theory (Elaboration Likelihood Model [32]). Each module began with a graph including bars to represent participants' current physical activity, their physical activity during the previous modules, as well as the minimum and optimal physical activity recommendations. Module 1, titled "Are you active enough," explained the physical activity recommendations and health benefits of physical activity tailored to their body mass index (BMI), age, and level of physical activity. The module ended with a suggested goal (based on their current activity level) to work toward until the next module. Module 2, titled "Let's set some goals," provided participants with information on goal setting and action planning. Module 3, titled "Physical activity and your environment," delivered tailored information on using participants' social and physical environments to increase their physical activity. Module 4, titled "Staying active," addressed relapse prevention. Participants also received tailored advice on their perceived benefits and barriers to being active and self-efficacy to become more active throughout the modules. The modules and intervention website were adapted from an earlier 2-module Web-based intervention with computer-tailored advice for inactive adults [33]. Focus groups were conducted to inform development of this prior intervention [34]. Updates were conducted by a website developer who also created the original intervention website.

An action-planning tool became available to participants after module 2. The tool allowed participants to create an action plan for up to 4 activities (specifying where, when, for how long, and with whom they will be active over the following 2 weeks). Participants could print a calendar-based overview of their action plan (Multimedia Appendix 6). The coaching group's 10- to 15-minute biweekly video-coaching sessions were conducted through a Web-based video-calling program of participants' choice (eg, Skype, Google Hangouts, Yahoo Messenger, and FaceTime). During the session the activity coach commented on the tailored advice participants received in the module from the previous week, answered any questions participants had, and provided encouragement, support, and accountability.

Measures

Participants' demographics including sex, age, BMI, household income (less than AUD $31,200, AUD $31,200-$77,999, more than AUD $78,000), education (less than secondary, secondary, further education), and employment (full time, part time or casual, and not in paid employment) were assessed in the baseline survey. Participants were also asked if they used a video-calling program (Skype, Google Hangouts, FaceTime, other, none). Completion of the coaching sessions, the length of the coaching sessions, and reasons for missed coaching sessions were recorded by the coach. Intervention retention and adherence were assessed by recording participants' completion of the research surveys and intervention modules, respectively, and website engagement was measured through Google Analytics. Google Analytics recorded the number of website visits and time spent on the website for each participant. Intervention participants' satisfaction with the intervention was assessed at the end of the intervention (week 9). Satisfaction with module questions (4 items), computer-tailored advice (14 items), website usability (13 items), overall program (5 items), and coaching (14 items) were all assessed. The items were specifically developed for this study, although based on previous research [35]. The items were on a 5-point Likert scale where participants were asked to rate their agreement (1=strongly agree to 5= strongly disagree) to statements about the intervention. All positively framed questions were reverse scored. For each category participants with a mean rating of 3.6 or higher (maximum = 5) were categorized as "satisfied." Coaching participants were also asked if they completed a coaching session and if not, why not. All intervention participants were asked 4 open-ended questions about 3 topics: the advice, website, and overall program. Coaching participants who completed a coaching session were asked an additional 4 open-ended questions specifically relating to the coaching sessions. The 4 questions for each category (advice, website, program, and coaching) were (1) what did you like about the advice, website, program, or coaching; (2) what did you not like; (3) any recommendations for improvement; and (4) any other thoughts (Multimedia Appendix 7). Responses for all questions were thematically analyzed. The main outcome, weekly physical activity, was assessed at all time points via the Active Australia Survey, (AAS), which has a high percentage agreement with other physical activity measures (67%-75%) [36] and has a good test-retest reliability (kappa = .52) [37] including when self-administered [38]. Quality of life was measured at all 3 time points by the SF-12 version 2, which is valid [39] and reliable [40] including when self-administered on the Web [41]. Physical health and mental health component scores were calculated from the SF-12 version 2 following manual instructions [42].

Data Analysis

Baseline demographics for participants in each trial arm (tailoring + video-coaching, tailoring-only, and control) are presented (Table 1). The demographics of completers versus dropouts, as well as coaching participants who did versus did not complete a coaching session, were compared using chi-square and *t* tests. To test feasibility of the coaching sessions, completion of the coaching sessions, the length of the coaching sessions, and reasons for missed coaching sessions were presented. Next, the 2 intervention groups, as well as coaching participants who did versus did not complete a coaching session, were compared on retention (dropouts vs completers) and adherence (completed 1-2 modules vs 3-4 modules) using a chi-square test, number of website visits and time spent on the website using *t* tests, and satisfaction scores (satisfied vs neutral or not satisfied) using a chi-square test. To test effectiveness,

longitudinal data were analyzed using intention-to-treat principles. Physical activity, mental health score, and physical health score were each modeled using linear mixed models with time (baseline, week 9, and 6 months) as a repeated factor, fixed effects of time and group (control, tailoring-only, tailoring + video-coaching), and a time by group interaction (Table 2). Significance level was set to $P<.05$.

Sample Size

Sample size calculations demonstrated that a sample size of 300, or 100 in each study arm, was required to detect between-group differences in physical activity from baseline to postintervention using linear mixed models [43]. This calculation was based on the alpha level of ≤.05 (80% power) and a small effect size (0.43) and 25% attrition, which are common in similar interventions [10].

Results

Flow of Participants

Of the 239 randomly assigned participants, 154 completed the baseline questionnaire and at least one of the intervention modules. Of these, 84 participants completed the postintervention survey at week 9 (55% retention). A total of 59 participants completed the 6-month follow-up questionnaire (38% retention). There were no demographic differences between those who completed the week 9 survey and those who did not. The majority of participants were recruited through Facebook (63%), and small percentages were recruited through Google (8%), a newspaper article (6%), letterbox drops (5%), family or friend (5%), leaflets (5%), posters (4%), community websites (3%), and newspaper advertisements (1%). Figure 1 presents the flow of participants through the trial.

Table 1. Baseline characteristics, physical activity, and quality of life by group assignment

Variables		Tailoring + video coaching	Tailoring only	Control
Sex n=154, n (%)				
	Male	16 (30)	14 (25)	7 (16)
	Female	37 (67)	42 (75)	38 (84)
Employment n=151[a], n (%)				
	Full time	19 (37)	21 (38)	18 (40)
	Part time or casual	10 (19)	10 (19)	9 (20)
	Not in paid employment	23 (44)	23 (43)	18 (40)
Education n=151[a], n (%)				
	Less than secondary	2 (4)	0 (0)	0 (0)
	Secondary	8 (15)	10 (19)	6 (13)
	Further education	42 (81)	44 (81)	39 (87)
Income n=112[b], n (%)				
	More than AUD $78,000	21 (55)	14 (36)	21 (60)
	AUD $31,200-$77,999	11 (29)	17 (44)	6 (17)
	Less than AUD $31,199	6 (16)	8 (20)	8 (23)
Uses Web-based video-calling n=151[a], n (%)				
	Yes	51 (63)	53 (68)	63 (79)
	No	30 (37)	25 (32)	17 (21)
Age (years), n=154		55.26 (10.93)	52.18 (11.53)	55.18 (13.45)
BMI[c], n=150[d], M(SD)		32.08 (7.43)	31.58 (7.43)	29.97 (6.75)
Baseline total physical activity (minutes/week), n=151, M(SD)		189.52 (214.30)	152.87 (174.33)	160.44 (191.23)
Week 9 total physical activity (minutes/week), n=83, M(SD)		387.83 (264.89)	315.17 (264.39)	211.00 (164.16)
6-month total physical activity (minutes/week), n=59, M(SD)		419.52 (77.109)	319.71 (164.77)	305.00 (315.86)
Baseline mental health score, n=148, M(SD)		46.10 (1.30)	41.36 (12.27)	42.74 (1.78)
Week 9 mental health score, n=82, M(SD)		48.97 (10.27)	30.03 (6.12)	45.19 (10.50)
6-month mental health score, n=59, M(SD)		48.16 (12.11)	43.90 (10.34)	44.94 (11.84)
Baseline physical health score, n=146, M(SD)		46.90 (10.17)	51.92 (8.28)	48.62 (8.98)
Week 9 physical health score, n=78, M(SD)		46.88 (12.05)	51.48 (6.95)	47.32 (9.37)
6-month physical health score, n=59, M(SD)		46.60 (10.22)	52.38 (5.47)	48.18 (10.46)

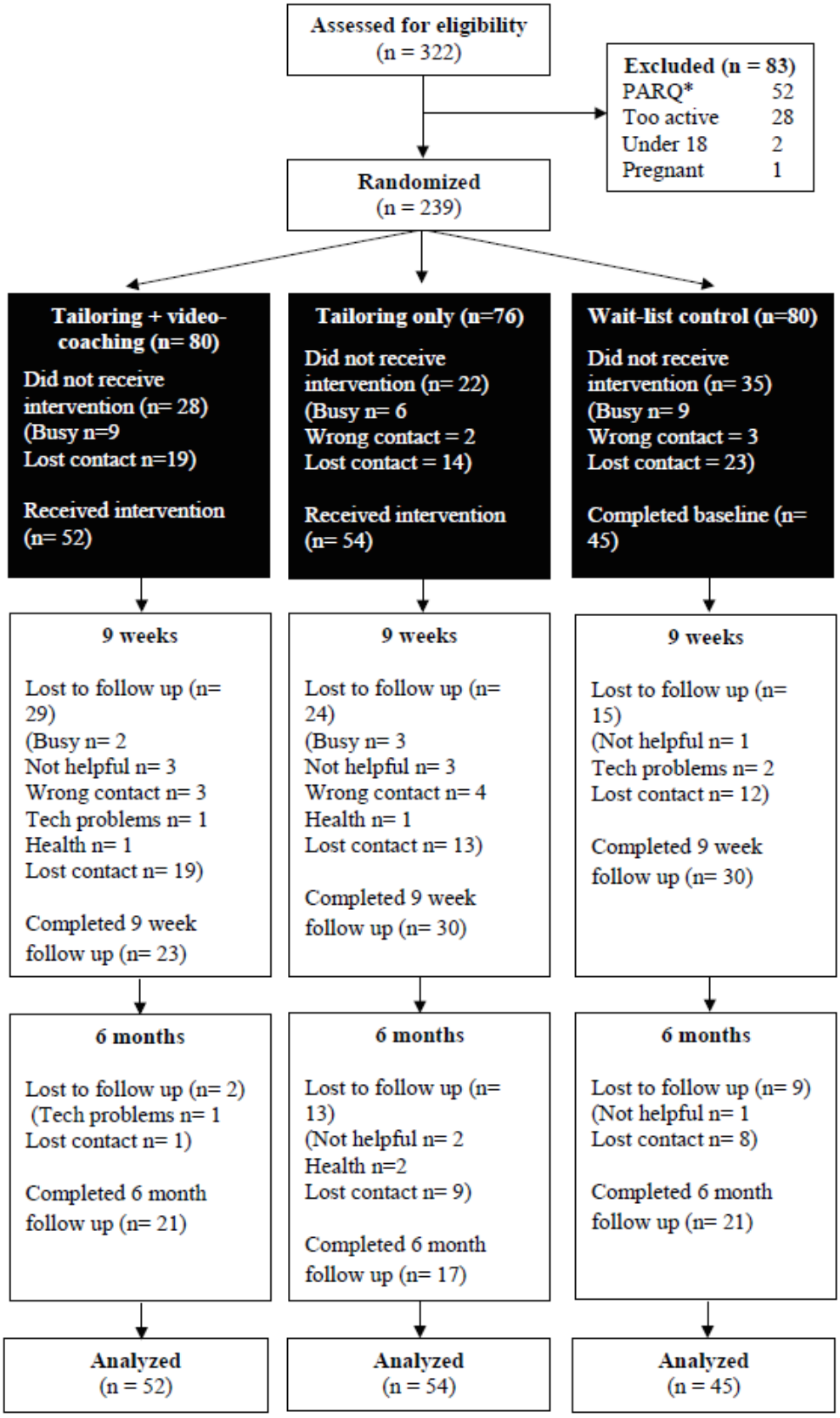
[a]Baseline data (employment, education, video calling use and BMI) were lost for 3 participants.

[b]A total of 39 respondents chose not to disclose their income.

[c]BMI: body mass index.

[d] BMI was lost for four participants.

Figure 1. Participant flow through the intervention trial. *PARQ: Physical Activity Readiness Questionnaire.



Sample Characteristics

The majority of participants were female (117/154, 76.0%) and were on average 54 years of age. Just under half (64/151, 43%) were not in paid employment and the majority (125/151, 82.8%) had completed a higher education course. Less than half (n=60/151, 39.7%) of participants were physically active and on average participated in 168 minutes of physical activity. The average BMI was 31, which is in the obese range. The majority (n=106/151, 70.2%) used a video-calling program.

Coaching Adherence and Satisfaction

Coaching participants had low adherence to the coaching sessions. Just under half of coaching group completers (11/23, 48%) and less than a quarter of all coaching participants (11/52, 21%) completed 1 coaching session. There were no demographic differences between those who participated in a coaching session and those who did not. Of the coaching participants who did not do a coaching session, 8 wanted a second chance to do a session but did not book or show up the second time, 7 refused to do one, 4 were too busy, 7 had technical difficulties, 2 had injuries, and contact was lost with 13. Of those who participated in at least 1 coaching session, an average of 2.4 sessions were completed, the average coaching session length was 10.4 minutes, and 15% of the coaching calls were interrupted with technical difficulties. A total of 8 coaching participants (8/52, 15%) completed the coaching satisfaction questions. The majority were satisfied with the coaching sessions (n=5/8, 71.4%). In response to open-ended coaching questions, participants said the sessions held them accountable (n=3), appreciated the support (n=1), appreciated the information (n=1), liked the structure (n=1), and liked talking about exercise that suits them (n=1). However, some had technical problems (n=2) and would have preferred to use a phone (n=1).

Intervention Adherence and Retention

Retention did not differ between intervention groups (χ^2_1= 4.7, *P*=.11). Participants who completed at least 1 coaching session had a higher percentage of week 9 survey completers (n=8/11, 73%), compared with other intervention participants (n=50/95, 53%), but this difference was not significant (χ^2_1= 1.6, *P*=.21). Just under half (n=50/106, 47%) of participants completed at least 3 of the 4 intervention modules. Intervention adherence was similar for the tailoring + video-coaching group and the tailoring-only group (χ^2_1= 2.1, *P*=.15); however, significantly more participants who participated in the coaching sessions completed at least 3 of the 4 intervention modules (n=9/11, 82%) compared with other intervention participants (n=41/95, 43%; χ^2_1= 6.0, *P*=.01).

Website Engagement

The average website visits and minutes spent on the website for the intervention groups were 7.53 (SD 7.14) and 87.07 (SD 77.33) minutes, respectively. The average number of website visits was similar for the tailoring + video-coaching group and the tailoring-only group ($t_{1,103}$= 0.05, *P*=.96). Average minutes spent on the website was higher for the tailoring + video-coaching group (mean 99.58 minutes, SD 95.71) than the tailoring-only group (mean 75.25 minutes, SD 52.90), but this was not significantly different (t_{1103}= 1.60, *P*=.11). Participants who completed the coaching sessions spent a significantly longer time (t_{1103}= 2.73, *P*=.02) on the intervention website (mean 174.64 minutes, SD 110.11) compared with other intervention participants (mean 77.84 minutes, SD 67.48). Participants who completed the coaching sessions also visited the website more frequently (mean 10.20, SD 3.85) compared with other intervention participants (mean 7.25, SD= 7.36), but this difference was not significant (t_{1103}= 1.24, *P*=.22).

Satisfaction

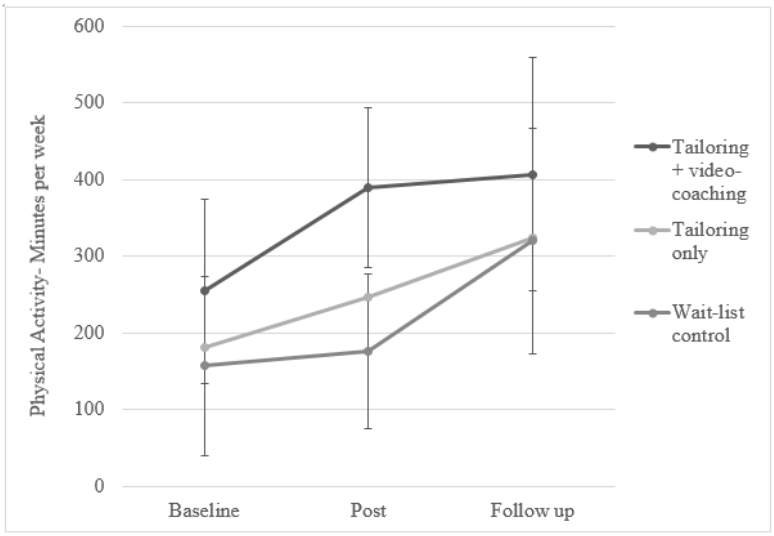
More than two-thirds of the participants were satisfied with the overall program (n=36/53, 68%), while the majority of participants were satisfied with the website usability (n=40/53, 77%), the computer-tailored advice (n=38/53, 76%), and the module questions (n=48/53, 91%). There was no difference between intervention groups on program satisfaction scores (χ^2_1= 0.14, *P*=.71). A higher percentage of those who participated in the coaching sessions (n=7/8, 88%) were satisfied with the program compared with other participants (n=29/45, 64%); however, this difference was not statistically significant (χ^2= 1.66, *P*=.20).

In response to open-ended questions on the overall program, participants mentioned that they liked the convenience (n=4), ease of use (n=4), the information (n=5), emails (n=2), found it motivating (n=5), and liked the accountability (n=2). Participants also mentioned that they would like more contact with a real person (n=7) and thought there were too many questions (n=2). In response to the questions on the advice, participants mentioned that it was easy to understand (n=12), was concise (n=4), laid out well (n=4), was nonjudgmental (n=2), and liked the personalization (n=2). However, some participants thought it was not personalized enough (n=11), did not like the Web-based format (n=4), and learned nothing new (n=2). In response to open-ended questions on the website, some participants mentioned that the website was easy to use (n=14), whereas others thought it was hard to use (n=5). Lastly, some participants thought the website could use more visuals and interesting links (n=6).

Physical Activity

Physical activity (minutes/week) improved from baseline to postintervention (week 9) and from baseline to follow-up (6 months) in all groups (Table 2,Figure 2). According to the linear mixed models analysis, the increase in physical activity from baseline to postintervention in the tailoring + video-coaching group in comparison with the control group was significant (Table 2). No significant difference was found between the intervention groups (Table 2). No significant differences were found between groups on physical activity changes from baseline to follow-up at 6 months.

Figure 2. Mean physical activity at baseline, week 9 (post intervention), and 6 months (follow-up), unadjusted. Baseline n=151, postintervention n=83, and follow-up n=59. Note: 95% confidence intervals presented for the coaching and control groups only.



Quality of Life

Physical health scores remained relatively constant across each time point (Table 2). Mental health scores remained relatively constant in the control and tailoring + video-coaching group; however, mental health scores dropped in the tailoring-only participants at postintervention (Table 2).

As such, the linear mixed models analysis indicated a significant difference in mental health between both intervention groups from baseline to postintervention (Table 2).

Table 2. Physical activity, mental health, and physical health changes by group: results of intention-to-treat analysis using linear mixed models

Outcome variables	Baseline to postintervention			Baseline to follow-up		
	Estimate[a]	*P* value	Cohen's *d*	Estimate[a]	*P* value	Cohen's *d*
Physical activity (n=151)						
Tailoring + video-coaching versus control	140.94 (−254.01 to −27.87)	.01	0.55	66.16 (−244.55 to 112.24)	.46	0.19
Tailoring + video-coaching versus tailoring-only	35.39 (−148.50 to 77.71)	.54	0.11	−25.16 (−211.74 to 161.43)	.79	0.18
Mental health (n=148)						
Tailoring + video-coaching versus control	0.04 (−5.87 to 5.94)	.99	0.22	0.36 (−5.65 to6.38)	.90	0.22
Tailoring + video-coaching versus tailoring-only	−13.18 (−19.10 to −7.26)	.00	0.91	1.43 (−4.91 to 7.77)	.65	0.34
Physical health (n=146)						
Tailoring + video-coaching versus control	−2.97 (−7.16 to 1.22)	.16	0.41	−2.07 (−6.34 to 2.20)	.34	0.43
Tailoring + video-coaching versus tailoring-only	−2.58 (−7.00 to 1.83)	.25	0.33	0.16 (−4.32 to 4.65)	.94	0.03

[a]Estimate of tailoring and video-coaching group's change in the dependent variable in comparison with control and tailoring-only. Higher scores represent greater improvements than the comparison group.

Discussion

Feasibility

The first aim of this study was to determine the feasibility of Web-based video-coaching sessions. The low participation and high satisfaction with the coaching sessions suggests that the majority of people participating in Web-based physical activity interventions are reluctant to talk to a coach using a video-calling program, but those who do find it worthwhile. The low participation in the coaching sessions could be explained by the high percentage of coaching participants (37%) who do not use video-calling software. This may be due to the older age of the participants who are less comfortable with technology [4]. Phone calls may have also been received better in this sample. The lower intervention adherence in the

video-coaching + tailoring group also suggests that increasing participant burden through an additional video-coaching component may reduce feasibility. The high satisfaction of the video-coaching sessions adds to past research findings of high satisfaction of coaching sessions conducted over the phone [44]. An important reason for the high satisfaction could be the accountability provided through the coaching sessions, as this was the most frequently given positive feedback for the sessions.

The two most common negative comments about the overall program were the lack of personal contact and that it was not personalized enough. Therefore, it is not surprising that those participants who participated in coaching sessions reported higher levels of program satisfaction. Participant satisfaction leads to improved engagement, which is important for intervention effectiveness [45]. In support of this, the participants who completed the coaching sessions were not only more satisfied with the program, but had significantly higher retention, higher adherence, and spent a significantly longer time on the website. This is in line with past research, demonstrating that personal contact in a Web-based intervention improves engagement [46,47]. However, it is possible that the coaching sessions are only effective in some people, and that the coaching sessions in this study may have retained a subsample of participants who were frequent Web users and therefore familiar with video-calling software or were more motivated to begin with. Due to the low participation in the coaching sessions, there were no significant differences in overall retention, adherence, engagement, and satisfaction between the 2 intervention groups. Web-based interventions with computer-tailored advice and coaching sessions may increase retention, adherence, engagement, and satisfaction, but only if they can convince participants to participate in the coaching sessions.

Effectiveness

The second aim of this study was to assess the effectiveness of the coaching sessions, by comparing physical activity and quality of life changes of the tailoring + video-coaching group with the tailoring-only and control groups. After the 8-week intervention period there was a significant treatment effect of the tailoring + video-coaching physical activity intervention on physical activity compared with no intervention. The improvement in physical activity compared with the control group (116 minutes per week, unadjusted) resulted in a moderate effect size and is considered clinically significant because of the large health effects seen from doing even small amounts of physical activity [48]. This finding is in line with findings that activity counseling over the phone [49], via email [50], and computer-tailored advice [13] improve participants' physical activity in comparison with a control group. Few studies have specifically tested the effectiveness of counseling through video calls in physical activity interventions. Pilutti et al [51] found that video-coaching to promote physical activity in patients with multiple sclerosis was effective. However, the coaching group's significant improvement in physical activity compared with the control group, in our study, could be due to the computer-tailored advice.

The tailoring + video-coaching group participants improved their physical activity 27 minutes per week (unadjusted) more in comparison with the tailoring-only group; however, no significant between-group differences were found. The availability of human support may have improved the overall physical activity of the tailoring + video-coaching group. Participants who needed to discuss their computer-tailored advice were able to. Limited studies have compared the effectiveness of coaching in addition to computer-tailored advice. Van Hoye et al [52] compared physical activity self-monitoring with and without additional face-to-face coaching sessions. They found that the coaching group had significantly greater physical activity improvements than the self-monitoring only group. An earlier study tested the effectiveness of email coaching in addition to a basic Web-based weight loss intervention [53]. The email-coaching group had significantly greater weight loss outcomes than the Web-based intervention only group. However, the effectiveness of coaching found in these studies may be due to the minimal nature of the comparison interventions (generic physical activity information + self-monitoring). The physical activity advice given in our study was highly tailored to participants' physical activity behavior, demographics, and psychosocial correlates of physical activity. This feedback might be enough to optimize physical activity outcomes (as demonstrated by a 172 minutes/week increase in physical activity in this group). The lack of significant differences between the intervention groups could also be due to low adherence to the coaching sessions, which may have reduced the effectiveness of the intervention in the tailoring + video-coaching group. The detailed computer-tailored advice may have discouraged participants to adhere to their coaching sessions as they were satisfied with the computer-tailored advice. Core intervention content may need to be delivered in the coaching sessions to promote higher adherence. Therefore, more research is needed to determine whether Web-based coaching is more effective than stand-alone computer-tailored advice.

The physical activity levels in the tailoring + video-coaching and tailoring-only groups were maintained at 6 months. There were, however, no significant between-group differences in physical activity changes from baseline to 6 months. This was due to the control group participants increasing their activity from 9 weeks to 6 months after the intervention. The absence of between-group physical activity changes at 6 months after the intervention is not uncommon in physical activity interventions; however, it is usually due to the intervention group's decline in physical activity rather than a physical activity increase in the control group [54].

There were no differences the physical health component of quality of life over time for any of the groups. This may be due to the sample not being large enough to detect subtle improvements or the overall high level of physical activity and physical health in the sample at baseline. Although previous studies have established a positive association between physical activity and quality of life [55], the effect of increased physical activity on quality of life is mainly seen in clinical samples that have a lower quality of life at baseline [56]. The significant reduction in quality of life in relation to mental health in the

tailoring-only group from baseline to postintervention in comparison was unexpected, as this group's participants did increase their physical activity at a clinically significant level and previous studies have associated this with an increase in quality of life [55].

Limitations

Limitations of the research include self-reported physical activity, which may be subject to social desirability bias. Participants and researchers were not blinded to group assignment, which may have biased results. The sample was predominantly female, white, and educated. Therefore, the results may not be generalizable to males, other cultures, and low socioeconomic groups. A high percentage of the sample was not in paid employment when compared with the Australian population data (43% vs 33%) [57]. This may be due to the high percentage of females and the older age of the participants. It could also be due to the people not in paid employment having more time to participate. Therefore, the results may not be generalizable to people in paid employment.

Furthermore, 40% of the participants were physically active at baseline despite the target group being inactive adults. This is not uncommon in physical activity studies [58], although results may not be generalizable to physically inactive Australians who are most in need of increasing their activity. The sample may have a high level of computer and Internet literacy as the majority of participants were reached through Facebook. Therefore, the findings may not be fully generalizable to people with a lower computer and Internet literacy. The additional phone calls the coaching participants received encouraging them to complete the coaching calls may have affected their physical activity levels. However, this is expected to be minimal due to the phone discussion being focused on scheduling the coaching call rather than their physical activity. Objective measurement of physical activity (eg, accelerometry) is needed to confirm the findings of this study. Although we were able to detect a difference in physical activity changes between the tailoring + video-coaching and tailoring-only groups, the analysis was underpowered to detect other differences between the groups as the required sample size of 100 per group was not met. The low retention is also a significant limitation. Low retention is common in Web-based programs, potentially due to the minimal face-to-face contact through either recruitment or the intervention. Similar retention rates have been observed in other Web-based health behavior interventions [59,60]. Retention is likely to be even lower outside of a randomized controlled trial setting that included reminder calls from researchers.

Conclusions

Combined Web-based video-coaching and computer-tailored advice was effective in comparison with a control group; however, only small nonsignificant improvements were seen when video-coaching was included in addition to computer-tailored advice in a Web-based physical activity intervention. Only a small percentage of participants adhered to Web-based video-coaching sessions, but those who participated were highly satisfied and more engaged in the intervention. Further research should investigate how adherence to Web-based coaching sessions can be improved.

Acknowledgments

Project and personal support for SA was provided by the National Health and Medical Research Council of Australia (ID 1055804). RCP is supported by a Salary Award from the National Health and Medical Research Council of Australia. CV (ID 100427) is supported by a Future Leader Fellowship from the National Heart Foundation of Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [29].

[PDF File (Adobe PDF File), 721KB - jmir_v18i8e223_app1.pdf]

Multimedia Appendix 2

Intervention landing page.

[PNG File, 158KB - jmir_v18i8e223_app2.png]

Multimedia Appendix 3

Participant information sheet.

[PDF File (Adobe PDF File), 34KB - jmir_v18i8e223_app3.pdf]

Multimedia Appendix 4

Tailored advice with graph.

[PNG File, 541KB - jmir_v18i8e223_app4.png]

Multimedia Appendix 5

Text-only tailored advice.

[PNG File, 76KB - jmir_v18i8e223_app5.png]

Multimedia Appendix 6

Action plan output.

[PNG File, 125KB - jmir_v18i8e223_app6.png]

Multimedia Appendix 7

Satisfaction survey.

[PDF File (Adobe PDF File), 103KB - jmir_v18i8e223_app7.pdf]

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
2. Department of Health, Australian Government. Canberra; 2014. Australia's Physical Activity & Sedentary Behaviour Guidelines for Adults (18-64 years) URL: http://www.health.gov.au/internet/main/publishing.nsf/content/health-pubhlth-strateg-phys-act-guidelines [accessed 2016-08-02] [WebCite Cache ID 6jTXNw9Zl]
3. Australian Bureau of Statistics. Canberra: ABS; 2013. Australian Health Survey: First Results, 2011-12 URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.001main+features12011-12 [accessed 2016-08-02] [WebCite Cache ID 6jTYteRKi]
4. Australian Communications and Media Authority. Canberra: ACMA; 2015. Australians embrace a digital life URL: http://www.acma.gov.au/theACMA/engage-blogs/engage-blogs/researchacma/Australians-embrace-a-digital-life [accessed 2016-08-02] [WebCite Cache ID 6jTZRtMmj]
5. Christopherson KM. The positive and negative implications of anonymity in Internet social interactions: "On the Internet, Nobody Knows You're a Dog". Computers in Human Behavior 2007 Nov;23(6):3038-3056. [doi: 10.1016/j.chb.2006.09.001]
6. McConnon A, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The Internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007;7:206 [FREE Full text] [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
7. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]
8. 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]
9. Joseph RP, Durant NH, Benitez TJ, Pekmezi DW. Internet-Based Physical Activity Interventions. Am J Lifestyle Med 2014;8(1):42-68. [doi: 10.1177/1559827613498059]
10. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9(1):52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
11. Lewis B, Williams D, Dunsiger S, Sciamanna C, Whiteley J, Napolitano M, et al. User attitudes towards physical activity websites in a randomized controlled trial. Prev Med 2008 Nov;47(5):508-513 [FREE Full text] [doi: 10.1016/j.ypmed.2008.07.020] [Medline: 18773915]
12. Short CE, Vandelanotte C, Dixon MW, Rosenkranz R, Caperchione C, Hooker C, et al. Examining participant engagement in an information technology-based physical activity and nutrition intervention for men: the manup randomized controlled trial. JMIR Res Protoc 2014;3(1):e2 [FREE Full text] [doi: 10.2196/resprot.2776] [Medline: 24389361]
13. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
14. Brouwer W, Kroeze W, Crutzen R, de Nooijer NJ, de Vries N, 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]

15. Broekhuizen K, Kroeze W, van Poppel M, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
16. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
17. Olsen JM. Health coaching: a concept analysis. Nurs Forum 2014 Jul;49(1):18-29. [doi: 10.1111/nuf.12042] [Medline: 24456550]
18. Barrera M, Toobert DJ, Angell KL, Glasgow RE, Mackinnon DP. Social support and social-ecological resources as mediators of lifestyle intervention effects for type 2 diabetes. J Health Psychol 2006 May;11(3):483-495. [doi: 10.1177/1359105306063321] [Medline: 16774900]
19. Winett RA, Anderson ES, Wojcik JR, Winett SG, Bowden T. Guide to health: nutrition and physical activity outcomes of a group-randomized trial of an Internet-based intervention in churches. Ann Behav Med 2007 Jun;33(3):251-261. [doi: 10.1080/08836610701358045] [Medline: 17600452]
20. Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
21. Skype. Does Skype use encryption? Security and blocked accounts URL: https://support.skype.com/en/faq/fa31/does-skype-use-encryption [accessed 2016-08-02] [WebCite Cache ID 6jTZwtap4]
22. Abbott JM, Klein B, Ciechomski L. Best Practices in Online Therapy. Journal of Technology in Human Services 2008 Jul 03;26(2-4):360-375. [doi: 10.1080/15228830802097257]
23. Fantus S, Mishna F. The Ethical and Clinical Implications of Utilizing Cybercommunication in Face-to-Face Therapy. Smith College Studies in Social Work 2013 Oct;83(4):466-480. [doi: 10.1080/00377317.2013.833049]
24. Wade SL, Oberjohn K, Burkhardt A, Greenberg I. Feasibility and preliminary efficacy of a web-based parenting skills program for young children with traumatic brain injury. J Head Trauma Rehabil 2009;24(4):239-247. [doi: 10.1097/HTR.0b013e3181ad6680] [Medline: 19625863]
25. Whittaker R, Dorey E, Bramley D, Bullen C, Denny S, Elley CR, et al. A theory-based video messaging mobile phone intervention for smoking cessation: randomized controlled trial. J Med Internet Res 2011;13(1):e10 [FREE Full text] [doi: 10.2196/jmir.1553] [Medline: 21371991]
26. Taylor A, Morris G, Pech J, Rechter S, Carati C, Kidd MR. Home Telehealth Video Conferencing: Perceptions and Performance. JMIR Mhealth Uhealth 2015;3(3):e90 [FREE Full text] [doi: 10.2196/mhealth.4666] [Medline: 26381104]
27. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
28. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. My Activity Coach - using video-coaching to assist a web-based computer-tailored physical activity intervention: a randomised controlled trial protocol. BMC Public Health 2014;14:738 [FREE Full text] [doi: 10.1186/1471-2458-14-738] [Medline: 25047900]
29. 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]
30. Alley S. My Activity Coach 2014. URL: http://www.myactivitycoach.org.au/eng/pages/23 [accessed 2016-02-17] [WebCite Cache ID 6fLvJMzJI]
31. Ajzen I. Attitudes, personality, and behaviour. Illinois: The Dorsey Press; 1988.
32. Cacioppo J, Petty RE. The elaboration likelihood model of persuasion. Advanced Consumer Research 1984;19:123.
33. Vandelanotte C, Duncan MJ, Plotnikoff RC, Mummery WK. Do participants' preferences for mode of delivery (text, video, or both) influence the effectiveness of a Web-based physical activity intervention? J Med Internet Res 2012;14(1):e37 [FREE Full text] [doi: 10.2196/jmir.1998] [Medline: 22377834]
34. Vandelanotte C, Mummery WK. Qualitative and quantitative research into the development and feasibility of a video-tailored physical activity intervention. Int J Behav Nutr Phys Act 2011;8(1):70 [FREE Full text] [doi: 10.1186/1479-5868-8-70] [Medline: 21718546]
35. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [Medline: 12641893]
36. Brown W, Bauman A, Chey T, Trost S, Mummery K. Comparison of surveys used to measure physical activity. Aust N Z J Public Health 2004 Apr;28(2):128-134. [Medline: 15233351]
37. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. J Sci Med Sport 2004 Jun;7(2):205-215. [Medline: 15362316]
38. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aust N Z J Public Health 2008 Dec;32(6):535-541. [doi: 10.1111/j.1753-6405.2008.00305.x] [Medline: 19076744]
39. Burdine JN, Felix MR, Abel AL, Wiltraut CJ, Musselman YJ. The SF-12 as a population health measure: an exploratory examination of potential for application. Health Serv Res 2000 Oct;35(4):885-904 [FREE Full text] [Medline: 11055454]
40. 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]

41. Whitehead L. Methodological issues in Internet-mediated research: a randomized comparison of internet versus mailed questionnaires. J Med Internet Res 2011;13(4):e109 [FREE Full text] [doi: 10.2196/jmir.1593] [Medline: 22155721]
42. Saris-Baglama RN, Dewey CJ, Chisholm GB, Plumb E, King J, Kosinski M, et al. Qualitymetric. Lincon, RI: QualityMetric Incorporated; 2010. Quality metric health outcomes scoring software 4.0. Installation guide URL: http://www.qualitymetric.com/download/InstallationGuide_ScoringSoftwareV4.pdf [accessed 2016-08-02] [WebCite Cache ID 6jTbXe8i3]
43. Lu K, Luo X, Chen P. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. Int J Biostat 2008;4(1):Article 9. [Medline: 22462117]
44. Nes Andréa AG, van Dulmen SB, Eide E, Finset A, Kristjánsdóttir OB, Steen IS, et al. The development and feasibility of a web-based intervention with diaries and situational feedback via smartphone to support self-management in patients with diabetes type 2. Diabetes Res Clin Pract 2012 Sep;97(3):385-393. [doi: 10.1016/j.diabres.2012.04.019] [Medline: 22578890]
45. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899]
46. Spittaels H, De BI. Implementation of an online tailored physical activity intervention for adults in Belgium. Health Promot Int 2006 Dec;21(4):311-319 [FREE Full text] [doi: 10.1093/heapro/dal030] [Medline: 16963783]
47. Meyer C, Ulbricht S, Gross B, Kästel L, Wittrien S, Klein G, et al. Adoption, reach and effectiveness of computer-based, practitioner delivered and combined smoking interventions in general medical practices: a three-arm cluster randomized trial. Drug Alcohol Depend 2012 Feb 1;121(1-2):124-132. [doi: 10.1016/j.drugalcdep.2011.08.019] [Medline: 21924563]
48. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
49. Hekler EB, Buman MP, Otten J, Castro CM, Grieco L, Marcus B, et al. Determining who responds better to a computer- vs. human-delivered physical activity intervention: results from the community health advice by telephone (CHAT) trial. Int J Behav Nutr Phys Act 2013;10:109 [FREE Full text] [doi: 10.1186/1479-5868-10-109] [Medline: 24053756]
50. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
51. Pilutti LA, Dlugonski D, Sandroff BM, Klaren RE, Motl RW. Internet-delivered lifestyle physical activity intervention improves body composition in multiple sclerosis: preliminary evidence from a randomized controlled trial. Arch Phys Med Rehabil 2014 Jul;95(7):1283-1288. [doi: 10.1016/j.apmr.2014.03.015] [Medline: 24699237]
52. Van HK, Boen F, Lefevre J. The Impact of Different Degrees of Feedback on Physical Activity Levels: A 4-Week Intervention Study. Int J Environ Res Public Health 2015 Jun;12(6):6561-6581 [FREE Full text] [doi: 10.3390/ijerph120606561] [Medline: 26067990]
53. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
54. 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]
55. Bize R, Johnson JA, Plotnikoff RC. Physical activity level and health-related quality of life in the general adult population: a systematic review. Prev Med 2007 Dec;45(6):401-415. [doi: 10.1016/j.ypmed.2007.07.017] [Medline: 17707498]
56. Fong DY, Ho J, Hui BP, Lee AM, Macfarlane DJ, Leung SS, et al. Physical activity for cancer survivors: meta-analysis of randomised controlled trials. BMJ 2012;344:e70 [FREE Full text] [Medline: 22294757]
57. Australian Bureau of Statistics. Canberra: ABS Education and Work URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/6227.0 [accessed 2016-08-02] [WebCite Cache ID 6jTbhIO3W]
58. Vandelanotte C, Stanton R, Rebar AL, Van Itallie AK, Caperchione CM, Duncan MJ, et al. Physical activity screening to recruit inactive randomized controlled trial participants: how much is too much? Trials 2015;16:446 [FREE Full text] [doi: 10.1186/s13063-015-0976-7] [Medline: 26449210]
59. Carr LJ, Dunsiger SI, Lewis B, Ciccolo JT, Hartman S, Bock B, et al. Randomized controlled trial testing an internet physical activity intervention for sedentary adults. Health Psychol 2013 Mar;32(3):328-336. [doi: 10.1037/a0028962] [Medline: 22823069]
60. Skår S, Sniehotta FF, Molloy GJ, Prestwich A, Araújo-Soares V. Do brief online planning interventions increase physical activity amongst university students? A randomised controlled trial. Psychol Health 2011 Apr;26(4):399-417. [doi: 10.1080/08870440903456877] [Medline: 20830646]

Abbreviations

BMI: body mass index

Edited by G Eysenbach; submitted 29.02.16; peer-reviewed by L Whitehead, B Irwin, B Bock, J Torous; comments to author 19.04.16; revised version received 01.06.16; accepted 13.06.16; published 12.08.16.

Please cite as:
Alley S, Jennings C, Plotnikoff RC, Vandelanotte C
Web-Based Video-Coaching to Assist an Automated Computer-Tailored Physical Activity Intervention for Inactive Adults: A Randomized Controlled Trial
J Med Internet Res 2016;18(8):e223
URL: http://www.jmir.org/2016/8/e223/
doi:10.2196/jmir.5664
PMID:27520283

Original Paper

Efficacy of Internet-Based Self-Monitoring Interventions on Maternal and Neonatal Outcomes in Perinatal Diabetic Women: A Systematic Review and Meta-Analysis

Ying Lau[1], RN, PhD; Tha Pyai Htun[1], MBBS, MPH; Suei Nee Wong[2], BSc (Hons), MSc; Wai San Wilson Tam[1], BSC, MPhil, PhD; Piyanee Klainin-Yobas[1], BSN, MSN, PhD

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

[2]National University of Singapore, Medical Resource Team, National University of Singapore Libraries, Singapore, Singapore

Corresponding Author:
Ying Lau, RN, PhD
National University of Singapore
Alice Lee Centre for Nursing Studies
Level 2, Clinical Research Centre
Block MD11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011603
Fax: 65 67767135
Email: nurly@nus.edu.sg

Abstract

Background: Self-monitoring using the Internet offers new opportunities to engage perinatal diabetic women in self-management to reduce maternal and neonatal complications.

Objective: This review aims to synthesize the best available evidence to evaluate the efficacy of Internet-based self-monitoring interventions in improving maternal and neonatal outcomes among perinatal diabetic women.

Methods: The review was conducted using Cochrane Central Register of Controlled Trials, PubMed, EMBASE, Cumulative Index to Nursing and Allied Health Literature, PsyINFO, Scopus, and ProQuest Dissertations and Theses to search for English-language research studies without any year limitation. A risk of bias table was used to assess methodological quality. Meta-analysis was performed with RevMan software. Cochran Q and I^2 tests were used to assess heterogeneity. The overall effect was assessed using z tests at $P<.05$. Of the 438 studies identified through electronic searches and reference lists, nine experimental studies from 10 publications were selected.

Results: Half of the selected studies showed low risk of bias and comprised 852 perinatal diabetic women in six countries. The meta-analysis revealed that Internet-based self-monitoring interventions significantly decreased the level of maternal glycated hemoglobin A1c (z=2.23, P=.03) compared to usual care among perinatal diabetic women at postintervention. Moreover, Internet-based self-monitoring interventions significantly decreased the cesarean delivery rate (z=2.23, P=.03) compared to usual care among the mixed group at postintervention.

Conclusions: This review shows neonatal or other maternal outcomes are similar between Internet-based self-monitoring interventions and usual diabetes care among perinatal diabetic women. The long-term effects of the intervention must be confirmed in future studies using randomized controlled trials and follow-up data.

(J Med Internet Res 2016;18(8):e220) doi:10.2196/jmir.6153

KEYWORDS

Internet; pregnancy in diabetics; interventions; meta-analysis

Introduction

Diabetes mellitus (DM) is one of the most common complications of pregnancy; preexisting diabetes mellitus (type 1 or type 2) and gestational diabetes mellitus (GDM) affect approximately 2.5% to 2.7% and 4.6% to 8.0% of all pregnant women, respectively [1]. Both GDM and preexisting diabetes are associated with increased medical costs and perinatal morbidity [1]. Existing interventions must be improved considering the increasing global incidence of diabetic pregnancy with serious perinatal outcomes [2]. Self-monitoring intervention is important in reducing maternal and neonatal complications related to diabetic pregnancies, both in cases of preexisting diabetes [3] and GDM [4]. Self-monitoring refers to systematic observation and recording of ongoing goal-directed activities [5] based on self-regulation theory [6]. Self-regulation involves self-awareness of the current condition of an individual [7]. Awareness could trigger a self-evaluation response involving the interpretation of one's condition against a goal or standard; after self-evaluation, a series of responses could be determined through self-adjustment and self-reinforcement [1,6]. Self-monitoring capitalizes on this motivation to achieve glycemic control [8], improve weight management [9], and reduce hospitalization and readmission rates [10].

Self-monitoring using the Internet offers new opportunities to engage participants in self-management. A previous study [11] suggested that self-monitoring using Internet-based interventions and face-to-face interventions elicited similar outcomes among the patients. Development of Internet-based interventions by using theory-based methods could promote substantial changes in the health behavior of a patient [12]. The Internet offers a diverse range of strategies for exchanging information and gaining knowledge [13] and thus can provide interactive ways to integrate communication with sensor-based systems (glucometer and pedometer) for transmitting information to a device or computer [14,15]. Sensors are used to record and transmit data to a computer, which then transmits the data to the provider and provides personalized/tailored feedback to the individual [14,15] regarding self-monitoring compliance with treatments and self-adjustment to diet, activity, and weight management.

Internet-based interventions employ a tracking system to improve self-reinforcement by using reminders (cues to action) [16], alerts [14], or graphic progress [17] through text messages (short message service, SMS) and email. Asynchronous and synchronous interactions generate identical interactional benefits [18]. Peer-support interactivity allows women to interact with one another with a pseudonym [15]; this process could empower women to take ownership of their well-being and initiate resolutions for issues they are encountering, thereby contributing to a sense of increased self-efficacy among perinatal diabetic women [19]. A longitudinal follow-up is important to test the sustainability of self-monitoring patterns over an extended period [20]. The advantages of using the Internet to deliver interventions include low cost, easy distribution, and convenient delivery to multiple locations at specific times [4,21]. Internet access is increasingly used as an educational and supportive source of information for perinatal women [22,23]. Internet-based interventions are rapidly developed with increased access to instant cyber connectivity; however, the effect of Internet-based self-monitoring on improving maternal and neonatal outcomes among perinatal diabetic women remains unclear.

Meta-analysis is used to document the application of Internet-based self-monitoring interventions among general diabetic population [24-26]. However, only a few studies were conducted on perinatal diabetic women. Four reviews focused on the use of technologies to evaluate healthy pregnant women in terms of maternal outcomes [27], women with complicated pregnancies in terms of cost effectiveness [28], a mixed group of patients (with type 1 DM and GDM) in terms of maternal-neonatal outcomes [29], and patients with GDM in terms of maternal outcomes [30]. These studies reported mixed results, did not include ongoing studies without outcomes [27], lacked systematic searching strategies [28,29], and evaluated limited studies (n=3) [30]. None of the studies focused on Internet-based self-monitoring approaches. Hence, further research must be performed, particularly in light of the rapid improvements in technologies worldwide. This review aims to systematically assess studies that examined Internet-based self-monitoring interventions for improving maternal and neonatal outcomes among perinatal diabetic women.

Methods

This study was performed in accordance with the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [31]. The protocol is registered to the PROSPERO database (CRD42016034142).

Eligibility Criteria

The full inclusion and exclusion criteria for the systematic review are described in Multimedia Appendix 1. Studies were included if they met the following criteria:

Population: perinatal women aged 18 years and older with GDM, type 1 DM, and/or type 2 DM;

Interventions: interact with perinatal diabetic women to undertake one or more of the following components associated with self-awareness, self-interpretation, self-adjustment, or self-reinforcement of glycemic level, physical activities, dietary intake, weight management, or medication adherence [7,10] by using the Internet;

Comparison: usual diabetes care as control group;

Outcomes: primary outcomes included glycated hemoglobin A_{1c} (HbA_{1c}) level, cesarean delivery, neonatal birth weight, and neonatal hypoglycemia at postintervention; secondary outcomes included biological outcomes (fasting blood glucose, weight gain, and change in body mass index [BMI] or weight), cognitive outcomes (satisfaction rate, empowerment, self-efficacy, or health-related quality of life), behavioral outcomes (insulin treatment rate or compliance rate with self-monitoring), emotional outcomes (depression or stress), and neonatal outcomes (large for gestational age or macrosomia) at postintervention; and

Type of design: experimental studies that were either a randomized controlled trial (RCT) or controlled clinical trial (CCT). We excluded studies if they were nonexperimental, qualitative, protocol, or conference papers regarding general diabetic populations.

Search Strategy

The search strategy aimed to find published or unpublished studies written in English. No restriction was applied to the search performed from inception until February 16, 2016 in the following electronic databases: Cochrane Central Register of Controlled Trials (CENTRAL), PubMed, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Scopus, and ProQuest Dissertations and Theses. Index and keyword terms were used (Multimedia Appendix 2). The keywords were exploded and truncated following the syntax rules of each database. Unpublished trials of relevance to the review were searched from the Clinical Trials Registry (www.clinicaltrials.gov). Unpublished data were requested if eligible trials maximized the scope of the search. Finally, we searched the reference lists of the included studies and relevant previous reviews to check for additional eligible studies.

Study Selection

Two authors (LY and TP) independently screened the titles and abstracts of the identified references from the literature search to identify potentially eligible studies. The full texts of the remaining references were evaluated. Ineligible reports were excluded based on inclusion criteria, and the reasons for exclusion were recorded. A third reviewer (KY) resolved disagreements between the two reviewers regarding inclusion of a study.

Quality Assessment

After identifying studies that fulfilled the selection criteria and verifying their eligibility by reading the completed articles, the studies were subjected to quality assessment. The quality of the studies was independently judged using criteria for determining bias in intervention studies recommended by the Cochrane Handbook for Systematic Reviews of Interventions [32]. The following indicators of internal validity specific to the methodology of RCT were collected: (1) random sequence generation (selection bias), (2) allocation concealment (selection bias), (3) blinding of participants and personnel (performance bias), (4) blinding of outcome assessment (detection bias), (5) incomplete outcome data (attrition bias), and (6) selective reporting (reporting bias) [32]. Assessment related to risk biases was assigned a judgment of “low risk,” “high risk,” or “unclear risk” of bias. One reviewer (LY) reviewed all studies with a subset reviewed by a second reviewer (TP). Disagreements were settled through discussion or consulting a third reviewer (KY).

Data Extraction

Two of the authors (LY and TP) extracted relevant data from all included articles. The following variables were obtained using structured data extraction items based on setting, country, design, population, gestation, age, intervention, control, sample size, outcomes, attrition, and intention-to-treat (ITT) analysis. The details of self-monitoring interventions were extracted based on components (glycemic, diet, weight gain control, physical activities, or/and medication adherence), transmission (asynchronous or asynchronous), functionality, facilities, interactivity, provider, peer support, duration, and follow-up. The two authors (LY and TP) thoroughly reviewed the summary tables for accuracy and relevance. When relevant data were missing or questionable in both published and unpublished trials, the authors were contacted for verification and to obtain additional information. Among 59 full-text articles, 10 were not clear because they had insufficient details (n=2) or no (n=8) outcomes. Although 10 authors were approached, none responded to our queries. Therefore, we excluded these 10 studies in the review.

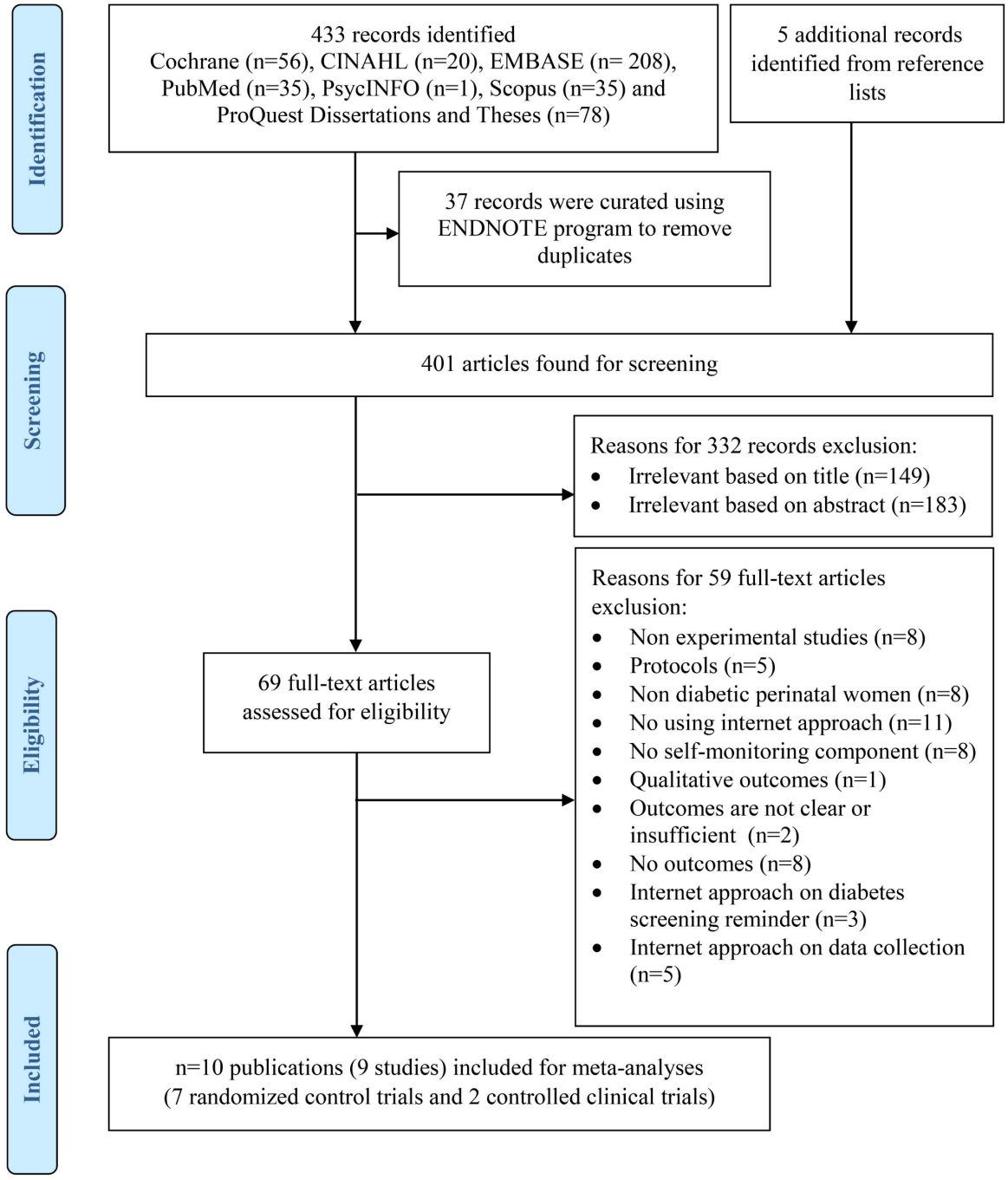
Statistical Analysis

RevMan software (Review Manager version 5.3 for Windows from the Nordic Cochrane Center, the Cochrane Collaboration, 2014) was used for meta-synthesis. Risk ratio (RR) was used as the effect measure for dichotomous outcome with Mantel-Haenszel method. Mean difference was used for continuous outcomes with inverse-variance method. Heterogeneity between studies was evaluated using Cochran Q (chi-square test) and I^2 statistics. The statistical significance for heterogeneity using the chi-square test was set as $P<.10$. The I^2 statistic was applied to describe total variations in study estimates because of heterogeneity. Heterogeneity degree was estimated using I^2 by setting 0%, 25%, 50%, or 75% for no, low, moderate, and high heterogeneity, respectively [33]. The fixed-effect model was used in cases without significant heterogeneity ($P>.10$), and the DerSimonian and Laird random-effects model was used in cases with heterogeneity among the studies ($P<.10$) and I^2 values of more than 50% [33]. Subgroup analysis was performed to explore the source of heterogeneity, and the predefined subgroup was the type of DM.

Results

Figure 1 shows the selection process (PRISMA flow diagram). A total of 438 studies were identified from the initial database search and reference lists. Of these studies, 37 articles were curated using Endnote to remove duplicates. Subsequently, 401 studies were included for screening and 332 articles were excluded based on analysis of text words in titles and abstracts. In all, 69 full-text articles were retrieved, reviewed, and selected based on relevance and quality for eligibility. Of these, 59 articles were excluded because of the following: nonexperimental nature; type of protocol; nondiabetic perinatal women as subject; not using Internet approach; lack of self-monitoring component; reported qualitative, unclear, insufficient, or no outcomes; and Internet approach employed on diabetes screening, reminder, data collection. Finally, nine studies from 10 publications were identified for inclusion in this systematic review.

Figure 1. PRISMA flow diagram of article selection procedure.



Study Characteristics

This meta-analysis included nine studies with 852 participants conducted across four countries (Table 1), which included the United States (n = 5) [15,16,34-36], Spain (n=2) [14,37,38], Italy (n=1) [39], and Ireland (n=1) [40]. All these studies were published articles. Research was conducted between 2007 [36] and 2015 [16]; 2015 had the highest number of publications (n=3) [14,16,40]. Seven of the studies used RCT designs and two used CCT designs [14,39]. The target populations were perinatal women with GDM or impaired glucose tolerance (n=5) [15,35,36], mixed group (n=4) [14,16,39], type 1 DM (n=0), and type 2 DM (n=0). The sample sizes varied among the nine studies and ranged from 19 [41] to 235 [39]. Nine studies reported more than one outcome. Attrition rates ranged from 2% [14,37,38] to 32% [16]. None of the studies used ITT analysis, and eight studies were supported by grants.

Table 1. Characteristics of the nine selected studies (10 publications).[a]

Author, year [ref]	Setting/Country	Design	Population/gestation/age, mean (SD)	Intervention	N	Duration (weeks)	Outcomes	Attrition rate, %
Bartholomew et al 2015 [16][b]	Antenatal clinic in Hawaii, USA	RCT	GDM or type 2 DM; <30 w; 33.2 (5.4)	Mobile phone, Internet technology (CIT)	I: 50; C: 50	3	Fasting and 2-hour postprandial blood compliance rate with SMBG; satisfaction rate	I: 20; C: 32
Carral et al 2015 [14][b]	GDM unit in Cadiz, Spain	CCT	GDM, type 1 or 2 DM; <30 w; 33.8 (4.6)	Web-based telemedicine system	I: 40; C: 64	—	HbA_{1c} (%); weight gain; cesarean delivery rate; insulin treatment rate; neonatal birth weight; large for gestational age; neonatal hypoglycemia	I: 5; C: 14
Dalfra et al 2009 [39]	12 Diabetes clinics in Italy	CCT	GDM or type 1 diabetes; <30 w; 33.8 (4.6)	Telemedicine with Glucobeep server	I: 105; C:130	10	HbA_{1c} (%); weight gain; cesarean delivery rate; insulin treatment rate; neonatal birth weight; macrosomia; SF36; CES-D; DSS; DHDS	Total: 15; I:—; C:—
Given et al 2015 [40][b]	2 Diabetes clinics in Ireland	RCT	GDM or IGT; 24-28 w; I: 33.5 (4.2), C: 30.1 (5.5)	Web-based telemedicine system	I: 24; C: 26	12	HbA_{1c} (%); cesarean delivery rate; insulin treatment rate; neonatal birth weight; macrosomia; neonatal hypoglycemia; satisfaction rate	I: 12.5; C:15.4
Homko et al 2007 [36][b]	Antenatal clinic or one of its satellites in Philadelphia, PA	RCT	GDM; <33 w; 18-45, I: 29.8 (6.6), C: 29.2 (6.7)	Internet-based telemedicine system using ITSMyHealthfile and Lassoweb data engine	I: 34; C: 25	—	HbA_{1c} (%); FBS (mg/dL); cesarean delivery rate; DES; neonatal birth weight; large for gestational age; neonatal hypoglycemia	I: 5.8; C:13.8
Homko et al 2012 [35][b]	Antenatal clinics (2) in Philadelphia, PA	RCT	GDM; <33 w; 18-45, I: 30.3 (6.0), C: 30.0 (7.5)	Internet-based telemedicine system with automatic telephone option	I: 40; C: 40	—	FBS (mg/dL); cesarean delivery rate; neonatal birth weight; large for gestational age; neonatal hypoglycemia	I: 10; C: 5
Kim et al 2012 [15][b]	University health system in Michigan	RCT	GDM within 3 years; >18 years (—)	Web-based pedometer program	I: 28; C: 21	13	Change in weight; change in BMI; change in self-efficacy for weight and activity	I: 9.5; C:17.9
Nicklas et al 2014 [34][b]	Hospital in Boston, MA	RCT	GDM; postnatal; 18-45 (—)	Web-based lifestyle intervention	I: 36; C: 39	24-40	Change in weight; change in BMI	I: 8.3; C:10.3
Pérez-Ferre et al 2010a,b [37,38][b]	Diabetes unit of a hospital in Madrid, Spain	RCT	GDM; <28 w; I: 33.3 (5.6), C: 34.2 (5.2)	Web-based telemedicine system	I: 50; C: 50	12	HbA_{1c} (%); weight; weight gain; cesarean delivery rate; neonatal birth weight; large for gestational age; neonatal hypoglycemia	I: 2.0; C: 4.0

[a] All studies had a usual treatment control group and none used ITT. —: Information not mentioned in article; BMI: body mass index; C: control group; CCT: controlled clinical trial; CES-D: Center for Epidemiologic Studies Depression Scale; DES: Diabetes Empowerment Scale; DHDS: Diabetes Health Distress Scale; DSS: Diabetes-related Stress Scale; FBS: fasting blood sugar; GDM: gestational diabetes mellitus; HbA_{1c}: glycated hemoglobin A_{1c}; I: intervention group; IGT: impaired glucose tolerance; ITT: intention-to-treat analyses; OGTT: Oral Glucose Tolerance Test; RCT: randomized controlled trial; SF36: SF-36 Health Survey; SMBG: self-monitoring of blood glucose.

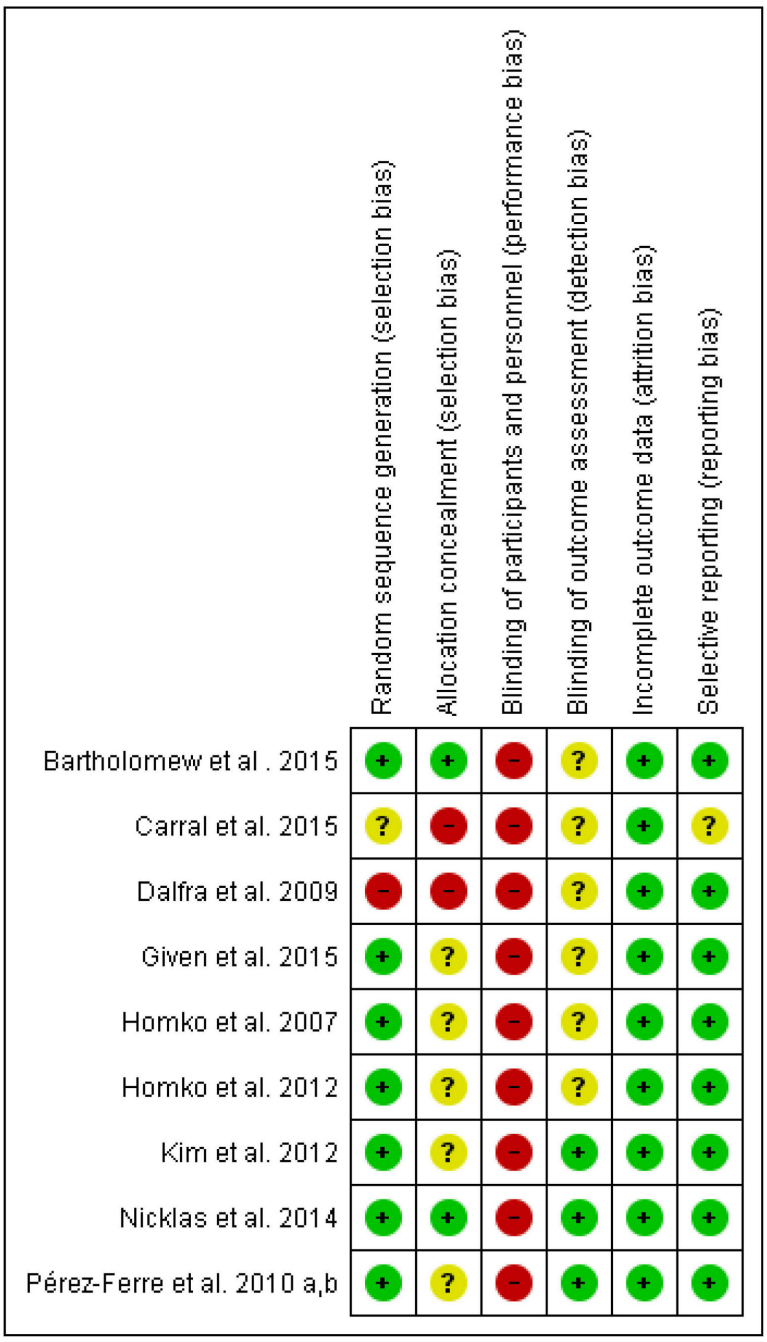
[b] These studies had grant support.

Study Quality

The summary of risk of bias is presented in Figure 2, and the risk of bias graph is shown in Multimedia Appendix 3. Seven of nine studies had adequate sequence generation for randomization. Two studies [16,34] had adequate allocation concealment. None of the studies implemented blinding of participants. Three studies [15,34,37,38] implemented blinding

of outcome assessment. All studies addressed low-risk bias concerning incomplete outcome data. Eight had low-risk bias for selective reporting.

Figure 2. Risk of bias summary.



Description of Internet-Based Self-Monitoring Interventions

Detailed elements of the Internet-based self-monitoring interventions are presented in Multimedia Appendix 4. The components of the self-monitoring interventions included glycemic control (n=8), diet control (n=7), physical activities (n=5), weight control (n=3) [15,34,39], and medication adherence (n=7). Functionalities of the interventions included system alert and reminder (n=4) [14,16,35,40], graphical progress (n=2) [15,37,38], and uploading, entering, and tracking own information (n = 3) [16,34,36] using website (n = 9), phone (n = 7), SMS text message (n = 5), email (n = 6), and animated video (n = 1) [34] that integrated communication with glucometer (n = 4) and pedometer (n = 1) [15]. The majority of the interventions used asynchronous communication (n=6), and three used synchronous communication [15,39,40] through two-way (n=9) feedback communication. The providers of the intervention were physicians (n = 7), nurses (n = 4), dietitians (n=1) [34], telemedicine service provider (n=1) [40], and study staff (n = 1) [15]. Only one intervention consisted of peer support using an online forum [15]. The duration of the intervention varied among the nine studies and ranged from 3 weeks [16] to 40 weeks [34]. Three of the studies [14,16,34] had follow-up after intervention. None of the studies reported

using theoretical or conceptual framework to design their interventions.

Efficacy of Internet-Based Self-Monitoring Interventions on Maternal Outcomes

Five studies [14,36-39,40] assessed the efficacy of interventions among 508 perinatal women by using HbA_{1c} levels as the outcome. The meta-analysis revealed that the intervention significantly improved HbA_{1c} levels (mean difference −0.12, 95% CI −0.22 to −0.02), as determined using inverse-variance method and fixed-effects model (I^2=0%, P=.69; Figure 3). A nonsignificant P value for the Cochran Q statistic indicated that the selected studies were homogeneous. The overall effect of intervention on HbA_{1c} was significant (z=2.39, P=.02). Subgroup analyses were performed to compare the effects of the interventions on HbA_{1c} between the GDM (n = 3) [36-38,40] and mixed groups (n = 2) [14,39]. However, no significant effect was found for subgroup differences (P=.73).

Six studies [14,35-39,40] assessed cesarean delivery rate as outcomes of interventions among 526 perinatal women, and the meta-analysis showed low heterogeneity (I^2= 20%, P=.28) (Figure 4). Moreover, the interventions did not significantly improve cesarean delivery rate for overall effect (RR=0.84, 95% CI 0.68-1.05; z=1.55, P=.12). Two subgroup analyses using the Mantel-Haenszel method and fixed-effects model revealed that the interventions significantly decreased the cesarean delivery rate among the mixed group (RR=0.73, z=2.23, P=.03) in two studies [14,39], but had no effect among the GDM group (RR=1.05, z=0.30, P=.77) in four studies [35-38,40]. No significant subgroup differences were found (P=.10). None and low heterogeneity were found between subgroups of women with GDM (I^2= 0%, P=.97) and the mixed group (I^2= 23%, P=.27).

Figure 3. Forest plot of mean difference (95% CI) in change of HbA1c (%) for the Internet-based self-monitoring intervention and control groups. IV: inverse variance.

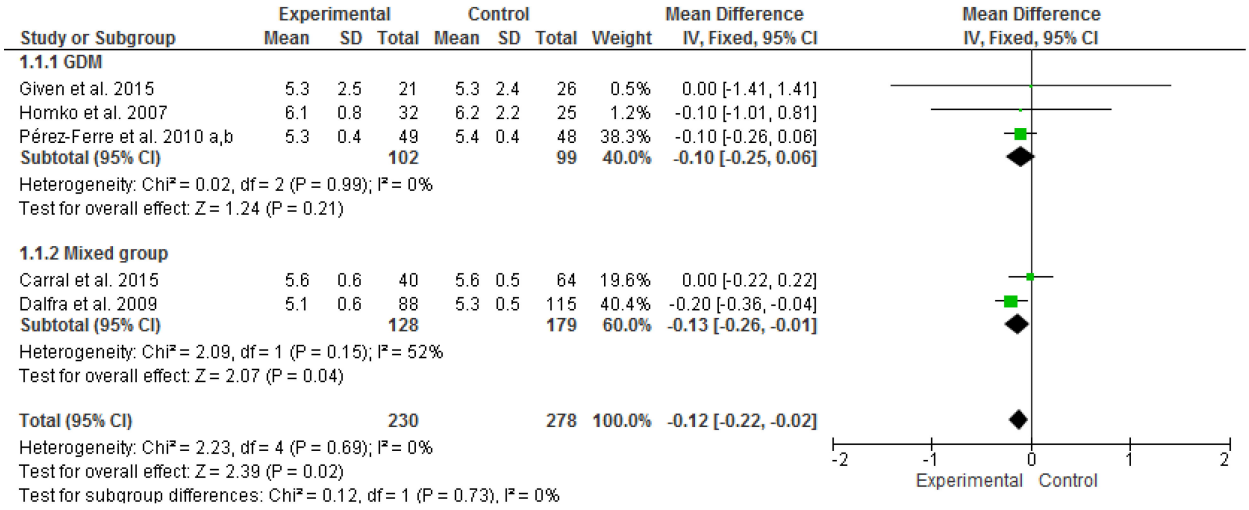
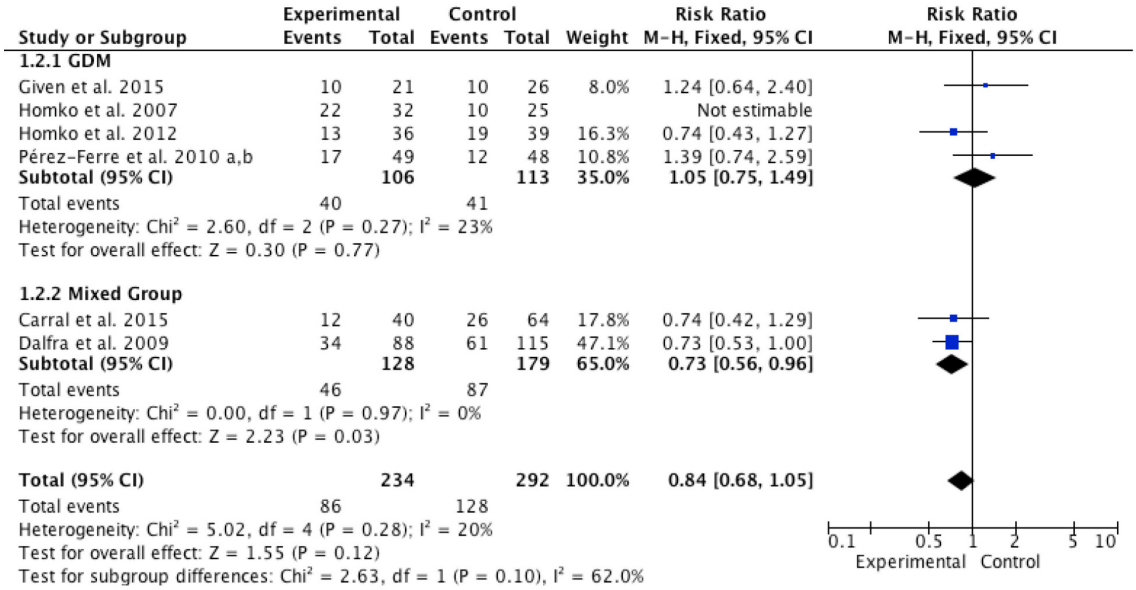


Figure 4. Forest plot of risk ratio in change of cesarean delivery rate for the Internet-based self-monitoring intervention and control groups.



Efficacy of Internet-Based Self-Monitoring Interventions on Neonatal Outcomes

Figure 5 shows the pooled meta-analysis results of six articles that determined the effect of interventions on neonatal body weight among 582 perinatal women. The meta-analysis showed low to moderate heterogeneity (I^2= 41%, P=.13). Four studies [35-38,40] of GDM group and two studies [14,39] of mixed group revealed similar neonatal weight (mean difference=27.30, z=0.62, P=.54) between the Internet-based self-monitoring intervention and control groups. Two subgroup analyses were performed and no significant differences were found between intervention and control groups either in the GDM group (mean difference=92.21, z=1.47, P=.14) or the mixed group (mean difference=–36.42, z=0.59, P=.56). The heterogeneity of GDM group (I^2=39%, P=0.18) and mixed group (I^2=30%, P=.23) ranged from low to moderate. The test for subgroup differences was not significant (P=.14).

Figure 6 shows the pooled meta-analysis results of five studies on neonatal hypoglycemia among 379 women. The intervention group demonstrated no significant difference on the overall effect (RR=1.09, z=0.24, P=.81) compared with the control group, as assessed using the Mantel-Haenszel method and fixed-effects model. No heterogeneity was found in the mixed group (I^2= 0%, P=.85) and overall result (I^2= 0%, P=.93). The result of subgroup analysis was not different (P=.79) between the mixed and GDM groups.

Table 2 summarizes the efficacy of the intervention on maternal outcomes including fasting blood sugar [35,36], weight gain [14,37,38], changes in BMI and weight [15,34], insulin treatment rate [14,37,38], satisfaction rate [16,40], compliance rate with self-monitoring of blood glucose [16], health-related quality of life [39], depressive symptoms [39], diabetic-related stress [39], diabetes health distress [39], diabetes empowerment [36], and change in self-efficacy for weight and activity [15], as well as neonatal outcomes including large for gestational age [14,35-38] and macrosomia [39,40]. The outcomes were not significantly different between intervention and control groups. Although the effects of diabetes-related stress and diabetes empowerment significantly differed in the Diabetes-related Stress Scale scores (P=.02) [39] and Diabetes Empowerment Scale scores (P=.003) [36], the findings of the single study could not contribute sufficient evidence to draw conclusions. The heterogeneity (I^2) ranged from 0% in the pooled meta-analysis of three studies on weight gain [14,37,38] to 95% from the pooled meta-analysis of two studies on satisfaction rate [16,40] by using fixed- and random-effect models, respectively. Although we identified substantial heterogeneity (I^2>50%), we encountered difficulty in investigating the result by using subgroup and sensitivity analyses for the two to three studies that indicated changes in BMI or weight [15,34], insulin treatment rate [14,37-39], and satisfaction rate [16,40].

Figure 5. Forest plot of mean difference (95% CI) in change of neonatal body weight (grams) for the Internet-based self-monitoring intervention and control groups. IV: inverse variance.

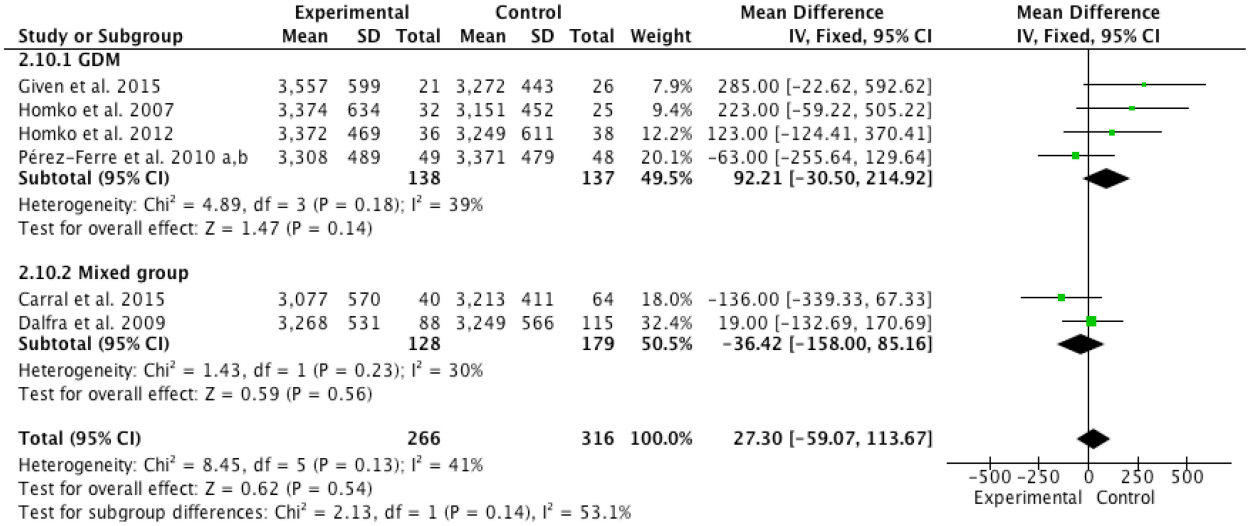


Figure 6. Forest plot of risk ratio for change in neonatal hypoglycemia rate for the Internet-based self-monitoring intervention and control groups.

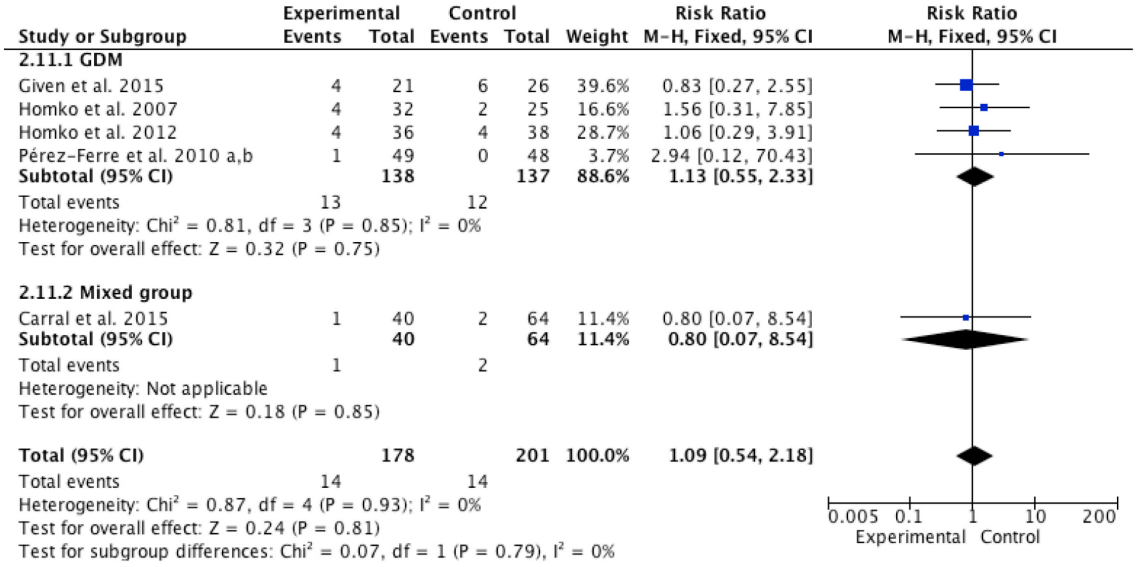


Table 2. Efficacy of Internet-based self-monitoring interventions on other maternal and neonatal outcomes.[a]

Outcomes	Studies included, n	RR[b]/MD[c] (95% CI)	Overall effect		Heterogeneity		Model
			z	*P*	I^2	*P*	
Maternal outcomes							
Fasting blood sugar	2 [35,36]	–0.66[c] (–4.28, 2.96)	0.36	.72	44%	.72	Fixed
Weight gain	3 [14,37,38]	–0.48[c] (–1.44, 0.47)	0.99	.32	0%	.98	Fixed
Change in BMI	2 [15,34]	–0.91[c] (–1.91, –0.09)	1.77	.08	64%	.09	Random
Change in weight	2 [15,34]	–2.53[c] (–5.10, –0.04)	1.93	.05	65%	.09	Random
Insulin treatment rate	3 [14,37-39]	1.06[b] (0.56, 2.02)	0.19	.85	71%	.03	Random
Satisfaction rate	2 [16,40]	1.75[b] (0.40, 7.58)	0.74	.46	95%	<.001	Random
Compliance rate with self-monitoring of blood glucose	1 [16]	1.02[b] (0.87, 1.20)	0.24	.81			
SF-36 Physical component	1 [39]	–2.2[c] (–4.50, 0.10)	1.88	.06	NA	NA	NA
SF-36 Mental component	1 [39]	2.10[c] (0.75, 4.95)	1.44	.15	NA	NA	NA
CES-D	1 [39]	1.50[c] (–1.35, 4.35)	1.03	.30	NA	NA	NA
DSS	1 [39]	4.10[c] (0.75, 7.45)	2.40	.02	NA	NA	NA
DHDS	1 [39]	4.90[c] (–0.20, 10.00)	1.88	.06	NA	NA	NA
DES	1 [36]	0.40[c] (0.14, 0.66)	3.00	.003	NA	NA	NA
Change in self-efficacy for weight	1 [15]	2.79[c] (–2.57, 8.15)	1.02	.31	NA	NA	NA
Change in self-efficacy for activity	1 [15]	–1.40[c] (–5.02, 2.22)	0.76	.45	NA	NA	NA
Neonatal outcomes							
Large for gestational age	4 [14,35,36,37,38]	1.39[b] (0.81, 2.40)	1.19	.23	0%	.68	Fixed
Macrosomia	2 [39,40]	1.46[b] (0.27, 7.98)	0.44	.66	69%	.07	Random

[a] CES-D: The Center for Epidemiologic Studies Depression Scale; DES: Diabetes Empowerment Scale; DHDS, Diabetes Health Distress Scale; DSS, Diabetes-related Stress Scale; SF36: SF-36 Health Survey. NA: not applicable.

[b] RR: risk ratio.

[c] MD: mean difference.

Discussion

This meta-analysis includes data from nine experimental studies, which included 852 women from four countries. The results revealed that the Internet-based self-monitoring interventions significantly decreased maternal HbA_{1c} levels compared with usual care among perinatal diabetic women at postintervention. Internet-based self-monitoring interventions significantly decreased the cesarean delivery rate compared to usual care among the mixed group at postintervention.

Internet-Based Self-Monitoring Interventions

The major components of the interventions included self-monitoring glycemic control, medication adherence, physical activity, and diet control. Most of the interventions used websites, phone devices, and/or a glucometer through an asynchronous two-way feedback system. None of the selected studies developed interventions by using theoretical frameworks. Nevertheless, the hypothesized mechanism of action of the interventions should be described according to the Template for Intervention Description and Replication checklist and guide [42]. Theory can explain the rationale of the elements essential to the intervention and how the intervention really worked [43]. Theory can inform interventions in different ways, from identifying theoretical constructs to be targeted or mechanisms underlying particular behavior change techniques to selecting for women the approach that could most likely benefit them toward the right direction [12]. However, the sustainability of the positive findings from these studies is questionable because only three interventions [14,16,34] had follow-up mechanisms. Evidence demonstrated a gradual decline in adherence to self-monitoring of diet, exercise, medication adherence, and weight management [20]. Thus, future studies need to report the long-term effects of the intervention over an extended period. Only one study used a peer-support approach that provided diabetic women with opportunities to discuss problems with others experiencing the same issues [15]; this limitation suggests further research is warranted to determine whether peer-based online forums are effective in improving neonatal or maternal outcomes [19].

Quality of the Evidence and Potential Biases

A high range of heterogeneity occurred between none (0%) to high (95%). The overall methodological quality of the studies included in the review was mixed and 78% (7/9) of the studies used methods to randomly assign women to either the intervention or the usual-care group using methods that we judged were at low risk of bias. This result was due to our selection criteria for either RCTs or CCTs. Thus, the majority prevented selection bias and insured against accidental bias. Only 22% (2/9) of the studies achieved adequate allocation concealment. Therefore, participants or providers could possibly foresee assignments to introduce selection bias. A potentially important source of bias in this meta-analysis was that none of the studies (0/9) achieved blinding of participants and personnel. Support intervention studies face considerable difficulties in blinding providers and women to an Internet-based group. Thus, all women would have performance bias. Only 33% (3/9) of the studies achieved an effective blinding of outcomes, perhaps owing primarily to the nature of the interventions. Even during an attempt made to blind outcome assessment, a high risk of response bias remained possible for outcomes relying on self-report and objective outcomes. Hence, the majority of women might harbor favorable expectation or increased apprehension in the Internet-based group or they might feel deprived or relieved in the usual-care group. The overall impact of sample attrition had a low-risk bias in all studies (9/9), which could improve the generalizability of findings and reduce attrition bias. Approximately 90% (8/9) of the studies reported primary and secondary outcomes that were reported in prespecified methods. Consequently, the selected studies did not obtain misleading results. None of the studies used ITT analysis, which is a method designed to solve problems of noncompliance and missing outcomes to maintain prognostic balance generated from the original random treatment allocation [44]. Therefore, all trials indicated overoptimistic estimates of the efficacy of the intervention on outcomes [44].

Glycated Hemoglobin A1c

The results of this meta-analysis suggest that Internet-based self-monitoring interventions elicit significant effects on helping perinatal diabetic women to reduce their HbA_{1c} levels, which is consistent with the previous meta-analytic review among adults with type 2 DM [24,25]. A previous review identified 11 studies that analyzed HbA_{1c} levels and found that eight of these studies demonstrated a small significant decline in HbA_{1c} because of substantial heterogeneity (I^2= 58%) in the effect interventions [24]. Although our review had no heterogeneity (I^2= 0%) in the five identified studies, the small effect might be explained by different intensities of in-person contact between the intervention and control groups. We found the same in-person follow-up interval in both groups of two studies [36-38], but different intervals between the intervention and control group were indicated in three other studies [14,39,40]. A previous review [24] suggested that the intensity of in-person contact between consultation visits might relate to the efficacy of an Internet-based approach. We could not find the significant effect among subgroups of GDM [36-38,40] because of the small sample size, which had lower statistical power to select the true effect [45].

According to self-regulation theory [6], perinatal diabetic women could review their own data to obtain better understanding of their medical condition for self-awareness. The Internet could provide increased ease and convenience of self-monitoring because processing power and connectivity could allow remote access to information, and algorithms can target most of the components of existing face-to-face interventions [13]. Two-way personalized/tailored feedback with recommendations via email, online, or text message [14,36-39,40] helped gain diabetic knowledge and information for self-adjustment of glycemic control [14,36-39,40], diet control [14,36-39,40], appropriate activities control [36,39], weight gain control [39], and medication adherence control [14,36-39,40]. Sending automated alerts and reminders [14,40], voice messages [39], and visualizing data using graphs [37,38] encouraged engagement to the intervention to reinforce

self-monitoring. Therefore, perinatal diabetic women capitalized on this motivation to improve HbA_{1c} levels.

Cesarean Delivery Rate

Internet-based self-monitoring interventions were found to significantly decrease the cesarean delivery rate for a pool of 307 women in the mixed group [14,39], but no significant difference was found for a pool of 219 women with GDM [35-38,40]. The results of the meta-analysis are consistent with a previous meta-analytic review among women with GDM [30]. The study reported nonstatistically significant differences were found in cesarean delivery rates between telemedicine and a usual-care group; however, cesarean delivery rate analysis included only three studies [35,36,38]. This analysis includes an additional three studies [14,39,40]. The reason behind the significant decrease in the cesarean delivery rate in the mixed group but not in the GDM group remains unclear. Small sample size possibly underpowered the detection of any difference in cesarean delivery rate [45] among the GDM group, which suggests additional research is needed.

Other Maternal and Neonatal Outcomes

This review showed similar neonatal or other maternal outcomes between the Internet-based self-monitoring interventions and usual diabetes care. However, the question remains as to whether Internet-based interventions may offer cost-effective service compared to usual care [28]. Interventions delivered over the Internet are likely to cost less than face-to-face services requiring frequent contact with health care personnel, and their relatively low delivery cost could result in an Internet-based intervention being more cost effective [4,26]. Currently, a dearth of evidence was detected regarding the effects of intervention on cognitive, behavioral, and emotional outcomes among perinatal diabetic women. Despite the identified nine individual cognitive, behavioral, and emotional outcomes in this review, evidence was too limited to draw any conclusion. Thus, additional good quality trials in this area are needed before firm conclusions can be made regarding the efficacy of Internet-based self-monitoring interventions on cognitive, behavioral, and emotional outcomes.

Limitations

This review has several limitations. First, this review included only studies published in English, all of which were conducted in developed regions with high access to the Internet or mobile devices. Therefore, the results may not be applicable to marginalized groups in developing regions. Second, the subgroup analyses we performed prevented drawing definitive conclusions on the efficacy of Internet-based self-monitoring interventions. Subgroup analyses may pose significant interpretation problems, such as false positive or false negative outcomes [46,47]. The false positive outcomes were found for subgroup analyses when no true outcome exists, and have been estimated at 5% per subgroup [46,47]. The false negative outcomes were found because of the small number of outcome events in each subgroup. Therefore, limited statistical power minimized the random error among the estimates of event rates. Third, the small sample size is another limitation given that five of them used a small sample size from 49 [15] to 50 [40], and we found a lack of studies with type 1 or type 2 DM during pregnancy. Fourth, HbA_{1c} is known to be a 3-month mean measure of glycemic control, but the duration of intervention was not mentioned [14,36] or was less than 3 months [39] in three selected trials. Therefore, the validity of this measure as an outcome at postintervention might be questionable. Fifth, a nonsignificant effect was found in the GDM subgroup, but a significant effect was detected in the mixed group; thus, the effect of the type of diabetes rather than the true intervention effect was contentious. Finally, two studies [14,39] had CCT designs with insufficient control of extraneous variables, which diminished the internal validity of their findings.

Implications for Future Research

Continuing research in this area is needed to develop effective Internet-based self-monitoring interventions to improve maternal and neonatal outcomes. Future studies should consider the theoretical base of the interventions [12] with a peer-support component [19] and long-term follow-up [20] to improve the efficacy and sustainability using a RCT design with ITT analyses [44]. However, determining the effective elements of Internet-based application is necessary. Further investigations are needed to divide these applications into specific components, features, transmission, functionality, facilities, interactivity, duration, and mode of delivery to differentiate the distinct effects of different functions [12]. This requirement is especially true in view of the lack of current research that explores the mechanism of effective interventions in different types of perinatal DM.

Clinical Implications

Internet-based self-monitoring interventions may function as important extensions of the range of services to enhance the access of diabetic women to support with self-monitoring especially between consultation visits. Based on the findings of this study, websites that integrate communication with sensor-based systems and a tracing system should be considered high priority in designing self-monitoring interventions to improve maternal glycemic control and cesarean delivery rates. The ubiquity of the Internet facilitates dissemination of information and support to a broader audience and allows information and support to be tailored according to individual characteristics and experiences [26]. Perinatal diabetic women could access and review content at any time and place. Multimedia features and interactivity could accommodate different learning styles [48]. Data visualization capabilities and cloud computing offer accessible display of outcome information, flexible dissemination channels within and between service settings, and ready access to collaborative communication and shared resources for perinatal women and health care providers [13]. Furthermore, gaming technology, Bluetooth technology, interactive voice response, virtual reality, Facebook presence, as well as blogs and Global Positioning System navigation systems are another advancing wave of technological development that might potentially help map out new avenues to promote and support Internet-based self-monitoring among perinatal diabetic women.

Conclusion

The rising popularity of the Internet might result in a shift from the traditional model of care toward an Internet-based health model. Internet-based self-monitoring interventions may introduce new approaches of improving maternal HbA_{1c} and cesarean delivery rates to perinatal diabetic women. Despite the limitations of this review and analysis, our findings have identified a need for future research to employ RCT designs with follow-up data to confirm the long-term effects of Internet-based self-monitoring interventions on maternal and neonatal outcomes among perinatal diabetic women.

Acknowledgments

The research was funded by a start-up grant (Ref: NUHSRO/2013/147/SU/01, WBS No: R545000055133) and HSS Seed Fund (1/2016) from the National University of Singapore.

Authors' Contributions

LY contributed to the conception and design of this study. LY and SN designed the search strategies and performed the literature search. LY, TP, and KY performed the review selection and LY contributed to the analysis and interpretation of data. LY contributed to drafting the article incorporation with all authors. All authors approved the final submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Selection criteria for systematic review.

[PDF File (Adobe PDF File), 24KB - jmir_v18i8e220_app1.pdf]

Multimedia Appendix 2

Index and keyword terms for searching in seven databases.

[PDF File (Adobe PDF File), 35KB - jmir_v18i8e220_app2.pdf]

Multimedia Appendix 3

Risk of bias graph.

[PDF File (Adobe PDF File), 22KB - jmir_v18i8e220_app3.pdf]

Multimedia Appendix 4

Description of Internet-based self-monitoring interventions in 9 selected studies.

[PDF File (Adobe PDF File), 43KB - jmir_v18i8e220_app4.pdf]

References

1. Son KH, Lim N, Lee J, Cho M, Park H. Comparison of maternal morbidity and medical costs during pregnancy and delivery between patients with gestational diabetes and patients with pre-existing diabetes. Diabet Med 2015 Apr;32(4):477-486 [FREE Full text] [doi: 10.1111/dme.12656] [Medline: 25472691]
2. Feig DS, Hwee J, Shah BR, Booth GL, Bierman AS, Lipscombe LL. Trends in incidence of diabetes in pregnancy and serious perinatal outcomes: a large, population-based study in Ontario, Canada, 1996-2010. Diabetes Care 2014 Jun;37(6):1590-1596. [doi: 10.2337/dc13-2717] [Medline: 24705609]
3. Balkhi AM, Reid AM, Westen SC, Olsen B, Janicke DM, Geffken GR. Telehealth interventions to reduce management complications in type 1 diabetes: A review. World J Diabetes 2015 Apr 15;6(3):371-379 [FREE Full text] [doi: 10.4239/wjd.v6.i3.371] [Medline: 25897348]
4. Wei Q, Sun Z, Yang Y, Yu H, Ding H, Wang S. Effect of a CGMS and SMBG on maternal and neonatal outcomes in gestational diabetes mellitus: a randomized controlled trial. Sci Rep 2016;6:19920 [FREE Full text] [doi: 10.1038/srep19920] [Medline: 26814139]
5. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]

6. Kirschenbaum DS. Self-regulatory failure: a review with clinical implications. Clin Psychol Rev 1987 Jan;7(1):77-104. [doi: 10.1016/0272-7358(87)90005-5]
7. Song M, Lipman TH. Concept analysis: self-monitoring in type 2 diabetes mellitus. Int J Nurs Stud 2008 Nov;45(11):1700-1710. [doi: 10.1016/j.ijnurstu.2008.04.005] [Medline: 18571173]
8. Greenwood DA, Young HM, Quinn CC. Telehealth remote monitoring systematic review: structured self-monitoring of blood glucose and impact on A1C. J Diabetes Sci Technol 2014 Feb 21;8(2):378-389. [doi: 10.1177/1932296813519311] [Medline: 24876591]
9. Shieh C, Knisely MR, Clark D, Carpenter JS. Self-weighing in weight management interventions: a systematic review of literature. Obes Res Clin Pract 2016 Feb 17:1. [doi: 10.1016/j.orcp.2016.01.004] [Medline: 26896865]
10. McBain H, Shipley M, Newman S. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. BMC Health Serv Res 2015 Dec;15:565 [FREE Full text] [doi: 10.1186/s12913-015-1221-5] [Medline: 26684011]
11. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
12. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
13. Davidesko S, Segal D, Peleg R. Online medicine for pregnant women. Int J Telemed Appl 2014;2014:379427 [FREE Full text] [doi: 10.1155/2014/379427] [Medline: 25132848]
14. Carral F, Ayala M, Fernández JJ, González C, Piñero A, García G, et al. Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. Diabetes Technol Ther 2015 May;17(5):349-354. [doi: 10.1089/dia.2014.0223] [Medline: 25629547]
15. Kim C, Draska M, Hess ML, Wilson EJ, Richardson CR. A web-based pedometer programme in women with a recent history of gestational diabetes. Diabet Med 2012 Feb;29(2):278-283 [FREE Full text] [doi: 10.1111/j.1464-5491.2011.03415.x] [Medline: 21838764]
16. Bartholomew ML, Soules K, Church K, Shaha S, Burlingame J, Graham G, et al. Managing diabetes in pregnancy using cell phone/Internet technology. Clin Diabetes 2015 Oct;33(4):169-174. [doi: 10.2337/diaclin.33.4.169] [Medline: 26487790]
17. Frost D, Beischer W. Telemedicine in the management of pregnancy in type 1 diabetic women. Diabetes Care 2000 Jun;23(6):863-864 [FREE Full text] [Medline: 10841015]
18. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
19. Dale JR, Williams SM, Bowyer V. What is the effect of peer support on diabetes outcomes in adults? A systematic review. Diabet Med 2012 Nov;29(11):1361-1377. [doi: 10.1111/j.1464-5491.2012.03749.x] [Medline: 22804713]
20. 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]
21. Poorman E, Gazmararian J, Parker RM, Yang B, Elon L. Use of text messaging for maternal and infant health: a systematic review of the literature. Matern Child Health J 2015 May;19(5):969-989. [doi: 10.1007/s10995-014-1595-8] [Medline: 25081242]
22. Gallo S, Phan A, Vanstone CA, Rodd C, Weiler HA. The change in plasma 25-hydroxyvitamin D did not differ between breast-fed infants that received a daily supplement of ergocalciferol or cholecalciferol for 3 months. J Nutr 2013 Feb;143(2):148-153 [FREE Full text] [doi: 10.3945/jn.112.167858] [Medline: 23256143]
23. Cormick G, Kim NA, Rodgers A, Gibbons L, Buekens PM, Belizán JM, et al. Interest of pregnant women in the use of SMS (short message service) text messages for the improvement of perinatal and postnatal care. Reprod Health 2012;9:9. [doi: 10.1186/1742-4755-9-9] [Medline: 22866753]
24. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
25. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, 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]
26. El-Gayar O, Timsina P, Nawar N, Eid W. A systematic review of IT for diabetes self-management: are we there yet? Int J Med Inform 2013 Aug;82(8):637-652. [doi: 10.1016/j.ijmedinf.2013.05.006] [Medline: 23792137]
27. O'Brien OA, McCarthy M, Gibney ER, McAuliffe FM. Technology-supported dietary and lifestyle interventions in healthy pregnant women: a systematic review. Eur J Clin Nutr 2014 Jul;68(7):760-766. [doi: 10.1038/ejcn.2014.59] [Medline: 24781682]
28. Ivey TL, Hughes D, Dajani NK, Magann EF. Antenatal management of at-risk pregnancies from a distance. Aust N Z J Obstet Gynaecol 2015 Feb;55(1):87-89. [doi: 10.1111/ajo.12254] [Medline: 25355283]

29. Chilelli NC, Dalfrà MG, Lapolla A. The emerging role of telemedicine in managing glycemic control and psychobehavioral aspects of pregnancy complicated by diabetes. Int J Telemed Appl 2014;2014:621384 [FREE Full text] [doi: 10.1155/2014/621384] [Medline: 25295059]
30. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2015 Oct;110(1):1-9. [doi: 10.1016/j.diabres.2015.07.007] [Medline: 26264410]
31. Moher D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med 2009 Aug 18;151(4):264. [doi: 10.7326/0003-4819-151-4-200908180-00135]
32. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions.: The Cochrane Collaboration; 2011 Mar. URL: http://handbook.cochrane.org/ [accessed 2016-06-03] [WebCite Cache ID 6jSMWu7Ni]
33. Higgins JP, 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]
34. Nicklas JM, Zera CA, England LJ, Rosner BA, Horton E, Levkoff SE, et al. A web-based lifestyle intervention for women with recent gestational diabetes mellitus: a randomized controlled trial. Obstet Gynecol 2014 Sep;124(3):563-570. [doi: 10.1097/AOG.0000000000000420] [Medline: 25162257]
35. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
36. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
37. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, de la Cruz MJ, Martín P, et al. A telemedicine system based on Internet and short message service as a new approach in the follow-up of patients with gestational diabetes. Diabetes Res Clin Pract 2010 Feb;87(2):e15-e17. [doi: 10.1016/j.diabres.2009.12.002] [Medline: 20044162]
38. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
39. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. J Telemed Telecare 2009;15(5):238-242. [doi: 10.1258/jtt.2009.081213] [Medline: 19590029]
40. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: A feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
41. di Biase N, Napoli A, Sabbatini A, Borrello E, Buongiorno AM, Fallucca F. Telemedicine in the treatment of diabetic pregnancy. Ann Ist Super Sanita 1997;33(3):347-351. [Medline: 9542261]
42. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
43. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: developing an ex post theory of a quality improvement program. Milbank Q 2011 Jun;89(2):167-205 [FREE Full text] [doi: 10.1111/j.1468-0009.2011.00625.x] [Medline: 21676020]
44. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
45. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013 May;14(5):365-376. [doi: 10.1038/nrn3475] [Medline: 23571845]
46. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med 2007 Nov 22;357(21):2189-2194. [doi: 10.1056/NEJMsr077003] [Medline: 18032770]
47. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005;365(9454):176-186. [doi: 10.1016/S0140-6736(05)17709-5] [Medline: 15639301]
48. Bert F, Gualano MR, Brusaferro S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 1;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]

Abbreviations

BMI: body mass index
CCT: controlled clinical trial
CINAHL: Cumulative Index to Nursing and Allied Health Literature
DM: diabetes mellitus
GDM: gestational diabetes mellitus

HbA1c: glycated hemoglobin A1c
ITT: intention-to-treat
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
RCT: randomized controlled trial
RR: risk ratio
SMS: short message service

Edited by G Eysenbach; submitted 05.06.16; peer-reviewed by T Rasekaba, J Furler; comments to author 30.06.16; revised version received 11.07.16; accepted 20.07.16; published 15.08.16.

Please cite as:
Lau Y, Htun TP, Wong SN, Tam WSW, Klainin-Yobas P
Efficacy of Internet-Based Self-Monitoring Interventions on Maternal and Neonatal Outcomes in Perinatal Diabetic Women: A Systematic Review and Meta-Analysis
J Med Internet Res 2016;18(8):e220
URL: http://www.jmir.org/2016/8/e220/
doi:10.2196/jmir.6153
PMID:27526637

Original Paper

Effectiveness of mHealth Interventions Targeting Health Care Workers to Improve Pregnancy Outcomes in Low- and Middle-Income Countries: A Systematic Review

Mary Amoakoh-Coleman[1,2], MD, PhD; Alexander Berend-Jan Borgstein[1], BSc; Stephanie FV Sondaal[1], BSc; Diederick E Grobbee[1], MD, PhD; Andrea Solnes Miltenburg[3], MSc, MD; Mirjam Verwijs[4,5], MSc; Evelyn K Ansah[6], MD, PhD; Joyce L Browne[7], MSc, MD; Kerstin Klipstein-Grobusch[1,8], PhD

[1]Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht University, Utrecht, Netherlands, Utrecht, Netherlands

[2]School of Public Health, Epidemiology and Disease Control, University of Ghana, Accra, Ghana

[3]Institute of Health and Society,, Department of Community Medicine,, University of Oslo, Norway, Oslo, Norway

[4]International Institute for Communication and Development, The Hague, Netherlands

[5]Afya (4) Connect, Change Lake Zone, Mwanza, United Republic Of Tanzania

[6]Ghana Health Sevice, Research and Development Division, Accra, Ghana

[7]Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center,, Utrecht University, Utrecht, Netherlands, Utrecht, Netherlands

[8]School of Public Health, Faculty of Health Sciences, Division of Epidemiology and Biostatistics, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Mary Amoakoh-Coleman, MD, PhD
Julius Global Health, Julius Center for Health Sciences and Primary Care
University Medical Center
Utrecht University, Utrecht, Netherlands
P.O. Box 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 244 623553
Fax: 31 302670511
Email: menba19@yahoo.com

Abstract

Background: Low- and middle-income countries (LMICs) face the highest burden of maternal and neonatal deaths. Concurrently, they have the lowest number of physicians. Innovative methods such as the exchange of health-related information using mobile devices (mHealth) may support health care workers in the provision of antenatal, delivery, and postnatal care to improve maternal and neonatal outcomes in LMICs.

Objective: We conducted a systematic review evaluating the effectiveness of mHealth interventions targeting health care workers to improve maternal and neonatal outcomes in LMIC.

Methods: The Cochrane Library, PubMed, EMBASE, Global Health Library, and Popline were searched using predetermined search and indexing terms. Quality assessment was performed using an adapted Cochrane Risk of Bias Tool. A strength, weakness, opportunity, and threat analysis was performed for each included paper.

Results: A total of 19 studies were included for this systematic review, 10 intervention and 9 descriptive studies. mHealth interventions were used as communication, data collection, or educational tool by health care providers primarily at the community level in the provision of antenatal, delivery, and postnatal care. Interventions were used to track pregnant women to improve antenatal and delivery care, as well as facilitate referrals. None of the studies directly assessed the effect of mHealth on maternal and neonatal mortality. Challenges of mHealth interventions to assist health care workers consisted mainly of technical problems, such as mobile network coverage, internet access, electricity access, and maintenance of mobile phones.

Conclusions: mHealth interventions targeting health care workers have the potential to improve maternal and neonatal health services in LMICs. However, there is a gap in the knowledge whether mHealth interventions directly affect maternal and neonatal outcomes and future research should employ experimental designs with relevant outcome measures to address this gap.

(J Med Internet Res 2016;18(8):e226) doi:10.2196/jmir.5533

KEYWORDS

maternal; mHealth; neonatal; providers of care; low- and middle-income countries

Introduction

The risk for maternal or newborn death is considerably higher in low- and middle-income countries (LMICs) as compared with high-income countries. Despite progress with global decline in maternal mortality, many LMICs still have high maternal mortality rates [1] in particular in LMICs in sub-Saharan Africa and Asia, where the majority of maternal deaths occur [2]. Between 1990 and 2010, globally the under-five mortality rate was reduced by only 28% instead of the targeted 67% [3,4]. Neonatal mortality rate counts toward 41% of the total under-five mortality rate and plays an important role in the slow reduction of under-five mortality rate.

High neonatal mortality rate particularly persists in LMICs [4]. This high burden of maternal and neonatal deaths is compounded with low numbers of physicians and midwives [4]. These human resource challenges are worsened by factors including migration of qualified health workers to better resourced countries, inadequate investment in national health systems, and devastation by major epidemics such as human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS), and malaria [5]. Task shifting as well as innovation in service delivery using available technology provides a promising opportunity to improve maternal and neonatal outcomes [6].

A potential tool to address maternal and neonatal outcome in LMICs is provided by the global increase in mobile technology. The International Telecommunication Union reported that in 2013, global mobile-phone subscriptions reached 6.8 billion and that the mobile-cellular penetration rate or the number of active mobile phone users within a specific population reached 89% in developing countries [7]. This has facilitated the development of a new component of electronic health, namely mobile health (mHealth). The main feature of mHealth is the exchange of health-related information in the form of coded data, text, images, audio, and video using mobile devices. This technology can be used to address challenges such as access, quality, affordability, matching of resources, and behavioral norms [8]. mHealth can be used by health care workers in LMICs to improve affordability of interventions for health promotion, increase health education, and address disease prevention [9-12]. mHealth could also play a prominent role in task shifting, allowing health care workers closely related to the community, such as community health workers (CHWs), to become an important intermediate between higher health institutions and the community. Although the global use of mHealth interventions is increasing, evidence of the effectiveness of mHealth apps is mostly limited to high-income countries, with a focus on the prevention and management of classic chronic diseases, such as diabetes and hypertension [13]. In LMICs, mHealth interventions have been successfully implemented for the management of HIV and tuberculosis [14-16]. Less evidence exists on the effectiveness of mHealth interventions aimed at health care workers providing maternal and neonatal services in LMICs. Therefore, the main objective of this study was to perform a systematic review to assess the effectiveness of mHealth interventions aimed at health care workers providing maternal and neonatal services in improving maternal and neonatal outcomes in LMIC.

Methods

Protocol and Registration

The current systematic review is based on the guidelines provided by PRISMA [17] and was registered in the PROSPERO registry for systematic reviews: (CRD42014010292). This review is part of a larger systematic review that investigated the potential of mHealth interventions targeting both health workers and pregnant women in LMICs to improve maternal and neonatal outcomes.

Information Sources and Search

An electronic systematic literature search was conducted within the following 5 databases: The Cochrane Library (Cochrane Database of Systemic Reviews), PubMed or MEDLINE, EMBASE, Global Health Library, and POPLINE using predefined search terms (Title or Abstract) and indexing terms (MeSH, Emtree) during the period of June 1, 2014, and August 31, 2014. In addition, Grey literature search was performed between October 2014 and April 2015 because many studies focusing on mHealth interventions are not published in peer-reviewed journals. A list was created of organizations working with mHealth interventions. These organizations consisted of nongovernmental organizations, governments' agencies, and the World Health Organization working group on mHealth (Multimedia Appendix 1). The websites of these organizations were searched for publications fitting the eligibility criteria. Furthermore, personal contacts (met through working in the field or at conferences) were approached for papers or documents to be included. Additional papers were found via the snowballing effect, using the reference list of included papers.

Eligibility Criteria

Studies focusing on the domain health care workers in combination with maternal and neonatal care in LMICs were eligible for inclusion. The list of LMICs was created according to the World Bank Classification [12]. The determinant mHealth was defined as a medical and public health practice supported

by mobile devices, such as mobile phones, tablets, and other wireless devices [12,18,19]. It makes use of voice messaging, short messaging service (SMS) text messaging, and apps that can be accessed via general packet radio service, third and fourth generation mobile telecommunications (3G and 4G systems), global positioning system, and Bluetooth technology. The outcomes were not prespecified because of the interest for any outcome related to our domain and determinant. Keywords used in these searches included pregnancy, pregnant, midwife, midwives, traditional midwives, traditional birth attendants (TBAs), CHW, maternal, antenatal, delivery, postnatal, neonatal, perinatal, baby, low resource setting, constrained resource, mHealth, mobile phone, smartphone, mobile app, tablet computer, SMS, short messaging, and telemedicine. The full search strategy can be found in Multimedia Appendix 2.

Included papers were all peer-reviewed, written in English, Dutch, French, German, or Spanish, and primary study papers. Papers were excluded when they did not match the domains and determinants, or were reports of proceedings, project protocols, secondary analysis, animal, biomolecular, or genetic studies. Citations of secondary analysis were reviewed for relevant citations. Interventions relating to the termination of pregnancy were excluded when they targeted the termination of pregnancy before 26-week gestation, as the fetus is then not yet regarded as viable. Interventions making use of a radio were excluded because these interventions fell outside the scope of our definition of mHealth.

Study Selection

The database searches were carried out by ABB and SFS. Subsequent review of search results was undertaken by ABB, MAC, SFS, JB, and KKG. Three reviewers (ABB, ASM, and MV) screened the papers found in the grey literature search. There were no disagreements on paper inclusion.

Data Extraction

Data extraction was done according to a standardized data extraction form based on: the study, study design, location, target population or size, form of mHealth, focus of evaluation measure (whether maternal or neonatal), mHealth function, relevant study findings with respect to outcome used in the study, role of mHealth, and the strengths, weaknesses, opportunities, and threats of the intervention.

Extraction of the data from database papers was done by a single reviewer (ABB) who was not blinded for journal or author details. Lack of clarity during the extraction process was resolved by consulting the second reviewer (MAC). Data extraction of the grey literature was done by 4 reviewers (ABB, ASM, MV, and MAC). In case of incomplete data, one attempt was made to contact the corresponding author by email.

Quality Assessment

The quality of the included papers was assessed according to an adapted Cochrane Risk of Bias Tool [20]. mHealth interventions as well as the target populations differed between the studies. This tool was used because it gives more guidance on details for classifying the risk of bias and therefore enhances uniformity of assessment (Multimedia Appendix 3). Bias was assessed on the selection process of the study population, completeness of data (example number of dropouts), origin of the data (measurements performed by authors or database research), blinding of the researchers or clinicians, the presence of a clear definition of the outcomes that were used, and whether confounders were taken into account in analysis. Risk of bias was assigned as either low risk, high risk, or unclear risk. The quality assessment tool can be found in Multimedia Appendix 3. Validity of the papers was taken into account in the Discussion section.

Synthesis of Results

Studies were grouped into 2 types: intervention and descriptive. Intervention studies employed more rigorous nonrandomized study designs used for evaluating interventions [21], whereas descriptive studies used mainly cross-sectional designs or were case studies. Data synthesis aims to give a narrative analysis. First, an overview of the scope of mHealth interventions is provided. The scope of the studies consists of the year of publication, region of the world, form of mHealth intervention, the mHealth function, topic addressed, and target population.

Narrative synthesis of the intervention studies are presented in an evidence table, in which the studies are analyzed according to their year of publication, study design, location or setting, target population, whether evaluation measures are maternal or neonatal, form of mHealth, mHealth functions related to data collection, educational, and communication and finally relevant findings. A similar evidence table was used to summarize the findings of the descriptive studies. Heterogeneous outcomes, settings, and varying study designs limited our ability to group the results of 2 or more papers together to conduct a meta-analysis for an overall quantitative conclusion. A strengths, weaknesses, opportunities, and threat analysis was also performed for all the included studies, as well as for mHealth as an intervention.

Results

Overview of Included Studies

A total of 3725 papers were identified in the database and grey literature searches. After removal of duplicates using Endnote (version 11), 2965 articles remained and were screened by title and abstract. This resulted in exclusion of 2909 articles, leaving 56 articles to screen for eligibility. Thirty-seven articles were further excluded. Reasons for exclusion included unavailability of full text (n=17), language (n=2), secondary analysis (n=8), reports (n=7), and unavailability of records providing additional but key information on studies (n=3). A total of 19 articles were included in our study, 10 intervention studies and 9 descriptive studies. Figure 1 illustrates the study screening and selection process, whereas Table 1 provides an overview of the scope of mHealth interventions in the included studies. Overall, 73.7% (14 of 19) of the studies were conducted in Africa and 26.3% (5 of 19) in Asia. Most interventions (68.4%) used SMS text messaging, but the form of text messaging varied between the studies. Text messaging forms included unidirectional and multidirectional text messages. Furthermore, 10.5% (2 of 19) of studies combined SMS text messages with another form of

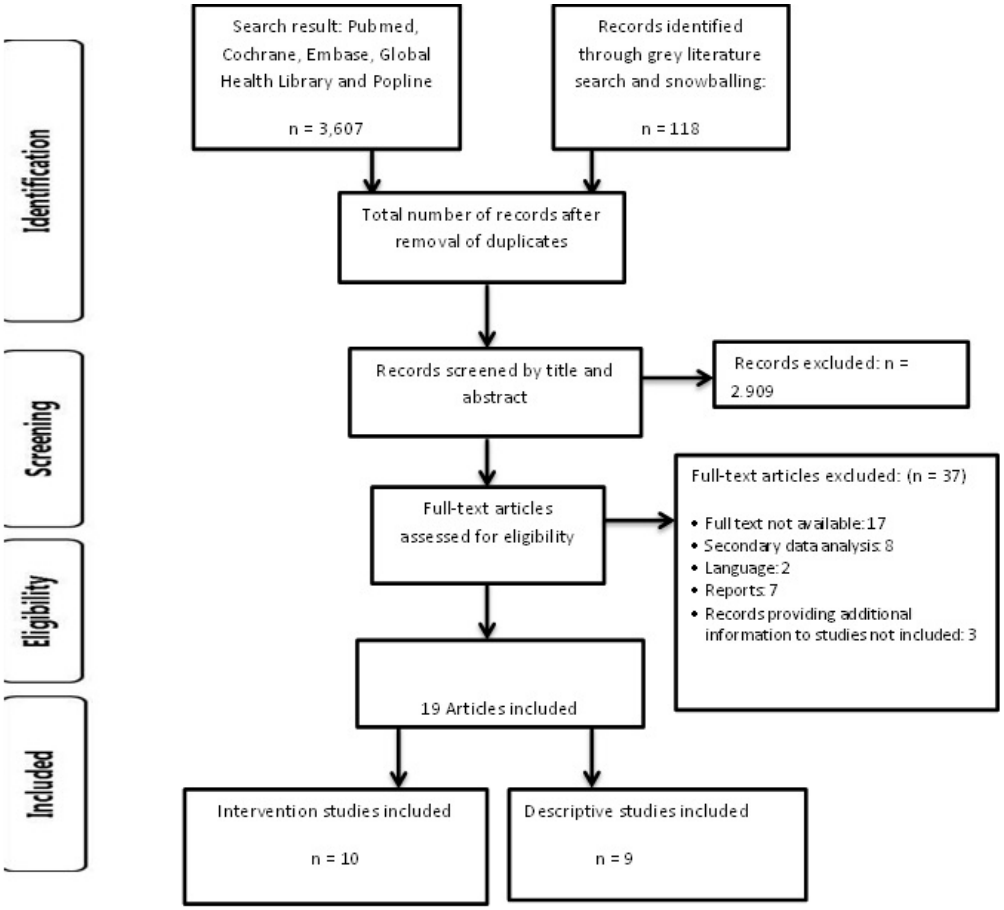
mHealth. mHealth function as well as target populations differed between the studies.

Regarding the quality of mHealth evaluations in the studies, only one of the intervention studies was a randomized controlled trial (RCT) [22]. The others [23-31] had quasi-experimental nonrandomized designs. Table 2 provides an overview of the characteristics of the intervention studies, respectively by RCT and nonrandomized controlled trial status. Study designs reported for the descriptive studies ranged from case studies to cross-sectional study designs [32-40]. Table 3 provides an overview of the characteristics of the descriptive studies and their key findings.

Table 1. Scope of studies included in the review.

Category	Subcategory	Intervention studies (N=10)		Descriptive studies (N=9)	
		Number of studies	% of studies	Number of studies	% of studies
Region	Africa	8	80.0	6	66.7
	Asia	2	20.0	3	33.3
Form of mHealth	Unidirectional text messaging	3	30.0	3	33.3
	Multidirectional text messaging	2	20.0	1	11.1
	Multidirectional text and voice messages	1	10.0	1	11.1
	Unidirectional text messaging and Web-based technology	1	10.0	1	11.1
	Mobile phone health apps or surveys	2	20.0	2	22.2
	Mobile phone software	-	-	1	11.1
	Mobile phone recording	1	10.0	-	-
mHealth function (certain studies contain multiple forms)	Data collection	6	60.0	4	44.4
	Educational	1	10.0	3	33.3
	Communication or information sharing	5	50.0	2	22.2
Topic (certain studies contain multiple topics)	Postpartum hemorrhage	1	10.0		
	Skilled maternal and newborn care	4	40.0	2	22.2
	Training or educating midwives and nurses	2	20.0	4	44.4
	Reproductive health	1	10.0	-	-
	HIV and pregnancy	2	20.0	1	11.1
	Malaria in pregnancy	-	-	1	11.1
	Postnatal depression	-	-	1	11.1
	Infant feeding	1	10.0	-	-
Target population (certain studies target multiple populations)	Traditional birth attendants	2	20.0	1	11.1
	Health extension workers	1	10.0	1	11.1
	Midwives	2	20.0	2	22.2
	Health care staff	2	20.0	2	22.2
	Medical students	1	10.0	-	-
	Community health workers	3	30.0	4	44.4
	Health surveillance assistants	1	10.0	-	-

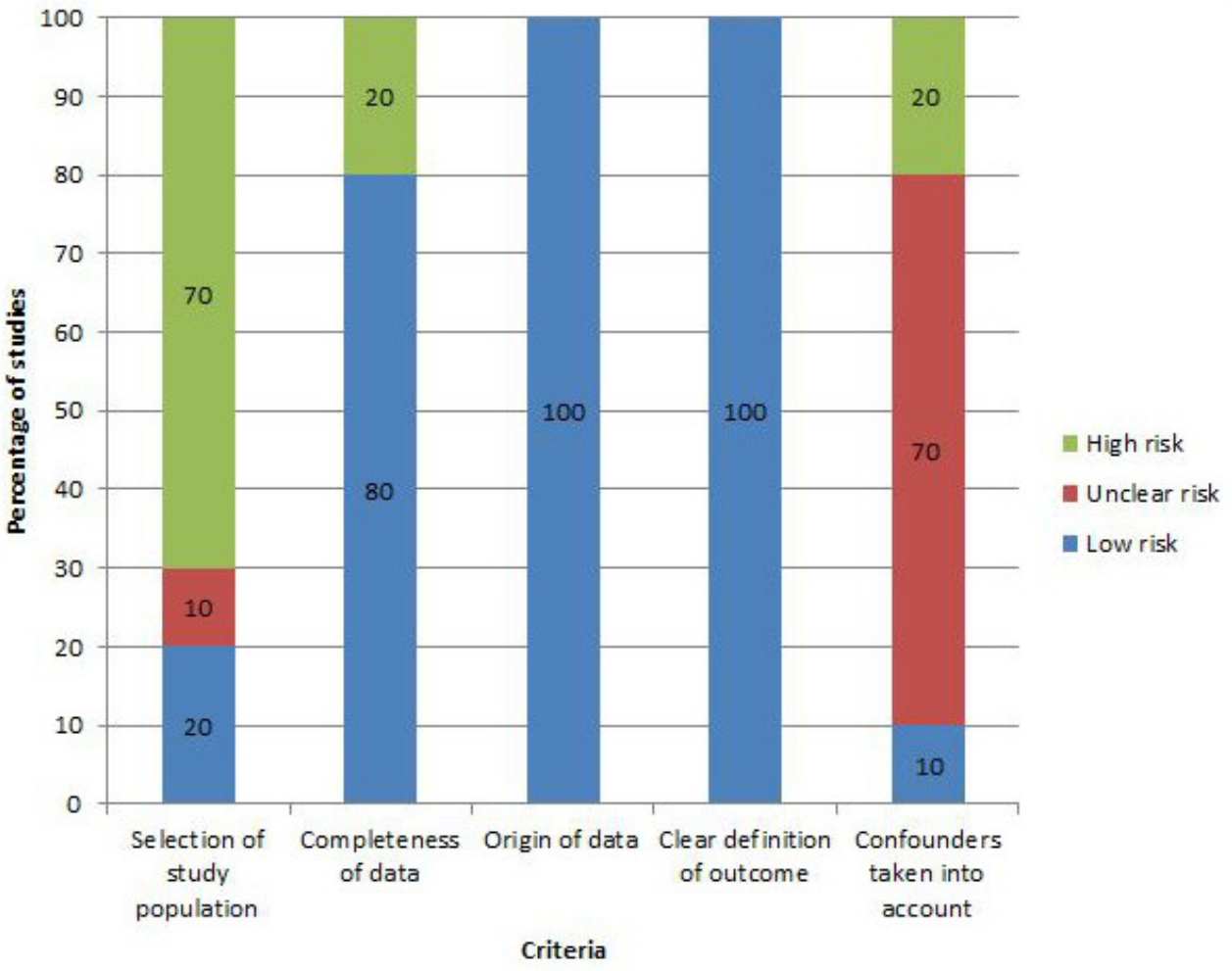
Figure 1. PRISMA flow diagram of studies included in this review.



Overall Risk of Bias Assessment of Intervention Studies

The overall risk of bias assessment is reported in Figure 2 and Multimedia Appendix 4. All included studies provided a clear definition of the outcomes and made use of self-measured data. Only one study accounted for confounders in the analysis [27], 2 studies did not [24,29], and the remaining 8 of the included studies did not mention whether or not this was the case [22,23,25,26,28,30,31,40]. Most studies had a high risk of bias for selection of study population, with one study with unclear risk of bias [25]. Two studies scored a high-risk bias for completeness of data, one study did not make use of a control group [24], and another paper performed the study without secondary monitoring [23].

Figure 2. Graphical presentation of risk of bias assessment for intervention studies included in the review.



Narrative Synthesis of Results of Intervention Studies

The intervention studies distinguish the use of mHealth directed at health care providers for data collection, communication, or educational purposes.

mHealth as a Data Collection Tool

Six papers described an mHealth intervention used as a data collection tool [22-24,26-28].

Two of these studies assessed the knowledge and skill retention of midwives and TBAs after training sessions on how to use mobile phones as data collection tool [27,28]. In rural Liberia, TBAs were trained in SMS texting data collection protocol and performed a 1-year posttest evaluation that showed that 63.6% of them displayed evidence of statistically significant knowledge and skill retention [28]. However, only 11.0% of the TBAs were able to perform more complex tasks, such as adding credit to the mobile phone to increase mobile phone usage time. This task proved to be difficult due to low literacy and poor mobile phone reception and 69.8% of TBAs relied on people with higher education to support them with mobile functions. In rural Liberia, traditional midwives were trained to use SMS texting for real-time remote data collection using a pregnancy reporting protocol. They showed a significant mean increase in mobile phone knowledge and skill acquisition afterward [27].

In a study conducted in Ethiopia, most health workers were able to use mobile phone health apps that are appropriate for their technical needs in terms of maternal health data collection [26], suggesting that mobile technologies allow health managers to more quickly and reliably have access to data which can help them identify whether and where there are issues in service delivery. Evaluation of mobile phone use by TBAs to report incidence, management, and outcome of postpartum hemorrhage, showed that 90.0% of the TBAs used their mobile phones to send SMS text messages in the 90 days following training and were able to use the protocols to report clinical outcomes [23]. In 4 Chinese villages, the use of mobile phone questionnaire for data collection on infant feeding practices was compared with a pen-and-paper method. The pen-and-paper method was observed to be more error prone [22] as 65.0% of the pen-and-paper records did not match and needed additional checking. Cost-wise, however, the mobile phone method was considerably more expensive at US $23 per questionnaire, compared with the pen-and-paper method that cost US $13 per questionnaire. In a study in Myanmar, a combination of Web-based and mobile technology was used to generate antenatal care and expanded program on immunization visit dates, in which the health worker was able to cross-check, identify, and update the mother and child status at the health facility or during home visits [24]. This tool improved antenatal

care and expanded program on immunization coverage in the study area with less delay in antenatal and immunization visits.

mHealth as Communication Tool

Three studies assessed the use of mHealth interventions as a communication tool [25,29,30]. In Malawi, it was shown that mobile phones can be used to reduce the communication gap between health workers and their district teams [25]. The intervention consisted of SMS text messages used to report stock-outs, asking for general information, reporting emergencies, confirming meetings, and requesting technical support. For CHWs in the intervention group, an average of 9 minutes and at an average cost of $0.6 to report issues and receive feedback per contact was estimated. The most common modes of communication among this group were phone calls (94.4%) and SMS text messages (100.0%). Face-to-face communication was only reported by 8.0% of the participants. For CHWs in the control district, 1681 minutes (28 hours) and an average cost of US $4.6 to report and receive feedback per contact were estimated. The most common modes of communication amongst this group were use of face-to-face communication with their supervisors at district level (92.0%) and phone calls (6.0%) with none of them using SMS text messages. The mHealth intervention was at least 4 times cheaper and 134 times more efficient, compared with traditional and most common methods of walking, biking, or using public transport to reach supervisors face to face.

SMS texting between CHWs and either ambulance, health facility staff, district hospital, and central level, enabled an effective and real-time 2-way communication alert system to reduce maternal and child health deaths in Rwanda [29]. The mHealth intervention resulted in a 20.0% increase in facility-based delivery, from 72.0% to 92.0%. In addition, CHWs became more proactive in identifying pregnant women and following up on registered pregnant women for appropriate care, by sending reminders to their mobile phones. In a study conducted in Zambia, SMS text messaging was used for transmission of test results between health facilities and caregivers to reduce the time needed for diagnosis of infant HIV infection [31]. The mean turnaround time for delivery of a test result to the relevant health facility decreased from 44.2 days preimplementation to 26.7 days postimplementation. The mean turnaround time for delivery of test results to the caregiver of the tested child decreased from 66.8 days to 35 days [31].

Table 2. Characteristics of the intervention studies included in the review and the relevant findings.

Study/ (Focus of evaluation)	Study design	Target population or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2014, Munro et al [28] (Maternal)	Nonrandomized design (preintervention and postintervention evaluation)	99 TBAs[a] trained; 63 retained 1-year post-training for complete evaluation (Rural Bong county in Liberia)	Utilization of mobile phone functions, coded SMS text messaging	Data collection and transmission (locating pregnant women; take data on age and referring them for antenatal care)	Participants demonstrated an increase in the mean number of skills that they were able to perform between pretest and both the immediate posttest and 1-year posttest. Mean number of skills that the participants were able to complete did decrease slightly between immediate and posttest. Individual skills verified a significant retention of knowledge between the pretest and posttest. Many TBAs continued to have trouble with the more complex skills of adding credit to a mobile phone (11% was able to do this at posttest). 70% of TBAs relied primarily on others with higher education to assist them with mobile functions; 24% used their phone to communicate with the certified midwife; 14.3 % wanted to communicate, but had poor reception.
2014, Pathfinder (Grey Literature) [30] (Maternal)	Nonrandomized design (preintervention and postintervention evaluation)	258 participants in an infant feeding health education program (Abuja and Nasawara in Nigeria)	150 CHW[c] in 10 primary health clinics	Education and communication (ANC[d] protocols, and client follow-up)	CHWs increased HIV[e] testing from 68% to 82%. Blood pressure measurement increased from 87% to 97%. The quality of care score from client interviews increased from 13.33 at baseline to 17.15 at end line, with the most significant improvements related to health counseling. Individual and group health counseling sessions became structured.
2013, Little et al [26] (Maternal)	Nonrandomized design (preintervention and postintervention evaluation	20 HEW[f], 12 midwives, 5 supervisors (Kilte and Awelalo districts in the Tigray region of Ethiopia)	Mobile phone app using open source components	Data collection (using appropriate technologies to meet needs of HEW and midwives	GPRS[g] connection was available in 35 health posts and centers (74%) of the study districts. There were very few instances of the mobile data network being unavailable for a substantial period of time. 34 of the 36 phones were retained; 2 were stolen with one later recovered; 3 phones had issues with insensitive screens and were replaced. Most health workers rapidly learned how to use and became comfortable with the touch screen devices so only limited technical support was needed.
2012, Zhang et al [22] (Maternal and Neonatal)	Randomized controlled trial	10 students of the Hebei Union School of Public Health (Zhaozhou Township, Hebei Province, China)	Mobile phone data collection	Data collection (Use of mobile phones for data collection on infant feeding practices compared with use of pen and paper)	In 120 copies of pen-and-paper questionnaires, 55 questionnaires contained errors. 65% of the pen-and-paper records did not match and needed to be checked. There was no significant difference between duration of pen-and-paper method versus mobile phone method. The mean cost per questionnaire was higher for the mobile phone questionnaire (US $23) than for the pen-and-paper questionnaire (US $13). The mobile phone method was acceptable to interviewers, with only minor problems that did not result in data loss.

Study/ (Focus of evaluation)	Study design	Target population or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2012, Lori et al [27] (Maternal)	Nonrandomized design (preintervention and postintervention evaluation)	99 TBAs (Rural Liberia)	SMS text messaging	Data collection (using a pregnancy reporting protocol)	Mean increase in mobile phone knowledge scores was 3.67 (95% CI 3.39-3.95). Data collectors also demonstrated a significant increase in their ability to perform each individual mobile phone task. Participants with a mobile phone in the family did significantly better on 3 of the 7 tasks in pretest.
2012, Seidenberg et al [31] (Infant)	Nonrandomized design (preintervention and postintervention evaluation)	At least 2 health workers from each facility (2 districts in Southern Zambia)	SMS text messaging	Data collection and transmission (to reduce the time between blood sampling for the detection of infant HIV infection and notification of the test results to the relevant point-of-care health facility by using SMS-based system	Mean turnaround time for delivery of a test result to the relevant health facility fell from 44.2 days (SD[h]:28) preimplementation to 26.7 days (SD:31.8) postimplementation. Mean turnaround times for delivery of a test result to a caregiver of the tested infant were 66.8 (SD: 38.8) preimplementation and 35 days (SD: 31.2) postimplementation.
2012, Lemay et al [25] (Maternal)	Nonrandomized controlled trial (staged design)	Health surveillance assistants and community health workers. 95 SMS users in Salima. 95 nonusers in Salima. 95 nonusers in Kasungu (Salima, Nkhotakota, and Kasungu Districts in Malawi)	SMS text messaging	Communication (reducing communication gaps between health workers and their district teams; increasing access to information and improve quality of services)	SMS used to report stock-outs, asking general information, reporting emergencies, confirming meetings, and requesting technical support. Among respondents who received phones, the most common form of communication was SMS (100%), phone calls (94%), public transport to travel (8%). Among respondents who did not receive phones: 92% used transport and only 6% used phone calls. None used SMS for communication. SMS participants needed an average of 9 minutes to report issues and receive feedback at an average of USD 0.61$. Health workers with no access to SMS spend an average of 1445 minutes (24 hours) to report and receive feedback at an average of USD 2.70$. In control district, it took 1681 minutes to report and receive feedback at an average of USD 4.56$.
2012, Ngabo et al [29] (Maternal)	Nonrandomized design (pre intervention and postintervention evaluation)	432 community health workers and the rest of the health system (Musanze, Rwanda)	SMS Text messaging (Rapid SMS-MCH[i] system)	Communication (SMS-based platform, enabling effective and real time 2-way communication for action, between CHWs at community level, and the rest of the health system. Used to improve access to antenatal, postnatal care, institutional delivery, and emergency obstetric care)	5734 SMS were sent. 11,502 pregnancies (81% of the 14,200 estimated annual pregnancies in district) were monitored. A 20% increase in facility-based delivery from 72% 12 months before to 92% at the end of pilot phase.
2011, Andreatta et al [23] (Maternal)	Nonrandomized design (posttraining evaluation)	8 TBA, 2 professional nurse midwives (Sene District in Ghana)	SMS text messaging	Data collection and communication (reporting postpartum hemorrhage occurrence, management, and outcome)	Both professionals and TBAs were able to use the specified reporting and text messaging protocols to report clinical outcomes. 425 births were reported during study period, with 13 cases of PPH[j]occurring (3.1%) cases.

Study/ (Focus of evaluation)	Study design	Target population or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2010, Kwaewkungwal et al [24] (Maternal and child, including neonatal)	Nonrandomized design (preintervention and postintervention evaluation)	Health workers in charge of ANC or EPI[k] services (sample size not indicated in paper) (Phung district, Thai-Myanmar)	SMS text messaging and Web-based apps	Data collection, automated generation of list, and update information regarding the antenatal care and child's immunization status on mobile phone when performing ANC or EPI activities off health care clinic	59% come on time as per scheduled dates after implementation compared with 44% before implementation. 44% of children who came to receive scheduled vaccines on time on the preset monthly immunization date after intervention compared with 35% before. Updating immunization data on mobile phone increased odds of EPI on time by 2.04.

[a]TBA: traditional birth attendant.

[b]SMS: short message service.

[c]CHW: community health worker.

[d]ANC: antenatal care visit.

[e]HIV: human immunodeficiency virus.

[f]HEW: health extension worker.

[g]GPRS: general packet radio service.

[h]SD: standard deviation.

[i]MCH: maternal and child health.

[j]PPH: postpartum hemorrhage.

[k]EPI: expanded program on immunization.

Narrative Synthesis of Results of the Descriptive Studies

In Indonesia, a theoretical model on the use of mobile phones to enhance the capacity of health workers was developed and tested among 223 midwives [35]. Mobile phone use was positively associated with midwives' access to institutional and peer information resources. SMS text messaging was used to educate midwives in a study conducted in South Africa and improved clinical practice was reported by 72.0% of participants. More than two-thirds (68.0%) of the midwives commonly shared and discussed the messages with their colleagues. All participants requested to receive more text messages on other important topics [40]. Again, in rural South Africa, the feasibility and acceptability of using mobile phone app for data collection amongst pregnant women living with HIV was explored [38]. Acceptability was high, as well as perceived usefulness and ease of use. Feasibility of conducting the interviews in the setting was also high, with no significant challenges with respect to network coverage, cost of hardware and software, and secure transmission of data. Among nurses receiving midwifery education, mobile phone usage facilitated authentic problem solving, reflective practice, and life-long learning [36]. In Afghanistan, World Vision rolled out an mHealth intervention, using the open source CommCare platform [39]. This was aimed at improving the quality of counseling for pregnant women, promote facility delivery, and facilitate timely referral of women and newborns to facilities and hopefully result in decreased maternal and newborn deaths following increase in utilization of services [39]. The study does not report on how these were achieved but mentions that it was successful, and the project was further expanded to deal with childhood nutrition. In one study, 9 health workers were provided with mobile phones with an installed algorithm that allowed for real-time access to data. This reduced time lag in patient data transmission and allowed for pregnant women to be categorized based on risk for treatment [32]. The Safer Deliveries project in Zanzibar, which employed an open source mobile app enrolled TBAs and resulted in an increase in registration of pregnant women as well as increase of facility delivery from 33.6% to 71.0% [33]. Finally, the International Institute for Communication and Development in 2013 piloted a project in Mali in which CHWs and specialists used a mobile app to report and monitor cases of malaria in pregnancy and among children aged younger than 5 years, as well as detect and respond promptly to any outbreak [34]. This resulted in a 31% reduction in malaria in pregnancy and 33% reduction in children under 5 years with malaria. Table 3 presents an overview of the characteristics and relevant findings of the descriptive studies.

Strengths, Weaknesses, Opportunities, and Threats Analysis of Included Studies

A Strengths, Weaknesses, Opportunities, and Threat analysis was conducted for all included studies. All studies included are relatively current studies published from 2010 onward providing up-to-date information. There is also a good variation of settings within the domain of LMIC, with studies conducted in West, East, and Southern Africa, and different parts of Asia. All forms of mHealth interventions as well as different functions that can be served by mHealth are represented in these studies. In all but one study, standardized phones were procured for participants. Strength of the included studies is the broad range of health worker categories considered, allowing for easy assessment of feasibility of mHealth apps for the daily work of health care workers. Weaknesses in the studies related mainly to their study design. Only one of the intervention studies was a RCT [22], the others combined various nonrandomized study designs, such

as preintervention and postintervention comparison, and in this regard, lack the specific rigor associated with intervention studies. The descriptive studies ranged from surveys to case studies. For each broad mHealth function, various apps were reported. For example, under communication function, one study reported mHealth used to report stock-outs to higher levels [25], whereas another study reported mHealth used to follow-up women and report clinical outcomes [23]. This could have been an advantage if more than one study reported specific apps to allow for comparison of results. Most of the studies report on process rather than outcome measures, and thus, this systematic provides an indication for opportunity for more rigorous intervention studies that focus on both specific maternal and neonatal outcome measures in the future.

Strengths, Weaknesses, Opportunities, and Threats Analysis of mHealth as an Intervention

Multiple studies mention low costs to be strength of the mHealth interventions compared with traditional methods [24,25,31,40]. mHealth interventions were found to be considerably more efficient than traditional methods used for communication [25], and to improve the effectiveness of community health services in terms of managing logistics, reporting events, and addressing emergencies. mHealth allowed for integration of all levels of health workers, including TBAs, to expedite emergency referrals and communicate with skilled providers like midwives [28].

Weaknesses included that the information in the text messages was too simple and needed additional detailed information [40]. In addition, remoteness of study sites was a limitation to Web-based education due to lack of or poor access to the Web [40]. Technological problems such as poor reception, lost and damaged phones, and difficulty with certain mobile phone models were also identified [40]. mHealth interventions that are dependent on existing information systems, with modular systems that are not interoperable, cannot be linked to other settings and data structures [24]. Other weaknesses include poor telephone maintenance and lack of or limited access to electricity in a number of communities.

Clear opportunities exist for utilization of mHealth. This includes the additional functions of the technology, such as global positioning system, taking and storing pictures and videos, as well as the ability to record sound. These can facilitate data collection tools in the future [29]. Another opportunity offered by mHealth is for a broader mobile network coverage that could expand the reach of health information to frontline health workers in remote areas and accelerate knowledge exchange between health workers and higher levels in the health system [25]. mHealth offers the opportunity for strengthening of health care infrastructure with the requisite financial support, and the technology can be applied to a broader scope of public health care [27]. It is also possible to address various health system issues using one mHealth program or intervention [37,39], and there are various software available that could be adapted to suit specific needs of health care [34]. Opportunities for public–private partnership also exist [29].

Factors that threaten mHealth implementation included lack of reliable Web coverage, which limits the potential of mHealth in the public sector [28,40], limited capacity to manage damaged phones, low literacy levels [28], and lack of appreciation by health workers of the need to use data where it is generated [24]. Several of the reported interventions were conducted by foreign agencies that could potentially result in limited sustainability [30,33,34,39]. This, however, could be overcome if local stakeholders and institutions engaged in the program appreciate the value of the capacity building offered. Finally, it is important that any tool adapted to be used through mHealth be itself efficient, to maximize the benefits of mHealth functions [37].

Table 3. Characteristics of the descriptive studies included in the review and the relevant findings.

Study (Focus of Evaluation Measures)	Study design	Target health workers or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2014, Lee et al [35] Maternal and Neonatal)	Cross-sectional	223 midwives (15 sub-districts of Aceh Besar, Indonesia)	One-way mobile phone use	Improving access to health-related resources: formal (medical professionals) and informal (peer workers) resources	Mobile phone use was positively associated with midwives' access to institutional and peer information resources. Access to institutional resources was positively associated with midwives' health knowledge, whereas access to peer resources was not. Access to peer resources was associated with higher self-efficacy, which was positively associated with health knowledge. Implications for technology interventions strategies targeted to community health workers in rural communities provided.
2014, Tsai et al [37] (Maternal)	2 Cross-sectional studies	Community health workers (sample size not stated in paper) (Khayelitsha Cape Town, South Africa)	Mobile phone program	Case finding (use of mobile phones for administering the EPDS[a] during the routine course of their community-based outreach and wellness work	CHWs[b] were able to detect probable antenatal depression using the scale during their routine outreaches with excellent discrimination, with area under the receiver-operating characteristic curve (AUC) values ranging from 0.91 to 0.99; 0.97 sensitivity and 0.76 specificity.
2014, Pimmer et al [36] (Maternal)	Case study	16 nurses attending an advanced midwifery course (Rural South Africa)	Mobile phones	Nurse education (mobile phones as educational tools)	Nursing students in resource poor settings use mobile technology as educational tools. These learning practices involve sociocognitive processes, learning in the form of joint problem solving and reflection, as well as more intensive forms of sociocultural participation. In order for educational institutions to more fully and more systematically harness the potentials of these media, a number of ethical and practical issues need to be addressed.
2014, D-Tree International [33] (Grey Literature) (Maternal and neonatal)	Cross-sectional	24 TBAs in phase I and 223 CHWs in phase II (Zanzibar, Tanzania)	Mobile phone with open source mobile app	Data collection and communication and information sharing	There was an increase in access to skilled care during pregnancy, childbirth, and post-partum care. 71% facility delivery compared with 33.6% in DHS.[c] 77% facility delivery in phase II (DHS range between 25% and 41%). Increase in primary care level deliveries (44% compared with 4% in 2011). Geographical differences in delivery habits highlighted. Increased self-efficacy and capability of frontline workers.
2014, World Vision [39] (Maternal and neonatal)	Cross-sectional	CHWs (sample size not provided in paper) (Afghanistan)	Mobile phone	Counseling and referrals	Promotion of health facility deliveries. Timely referrals of women and newborns. Decrease in maternal and new-born deaths (worth noting that exact measures of these are not stated in the paper).

Study (Focus of Evaluation Measures)	Study design	Target health workers or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2013, van Heerden et al [38] (Maternal)	Cross-sectional	16 data collectors; (Rural South Africa)	MPAPI[d]	Data collection (feasibility of face-to-face maternal health data collection from pregnant women living with HIV using a mobile phone survey app)	Perceived usefulness was reported to be slightly higher than perceived ease of use. After 3 months of field use, interviewer perceptions of both perceived ease of use and perceived usefulness were found to be higher than before training. High feasibility of conducting MPAPI interviews in this setting. Network coverage was available in all clinics and hardware, software, cost, and secure transmission to the data center presented no significant challenges over the 21-month period. For the 12 MLH[f]participants in group 2, anxiety about the multimedia capabilities of the phone was evident. Their concern centered on the possibility that their privacy may be invaded by interviewers using the mobile phone camera to photograph them. For participants in group 3, having the interviewer sit beside versus across from the interviewee during the MPAPI interview was received positively by 95% of MLH. Privacy (6%) and confidentiality (5%) concerns were low for group 3 MLH. Mobile phones were found both to be acceptable and feasible in the collection of maternal and child health data from women living with HIV in South Africa.
2013, IICD[g] [34] (Grey Literature) (Maternal and children)	Cross-sectional	50 CHWs and 10 health specialists (Yirimodjo, Mali)	MAMMA[h]	Data collection and monitoring (questionnaire with malaria indicators and monitoring of disease evolution)	31% reduction in malaria in pregnancy cases. Faster treatment response (65%). 32% reduction in malaria in children under 5 years. 20% increase in pregnant women sleeping under bed net. 42% increase in pregnant women receiving preventative medication.
2012, Woods et al [40] (Maternal and neonatal)	Cross-sectional	50 midwives out of 2500 midwives from public and private sectors (South Africa)	SMS text messaging; with link to a website with additional information	Education	86% enjoyed and learned from weekly text messages. 72% felt that the messages improved clinical practice. 68% shared and discussed the messages with colleagues.

Study (Focus of Evaluation Measures)	Study design	Target health workers or size (Setting)	Form of mHealth	mHealth function	Relevant findings
2010, Alam et al [32] (Grey Literature) (Neonatal)	Case study	9 BRAC[i] health workers (3 urban slums of Dhaka, Bangladesh)	Mobile phones with smart algorithms (The Click Module)	Data collection (real-time access to data)	Health workers could send data directly to the central MIS[j] system. Reduced time lag in data transmission. Established a secured Web page containing all patient data. Established an automated decision tree that categorizes patients depending on their risk levels.

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]CHWs: community health workers.

[c]DHS: demographic and health survey.

[d]MPAPI: mobile phone-assisted personal interviewing.

[e]HIV: human immune deficiency virus.

[f]MLH: mothers living with HIV.

[g]IICD: International Institute for Communication and Development.

[h]MAMMA: Mamans Mobiles contre le Malaria au Mali.

[i]BRAC: building resources across community (a nongovernmental organization).

[j]MIS: management information system.

Discussion

Principal Findings

This systematic review shows effective use of mHealth interventions as communication, educational, and data collection tools by health workers to report on medical events related to maternal and child health within their community. These constitute health systems strengthening app tools [41]. The specific mHealth functions have enabled health workers to track pregnant women in their care, as well as facilitate referrals. Such strategies, targeted at data collection and reminders for antenatal visits, directly impact on service utilization such as antenatal coverage, whereas those aimed at improving skilled attendance at delivery and facility delivery impact more directly on mortality [42]. Unfortunately, only one of the studies included in our review directly reports the effect of mHealth on maternal and neonatal mortality, without exact details of coverages. Although other studies have shown that mHealth interventions targeted at clients can reduce perinatal mortality by 50% [43], we could not convincingly demonstrate this effect for health worker-targeted interventions.

The fact that most of the studies included in this review targeted health workers at community level provides insights into the possibility of creating an intermediate layer in which health workers form an important linkage between higher health institutions and the community in harnessing the befits of mHealth.

Some challenges of mHealth were identified, and these were mainly technological problems, such as mobile network coverage, Web-based access, electricity access, and maintenance of mobile phones [11,29,40]. These could negatively affect the expansion of the mHealth interventions in LMIC if not addressed. However, technological improvement comes with associated costs. Engaging the private sector in a public-private partnership can reduce such cost [29] and facilitate the expansion of mHealth interventions in LMIC in the future. Decreasing costs of handsets will potentially reduce further the cost of mHealth interventions.

Given the large investments in mHealth [44,45], with the highest cost of service provision in LMIC [6], experimental evaluations which will thoroughly assess its impact [45], more specifically on maternal and neonatal health outcomes, will be most beneficial. Future studies should investigate the effectiveness of the interventions by measuring similar outcomes. Health workers' level of literacy affected their ability to perform complex tasks on mobile devices [27]. It is important to address this challenge if the full complement, respectively potential of mHealth is to be deployed at all levels of service delivery, including the community level. Lower cadre of health staff will need adequate training to ensure their optimization of such interventions.

Limitations

Limitations of our review include a high risk of bias observed for some of the intervention studies, mainly relating to limited consideration of confounding. Only one study was a RCT. Most studies were pilot or implementation studies. A further limitation of the current systematic review is the domain limitation of LMIC. This affects the generalizability of our results to other settings, as we are aware that some studies in high-income countries or low-income mothers in high-income countries could provide informative insight in the effectiveness of mHealth interventions to improve health outcomes [13].

The strength of the current systematic review is the comprehensive search conducted including available grey literature reflecting current activities of nongovernmental organizations, which are often not published in peer-reviewed journals. This paper thus provides a comprehensive overview of the available literature on the effectiveness of mHealth

interventions to date and narratively assesses the broad function of mHealth. The methodology used in the narrative synthesis looks at the broad function of mHealth as used in the study, the targeted frontline providers, and the effectiveness of the mHealth intervention, an approach that facilitated easy assessment of the usefulness of the various mHealth functions.

Conclusions

This systematic review indicates that mHealth interventions targeting health care workers have the potential to materially improve maternal and neonatal health services in LMICs. There is, however, a gap in the knowledge of how mHealth interventions directly affect maternal and neonatal outcomes and future research should employ experimental designs to address this gap.

Acknowledgments

MAC gratefully acknowledge technical support from the Julius Center for Health Sciences and Primary Care. The authors also thank the Netherlands Organization for Scientific Research (NWO) Global Health Policy and 396 Health Systems Research Program, Netherlands, for providing funds for the conduct of the study (Grant number: 07.45.102.00) and Verena Schier for the use of her adapted Cochrane Bias Tools for assessing risk of bias.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of organizations contacted for grey literature.

[PDF File (Adobe PDF File), 35KB - jmir_v18i8e226_app1.pdf]

Multimedia Appendix 2

Full search strategy.

[PDF File (Adobe PDF File), 63KB - jmir_v18i8e226_app2.pdf]

Multimedia Appendix 3

Adapted quality assessment tool.

[PDF File (Adobe PDF File), 57KB - jmir_v18i8e226_app3.pdf]

Multimedia Appendix 4

Table showing Risk of Bias assessment for included intervention studies.

[PDF File (Adobe PDF File), 55KB - jmir_v18i8e226_app4.pdf]

References

1. World Health Organization. 2015 Nov. Maternal mortality: Facts sheets URL: http://www.who.int/mediacentre/factsheets/fs348/en/ [accessed 2016-08-01] [WebCite Cache ID 6jR2ADhI3]
2. van Lonkhuijzen L, Stekelenburg J, van Roosmalen J. Maternity waiting facilities for improving maternal and neonatal outcome in low-resource countries. Cochrane Database Syst Rev 2009(3):CD006759. [doi: 10.1002/14651858.CD006759.pub2] [Medline: 19588403]
3. Countdown Coverage Writing Group, Countdown to 2015 Core Group, Bryce J, Daelmans B, Dwivedi A, Fauveau V, et al. Countdown to 2015 for maternal, newborn, and child survival: the 2008 report on tracking coverage of interventions. Lancet 2008 Apr 12;371(9620):1247-1258. [doi: 10.1016/S0140-6736(08)60559-0] [Medline: 18406859]
4. Lawn JE, Kerber K, Enweronu-Laryea C, Cousens S. 3.6 million neonatal deaths--what is progressing and what is not? Semin Perinatol 2010 Dec;34(6):371-386. [doi: 10.1053/j.semperi.2010.09.011] [Medline: 21094412]
5. Chen L, Evans T, Anand S, Boufford JI, Brown H, Chowdhury M, et al. Human resources for health: overcoming the crisis. Lancet 2004;364(9449):1984-1990. [doi: 10.1016/S0140-6736(04)17482-5] [Medline: 15567015]
6. Sanou B. e-agriculture. The World in 2013: ICT Facts and Figures URL: http://www.e-agriculture.org/news/itu-world-2013-ict-facts-and-figures [accessed 2016-08-01] [WebCite Cache ID 6jR4XTMTq]
7. Qiang C, Yamamichi M, Hausman V, Altman D. World Bank. 2011 Dec. Mobile Applications for the Health Sector URL: http://siteresources.worldbank.org/INFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/mHealth_report.pdf [accessed 2016-08-01] [WebCite Cache ID 6jR8lLHqG]

8. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
9. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251 [FREE Full text] [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
10. Kouris I, Mougiakakou S, Scarnato L, Iliopoulou D, Diem P, Vazeou A, et al. Mobile phone technologies and advanced data analysis towards the enhancement of diabetes self-management. Int J Electron Healthc 2010;5(4):386-402. [doi: 10.1504/IJEH.2010.036209] [Medline: 21041177]
11. Center for Global Health and Economic Development Earth Institute, Columbia University. mHealth Knowledge. 2010 May. Barriers and Gaps Affecting mHealth in Low and Middle Income Countries: Policy White Paper URL: http://www.mhealthknowledge.org/sites/default/files/28_barriers_and_gaps.pdf [accessed 2016-08-02] [WebCite Cache ID 6jSNJ6sdb]
12. World Health Organization. 2010. Telemedicine: Opportunities and Developments in Member States: Report on the second Global Survey on eHealth URL: http://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2016-08-02] [WebCite Cache ID 6jSNnjhm4]
13. Chi BH, Stringer JS. Mobile phones to improve HIV treatment adherence. Lancet 2010 Nov 27;376(9755):1807-1808. [doi: 10.1016/S0140-6736(10)62046-6] [Medline: 21071073]
14. Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010 Jul;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
15. Kunutsor S, Walley J, Katabira E, Muchuro S, Balidawa H, Namagala E, et al. Using mobile phones to improve clinic attendance amongst an antiretroviral treatment cohort in rural Uganda: a cross-sectional and prospective study. AIDS Behav 2010 Dec;14(6):1347-1352. [doi: 10.1007/s10461-010-9780-2] [Medline: 20700644]
16. Lester R, Ritvo P, Mills E, Kariri A, Karanja S, Chung M, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. The Lancet 2010 Nov 27;376(9755):1838-1845.
17. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 2009 Jul 21;339:b2700.
18. The World Bank. 2015. World Bank Country and Lending Groups URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 [accessed 2016-07-28] [WebCite Cache ID 6jLbvRtcM]
19. Global Observatory for eHealth, World Health Organization. 2011 Jun. mHealth - New horizons for health through mobile technologies. Global Observatory for eHealth Series - Volume 3 URL: http://www.who.int/goe/publications/ehealth_series_vol3/en/ [accessed 2016-08-02] [WebCite Cache ID 6jSPCveYT]
20. Higgins J, Green S. The Cochrane Collaboration. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] URL: http://training.cochrane.org/handbook [accessed 2016-08-02] [WebCite Cache ID 6jSPUgy9B]
21. Cochrane Childhood Cancer. Non-randomised controlled study (NRS) designs URL: http://childhoodcancer.cochrane.org/non-randomised-controlled-study-nrs-designs [accessed 2016-08-02] [WebCite Cache ID 6jSPfd7rt]
22. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. J Med Internet Res 2012;14(5):e119 [FREE Full text] [doi: 10.2196/jmir.2183] [Medline: 22989894]
23. Andreatta P, Gans-Larty F, Debpuur D, Ofosu A, Perosky J. Evaluation of simulation-based training on the ability of birth attendants to correctly perform bimanual compression as obstetric first aid. Int J Nurs Stud 2011 Oct;48(10):1275-1280. [doi: 10.1016/j.ijnurstu.2011.03.001] [Medline: 21450290]
24. Kaewkungwal J, Singhasivanon P, Khamsiriwatchara A, Sawang S, Meankaew P, Wechsart A. Application of smart phone in “Better Border Healthcare Program”: a module for mother and child care. BMC Med Inform Decis Mak 2010;10:69 [FREE Full text] [doi: 10.1186/1472-6947-10-69] [Medline: 21047412]
25. Lemay NV, Sullivan T, Jumbe B, Perry CP. Reaching remote health workers in Malawi: baseline assessment of a pilot mHealth intervention. J Health Commun 2012;17 Suppl 1:105-117. [doi: 10.1080/10810730.2011.649106] [Medline: 22548604]
26. Little A, Medhanyie A, Yebyo H, Spigt M, Dinant GJ, Blanco R. Meeting Community Health Worker Needs for Maternal Health Care Service Delivery Using Appropriate Mobile Technologies in Ethiopia. PLOS ONE 2013 Oct 29;8(10):e77563.
27. Lori J, Munro ML, Boyd CJ, Andreatta P. Cell phones to collect pregnancy data from remote areas in Liberia. J Nurs Scholarsh 2012 Sep;44(3):294-301 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01451.x] [Medline: 22672157]
28. Munro ML, Lori JR, Boyd CJ, Andreatta P. Knowledge and skill retention of a mobile phone data collection protocol in rural Liberia. J Midwifery Womens Health 2014;59(2):176-183 [FREE Full text] [doi: 10.1111/jmwh.12155] [Medline: 24655593]

29. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR, et al. Designing and Implementing an Innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. Pan Afr Med J 2012;13:31 [FREE Full text] [Medline: 23330022]
30. Pathfinder International. 2014. Assessment of the Quality of Antenatal Care Services Provided by Health Workers using a Mobile Phone Decision Support Application in Northern Nigeria: A Pre/Post-intervention Study URL: http://www.pathfinder.org/publications-tools/antenatal-care-services-northern-nigeria.html [accessed 2016-08-02] [WebCite Cache ID 6jSQlNVxD]
31. Seidenberg P, Nicholson S, Schaefer M, Semrau K, Bweupe M, Masese N, et al. Early infant diagnosis of HIV infection in Zambia through mobile phone texting of blood test results. Bulletin of the World Health Organization 2012;90:348-356.
32. Alam M, Khanam T, Khan R. Assessing the scope for use of mobile based solution to improve maternal and child health in Bangladesh: A case study. 2010 Dec Presented at: 4th ACM/IEEE International Conference on Information and Communication Technologies and Development; December 13-16, 2010; London, UK.
33. D-Tree International. Safer Deliveries in Zanzibar URL: http://www.d-tree.org/saving-lives/womens-lives/safer-deliveries/ [accessed 2016-07-28] [WebCite Cache ID 6jLfG0zzV]
34. Eugelink H. IICD. 2013. Combating Mother and Child Malaria Mortality with Mobiles URL: http://www.iicd.org/articles/combating-mother-and-child-malaria-mortality-with-mobiles [accessed 2016-08-03] [WebCite Cache ID 6jU34x964]
35. Lee S, Chib A, Kim JN. Midwives' cell phone use and health knowledge in rural communities. J Health Commun 2011 Oct;16(9):1006-1023. [doi: 10.1080/10810730.2011.571344] [Medline: 21728781]
36. Pimmer C, Brysiewicz P, Linxen S, Walters F, Chipps J, Gröhbiel U. Informal mobile learning in nurse education and practice in remote areas--a case study from rural South Africa. Nurse Educ Today 2014 Nov;34(11):1398-1404. [doi: 10.1016/j.nedt.2014.03.013] [Medline: 24745478]
37. Tsai A, Tomlinson M, Dewing S, le Roux IM, Harwood J, Chopra M, et al. Antenatal depression case finding by community health workers in South Africa: feasibility of a mobile phone application. Arch Womens Ment Health 2014 Oct;17(5):423-431 [FREE Full text] [doi: 10.1007/s00737-014-0426-7] [Medline: 24682529]
38. van Heerden A, Norris S, Tollman S, Richter L, Rotheram-Borus MJ. Collecting maternal health information from HIV-positive pregnant women using mobile phone-assisted face-to-face interviews in Southern Africa. J Med Internet Res 2013;15(6):e116 [FREE Full text] [doi: 10.2196/jmir.2207] [Medline: 23748182]
39. World Vision International. Afghanistan mHealth URL: http://www.wvi.org/health/afghanistan-mhealth [accessed 2016-08-02] [WebCite Cache ID 6jSSPPL5Z]
40. Woods D, Attwell A, Ross K, Theron G. Text messages as a learning tool for midwives. S Afr Med J 2012 Feb;102(2):100-101. [Medline: 22310443]
41. Labrique A, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening toolscommon applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171.
42. Jo Y, Labrique A, Lefevre A, Mehl G, Pfaff T, Walker N, et al. Using the lives saved tool (LiST) to model mHealth impact on neonatal survival in resource-limited settings. PLoS One 2014;9(7):e102224. [doi: 10.1371/journal.pone.0102224] [Medline: 25014008]
43. Lund S, Rasch V, Hemed M, Boas I, Said A, Said K, et al. Mobile phone intervention reduces perinatal mortality in zanzibar: secondary outcomes of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2014;2(1):e15. [doi: 10.2196/mhealth.2941] [Medline: 25098184]
44. Brettlecker G, Cáceres C, Fernández A, Fröhlich N, Kinnunen A, Ossowski S, et al. Technology in Healthcare. In: Schumacher M, Helin H, Schuldt H, editors. CASCOM: Intelligent Service Coordination in the Semantic Web. Basel: Birkhäuser; 2008:125-139.
45. Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun 2012;17 Suppl 1:73-81. [doi: 10.1080/10810730.2011.649158] [Medline: 22548602]

Abbreviations

CHW: community health worker
HIV: human immunodeficiency virus
LMIC: low- and middle-income country
RCT: randomized controlled trial
SMS: short messaging service
TBA: traditional birth attendant

Edited by G Eysenbach; submitted 19.01.16; peer-reviewed by M Larsen, I Ibiebele, J Watterson, C Hawke; comments to author 03.03.16; revised version received 27.04.16; accepted 04.06.16; published 19.08.16.

Please cite as:
Amoakoh-Coleman M, Borgstein ABJ, Sondaal SFV, Grobbee DE, Miltenburg AS, Verwijs M, Ansah EK, Browne JL, Klipstein-Grobusch K
Effectiveness of mHealth Interventions Targeting Health Care Workers to Improve Pregnancy Outcomes in Low- and Middle-Income Countries: A Systematic Review
J Med Internet Res 2016;18(8):e226
URL: http://www.jmir.org/2016/8/e226/
doi:10.2196/jmir.5533
PMID:27543152

Original Paper

Use and Appreciation of a Tailored Self-Management eHealth Intervention for Early Cancer Survivors: Process Evaluation of a Randomized Controlled Trial

Iris Maria Kanera[1], MSc; Roy A Willems[1], MSc; Catherine A W Bolman[1], MSc, PhD; Ilse Mesters[2], MSc, PhD; Victor Zambon[3], MD; Brigitte CM Gijsen[4], MSc; Lilian Lechner[1], MSc, PhD

[1]Faculty of Psychology and Educational Sciences, Open University of the Netherlands, Heerlen, Netherlands

[2]Research Institute CAPHRI, Department of Epidemiology, Optimizing Patient Care, Maastricht University, Maastricht, Netherlands

[3]Department of Urology, Zuyderland Medical Center, Heerlen and Sittard-Geleen, Netherlands

[4]Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, Netherlands

Corresponding Author:
Iris Maria Kanera, MSc
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
Valkenburgerweg 177
Heerlen, 6401 DL
Netherlands
Phone: 31 455762448
Email: iris.kanera@ou.nl

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2016/9/e242/

Abstract

Background: A fully automated computer-tailored Web-based self-management intervention, Kanker Nazorg Wijzer (KNW [Cancer Aftercare Guide]), was developed to support early cancer survivors to adequately cope with psychosocial complaints and to promote a healthy lifestyle. The KNW self-management training modules target the following topics: return to work, fatigue, anxiety and depression, relationships, physical activity, diet, and smoking cessation. Participants were guided to relevant modules by personalized module referral advice that was based on participants' current complaints and identified needs.

Objective: The aim of this study was to evaluate the adherence to the module referral advice, examine the KNW module use and its predictors, and describe the appreciation of the KNW and its predictors. Additionally, we explored predictors of personal relevance.

Methods: This process evaluation was conducted as part of a randomized controlled trial. Early cancer survivors with various types of cancer were recruited from 21 Dutch hospitals. Data from online self-report questionnaires and logging data were analyzed from participants allocated to the intervention condition. Chi-square tests were applied to assess the adherence to the module referral advice, negative binominal regression analysis was used to identify predictors of module use, multiple linear regression analysis was applied to identify predictors of the appreciation, and ordered logistic regression analysis was conducted to explore possible predictors of perceived personal relevance.

Results: From the respondents (N=231; mean age 55.6, SD 11.5; 79.2% female [183/231]), 98.3% (227/231) were referred to one or more KNW modules (mean 2.9, SD 1.5), and 85.7% (198/231) of participants visited at least one module (mean 2.1, SD 1.6). Significant positive associations were found between the referral to specific modules (range 1-7) and the use of corresponding modules. The likelihoods of visiting modules were higher when respondents were referred to those modules by the module referral advice. Predictors of visiting a higher number of modules were a higher number of referrals by the module referral advice (β=.136, *P*=.009), and having a partner was significantly related with a lower number of modules used (β=-.256, *P*=.044). Overall appreciation was high (mean 7.5, SD 1.2; scale 1-10) and was significantly predicted by a higher perceived personal relevance (β=.623, *P*=.000). None of the demographic and cancer-related characteristics significantly predicted the perceived personal relevance.

Conclusions: The KNW in general and more specifically the KNW modules were well used and highly appreciated by early cancer survivors. Indications were found that the module referral advice might be a meaningful intervention component to guide the users in following a preferred selection of modules. These results indicate that the fully automated Web-based KNW provides personal relevant and valuable information and support for early cancer survivors. Therefore, this intervention can complement usual cancer aftercare and may serve as a first step in a stepped-care approach.

Trial Registration: Nederlands Trial Register: NTR3375; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3375 (Archived by WebCite at http://www.webcitation.org/6jo4jO7kb)

(J Med Internet Res 2016;18(8):e229) doi:10.2196/jmir.5975

KEYWORDS

eHealth; web-based intervention; computer tailoring; cancer survivorship; intervention usage; appreciation; multiple behavior intervention; process evaluation; self-management

Introduction

Recovery from cancer and its treatment can be challenging for cancer survivors. A variety of physical, psychosocial, and lifestyle difficulties might impede the resumption of previous daily life activities [1]. Cancer aftercare guidelines for oncology professionals recommend paying attention to the early detection and recognition of psychological distress, fatigue, pain, problems with daily activities, lifestyle risks, and also to stimulating self-care within the first year after completing the primary curative cancer treatment [2,3]. Further, due to the aging population and improved cancer care, the population of cancer survivors is growing while complaints, needs, and preferences of cancer survivors can vary individually over the different subjects and time [4-7]. For these reasons, fully automated, computer-tailored Web-based cancer aftercare interventions may be suitable for providing a large number of cancer survivors with personalized advice at relatively low costs [8]. Moreover, online solutions fit well with the increasing numbers of cancer survivors who search the Internet for health-related information, especially with those survivors who do not seek face-to-face guidance or treatment [9,10]. Web-based interventions might be appropriate to be integrated as a first step in a stepped-care approach as it offers a low-intensive intervention first before referring to interventions that are more intensive. Such first-step, low-intensive interventions might be sufficient to meet the personal needs of a large proportion of survivors with relatively mild complaints and are less costly [11]. In addition, Web-based interventions can comprise relevant information as written text, videos, animations, interactive features, hyperlinks, while personalization of the content is possible by applying computer tailoring [12-14].

The Web-based intervention *Kanker Nazorg Wijzer* (Cancer Aftercare Guide, KNW) is a fully automated intervention that aims to increase survivors' quality of life (QoL) by providing psychosocial support as well as promoting positive lifestyle changes, and it targets cancer survivors of any type of cancer [15]. The KNW consists of seven self-management training modules covering the topics return to work, fatigue, anxiety and depression, social relationship and intimacy issues, physical activity, diet, and smoking cessation (see Figure 1), supplemented with one general information module on residual symptoms. Based on the responses to a screening questionnaire, cancer survivors receive personalized advice on which KNW modules are most relevant for them to use. This Module Referral Advice (MRA) is designed in a fashion analogous to traffic lights as displayed in Figure 2. This MRA aims to guide participants through the wide-ranging KNW portal, based on experienced complaints and identified needs, as assessed by the screening questionnaire. The KNW has been shown to be effective in reducing fatigue and depressive symptoms and in improving quality of life domains (ie, emotional and social functioning) [16]. In addition, strong indications were found that KNW users are engaged in more moderate physical activity and have a higher intake of vegetables, fruits, and fish 6 months after they started using the KNW [17]. Besides assessing the effects of the KNW, it is important to understand how this complex intervention was used and appreciated by the participants, whether use and appreciation was predicted by certain user characteristics, and to evaluate relevant key intervention components [8,18-20]. Moreover, it is essential to examine specifically whether the provided information was perceived as personally relevant in order to evaluate the computer tailoring.

Previously published Web-based interventions in the areas of lifestyle, mental health, and chronic conditions differ with regard to the number of (cancer-related) topics, the composition of the target group, the intervention components, and the delivery mode [8,21-25]. Generally, typical Web-based interventions are modular in set-up, are updated weekly, require weekly visits, last for about 10 weeks, and include interaction with the system, peers, or a counselor [26]. The actual use of most interventions was low, or data on the use have been poorly reported [8,26]. The extent of use might be influenced by differences in participant and intervention characteristics [27]. Prior studies among cancer survivors have shown that different user characteristics were related to different user patterns: for example, a higher usage was found among those with low levels of self-reported social support and a high illness burden, and among survivors who were working and who received radiotherapy [28,29]. Being female, middle aged or older, having mid to high levels of education, a healthy body mass index (BMI), a healthier lifestyle, and having a low quality of life were predictors for a higher use of (multiple behavior) eHealth interventions among the general population [30,31]. Reported intervention characteristics that might predict usage were peer or counselor support, in-person contact, updates of the intervention, and sending reminders [20,26,27]. According to

previously published studies, mixed results were found on the relationship between intervention usage and outcomes, such as symptom distress, depression, and lifestyle behaviors [29,32,33]. With regard to appreciation, prior studies reported that Web-based interventions were positively evaluated by cancer survivors, and a higher use was associated with a higher appreciation in a generic Web-based intervention for breast cancer survivors [24,34,35].

The design of the KNW portal differs from most of the existing Web-based interventions for cancer survivors by providing personalized self-management training on seven topics and by allowing users to choose which modules they want to use during an intervention period of 6 months. Previously identified effective intervention characteristics of Web-based lifestyle interventions were tailored feedback, the use of theory, interactivity, goal setting, and online or in-person contact [8,26]. The KNW comprises all these elements, except for in-person contact. However, the MRA provides automated personalized guidance through the KNW modules. Given the large scope and the varied target group of the KNW portal, it is important to assess how the intervention was used, appreciated, whether the content was sufficiently tailored to be perceived as personal relevant, and what possible factors, including personal relevance, might predict the module use and its appreciation. In addition, the MRA might be a meaningful intervention component; therefore, the association between the MRA and the KNW module use also needs to be evaluated.

The main objective of this study is threefold: (1) to describe the use of the KNW modules and to identify predictors of a higher number of modules used, (2) to investigate the adherence to the provided MRA, and (3) to describe the appreciation of the KNW and its predictors. Additionally, to explore how well the tailoring worked and whether the perceived personal relevance might be different among subgroups, we explored possible predictors of personal relevance.

Figure 1. Overview of the scope and sequence of the modules. From Willems et al (2015). Used with permission.

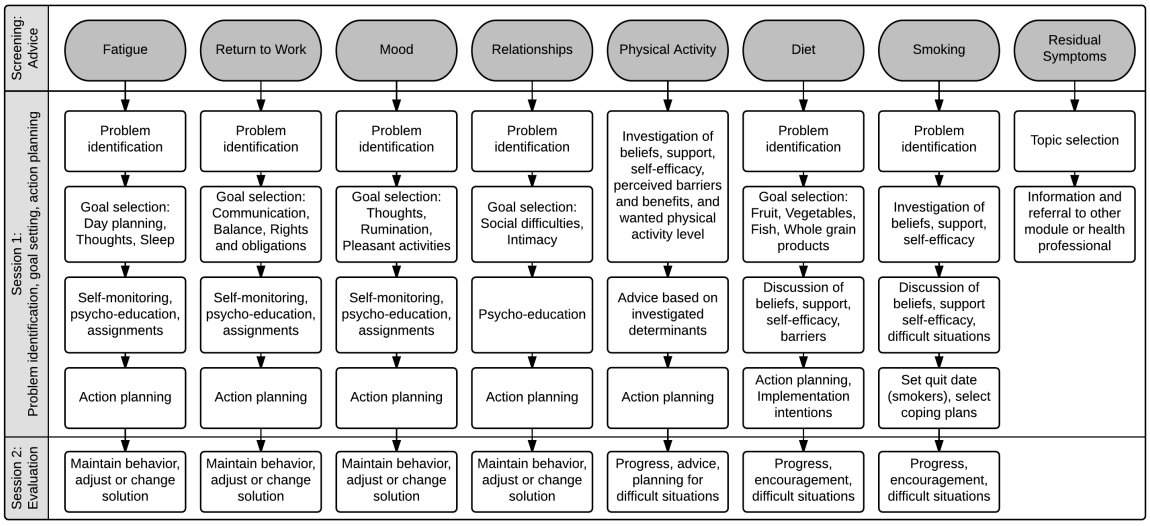
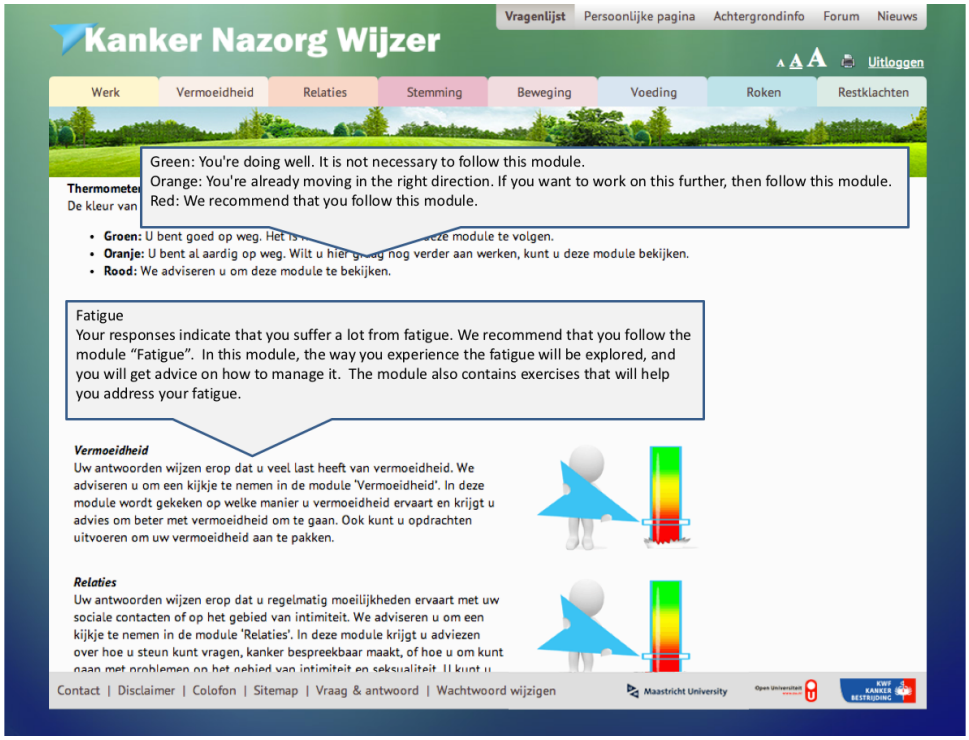


Figure 2. Module Referral Advice that encourages participants to follow relevant KNW modules. Adapted from Willems et al (2015). Used with permission.



Methods

This process evaluation was conducted as part of a two-armed randomized controlled trial (RCT) that evaluates the effects of the KNW portal. For the purpose of this report, all respondents of the intervention condition were included in the analyses. The details of the trial design, sample size calculation, participant eligibility, recruitment procedures, and the intervention have been published elsewhere [15-17]. Ethical approval for this trial (Dutch Trial Register NTR3375) was obtained from the Medical Research Ethics Committee, METC Z (NL41445.096.12, 12-T-115). All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments of comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Specific Intervention Elements: Module Referral Advice and Module Principles

A comprehensive description of the intervention, including the eight KNW modules, the underlying theoretical frameworks, and technical features are published in detail elsewhere [15,17]. This section describes the details of the MRA that was based on personal scores from the baseline questionnaire and that can refer to the seven self-management modules of the KNW (see Figure 2). The classification criteria for green, orange, and red MRA are summarized in Table 1 [36-45]. A green MRA signifies that the respondent reported no complaints, or minor complaints or needs, concerning the specific topic. Therefore, following the correspondent module is not a high priority. An orange MRA was provided when the respondents reported elevated but not severe complaints, or when respondents partially adhered to the lifestyle recommendations of the World Cancer Research Fund/American Institute for Cancer Research and the American Cancer Society [46,47]. The orange advice praises respondents' reasonably positive scores; however, it is recommended that they follow the corresponding module for further improvement. This orange category includes a wide coverage of score ranges, allowing for participants with higher, but not severe scores to still receive some positive and encouraging feedback and not lose their motivation to follow a module due to feedback that might be perceived as too stringent. A red MRA was provided only when severe psychosocial complaints, problematic functioning, or low/no adherence to lifestyle recommendations was reported, thus indicating that the respondent might be in high need of support concerning the specific topic. In that case, it was strongly recommended to follow the corresponding module. More detailed information on the underlying measures and cut-off points is included in Multimedia Appendix 1.

Table 1. Classification of the green, orange, and red MRA.

	Measurements and classification criteria[a]	MRA categories		
		Green	Orange	Red
Fatigue	CIS, subscale subjective fatigue (1-56) [36]	<27	27-35	>35
Return to work	Extended CaSUN [37,38]: Needs to adjust/ find a job (0-5); Needs to receive financial support (0-5); Needs support up on returning to work (0-5); Needs legal information (0-5)	No needs	Score on needs 3-12	Score on needs ≥13
Mood	HADS-A (0-21); HADS-D (0-21) [39]; MAC: dimension negative adjustment to cancer (16-64) [40]	HADS-A<8 and HADS-D<8 and MAC ≤36	HADS-A < 8 and HADS-D 8-15 and/ or MAC > 36; HADS-A 8-15 and HADS-D <8 or 8-15	HADS-A < 8 or 8-15 and HADS-D >15; HADS-D < 8 or 8-15 and HADS-A >15
Relationships	SSL-D (6-24) [41]/ CaSUN (2 items[b])	SSL-D ≤7	SSL-D=8 or 9 & needs CaSUN	SSL-D ≥10 & needs CaSUN
Physical activity	SQUASH [42,43]: Weekly ≥150 min moderate to vigorous PA; Daily ≥30 min of moderate PA on ≥5 days p/w	Meeting both conditions	Meeting 1 out of 2 conditions	Meeting no conditions
Diet	Dutch Standard Questionnaire on Food Consumption [44]: Daily ≥200g vegetables; Daily ≥2 pieces of fruit; Weekly ≥2 servings of fish; Daily ≥15g whole grains[c]; Daily ≥4 servings of potatoes/ whole-grain rice/ whole-grain pasta	Meeting at least 4 out of 5 conditions	Meeting 2 or 3 out of 5 conditions	Meeting 1 or 0 out of five conditions
Smoking	Smoking, not smoking, time point of quitting [45]	Never/formersmokers, quit prior to cancer diagnosis	Quit smoking after cancer diagnosis	Current smokers

[a]CIS: Checklist Individual Strength; PA: physical activity; HADS: Hospital Anxiety and Depression Scale, HADS-A: subscale anxiety, HADS-D: subscale depression; MAC: Mental Adjustment to Cancer Scale; SSL-D: Social Support List discrepancy subscale; SQUASH: Short Questionnaire to Assess Health Enhancing Physical Activity

[b]Needs related to sexuality and fertility.

[c]Whole-grain bread, oatmeal, cereals.

Throughout the different KNW intervention modules, principles of problem-solving therapy, cognitive behavioral therapy, social cognitive theories, and self-regulation theories were applied [48-51]. According to the I-Change Model [50], awareness factors such as knowledge, cues to action, and risk perception might be important determinants in the dynamic process of behavior change by influencing motivation and intention. By applying the MRA, participants were made aware of their current psychosocial status and lifestyle behaviors in relation to the norms and guidelines, with the aim of guiding the participants toward the appropriate self-management modules. When using the modules, self-management skills training was provided by encouraging respondents to observe their current behavior more in detail, choose themes to work on, set goals, and to prepare action and coping plans, followed by monitoring their experiences and possible progress in the changed strategies and behaviors. Within the modules, the information and support was tailored to the current emotional status, lifestyle behavior, and motivational determinants (attitude, self-efficacy, intention) by the application of computer tailoring. Furthermore, the feedback was tailored to personal characteristics (gender, age, marital status, children, education level), and cancer-related and medical issues (type of cancer, comorbidities). Four weeks after completing (parts of) one module, the participants were invited to reflect on their behavioral change plans and experiences in a brief personalized evaluation session. They were also encouraged to continue applying the previously recommended self-management skills. Furthermore, valuable generic information about lifestyle and psychosocial issues was accessible when visiting the user forum and the monthly news items.

Measurements

All data were derived from online self-report questionnaires and logging details.

Module Use

Module use was assessed by using logging data. Actual use was dichotomized (yes/no) for each module separately (in total eight modules). Module use was categorized into “yes” when at least the first three pages of a module were used. These three pages comprised important key information after which participants followed personalized pathways through the modules. The individual pathways were based on the responses to the baseline questionnaire, own preferences and goals, and take into consideration that the amount of needed information and/or support can vary to initiate behavior change [33]. Additionally, by assessing login data (last day the separate modules were used), the number of weeks of module engagement was registered.

Appreciation

At 6-month follow-up, the overall rating of the KNW and separate ratings for each of the used module(s) were assessed on a scale ranging from 1 (very poor) to 10 (outstanding) (eg, “Overall, how do you rate the KNW? Select your rating (1-10)”;

"How do you rate module mood on a scale from 1 to 10"). Further, four separate items were measured to evaluate whether the provided information and support was understandable, useful, personally relevant, and recommendable to fellow patients, on a 5-point Likert-scale, ranging from 1 (low) to 5 (high). The perceived personal relevance ("Was the information from the Kanker Nazorg Wijzer of personal relevance for you?") was included in the analysis of this study to explore whether computer tailoring worked well within the KNW. These items correspond to items that were used in other studies to measure the appreciation of Web-based interventions [52-54].

Demographic and Cancer-Related Variables

Information about demographic and cancer-related characteristics was collected at baseline. Standard questions were used to measure age, gender, and marital status. Marital status was dichotomized into "with partner" (married, cohabiting partners) and "without partner" (single, divorced, widowed). Education level was categorized into "low" (lower vocational education, medium general secondary education), "medium" (secondary vocational education, higher general secondary education), and "high" (higher vocational education, university education). Employment status was dichotomized into "working" (self-employed, in paid employment) and "not working" (unemployed, retired, unable to work). Type of cancer was categorized into breast, colorectal, and other types of cancer (ie, bladder, esophageal, gynecologic, hematologic, kidney, liver, lung, prostate, stomach, testicular, and thyroid cancer). Type of treatment was categorized into surgery and chemotherapy and radiotherapy, surgery and chemotherapy, surgery and radiotherapy, and other types of treatment. Further, aftercare (yes/no) and comorbidities (yes/no) were measured, and height and weight were assessed to determine BMI. The time since completion of primary treatment in weeks was based on registry data from the hospitals.

Statistical Analyses

The analyses were performed using STATA version 13.1. Descriptive statistics were used to describe demographic and cancer-related characteristics of the module (non-) users and the number of weeks of module engagement among all participants of the intervention condition at baseline. To calculate the appreciation outcomes, participants who completed the relevant questions at the 6-month measurement and who used the corresponding modules were included. Chi-square tests were used to determine the relationships between the MRA and the subsequent module use with a two-sided alpha=.05 level of significance. Negative binominal regression analysis was used to identify the predictors of a higher number of modules used (0-8), due to overdispersed count data. Independent variables (hypothesized predictors) were demographic variables (gender, age, marital status, education, employment), cancer-related variables (cancer type, type of treatment, number of weeks after completing primary cancer treatment, aftercare, comorbidities, BMI), the number of red and orange MRA, ranging from 0-7, and the perceived personal relevance, ranging from 1-5. To examine the predictors of a higher overall appreciation of the KNW, multiple linear regression analysis was applied among participants who completed the follow-up questionnaire after 6 months. The dependent variable was the overall rating of the KNW, measured at 6-month follow-up, ranging from 1-10. The same independent variables as described above were counted as predictors. Furthermore, the number of used modules (sum score 0-8) was added to the multiple linear regression model. To explore possible predictors of perceived personal relevance, ordered logistic regression analysis was conducted, taking into consideration that the dependent variable, perceived personal relevance, was an ordinal variable, ranging from 1-5. Within this analysis, all demographic and cancer-related characteristics were added as independent variables. Dummy coding was used for categorical variables including more than two categories and the continuous and ordinal variables were standardized in all conducted regression analyses. Since filling out all computer-based questions was required, and respondents were reminded automatically if a question was not answered, there were no missing data at baseline. Missing data at 6-month follow-up due to dropout were not imputed when calculating appreciation outcomes.

Results

Baseline characteristics of the intervention participants are displayed in Table 2. The majority of the participants was female (79.2%, 183/231), mean age was 55.6 (SD 11.5) years, and 70.1% (162/231) had been treated for breast cancer. A detailed overview of cancer diagnoses among the sample is shown in Multimedia Appendix 2. Mean time since completing primary cancer treatment was 25.1 (SD 13.5) weeks.

Module Use

The majority (80-100%) of the module users continued after reading the first three compulsory pages of the different modules. The numbers and percentages of participants who used the separate modules are displayed in Table 2. The diet module (134/231, 58.0%) was used most often, and the smoking module was used least often (23/231, 10.0%). However, from all the smokers at baseline (n=27), 13 (48%) individuals used the module Smoking. Overall, the participants used on average 2.1 (SD 1.6) KNW modules; 14.3% (33/231) used no modules, 30.3% (70/231) used one module, 18.2% (42/231) used two modules, 21.2% (49/231) used three modules, 8.7% (20/231) used four modules, 3.9% (9/231) used five modules, and 3.4% (8/231) individuals used six or more modules. Module engagement was highest during the first 16 weeks after getting KNW access: around 80% of the users used the modules within this period.

Provided Module Referral Advice

Table 3 displays how the red, orange, and green MRA ranged among the participants and how the modules were used. For fatigue, diet, and smoking, more red compared to orange MRA was provided, and for return to work, mood, relationships, and PA, more orange compared to red MRA was given. Green MRA was most frequently given with regard to smoking, return to work, mood, and relationships. Module use after getting a red or orange MRA was 58.8% and 38.6% for module fatigue, 55.6% and 52.4% for module return to work, 25% and 30.3% for module mood, and 25.9% and 27.3% for module relationships. Concerning the lifestyle modules, module use

after receiving a red or orange MRA for PA was 25% and 35%, for diet 50.4% and 68.7%, and for smoking 48.2% and 42.9%. From the 231 participants, 173 (74.9%) received at least one red MRA, and 192 (83.1%) received at least one orange MRA. On average, the participants were referred to 2.9 (SD 1.5) relevant modules (either red or orange MRA, not displayed).

Adherence to the Provided Module Referral Advice

The relations between the color of MRA (respectively red, orange, green) and module use are shown in Table 4. In general, the likelihood that participants actually used a relevant module was higher when the MRA was red or orange compared to green. When comparing module use after receiving a red MRA versus an orange MRA for the modules return to work, mood, relationships, PA, smoking, the differences were small, meaning that both colors led to comparable module participation. Participants used modules Fatigue (X^2=4.599, P=.032, OR 2.262) more often when a red MRA was provided compared to an orange MRA. The diet module (X^2=7.553, P=.006, OR .463) was used more often when an orange MRA was provided compared to a red MRA.

Table 2. Overall baseline characteristics of the KNW participants and categorized for module use (N=231).

	Overall (N=231)	No module (n=33, 14.3%)	KNW Modules							
			Fatigue (n=82, 35.5%)	Return to work (n=53, 22.9%)	Mood (n=49, 21.2%)	Relationships (n=38, 16.5%)	Physical activity (n=51, 22.1%)	Diet (n=134, 58%)	Smoking (n=23, 10%)	Residual symptoms (n=47, 20.4%)
Female, n (%)	183 (79.2)	26 (78.8)	63 (76.8)	46 (86.8)	41 (83.7)	30 (79.0)	44 (86.3)	106 (79.1)	17 (73.9)	40 (85.1)
Age, mean (SD)	55.6 (11.5)	52.5 (10.7)	55.1 (11.6)	52.8 (9.5)	54.4 (11.7)	55.9 (12.1)	56.3 (9.7)	56.0 (11.1)	51.6 (8.7)	56.2 (9.0)
With partner, n (%)	193 (83.6)	27 (81.8)	65 (79.3)	43 (81.1)	37 (75.5)	31 (81.6)	42 (82.4)	109 (81.3)	16 (69.6)	36 (76.6)
BMI, mean (SD)	26.0 (5.0)	27.2 (7.3)	26.2 (4.3)	25.7 (5.0)	25.3 (4.0)	26.1 (3.5)	26.1 (3.6)	25.4 (4.7)	24.8 (3.1)	25.4 (3.9)
Education, n (%)										
Low	76 (32.9)	13 (39.4)	23 (28.1)	12 (22.6)	15 (30.6)	12 (31.6)	18 (35.3)	42 (31.3)	9 (39.1)	13 (27.7)
Medium	76 (32.9)	12 (36.4)	31 (37.8)	20 (37.7)	20 (40.8)	13 (34.2)	18 (35.3)	44 (32.8)	7 (30.4)	14 (29.8)
High	79 (34.2)	8 (24.2)	28 (34.2)	21 (39.6)	14 (28.6)	13 (34.2)	15 (29.4)	48 (35.8)	7 (30.4)	20 (42.6)
Working at baseline, n (%)	122 (52.8)	20 (60.6)	40 (48.8)	38 (71.7)	28 (57.1)	18 (47.4)	27 (52.9)	70 (52.2)	13 (56.5)	26 (55.3)
Type of cancer, n (%)										
Breast	162 (70.1)	24 (72.7)	55 (67.1)	40 (75.5)	36 (73.5)	27 (71.1)	41 (80.4)	94 (70.2)	18 (78.3)	32 (68.1)
Colon	29 (12.6)	4 (12.1)	10 (12.2)	4 (7.6)	6 (12.2)	5 (13.2)	2 (3.9)	19 (14.2)	3 (13.0)	9 (19.2)
Other	40 (17.3)	5 (15.2)	17 (20.7)	9 (16.9)	7 (14.3)	6 (15.8)	8 (15.7)	21 (15.7)	2 (8.7)	6 (12.8)
Had cancer before, n (%)	24 (10.4)	5 (15.2)	8 (9.8)	3 (5.7)	4 (8.2)	3 (7.9)	5 (9.8)	13 (9.7)	2 (8.7)	5 (10.6)
Treatment, n (%)										
Surgery, chemo, radio	86 (37.2)	11 (33.3)	37 (45.1)	20 (37.7)	20 (40.8)	18 (47.4)	22 (43.1)	53 (39.6)	11 (47.8)	22 (46.8)
Surgery, chemo	61 (26.4)	11 (33.3)	17 (20.7)	16 (30.2)	16 (32.7)	9 (23.7)	12 (23.5)	35 (26.1)	7 (30.4)	15 (31.9)
Surgery, radio	46 (19.9)	5 (15.2)	15 (18.3)	11 (20.8)	10 (20.4)	5 (13.2)	11 (21.6)	26 (19.4)	3 (13.0)	8 (17.1)
Other	38 (16.5)	6 (18.2)	13 (15.9)	6 (11.3)	3 (6.1)	6 (15.8)	6 (11.8)	20 (14.9)	2 (8.7)	2 (4.3)
Weeks since completion treatment, mean (SD)	25.1 (13.5)	27.1 (15.6)	24.1 (14.4)	22.3 (13.7)	25.3 (13.6)	26.5 (12.9)	23.7 (13.6)	25.0 (13.1)	22.1 (13.2)	25.4 (3.9)
Having comorbidities, n (%)	62 (26.8)	10 (30.3)	25 (30.5)	14 (26.4)	12 (24.5)	10 (26.3)	15 (29.4)	34 (25.4)	7 (30.4)	8 (17.0)
Using aftercare, n (%)	145 (62.8)	25 (75.8)	46 (56.1)	38 (71.7)	32 (65.3)	29 (76.3)	31 (60.8)	83 (61.9)	12 (52.2)	29 (61.7)

Table 3. Provided MRA and subsequent module use.

Module	Red			Orange			Green		
		Followed module, %			Followed module, %			Followed module, %	
	%	yes	no	%	yes	no	%	yes	no
Fatigue	34.6	58.8	41.3	19.1	38.6	61.4	46.3	16.8	83.2
Return to work	3.9	55.6	44.4	18.2	52.4	47.6	77.9	14.4	85.6
Mood	1.7	25	75	28.6	30.3	69.7	69.7	17.4	82.6
Relationships	11.7	25.9	74.1	19.1	27.3	72.7	69.3	11.8	88.1
Physical activity	5.2	25	75	35.9	37.4	62.7	58.9	12.5	87.5
Diet	53.3	50.4	49.6	42.9	68.7	31.3	3.9	44.4	55.6
Smoking	11.7	48.2	51.9	3.1	42.9	57.1	85.3	3.6	96.5

Table 4. Relationship between the MRA and module use (chi-square tests; df=1).

Module (yes/no)	Red compared to orange			Red compared to green			Orange compared to green		
	X^2	*P*	Odds ratio (95% CI)	X^2	*P*	Odds ratio (95% CI)	X^2	*P*	Odds ratio (95% CI)
Fatigue	4.599	.032[a]	2.262 (.99-5.16)	35.485	.000[a]	7.042 (3.12-14.69)	8.332	.004[a]	3.113 (1.30-7.37)
Return to work	0.030	.863	1.136 (.21-6.56)	10.565	.001[a]	7.404 (1.46-39.25)	28.920	.000[a]	6.515 (2.92-14.47)
Mood	0.050	.822	.767 (.01-10.27)	0.156	.693	1.583 (.03-20.50)	4.680	.031[a]	2.065 (1.00-4.21)
Relationships	0.016	.901	.933 (.26-3.11)	3.810	.051	2.597 (.81-7.49)	6.349	.012[a]	2.783 (1.11-6.73)
Physical activity	0.696	.404	.186 (.00-1.48)	1.474	.225	2.333 (.37-10.57)	18.60	.000[a]	4.173 (2.02-8.74)
Diet	7.553	.006[a]	.463 (.26-.83)	0.119	.730	1.27 (.26-6.71)	2.182	.140	2.742 (.54-14.67)
Smoking	0.063	.803	1.238 (.17-10.06)	58.075	.000[a]	25.204 (7.67-85.09)	22.400	.000[a]	20.357 (2.40-141.94)

[a]Statistically significant result.

Appreciation

From the 231 participants who had access to the KNW intervention, 182 responded to the questions concerning appreciation after 6 months. The overall appreciation of the KNW was high (mean 7.5, SD 1.2) (Table 5). In general, the overall KNW was rated more positively among module users compared to non-module users. Ratings of the separate modules ranged from 6.4 (satisfactory) for the residual symptoms module to 8 (good) for smoking module. Personal relevance ranged from 2.9 to 3.5 (a little bit relevant to relevant). The ratings for comprehensibility, usefulness, and recommendation to other cancer survivors were all positive and very uniform (Table 5).

Table 5. Appreciation of KNW after 6 months.

	Overall	No module	Fatigue	Return to work	Mood	Relationships	PA	Diet	Smoking	Residual symptoms
Overall KNW (1-10), mean (SD)	7.5 (1.2)	7.1 (2.0)	7.6 (1.1)	7.6 (1.1)	7.4 (1.0)	7.4 (1.0)	7.6 (1.1)	7.5 (1.0)	7.8 (1.2)	7.4 (1.1)
Modules (1-10)[a], mean (SD)			7.3 (1.3)	7.0 (1.3)	7.5 (1.2)	7.2 (0.8)	7.7 (1.1)	7.6 (1.0)	8 (1.3)	6.4 (1.9)
Subquestions on content (1-5) [b], mean (SD)										
Understandable?	4.3 (0.6)	4.1 (1.0)	4.4 (0.5)	4.4 (0.5)	4.3 (0.5)	4.5 (0.5)	4.4 (0.5)	4.4 (0.5)	4.3 (0.5)	4.4 (0.5)
Useful?	3.7 (0.8)	3.7 (1.1)	3.8 (0.8)	3.7 (0.8)	3.7 (0.8)	3.7 (0.8)	3.7 (0.7)	3.7 (0.8)	3.8 (0.9)	3.4 (0.9)
Personal relevant?	3.2 (0.9)	2.9 (1.2)	3.4 (0.8)	3.3 (0.7)	3.2 (0.9)	3.4 (0.9)	3.5 (0.7)	3.2 (0.8)	3.3 (0.9)	3.3 (0.9)
Recommendable to fellow survivors?	3.9 (1.0)	3.6 (1.1)	3.9 (1.0)	3.9 (1.0)	3.8 (1.0)	3.7 (1.0)	4 (1.0)	3.9 (1.0)	4.1 (0.9)	3.8 (1.0)

[a]No module n=18, fatigue n=47, return to work n=27, mood n=13, relationships n=11, PA n=28, diet n=77, smoking n=6, residual symptoms n=14.

[b]No module n=18, fatigue n=67, return to work n=46, mood n=45, relationships n=34, PA n=45, diet n=115, smoking n=18, residual symptoms n=39.

Predictors of a Higher Number of Modules Used

Using a higher number of modules was predicted by a higher number of red/orange MRA (β=.136, *P*=.009), and by a higher perceived personal relevance (β=.150, *P*=.014). Moreover, having a partner was significantly related with a lower number of modules used (β=-.256, *P*=.044) (Multimedia Appendix 3).

Predictors of a Higher Appreciation of KNW Overall

A higher appreciation with the overall KNW was significantly predicted by a higher perceived personal relevance (β=.623, *P*=.000) (Multimedia Appendix 4). None of the demographic and cancer-related variables, or the number of red/orange MRA, or number of modules used predicted a higher overall appreciation of the KNW intervention.

Predictors of a Higher Perceived Personal Relevance

None of the demographic and cancer-related characteristics significantly predicted the perceived personal relevance of the KNW content, indicating that the KNW content was rated comparably personal relevant among individuals with different demographic and cancer-related characteristics (Multimedia Appendix 5).

Discussion

Principal Findings

This process evaluation of the Web-based KNW evaluated the automated guidance toward the KNW modules and subsequent module use, and the appreciation of this intervention. Despite the noncommittal nature of the KNW, more than 85% of the participants used one or more of the eight modules, and there was clear interest in all eight modules. This result confirms the need for wide-ranging support among early cancer survivors. Interestingly, automated referrals to specific modules were related to a higher number of modules used. Moreover, the complex KNW was highly appreciated and perceived as personal relevant by early cancer survivors.

The MRA aimed to guide the respondents toward the appropriate modules by giving feedback about current problem areas and needs. Cancer survivors might not have noticed some of these needs, and the MRA may have raised awareness about these topics. The importance of increasing awareness is theoretically grounded as described by Weinstein and Sandman [55] in their Precaution Adoption Process Model. That model includes a sequence of five stages within behavior change: “unaware of the issue,” “aware of the issue but not personally engaged,” “engaged and deciding what to do,” “planning to act but not yet having acted,” and “acting.” Prior research confirmed that a considerable number of colorectal cancer survivors were unaware of healthy diet recommendations, and older cancer survivors reported being less aware of the beneficial effects of a healthy lifestyle [56,57]. In addition, research revealed that cancer survivors might be less aware of available psychosocial support and solutions to psychosocial problems, while, for example, addressing maladaptive illness perceptions and adopting a more adaptive self-management may lead to better health outcomes [58,59]. Consequently, curiosity about available self-management support needs to be encouraged [8]. In accordance with the I-Change Model, the MRA could increase knowledge about the current level of well-being, psychosocial conditions, and lifestyle behavior. Besides that, the MRA could elevate the risk perception and may serve as a cue to action with regard to the relevant topics, given that the solutions to the problems are provided (relevant self-management module) [50]. These awareness/solution triggers might positively influence the motivation and intention to perform desired behavior, which is in line with the findings of Walthouwer et al [60], who identified awareness as an important moderator in the relationship between psychosocial determinants and specific dietary behavior (eating in moderation) in the general population. Results in our study illustrate that these

awareness/solution triggers are most likely to be followed when a red or orange MRA was provided. Thus, the MRA successfully referred those respondents with elevated as well as severe complaints and/or needs. However, this did not apply for fatigue because highly fatigued respondents (red MRA) were more likely to use the fatigue module compared to participants with less fatigue (orange MRA). Additionally, with regard to diet, results might indicate that especially those who were already engaged more in a healthy diet were more likely to use the diet module. Furthermore, the topic diet could be of general interest to the participants, while the topic fatigue might be most interesting for participants with specific complaints. Consequently, the MRA may be a meaningful intervention component to increase motivation, subsequent module use, and problem-solution, while MRA adherence might be related to the specific behavior. Using topic-specific KNW modules has shown to be effective in decreasing fatigue, depressive feelings, and was beneficial in increasing moderate physical activity and fruit and fish consumption [16,17].

Within the KNW, participants were referred on average to 2.9 modules, while on average 2.1 modules were used. The appreciation rates were high, and the results showed that a higher number of modules used did not contribute to a higher appreciation. However, a higher perceived personal relevance did contribute to a higher appreciation. This is in line with Wilson et al [61] reporting that a moderate number of recommendations in multiple behavior interventions might produce the highest level of change, while engagement with a higher number of recommendations might be too demanding. Within the KNW, respondents were allowed to make their own choices, despite the provided MRA. Prior research confirms that the possibility to choose within multiple behavior interventions may prevent high attrition rates and could improve intervention outcomes [31,32,62]. Offering wide ranging support in combination with personalized referral to relevant topics and the possibility to choose might prevent overload. Donkin et al [33] support this suggestion by reporting that a certain level of usage might be needed to obtain benefit from an online intervention for depression. However, after reaching a point of therapy saturation, little or no additional program gains might be expected. This is in line with a Web-based study among cancer survivors and with another Web-based obesity prevention study among the general population, which reported that more intervention use did not result in better intervention outcomes [28,63]. Using a higher number of modules may not be necessary for all users to benefit most from the KNW. Our results revealed that having no partner was related to the use of a higher number of modules, and participants who were in greater need of support (higher number of red/orange MRA) indeed used a higher number of modules. This is consistent with the findings of Borosund et al [28], who reported that, in particular, cancer survivors with low levels of social support and a high illness burden used self-management components of a Web-based illness management support system. Furthermore, higher perceived personal relevance was related to using a higher number of modules, which might be explained by receiving a higher amount of computer-tailored content within the modules. The overall KNW was highly appreciated with an average grade of 7.5, indicating an appreciation from very satisfactory to good. The low variability (SD 1.2) indicates a considerably unanimous positive rating 6 months after getting access to the KNW. Results from our study indicate that perceived personal relevance might be a key component to explain a higher appreciation. Computer tailoring was applied within the KNW in order to create personal relevant feedback. Since perceived personal relevance could not be predicted by demographic and cancer-related characteristics, we can conclude that the tailoring of information worked well. In comparison, the overall satisfaction of a generic fully automated Web-based self-management intervention for breast cancer survivors was mean 7 (SD 1.2) [24]. In addition, the overall appreciation of a Web-based weight management intervention for overweight adults was mean 6.6, and the overall appreciation of a Web-based text- and video-tailored intervention for smoking cessation in the general population was mean 6.45 (SD 1.62; scales ranged from 1-10) [53,54]. The overall appreciation ratings of KNW module users were more positive than the ratings of module non-users, although the module non-users were still quite positive in their ratings. In addition to the modules, the KNW has a user forum and participants received monthly emails inviting them to visit generic monthly news items. Filling out the screening questionnaire and follow-up questionnaires, combined with receiving personalized feedback on problem areas (by the MRA), as well as the additional KNW features, might already have raised awareness and provided other valuable information to achieve benefits among module non-users. Overall, the high appreciation rate indicates that the broad design and tailored information of the KNW seem to fit well with the needs of early cancer survivors (in which, breast cancer survivors were overrepresented).

Limitations

Some limitations need to be addressed. First, providing data on completion of the separate themes and specific activities within the modules, and on completion of the evaluation sessions was not possible due to the module design. This information might be interesting for future studies; therefore, we recommend future interventions to study in more detail participation of intervention modules. Second, within our study, it was not possible to compare the relationships between the MRA and module use to a control group not receiving the MRA. Consequently, these associations need to be interpreted with caution, as it is conceivable that without the MRA, some of the same modules would have been used. Future experimental research might explore the specific effects of a similar automated referral system on subsequent choices. Third, this eHealth intervention requires respondents to have computer skills and health literacy, such as competence at accessing, understanding, appraising, and applying the health information provided [64]. However, since eHealth literacy was not assessed in this study, it is not possible to estimate the extent to which this might have influenced initial recruitment and the use and appreciation of the KNW. Fourth, mainly middle-aged, female breast cancer survivors who scored fairly well on QoL and depression participated, which might be too selective a group to represent the general cancer survivor population. During recruitment, mainly breast cancer outpatient clinics participated. Five-year survival rates of breast cancer are relatively high [6]. Unless

mostly females with higher socioeconomic status are reached in Web-based interventions in general, interpretations of these findings should be viewed with caution [8].

Conclusion

The general KNW and the KNW modules were substantially used and highly appreciated by early cancer survivors, thus confirming the need for wide-ranging support among this target group. Results indicate that the MRA may be seen as a meaningful key component of the fully automated KNW intervention by guiding users to follow a preferred selection of modules, given their current complaints and identified needs. Moreover, the overall intervention and separate modules were highly appreciated, which could be explained by a higher perceived personal relevance. We can conclude that computer tailoring worked well and that the range of topics, design, and personalized information suited the needs of early cancer survivors. This process evaluation adds meaningful information on the use and appreciation of Web-based cancer aftercare interventions and confirms that the KNW offers valuable and appropriate support for early cancer survivors to complement usual cancer aftercare and may serve as a first step in a stepped-care approach.

Acknowledgments

Our sincere thanks go to the medical staff of the Dutch hospitals, who were involved in recruitment of participants: Beatrix Hospital Gorinchem, Bernhoven Hospital Uden, Bronovo Hospital The Hague, Catharina Hospital Eindhoven, Diakonessenhuis Hospital Utrecht, Elkerliek Hospital Helmond, Flevo Hospital Almere, IJsselland Hospital Capelle aan den Ijssel, Ikazia Hospital Rotterdam, Laurentius Hospital Roermond, Lievensberg Hospital Bergen op Zoom, Maasstad Hospital Rotterdam, Sint Anna Hospital Geldrop, Sint Jans Gasthuis Hospital Weert, Slingeland Hospital Doetinchem, Spaarne Hospital Hoofddorp, University Medical Center Groningen, University Medical Center Maastricht, VieCuri Medical Center Venlo, Langeland Hospital Zoetermeer, Zuyderland Medical Center Heerlen, and Sittard-Geleen.

Special thanks to Audrey Beaulen who provided helpful assistance in the recruitment of the subjects, and Linda Küsters for her contribution in the development of the lifestyle modules.

This research project was funded by the Dutch Cancer Society (grant number NOU2011-5151).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Determination of the Module Referral Advice categories (red, orange, green).

[PDF File (Adobe PDF File), 23KB - jmir_v18i8e229_app1.pdf]

Multimedia Appendix 2

Overview of cancer diagnoses among the KNW sample.

[PDF File (Adobe PDF File), 16KB - jmir_v18i8e229_app2.pdf]

Multimedia Appendix 3

Predictors of a higher number of followed KNW modules (N=182).

[PDF File (Adobe PDF File), 31KB - jmir_v18i8e229_app3.pdf]

Multimedia Appendix 4

Predictors of a higher appreciation of KNW (N=182).

[PDF File (Adobe PDF File), 31KB - jmir_v18i8e229_app4.pdf]

Multimedia Appendix 5

Predictors of a higher perceived personal relevance of KNW content (N=182).

[PDF File (Adobe PDF File), 24KB - jmir_v18i8e229_app5.pdf]

References

1. Bluethmann SM, Basen-Engquist K, Vernon SW, Cox M, Gabriel KP, Stansberry SA, et al. Grasping the 'teachable moment': time since diagnosis, symptom burden and health behaviors in breast, colorectal and prostate cancer survivors. Psychooncology 2015 Jun 8:Epub ahead of print. [doi: 10.1002/pon.3857] [Medline: 26060053]
2. Netherlands Comprehensive Cancer Organisation. Guideline cancer survivorship care URL: http://oncoline.nl/index.php?language=en [accessed 2016-05-15] [WebCite Cache ID 6hWEoQwvG]
3. Runowicz CD, Leach CR, Henry NL, Henry KS, Mackey HT, Cowens-Alvarado RL, et al. American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. J Clin Oncol 2016 Feb 20;34(6):611-635. [doi: 10.1200/JCO.2015.64.3809] [Medline: 26644543]
4. Willems RA, Bolman CAW, Mesters I, Kanera IM, Beaulen AJM, Lechner L. Cancer survivors in the first year after treatment: the prevalence and correlates of unmet needs in different domains. Psychooncology 2016 Jan;25(1):51-57. [doi: 10.1002/pon.3870] [Medline: 26110652]
5. Kanera IM, Bolman CAW, Mesters I, Willems RA, Beaulen AJM, Lechner L. Prevalence and correlates of healthy lifestyle behaviors among early cancer survivors. BMC Cancer 2016;16:4 [FREE Full text] [doi: 10.1186/s12885-015-2019-x] [Medline: 26732757]
6. The Netherlands Cancer Registry. Cijfers over kanker (statistics on cancer). 2016. URL: http://www.cijfersoverkanker.nl [accessed 2016-05-13] [WebCite Cache ID 6hTR56yD0]
7. Given CW, Given BA. Symptom management and psychosocial outcomes following cancer. Semin Oncol 2013 Dec;40(6):774-783. [doi: 10.1053/j.seminoncol.2013.09.001] [Medline: 24331196]
8. 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]
9. Chou WS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
10. Ritterband LM, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]
11. Krebber AM, Leemans CR, de Bree R, van Straten A, Smit F, Smit EF, et al. Stepped care targeting psychological distress in head and neck and lung cancer patients: a randomized clinical trial. BMC Cancer 2012;12:173 [FREE Full text] [doi: 10.1186/1471-2407-12-173] [Medline: 22574757]
12. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
13. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
14. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015]
15. Willems RA, Bolman CAW, Mesters I, Kanera IM, Beaulen AAJM, Lechner L. The Kanker Nazorg Wijzer (Cancer Aftercare Guide) protocol: the systematic development of a web-based computer tailored intervention providing psychosocial and lifestyle support for cancer survivors. BMC Cancer 2015;15:580 [FREE Full text] [doi: 10.1186/s12885-015-1588-z] [Medline: 26260318]
16. Willems RA, Bolman CAW, Mesters I, Kanera IM, Beaulen AAJM, Lechner L. Short-term effectiveness of a web-based tailored intervention for cancer survivors on quality of life, anxiety, depression, and fatigue: randomized controlled trial. Psychooncology 2016 Mar 14:-. [doi: 10.1002/pon.4113] [Medline: 26988800]
17. Kanera IM, Bolman CAW, Willems RA, Mesters I, Lechner L. Lifestyle-related effects of the web-based Kanker Nazorg Wijzer (Cancer Aftercare Guide) intervention for cancer survivors: a randomized controlled trial. J Cancer Surviv 2016 Mar 17:-. [doi: 10.1007/s11764-016-0535-6] [Medline: 26984534]
18. Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. Qual Saf Health Care 2003 Feb;12(1):40-46 [FREE Full text] [Medline: 12571344]
19. Steckler A, Linnan L. Process evaluation for public health interventions and research. San Francisco, CA: Jossey-Bass; 2002.
20. 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]
21. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
22. Kim AR, Park HA. Web-based Self-management Support Interventions for Cancer Survivors: A Systematic Review and Meta-analyses. Stud Health Technol Inform 2015;216:142-147. [Medline: 26262027]

23. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
24. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MF, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. J Med Internet Res 2013;15(8):e170 [FREE Full text] [doi: 10.2196/jmir.2566] [Medline: 23958584]
25. McAlpine H, Joubert L, Martin-Sanchez F, Merolli M, Drummond KJ. A systematic review of types and efficacy of online interventions for cancer patients. Patient Educ Couns 2015 Mar;98(3):283-295. [doi: 10.1016/j.pec.2014.11.002] [Medline: 25535016]
26. 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]
27. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
28. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
29. Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J. Exposure to a patient-centered, Web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]
30. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who Follows eHealth Interventions as Recommended? A Study of Participants' Personal Characteristics From the Experimental Arm of a Randomized Controlled Trial. J Med Internet Res 2015;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]
31. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]
32. Schulz DN, Schneider F, de VH, van Osch LA, van Nierop PW, Kremers SP. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
33. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
34. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an Internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012 Jul;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]
35. De CK, Charlier C, Van Hoof E, Pauwels E, Lechner L, Bourgois J, et al. Development and usability of a computer-tailored pedometer-based physical activity advice for breast cancer survivors. Eur J Cancer Care (Engl) 2015 Sep;24(5):673-682. [doi: 10.1111/ecc.12225] [Medline: 25132146]
36. Vercoulen JHHM, Swanink CMA, Fennis JFM, Galama JMD, van der Meer JWM, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994 Jul;38(5):383-392. [Medline: 7965927]
37. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Lo SK, et al. The development and evaluation of a measure to assess cancer survivors' unmet supportive care needs: the CaSUN (Cancer Survivors' Unmet Needs measure). Psychooncology 2007 Sep;16(9):796-804. [doi: 10.1002/pon.1137] [Medline: 17177268]
38. Mesters I, Willems RA, Kanera IM, Bolman CAW, Keeman M, Lechner L. Psychometric Quality of the Extended Cancer information Needs Questionnaire (CaSUN-NL). In: Archives of Physical Medicine and Rehabilitation;. 2015 Presented at: American Congress of Rehabilitation Medicine; 25-30 October 2015; Dallas, TX p. e65.
39. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
40. Watson M, Homewood J. Mental Adjustment to Cancer Scale: psychometric properties in a large cancer cohort. Psychooncology 2008 Nov;17(11):1146-1151. [doi: 10.1002/pon.1345] [Medline: 18626853]
41. van Sonderen E. Het meten van sociale steun met de Sociale Steun Lijst Interacties (SSL-I) en Sociale Steun Lijst Discrepanties (SSL-D). Een handleiding. Groningen, The Netherlands: Research Institute SHARE, Rijksuniversiteit Groningen URL: https://www.umcg.nl/SiteCollectionDocuments/research/institutes/SHARE/assessment%20tools/handleiding_ssl_2e_druk.pdf [accessed 2016-05-15] [WebCite Cache ID 6hWVyHHJ7]
42. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]

43. de Hollander EL, Zwart L, de Vries SI, Wendel-Vos W. The SQUASH was a more valid tool than the OBiN for categorizing adults according to the Dutch physical activity and the combined guideline. J Clin Epidemiol 2012 Jan;65(1):73-81. [doi: 10.1016/j.jclinepi.2011.05.005] [Medline: 21840174]
44. van den Brink CL, Ocke M, Houben A, van NP, Droomers M. Rijksinstituut voor Volksgezondheid en Milieu. 2016. Validering van standaardvraagstelling voeding voor Lokale en Nationale Monitor Volksgezondheid URL: http://www.rivm.nl/Documenten_en_publicaties/Wetenschappelijk/Rapporten/2005/augustus/Validering_van_standaardvraagstelling_voeding_voor_Lokale_en_Nationale_Monitor_Volksgezondheid [accessed 2016-05-15] [WebCite Cache ID 6hWX6IMJP]
45. Mudde A, Willemsen M, Kremers S, De Vries H. Meetinstrumenten voor onderzoek naar roken en stoppen met roken [Measurement Instruments used in Research with Regard to Smoking and Smoking Cessation]. Den Haag: Stivoro; 2006.
46. Rock CL, Doyle C, Demark-Wahnefried W, Meyerhardt J, Courneya KS, Schwartz AL, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274 [FREE Full text] [doi: 10.3322/caac.21142] [Medline: 22539238]
47. World Cancer Research Fund International. Food, Nutrition, and Physical Activity: a Global Perspective. Washington, DC; 2016. Policy and Action for Cancer Prevention URL: http://www.wcrf.org/int/research-we-fund/continuous-update-project-cup [accessed 2016-05-15] [WebCite Cache ID 6hWS3inGz]
48. D'Zurilla TJ, Nezu AM. Problem-solving therapy: a positive approach to clinical intervention. New York: Springer; 2007.
49. Bleijenberg G, Berends T, Voskamp H, Bazelmans E, Verhagen S. Cognitive behaviour therapy for post-cancer fatigue: a treatment protocol. In: Gielissen MFM, editor. Fatigue in cancer survivors: from assessment to cognitive behaviour therapy (Doctoral dissertation). Nijmegen: Radboud University; 2007.
50. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626 [FREE Full text] [Medline: 14572020]
51. Baumeister RF, Heatherton TF, Tice DM. Losing control: how and why people fail at self-regulation. San Diego, CA: Academic Press; 1994.
52. Walthouwer MJ, Oenema A, Lechner L, de Vries H. Comparing a Video and Text Version of a Web-Based Computer-Tailored Intervention for Obesity Prevention: A Randomized Controlled Trial. J Med Internet Res 2015;17(10):e236 [FREE Full text] [doi: 10.2196/jmir.4083] [Medline: 26481772]
53. van Genugten L, van Empelen P, Boon B, Borsboom G, Visscher T, Oenema A. Results from an online computer-tailored weight management intervention for overweight adults: randomized controlled trial. J Med Internet Res 2012;14(2):e44 [FREE Full text] [doi: 10.2196/jmir.1901] [Medline: 22417813]
54. Stanczyk N, Bolman CAW, van Adrichem M, Candel M, Muris J, de Vries H. Comparison of text and video computer-tailored interventions for smoking cessation: randomized controlled trial. J Med Internet Res 2014;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]
55. Weinstein ND, Sandman PM. A model of the precaution adoption process: evidence from home radon testing. Health Psychol 1992;11(3):170-180. [Medline: 1618171]
56. Hawkins NA, Berkowitz Z, Rodriguez JL. Awareness of Dietary and Alcohol Guidelines Among Colorectal Cancer Survivors. Am J Prev Med 2015 Dec;49(6 Suppl 5):S509-S517 [FREE Full text] [doi: 10.1016/j.amepre.2015.08.012] [Medline: 26590646]
57. Niu C, Eng L, Qiu X, Shen X, Espin-Garcia O, Song Y, et al. Lifestyle Behaviors in Elderly Cancer Survivors: A Comparison With Middle-Age Cancer Survivors. J Oncol Pract 2015 Jul;11(4):e450-e459 [FREE Full text] [doi: 10.1200/JOP.2014.002287] [Medline: 26060227]
58. Mehnert A, Koch U. Psychological comorbidity and health-related quality of life and its association with awareness, utilization, and need for psychosocial support in a cancer register-based sample of long-term breast cancer survivors. J Psychosom Res 2008 Apr;64(4):383-391. [doi: 10.1016/j.jpsychores.2007.12.005] [Medline: 18374737]
59. Thong MS, Kaptein AA, Vissers PA, Vreugdenhil G, van de Poll-Franse LV. Illness perceptions are associated with mortality among 1552 colorectal cancer survivors: a study from the population-based PROFILES registry. J Cancer Surviv 2016 Mar 19:Epub ahead of print. [doi: 10.1007/s11764-016-0536-5] [Medline: 26995005]
60. Walthouwer MJ, Oenema A, Candel M, Lechner L, de Vries H. Eating in moderation and the essential role of awareness. A Dutch longitudinal study identifying psychosocial predictors. Appetite 2015 Apr;87:152-159. [doi: 10.1016/j.appet.2014.12.214] [Medline: 25544317]
61. Wilson K, Senay I, Durantini M, Sánchez F, Hennessy M, Spring B, et al. When it comes to lifestyle recommendations, more is sometimes less: a meta-analysis of theoretical assumptions underlying the effectiveness of interventions promoting multiple behavior domain change. Psychol Bull 2015 Mar;141(2):474-509 [FREE Full text] [doi: 10.1037/a0038295] [Medline: 25528345]
62. Kwak L, Kremers SP, Candel MJ, Visscher TL, Brug J, van Baak MA. Changes in skinfold thickness and waist circumference after 12 and 24 months resulting from the NHF-NRG In Balance-project. Int J Behav Nutr Phys Act 2010;7:26 [FREE Full text] [doi: 10.1186/1479-5868-7-26] [Medline: 20370934]

63. Walthouwer MJ, Oenema A, Lechner L, de Vries H. Use and Effectiveness of a Video- and Text-Driven Web-Based Computer-Tailored Intervention: Randomized Controlled Trial. J Med Internet Res 2015;17(9):e222 [FREE Full text] [doi: 10.2196/jmir.4496] [Medline: 26408488]
64. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12(1):80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]

Abbreviations

BMI: Body Mass Index
CaSUN: Cancer Survivors' Unmet Needs questionnaire
CIS: Checklist Individual Strength
HADS: Hospital Anxiety and Depression Scale
KNW: Kanker Nazorg Wijzer (Cancer Aftercare Guide)
MAC: Mental Adjustment to Cancer Scale
MRA: Module Referral Advice
PA: physical activity
QoL: quality of life
RCT: randomized controlled trial
SQUASH: Short Questionnaire to Assess Health Enhancing Physical Activity
SSL-D: Social Support List Discrepancy Subscale

Edited by G Eysenbach; submitted 17.05.16; peer-reviewed by C Short, E Børøsund; comments to author 02.06.16; revised version received 11.07.16; accepted 20.07.16; published 23.08.16.

Please cite as:
Kanera IM, Willems RA, Bolman CAW, Mesters I, Zambon V, Gijsen BCM, Lechner L
Use and Appreciation of a Tailored Self-Management eHealth Intervention for Early Cancer Survivors: Process Evaluation of a Randomized Controlled Trial
J Med Internet Res 2016;18(8):e229
URL: http://www.jmir.org/2016/8/e229/
doi:10.2196/jmir.5975
PMID:27554525

Original Paper

Effectiveness of Web-Delivered Acceptance and Commitment Therapy in Relation to Mental Health and Well-Being: A Systematic Review and Meta-Analysis

Menna Brown[1], BSc (Hons), MSc; Alexander Glendenning[1], BSc, MSc; Alice E Hoon[1], BA (Hons), MSc, PhD; Ann John[1], MBBS, MD

Swansea University, Medical School, Swansea, United Kingdom

Corresponding Author:
Menna Brown, BSc (Hons), MSc
Swansea University
Medical School
ILS2
Swansea, SA2 8PP
United Kingdom
Phone: 44 179260 ext 6213
Fax: 44 1792606213
Email: menna.brown@swansea.ac.uk

Abstract

Background: The need for effective interventions to improve mental health and emotional well-being at a population level are gaining prominence both in the United Kingdom and globally. Advances in technology and widespread adoption of Internet capable devices have facilitated rapid development of Web-delivered psychological therapies. Interventions designed to manage a range of affective disorders by applying diverse therapeutic approaches are widely available.

Objective: The main aim of this review was to evaluate the evidence base of acceptance and commitment therapy (ACT) in a Web-based delivery format.

Method: A systematic review of the literature and meta-analysis was conducted. Two electronic databases were searched for Web-delivered interventions utilizing ACT for the management of affective disorders or well-being. Only Randomized Controlled Trials (RCTs) were included.

Results: The search strategy identified 59 articles. Of these, 10 articles met the inclusion criteria specified. The range of conditions and outcome measures that were identified limited the ability to draw firm conclusions about the efficacy of Web-delivered ACT-based intervention for anxiety or well-being.

Conclusions: ACT in a Web-based delivery format was found to be effective in the management of depression. Rates of adherence to study protocols and completion were high overall suggesting that this therapeutic approach is highly acceptable for patients and the general public.

(J Med Internet Res 2016;18(8):e221) doi:10.2196/jmir.6200

KEYWORDS

acceptance and commitment therapy; systematic review; meta-analysis; depression; anxiety; quality of life; Internet-based; mobile-based

Introduction

The need for effective interventions to improve mental health and emotional well-being at a population level are gaining prominence globally [1]. These interventions have fast become a priority issue for policy makers in the United Kingdom [2,3] and Europe [4] who have recognized the importance of health and well-being across the lifespan.

Mental health and emotional well-being are a fundamental component of good health. A lack of emotional well-being underpins many physical diseases, unhealthy lifestyles, and social inequalities in health. Common mental disorders (CMDs) such as depression and anxiety are known to be associated with

the adoption of unhealthy lifestyle behaviors, including smoking, increased alcohol consumption (above recommended limits), limited physical exercise, and obesity [5,6]. As such, the huge estimated economic costs of these disorders on individuals and society [7] usually associated with lost productivity and burden of health and social services [1,8,9], are likely to be underestimated.

Thus, effective treatments and resources that support individuals to improve their mental health and well-being through psychological treatment programs or health behavior change are of increasing interest to government agencies, health services, commercial enterprises, and individuals themselves. This comes at a time when health and public policy agendas are increasingly encouraging and slowly shifting responsibility for both physical and psychological health, well-being, and lifestyle choice to the individual themselves, so called "self-care" [10,11]. This shift towards personal responsibility is being aided through widespread commercial and technological advancement. Web-based mobile apps that support and encourage healthy lifestyle choices and behavior changes are widely available and affordable [12] and are increasingly utilized for health information [13] and treatment (eg, *Moodgym*, *fear fighter*, *beating the blues*). Web-based treatment programs employing cognitive behavior therapy (CBT) are considered to be an effective treatment for a range of conditions, including Post-Traumatic Stress Disorder [14], obsessive compulsive disorder [15], depression [16], anxiety [17], and social phobia [18].

Acceptance and commitment therapy (ACT) has enjoyed a steady rise in interest as an alternative therapeutic intervention to CBT. ACT is considered a third wave CBT, philosophically rooted in functional contextualism [19,20] and relational frame theory [21]. ACT differs from traditional CBT in a number of ways, most notably in that it does not consider thoughts and beliefs as correct or incorrect; and symptom reduction is not the goal of treatment but is a by-product of the process [19]. ACT is based on the principles of self-acceptance and a commitment to one's personal values; and encourages the adoption of behaviors that are in agreement with those personal values. ACT aims to encourage individuals toward (1) acceptance of difficult and unwelcome thoughts or emotions, and (2) promotion and simultaneous adoption of actions and behaviors, into daily practice, which are in line with these individual core values and principle beliefs. ACT interventions commonly incorporate mindfulness and experiential exercises that promote contact with the present moment.

As interest and research into the application of ACT grows so too must the evaluation of its evidence base. Reviews and meta-analyses have examined the effectiveness of ACT across a range of disorders. Öst [22] concluded that "ACT is not yet well established for any disorder" but showed promise in the treatment of chronic pain and tinnitus with additional possible efficacy for depression, psychotic symptoms, drug abuse, and stress at work. Before this, Ruiz [20] evaluated face-to-face delivery of ACT as compared with traditional (CBT) across a range of conditions and reported a significant mean effect size in support of ACT, for depression and quality of life but not anxiety. Sharp [23] reported that the research base, although small, suggested ACT was effective for a range of anxiety disorders. Powers et al [24] reported a "clear effect and overall advantage of ACT compared to control conditions" but found no evidence to suggest it was more effective than established treatments. Others have also reported ACT to be effective across a range of conditions, including psychiatric disorders [25], chronic pain [26-28], tinnitus [29], multiple sclerosis [30], anxiety disorders [31,32], stress [33-35], and health behavior or lifestyle change together with smoking [36-39] and weight optimization [40]. Thus whilst there is some uncertainty of the effectiveness of ACT, it appears to be related more to establishing the evidence base rather than ACT being an ineffective intervention.

The recent surge in interest in Web-based interventions warrants further review of ACT in the context of a Web-based delivery format to manage CMDs. No previous review has focused exclusively on ACT as implemented in Web format. Although it is important to note that Öst [22] included 3 Web-based ACT interventions. The Association for Contextual Behavioral Science (ACBS) website lists 9 computerized versions of ACT since 2013. Thus, a review focusing solely on this application of ACT is required to assess the evidence base of its effectiveness in the treatment of CMDs.

This review aimed to examine the published, peer-reviewed evidence pertaining to the effectiveness of ACT in the treatment of CMDs and well-being in a Web-based delivery format.

Specific objectives include:

1. Identify randomized controlled trials (RCTs) of Web-based interventions that have employed ACT as the main therapeutic approach, for the treatment of a CMD or improvement of well-being in any population.

2. Appraise and synthesize the evidence on effectiveness for depression, anxiety, and quality of life.

The secondary aim was to report rates of adherence to the study protocol, calculated as a percentage of those randomized to the intervention and completed post assessment.

Methods

Search Process

Systemic searches of electronic databases Medline Complete (EBSCO interface) and PsychINFO (EBSCO interface) were conducted from (database inception) to February 10 2016. Standardized subject terms were utilized in each electronic database. The keywords search table can be viewed in Multimedia Appendix 1. The MEDLINE Strategy (EBSCO interface) was adapted for PsychINFO. To address grey literature the list published on the ACBS website of computerized ACT interventions, since 2013, was reviewed. Reference lists of identified studies were examined for additional articles.

Studies were required to meet the following criteria: (1) the study must be published in a peer-reviewed, English language journal (2) the intervention was based on ACT (3) RCT design (random allocation of participants into either intervention and control or intervention and active treatment arm) (4) the study

delivers the intervention via the Web (5) the intervention was designed to be accessed on more than one occasion (6) the intervention was designed to manage a CMD or improve well-being and (7) the study must report a measure of effectiveness of the intervention (ie, pre- and post-outcome measure) to enable the calculation of an effect size. Studies were excluded if (1) participants were under the age of 18 years and (2) reanalysis of data from a subsample of a previously published RCT.

Procedure

Three reviewers (MB, AJ, AG) independently reviewed the title and abstract against inclusion and exclusion criteria. Final decisions were triangulated. Studies were included for full text review where one reviewer indicated to include. The full text article was then assessed against inclusion and exclusion criteria. Studies were excluded when they did not meet a single criterion. The first instance where they did not meet eligibility was recorded and the study was not assessed for other inclusion criteria [41]. In instances where more than one study was retrieved by the same author (MB) checked that the data presented were from different populations. The final list was discussed with the expert reviewer (AJ) to ensure consensus was reached.

Data Extraction

A data extraction sheet was developed and piloted. The following data were extracted for analysis:

- Citation reference (authors, title and date, country)
- RCT characteristics: RCT design, total number of trial arms, type of comparator or control, method of randomization, allocation sequence concealment, blinding, and sample size and total number allocated to each trial arm
- Participant characteristics: setting (recruited from), condition, comorbidity, diagnostic criteria, age, sex, self-referred or clinician refereed
- Intervention characteristics: name of ACT intervention, specific elements of ACT included, guided or automated delivery, type of guide, type of communication with guide, additional support, intended duration and modules included, format of delivery (sequential or free navigation), features of the system (reminders, personalization)
- Effectiveness of intervention: primary and secondary outcome measures; effect size at post assessment and follow up where available
- Adherence: calculated as a percentage of those randomized to the intervention and completed post assessment.

Risk of Bias

The Cochrane Collaboration's tool for assessing risk of bias was applied [41]. A total of 6 risk domains were evaluated. Each domain generated a level of risk: low, high, and unclear, from which an overall level of bias was determined for each RCT. No contact with the publication authors was made to discuss or further clarify points relating to a particular study.

Analysis

Data were entered into MATLAB, Mathworks and Review Manager (RevMan) version 5.3 (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen), cleaned, and checked for missing values and errors.

Effect Size Computation

Two categories of effect size were calculated, one comprising between-group effects measured post-treatment and the other comprising within-group effects measured between pre- and post-treatment. In those studies that included more than one comparison condition, the active control was chosen as the comparison condition. For each of these 2 categories, summary effect sizes were then calculated for the following 3 categories of outcome measures: (1) depression (2) anxiety and (3) quality of life. All between-group effect sizes are signed so that a positive value is in favor of the Web-based ACT condition and all within-group effect sizes are signed so that a positive value is in favor of the post-treatment time point of the pre-treatment time point. Hedges' g was chosen for effect size, using means and standard deviations of the outcome measures of participants.

Meta-Analysis

Analyses were conducted using RevMan version 5.3.5 and MATLAB R2015a. The DerSimonian and Laird random-effects model [42] was adopted in each case, based on the assumption that variation of true effects exists between studies. Using this model, the summary effect sizes outlined in the previous section were calculated. Corresponding tests for statistical significance were computed in the form of both two-tailed P-values and 95% confidence intervals. The heterogeneity of true effects was assessed by the I^2-statistic. Corresponding P-values were computed to assess the extent of uncertainty in Q, following the assumption that Q follows a χ^2 (k-1) - distribution, with k-1 degrees of freedom.

Interpretation of effect sizes was based on Cohen's rule-of-thumb, that is, small effects were categorized as $0.2 \leq g<0.5$, medium effects as $0.5 \leq g<0.8$ and large effects as $g>0.8$ [43]. The proportion of dispersion due to true effects was categorized by the well-established scale of Higgins et al [41], that is, the intervals $25\% \leq I^2 < 50\%$; $50\% \leq I^2$ 75% and $I^2 >$ 75% and indicate a low, medium, and high proportion of dispersion due to true effects, respectively.

Where the outcome measure was dichotomized, it was not possible to calculate Hedges' g directly. In this instance, data were transformed using the method given in Figure 1 [44].

Figure 1. Expression relating the odds ratio and Hedges' g where df is the number of degrees of freedom.

$$g = \left(1 - \frac{3}{4df - 1}\right)\frac{\sqrt{3}}{\pi}\ln \mathrm{OR}$$

where OR is the odds ratio and df is the number of degrees of freedom associated with the problem.

Results

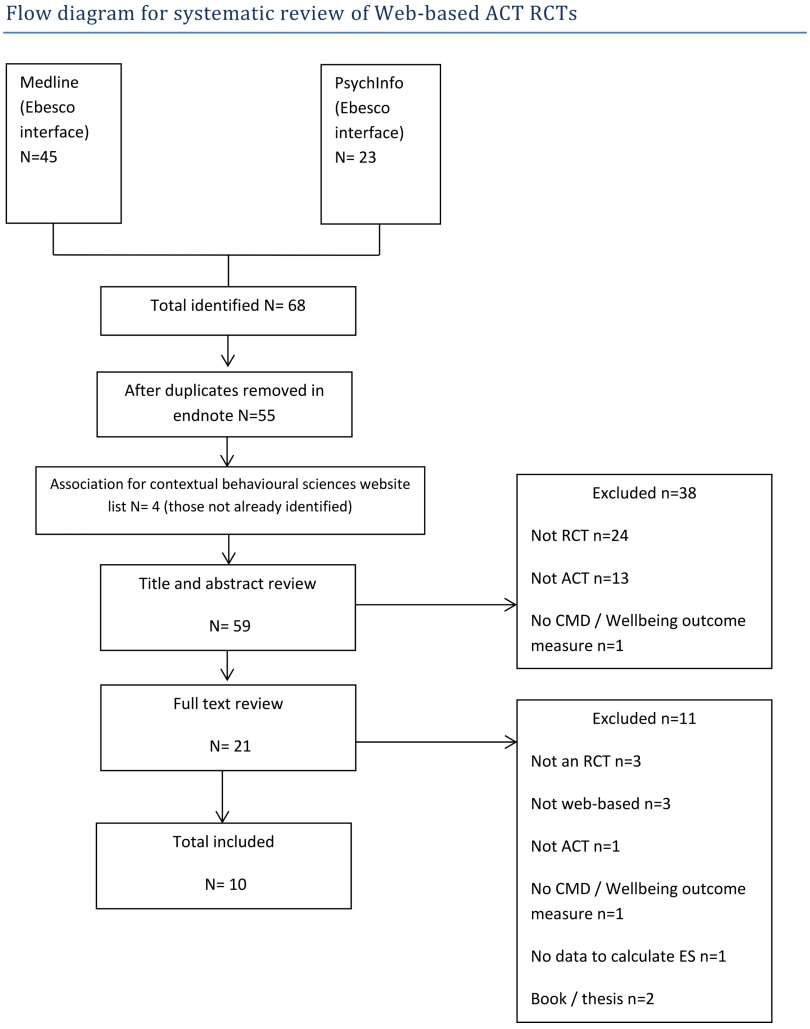
Principal Findings

The database searches identified 55 articles; a further 4 were identified through review of the ACBS website. This led to 59 articles being included for title and abstract review. A total of 38 articles were excluded at this stage leaving 21 articles that met inclusion criteria for full text review (Figure 2 PRISMA flowchart). Furthermore, 11 were excluded during full text review, 3 were not RCT design, 3 were not Web-based, 1 did not employ ACT, 1 did not report pre-post outcome for a CMD, 1 did not include data from which an effect size could be calculated, and 2 were not peer-reviewed. Ten RCTs met the full inclusion criteria (Multimedia Appendix 2) and were included.

Two instances were identified where RCTs reported data from the same author [45-48]. In each instance, the 2 reports were assessed. Lappalainen et al [45,46] specified different data collection points, 2011 and 2012 whereas Levin et al [47,48] indicated that participants received different compensation and rewards for taking part (USD$10 plus research credits; USD$60), which suggested different participants were included. For these reasons, each of these articles were included in the review.

Of the 10 RCTs, 7 included 2 trial arms, of which 3 included a wait list control (WLC), and 4 had active controls as the comparator arm. The remaining 3 RCTs included 3 armed trials, of which 2 used an active control plus a WLC and one used 2 active interventions as control (Multimedia Appendix 3). A total of 3 trials were undertaken in Sweden, 3 in the United States, 2 in Finland, and 2 in the Netherlands.

Figure 2. PRISMA flowchart of included studies.



Participants and Condition

All participants were self-selected. Of the participants, 2 RCTs were recruited from a clinical population (pain clinics) [49,50], 2 from an undergraduate student population [47,48] whereas the remainder were recruited from the general population [29,36,45,46,51]. Trials ranged in size from 38 to 236 participants. Three RCTs included more than 100 participants.

Five interventions were primarily designed to manage and reduce depression and depressive symptoms [36,45,46,51,52] of which one specifically focused on depression in smokers [36], one targeted psychological distress [45], one well-being [46], two chronic pain [49,50], and one tinnitus [29]. All included pre- and post-outcome measures for a CMD, specifically anxiety or depression.

Methods to confirm a diagnosis of primary condition varied. Medical examination and telephone screening [49], computerized screening interview combined with a structured telephone interview [50-52], computerized screening followed by telephone interview and face-to-face meeting plus a medical confirmation of tinnitus [29], self-assessment questionnaires [36], structured clinical telephone interview [45,46], and none [47,48].

Outcome Measures

A range of outcomes measures were utilized across each RCT which included, Hospital and Depression Scales (HADS), Beck Depression Inventory (BDI), Beck Anxiety Inventory (BAI), Anxiety and depression detector, Depression Anxiety Stress Scales (DASS), and Center for Epidemiologic Studies Depression Scale (CES-D). A range of secondary of measures were also used for quality of life and psychological distress, Quality of Life Inventory (QOLI), general health Questionnaire 12 items (GHQ-12), Symptom checklist 90 items (SCL-90), and the Mental health Continuum Short Form (MHC-SF). A range of ACT specific outcomes were measured in 7 of the included articles, Avoidance and Inflexibility Scale (AIS), Acceptance and Action Scale (AAQ-II), Five Facet Mindfulness Questionnaire (FFMQ), Psychological Inflexibility Scale (PIPS), Engaged Living Scale (ELS), plus an ACT knowledge questionnaire (Multimedia Appendix 4).

ACT Intervention

Eight separate interventions were identified (Depressionshjälpen, webQuit.org, ACT, Living to the Full, Living with pain, The Good Life compass, ACT-CL, and one unnamed), 2 interventions were utilized in 2 different RCTs (The Good Life compass and ACT-CL). Of which, one was a progression of the first and included additional ACT components [47]. Nine used ACT as the sole therapeutic approach and one [51] used behavioral activation in combination with ACT.

All provided details of the ACT characteristics employed. Four specifically stated that they included modules that attended to all of the 6 core principles of ACT [36,45,50,52]. Two studies used 4 core principles [46,49]. Three studies employed 3 core principles [29,47,51]. One study used 2 core principles [45]. The two Levin et al studies [47,48] ACT-CL were described as prototype interventions exploring feasibility and acceptance. All included mindfulness, experiential exercises, or metaphors, whereas 4 stated they included a maintenance plan for participants (Multimedia Appendix 5).

A total of 3 interventions were automated, that is, no guide or coach was involved in the delivery of the therapeutic program [36,47,48] and the remainder were guided interventions (therapists or coaches assisted and supported participants throughout the delivery of the intervention program). Of the 7 guided interventions, the “guide” was a combination of trained psychologists and psychology graduate students (n=2) or graduate psychology students (n=5). Two of the interventions provided additional support to the clinical guide. One included an administrator and the other a computer technician, both could be contacted if required. The guides provided clinical support in a variety of ways, written secure messages and feedback via the system (n=5), written feedback via email (n=2) and verbal communication over the telephone (n=3). Of which, one was to deliver support and guidance and two acted as reminders to complete the next module in the program. All contact was asynchronous. Of the 3 automated interventions, 2 stated that they provided automated feedback. One intervention provided an additional workbook and a CD, one included a face-to-face meeting at the start and end of the treatment, and one reported a face-to-face meeting prior to commencement of intervention. A total of 6 of the 7 RCTs identified the number of therapists involved in the delivery of the ACT intervention, therapist numbers ranged from 2 to 18.

Intended duration of ACT interventions varied between 3 and 12 weeks in length (M=7.4, SD=3.06) and interventions included between 2 and 9 modules (M=6.4, SD= 2.5).

System Features

In a Web-based context, the features incorporated into the design of the intervention and computerized system, are of interest as they have the potential to influence engagement and adherence [53]. Email reminders were included in 4 RCTs [46-48,52], short message service reminders [45,48,49,52] and homework tasks were incorporated into 6 RCTs [29,45,47,49-51]. Progression through the program was controlled by either the system or the guide, depending on successful completion of prior modules in 7 systems. Telephone reminders were used in 3 guided interventions to prompt use and encourage continued engagement with the program [46,47,49], personalized feedback on receipt of homework assignments was provided in 5 of the interventions [45-48,52], and the option to personalize the home page was available in one [52]. None of them included any type of social networking feature.

A total of 8 interventions were designed to be accessed in a sequential manor, in which modules were to be completed in a predetermined order. One intervention [46] allowed participant’s free navigation of the system, meaning participants could access and complete modules in any order they decided. However, the recommendation was to work through the modules in the suggested order. One intervention did not report format [36].

Adherence to Protocol, Usage Data and Satisfaction

Adherence was calculated as a percentage of those randomized to the intervention that also completed postassessment. Nine RCTs reported adherence data, adherence ranged between 48% and 100% (M=82.6%, SD=17.8) and adherence to control ranged between 53.1% and 100% (M=83.4%, SD=16.4%) the control group selected was the active treatment not WLC.

A total of 7 RCTs reported data usage. The data reported varied. Pots et al [52] reported the mean number of completed modules. Carlbring et al [51] reported the number of participants who did not complete any modules, 16% in the CBT condition, and 6% in the ACT condition. Jones et al [36] reported website usage; usage in the ACT condition was significantly higher compared with the control condition (21.7 minutes per login vs 9.4 minutes). Lappalainen et al [46] reported average time spent per week, 50% of participants spent less than one hour a week, 38.9% spent 1-2 hours and 11.1% spent more than 2 hours. Levin et al [47] reported the percentage of participants

completing each module, 85% completed lesson 1 and 55% completed lesson 2. Trompetter et al [50] reported module completion, 72% completed 6 modules and 66.2% completed all 9 modules. Levin et al [48] reported a summary of program usage across both intervention and control conditions, 92% completed both modules and spent on average 81.98 minutes using the program.

Satisfaction with treatment is also of importance for adherence and engagement within a Web context. Six RCTs reported a measure of participant satisfaction, of which, 2 focused on system usability [36,45-48,50].

Meta-Analysis

The number of studies used in the between-group meta-analyses were k=10, k=7, and k=8 for the depression, anxiety, and quality of life outcome measures, respectively. For the within-group meta-analyses, these were k=9, k=7, and k=8.

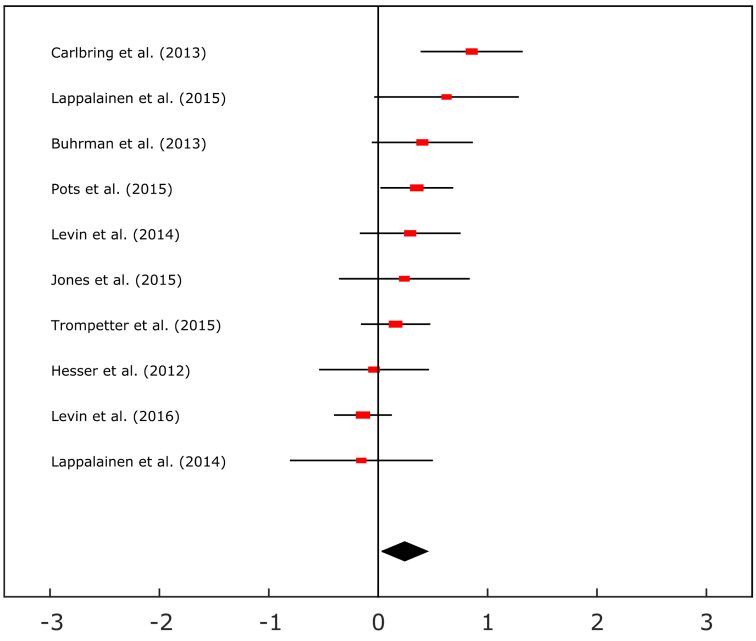
Studies generally presented the mean and standard deviation of participants' scores in each outcome measure. The one exception to this was Jones et al [36] whose data on the depression outcome measure was dichotomized, meaning it was not possible to calculate Hedges' *g* directly. Thus Jones et al [36] was included in the between-group summary effect size calculation but could not be included in the within-group category, as the odds ratio is undefined for this category.

With regard to the between-group summary effect sizes, the effect size for the depression outcome was small and in favor of ACT with g=0.24. This was also shown to be statistically significant with P=.02 as was the proportion of heterogeneity attributed to true effects, I^2= 55%. The effect size for anxiety was statistically significant. However, it fell short of the lower limit for small effect size with g=0.18, P=.03. Heterogeneity for this outcome measure did not reach statistical significance P=.49, with the same being true for both the effect size and heterogeneity for the quality of life outcome measure. Summary effect sizes belonging to the between-group category are shown in Multimedia Appendix 6.

However, the within-group category demonstrated that participants' scores improved greatly between time points over all outcomes, as displayed in Multimedia Appendix 7. Effect sizes for both the depression and anxiety outcomes were medium in magnitude with g=0.73 and 0.51, and the quality of life outcome attained a small effect size of g=0.44, all of which were statistically significant P=.001; P<.001 respectively. The depression and anxiety outcomes also indicated high proportions of heterogeneity between studies with $I^2 \geq 75\%$ whereas ,the quality of life outcome showed a medium proportion of heterogeneity, all of which were also statistically significant with P<.001. Figure 3 shows the forest plot for the depression summary effect size for ACT interventions versus comparison groups.

It was not possible to calculate any summary effect sizes at follow-up due to a lack of published data.

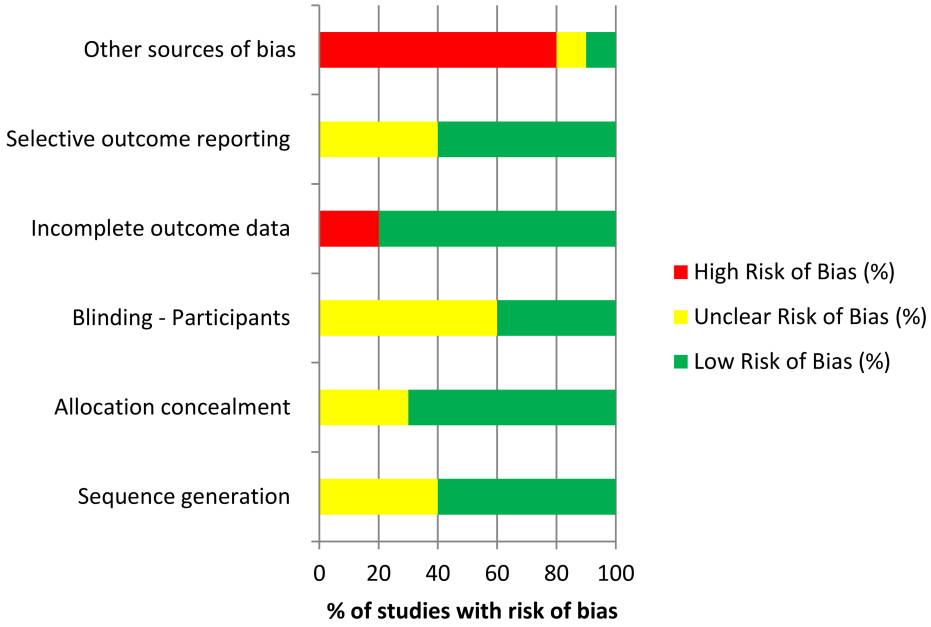
Figure 3. Forest plot produced in MATLAB for the depression outcome measure in the between-group meta-analysis. The x-axis units are in Hedges' g.



Risk of Bias

The risk of bias judgment for each study was made on the information provided in the publication only. Of the 10 studies included, 4 (40%) were judged to be of low risk of bias [29,46,49,52], one (10%) of high risk [36], and an unclear risk of bias was assigned to 5 (50%) [45,47,48,50,51] of the included studies. Figure 4 shows the risk of bias summary.

Figure 4. Risk of bias summary per domain.



Discussion

Overall Findings

The aim of this study was to examine the published, peer-reviewed literature pertaining to the effectiveness of ACT in the management of CMD and well-being in a Web-based delivery format. Ten RCTs met inclusion criteria. A total of 5 studies focused on depression, 1 on psychological distress, 1 on well-being, 2 on chronic pain, 1 one on tinnitus. All included pre- and post-outcome measures for anxiety, depression, or quality of life. The majority compared against an active comparator or active comparator plus wait list control.

Principal Results

ACT delivered via a Web-based format is effective for the management of depression (small effect size) and anxiety (which neared the threshold for a small effect size). Web-based ACT interventions were not found to be effective for improving quality of life. Due to lack of published data effect size was not calculated for follow-up effects.

Findings support ACT in a Web-based delivery format for the management of depression. Review of those RCTs focused on depression as the primary outcome [36,45,46,49,51] revealed that ACT was more effective than control, where control was either WLC or an active control. For example, Carlbring et al [51] reported a large effect size and 25% of participants experienced clinical recovery posttreatment. Low dropout and good engagement were reported. On average, participants completed 5 of the 7 modules. Lappalainen et al [45,46] reported Web-delivered ACT was effective over and above face-to-face delivery and WLC with differential outcomes identified in favor of the Web-delivered intervention. Furthermore, treatment effects were maintained at follow-up although they did level out in-line with the WLC. Pots et al [52] reported significant effect of ACT compared to both WLC and an expressive writing condition that were maintained at follow-up. Finally Jones et al [36], while not finding a statistically significant difference, reported positive benefit of ACT over attention control conditions coupled with good user satisfaction for the intervention in a depressed, smoking population. This nonstatistical finding was likely limited by the use of only one depression screening question. All of these studies identified a need for further research and each was limited by small sample sizes. Current findings are in line with those of prior reviews of ACT, Öst [22] reported that ACT was “possibly efficacious for depression.” Five studies primarily targeting depression were reviewed of which, none were Web-based, and three were compared with treatment as usual. Comparing the current effect size for depression against Öst’s [22] overall effect size for active comparators (across all subsamples) the current findings are in-line (g=0.24) however this compares less favorably for overall WLC comparison (g=0.63) but are in-line with those reported by Ruiz (g=0.27) [20].

The effect size for anxiety was statistically significant, although it only neared the threshold for a small effect size (g=0.18). There are a number of reasons that might contribute an explanation to this observation of lower effect size. None of the included interventions were primarily designed to treat anxiety; in each case the secondary outcome measure was included. In addition, where pretreatment anxiety scores were low, this limited potential improvement. Furthermore, only 7 of the RCTs included an outcome measure that could be included in the effect size calculation for anxiety. However, Pots et al [52] reported that ACT had positive outcomes for anxiety as well as depression. In prior review findings when ACT was compared with CBT, effect sizes for anxiety did not meet statistical significance [20]. Sharp [23] reviewed ACT specifically for use with anxiety disorders and concluded that data provided preliminary support.

ACT was not found to be effective in delivering improvements in quality of life. However, as with anxiety, none of the interventions specifically targeted quality of life and 2 did not

include an outcome measure that could be included in the meta-analysis calculation [36,47]. Both of the RCTs reported by Levin et al [47,48] were focused on feasibility of ACT in an undergraduate population and contained the fewest number of ACT components of all interventions reviewed (the intervention was detailed to be underdevelopment and acknowledged it contained fewer elements). The most recent one of these reported equivalence in outcomes for ACT to the attention control website and lower program usage. However, their analysis of usage patterns suggested that those who engaged more with ACT experienced more positive outcomes and increased psychological flexibility. This was in comparison with their earlier findings that suggested strong acceptability and feasibility. The limited focus on quality of life or well-being is in line with findings identified by Öst [22], where no included studies used them as the primary outcome measure. However in the review by Ruiz [20] an effect size (g=0.25) was reported which is higher than that found in the current data. This could be attributable to a few things arguably, while quality of life and well-being can be compared, they are not necessarily the best fit.

Variation in intervention features, delivery format (guided or automated), and study context warrants discussion. For example, The Levin et al studies [47,48] included fewer modules that were available for a shorter period of time, did not conduct preassessment screening and participants received a financial reward for their participation. Jones et al [36] unguided intervention included one preassessment depression screening question. Opportunities for improvement, detection of disorders and effect sizes may be underestimated, due to use of screening questionnaires that have not been validated at baseline. Lappalainen et al [45,46] in contrast utilized multiple therapists to guide participants. Guided interventions, using CBT, are associated with higher effect sizes.

The primary goal of ACT is to improve psychological flexibility. Psychological flexibility in the context of ACT has been defined as “the measure of how a person (1) adapts to fluctuating situational demands, (2) reconfigures mental resources, (3) shifts perspective, and (4) balances competing desires, needs, and life domains” [54]. Examination of ACT specific outcome measures identified improvements in psychological flexibility. Jones et al [36] reported significant improvements in willingness to experience physical triggers and a trend toward willingness to experience emotional triggers (AIS); Lappalainen et al [45,46] reported significant effect in mindfulness skills and psychological flexibility in both face-to-face and Web-based delivery of ACT (using AAQ-II) as did Pots et al [52] with the exception of improvements on the mindfulness facet. However Levin et al [47,48] did not find any significant effect on FFMQ measure or AAQ-II but did report improvements in ACT knowledge. Trompetter et al [50] reported significant improvements at three- and six-month follow-up (FFMQ-SF) and 6 months on PIPS. Although mixed, these findings suggest support for ACT in improving psychological flexibility.

It is important to consider Web-based delivery format in its own right, this is in light of the recent expansion of interest into this field (ACT), evidenced by the multitude of protocol and feasibility studies identified during this review and through the ACBS website.

The secondary aim of this review was to report rates of adherence to Web-based interventions employing ACT as the therapeutic approach. Poor adherence to Web-based mental health interventions is of widespread concern and is well documented in the literature [53,55-57]. Poor adherence has the potential to limit effectiveness [58] and reduce cost effectiveness [59] which is a key benefit of this delivery format. The mean rate of adherence to protocol (82.6%) was higher than published means for CBT based interventions where dropout rates range from 2% to 83% [60] and was comparable to the rates of adherence for the control groups (83.4%). Adherence was calculated as a percentage of those randomized to the intervention and completed postassessment. However reported rates of completion rates remained lower (68.4%). Four interventions [47,48,50,52] specified that they included stakeholders in the design and development process to address adherence, or specifically employed persuasive design features in a bid to encourage adherence and engagement. For example, Pots et al [52] incorporated persuasive technology (but did not specify which) in the design of their intervention, adherence and engagement was reported to be high (73% of participants completed all 9 modules); whereas Levin et al [47] utilized a “tunneled” format. Tunneling refers to using the computerized system to guide users through the therapeutic content in a predefined order, ensuring opportunity to persuade along the way [61]. Such initiatives may help to promote and increase adherence. Couper et al [62] report that higher engagement with the intervention program and Web-based materials is associated with increased likelihood of adherence to study protocols, follow-up data collection points and importantly, changes in health behavior (fruit and vegetable intake study); Strecther [63] reported similar findings in a smoking cessation study where quit rates increased for each additional webpage opened. Although these studies did not employ ACT, it is feasible that the same association could be important across all therapeutic interventions delivered via a Web-based format. Indeed, Trompetter et al [50] noted the increasing need to address this issue in a Web-based context.

Limitations

There are several limitations to note, such as the variety of outcome measures used in each RCTs may limit the usefulness of the findings. For example, on the quality of life effect size measure 2 studies did not report any outcome measures and a further 2 used a Symptom Checklist 90 (SCL-90) which is considered a measure of well-being as opposed to quality of life. Furthermore, a higher score on the SCL-90 represents lower well-being as opposed to Quality of Life Inventory (QOLI) and Mental Health Continuum Short Form (MHC-SF) where a higher score represents higher quality of life. However, this difference in scoring was adjusted for in the statistical analysis.

Due to lack of published data for follow-up measures, a follow-up effect size could not be calculated.

Due to the small number of included RCTs, meta-analysis by design (eg, WLC or active comparator) was not possible. Of the 10 RCTs reviewed, 3 included WLC, the remainder used

either an attention control website, alternative intervention (CBT, moderated discussion forum and expressive writing), or face-to-face delivery format. In the instance where more than one comparator group data was available [29,50,52] we compared against the active treatment. This decision was taken for a number of reasons. Firstly, the majority of included RCTs used active comparators so this was in line with the other comparisons drawn. Secondly, active comparators represent the most realistic real world alternative. For example, CBT as a Web-based therapeutic treatment has a strong and well-established evidence base [15,64] and interventions using this approach are freely available (eg, *MoodGYM*) for the treatment of depression and other CMD as are psycho-education websites and online discussion groups. We acknowledge that the diversity of comparators may limit the generalizability of our findings. However, we feel this is a practical and useful approach to adopt in the current review.

One study [51] was based on behavioral activation and ACT and as such, the effects of the treatment intervention cannot exclusively be attributed to ACT.

The interpretation of effect size magnitude (both between- and within-group) used in this paper adheres to Cohen's rule-of-thumb. However, concerning the interpretation of within-group effect sizes, some authors prefer to opt for an alternative classification [65]. Here small effects are classified as $0.5 \leq g < 0.8$, medium effects as $0.8 \leq g < 1.1$ and large effects as $g \geq 1.1$. When this classification is applied to the 3 within-group meta-analyses conducted in this paper then the effect sizes for the depression and anxiety outcomes of $g=0.73$ and 0.51 are recategorized as small, while the effect size for the quality of life outcome of $g=0.44$ does not reach the threshold for small effect size. However, statistical significance remains unaffected in all the 3 cases.

Usage data were not reported in all RCTs, in the future standardized reporting is recommended including agreement on completion rate. For example, Karyotaki et al [66] advocate intervention completion to be defined as, completion of 75% or more of the total modules. Finally, publication bias and the trend to report positive results over negative or neutral results must be taken into consideration when reviewing the results of the current meta-analysis. It is possible that our results are overestimated as a result. We reviewed all relevant ACT RCTs listed on the ACBS website to identify any further studies not cited in our database searches.

Implications for Practice

Web-based delivery of interventions has undergone a recent expansion in a range of health contexts, physical health, mental health, and lifestyle behavior change. This brings with it new considerations for the effective delivery of therapeutic interventions. Increased interest in this delivery format stems from the explosion in access to affordable personal mobile devices that offer easy access to the Internet from all locations with 3G or Wi-Fi coverage. Such easy and convenient accessibility is thus one of the key advantages of this new delivery format, coupled with the cost effectiveness associated with its ability to facilitate widespread reach access across the population.

Further research into the use of ACT via Web-delivery, is required to continue to explore its effectiveness and to understand the most effective components for this delivery context. Specifically those targeting anxiety and well-being would be of benefit as positive well-being continues to grow as an area of public interest and means to promote and prevent poor mental health. Studies should focus on recruiting larger populations to avoid concerns with lack of statistical power and to ensure wider generalizability of findings. Studies should also seek to examine the long-term effect of ACT through inclusion of follow-up periods in future RCTs. In this systematic search, 9 protocol documents and 5 feasibility studies were identified suggesting that this is a developing field of interest and that evidence will expand in coming years. Thus, we provide a first review of the evidence.

Risk of bias assessment concluded that the majority of studies had a low or unclear risk of bias and thus there is potential for future studies to ensure that they continue to report in line with Cochrane recommendations to ensure the availability of best quality evidence.

It is worth noting that the intervention used in Carlbring et al [51] has since been utilized in subsequent RCT [67] but where it is described predominately as iCBT. Thus, this study would not have been identified in the current search strategy and, if it had been identified through other sources, would have been excluded at title-abstract stage due to variations in reporting of the intervention components and its stated theoretical perspective. While a sensitivity analysis, where we included this study did not alter our findings (Multimedia Appendix 8), this did highlight a wider concern in this research field. Future research into Web-based interventions, using all types of therapy should consider the way in which they report and describe the intervention and treatment. Consistent reporting of the intervention across RCTs would allow assessment of any differences in effectiveness of these types of programs (CBT, ACT), which are clearly distinguished in the field of psychology in terms of their mechanisms of action. This would also facilitate the systematic review process as research in this area develops.

In addition, analyses of ACT-specific outcome measures are a potential for further exploration.

Conclusion

ACT is efficacious for the treatment of depression in a Web-based delivery format. Further research is required across all mental health and emotional well-being domains to continue to develop and review the evidence base in this delivery format. This review adds strength to this evidence base, across delivery formats.

Acknowledgments

Health and Care Research Wales [SCS-14-11] supported the realization of this review. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflicts of Interest

None Declared.

Multimedia Appendix 1

Keyword search terms.

[PNG File, 677KB - jmir_v18i8e221_app1.png]

Multimedia Appendix 2

Main results table.

[PNG File, 1MB - jmir_v18i8e221_app2.png]

Multimedia Appendix 3

Trial arms and comparators.

[PNG File, 732KB - jmir_v18i8e221_app3.png]

Multimedia Appendix 4

Outcome measures utilized in each RCT.

[PNG File, 935KB - jmir_v18i8e221_app4.png]

Multimedia Appendix 5

The various key statistics relating to the summary between-group effect sizes by outcome measure. From left to right these are: k number of studies; g summary effect size with confidence interval (CI); Z test statistic; p-value relating to g; Q-statistic for total dispersion between studies; I^2-statistic for proportion of dispersion due to true effects; and p-value relating to I^2.

[PNG File, 275KB - jmir_v18i8e221_app5.png]

Multimedia Appendix 6

The key statistics relating to the summary within-group effect sizes by outcome measure.

[PNG File, 281KB - jmir_v18i8e221_app6.png]

Multimedia Appendix 7

Table of interventions by core processes.

[PNG File, 659KB - jmir_v18i8e221_app7.png]

Multimedia Appendix 8

Meta-analysis results additional permutations.

[PNG File, 197KB - jmir_v18i8e221_app8.png]

References

1. World Health Organization. 2001. The World Health Report 2001: Mental Health: New Understanding, New Hope URL: http://www.who.int/whr/2001/en/whr01_en.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYmm7aln]
2. Prince M, Patel V, Saxena S, Maj M, Maselko J. No health without mental health. The Lancet 2007;370(9590):859-877. [Medline: 17804063]
3. Department of Health, London. 2014. Public Mental Health Priorities: Investing in the Evidence URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/413196/CMO_web_doc.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYn6fl9l]

4. Europa. 2008. European Pact For Mental Health And Well-Being URL: http://ec.europa.eu/health/ph_determinants/life_style/mental/docs/pact_en.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYnC0UxI]
5. Talen MR, Mann MM. Obesity and Mental Health. Primary Care: Clinics in Office Practice 2009 Jun;36(2):287-305. [doi: 10.1016/j.pop.2009.01.012]
6. Scott D, Happell B. The high prevalence of poor physical health and unhealthy lifestyle behaviours in individuals with severe mental illness. Issues in Mental Health Nursing 2011;32(9):589-597. [Medline: 21859410]
7. Thomas CM, Morris S. Cost of depression among adults in England in 2000. The British Journal of Psychiatry 2003;183(6):514-519. [Medline: 14645022]
8. McCrone P, Dhanasiri S, Patel A. Kingsfund. England: King's Fund; 2008. Paying the price: the cost of mental health care in England to 2026 URL: http://www.kingsfund.org.uk/sites/files/kf/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYnZMleT]
9. Kolappa K K, Henderson DC, Kishore SP. No physical health without mental health: lessons unlearned? Bull World Health Organ 2013;91(1):3-3A. [Medline: 23397342]
10. The Centre for Economic Performance's Mental Health Policy Group. Centre for Economic Performance. England: The London School of Economics And Political Science; 2012. How mental illness loses out in the NHS URL: http://cep.lse.ac.uk/pubs/download/special/cepsp26.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYnqDuJi]
11. Department of Health, London. England: DoH; 2006. Supporting People with Long term Conditions to Self care: A Guide to Developing Local Strategies and Good Practice URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4130868.pdf [accessed 2016-08-06] [WebCite Cache ID 6jYnxBEzZ]
12. Neuhauser L, Kreps GL. Rethinking Communication in the E-health Era. Journal of Health Psychology 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
13. Escoffery C, Miner K, Adame D. Internet use for health information among college students. Journal of American College Health 2005;53(4):183-188. [Medline: 15663067]
14. Knaevelsrud C, Maercker A. Long-term effects of an Internet-Based treatment for posttraumatic stress. Cognitive Behaviour Therapy 2010 Jan;39(1):72-77. [doi: 10.1080/16506070902999935] [Medline: 19675958]
15. Mataix-Cols D, Marks IM. Self-help with minimal therapist contact for obsessive-compulsive disorder: a review. European Psychiatry 2006 Mar;21(2):75-80. [doi: 10.1016/j.eurpsy.2005.07.003] [Medline: 16360307]
16. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. General Hospital Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
17. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychological medicine 2003 Feb;33(2):217-227. [Medline: 12622301]
18. Titov N, Andrews G, Choi I, Schwencke G, Johnston L. Randomized controlled trial of web-based treatment of social phobia without clinician guidance. Australian and New Zealand Journal of Psychiatry 2009 Jan;43(10):913-919. [doi: 10.1080/00048670903179160]
19. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. Behavior Therapy 2004;35(4):639-665. [doi: 10.1016/S0005-7894%2804%2980013-3]
20. Ruiz FJ. Acceptance and commitment therapy versus traditional cognitive behavioral therapy: A systematic review and meta-analysis of current empirical evidence. International Journal of Psychology and Psychological Therapy 2012;12(3):333-358 [FREE Full text]
21. Hayes S. In: Hayes S, Barnes-Holmes D, Roche B, editors. Relational frame theory: A post-Skinnerian account of human language and cognition. New York: Springer Science & Business Media; 2001.
22. Ost L. The efficacy of Acceptance and Commitment Therapy: an updated systematic review and meta-analysis. Behaviour Research and Therapy 2014 Oct;61:105-121. [doi: 10.1016/j.brat.2014.07.018] [Medline: 25193001]
23. Sharp K. A review of acceptance and commitment therapy with anxiety disorders. International Journal of Psychology & Psychological Therapy 2012;12(3):359-372 [FREE Full text]
24. Powers MB, Zum Vorde Sive Vording MB, Emmelkamp PM. Acceptance and commitment therapy: a meta-analytic review. Psychotherapy and psychosomatics 2009;78(2):73-80. [doi: 10.1159/000190790] [Medline: 19142046]
25. Bach P, Hayes SC. The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. Journal of Consulting and Clinical Psychology 2002 Oct;70(5):1129-1139. [Medline: 12362963]
26. Wicksell RK, Kemani M, Jensen K, Kosek E, Kadetoff D, Sorjonen K, et al. Acceptance and commitment therapy for fibromyalgia: a randomized controlled trial. European Journal of Pain 2013 Apr;17(4):599-611. [doi: 10.1002/j.1532-2149.2012.00224.x] [Medline: 23090719]

27. Alonso-Fernández M, López-López A, Losada A, González JL, Wetherell JL. Acceptance and Commitment Therapy and Selective Optimization with Compensation for Institutionalized Older People with Chronic Pain. Pain Med 2015 Aug 25;21(1):59. [doi: 10.1111/pme.12885] [Medline: 26304771]
28. Nasiri A, Kazemi-Zahrani H. The Effectiveness of Group Acceptance and Commitment Therapy on Pain Intensity, Pain Catastrophizing and Pain-Associated Anxiety in Patients with Chronic Pain. Asian Social Science 2015;11(26):112 [FREE Full text] [doi: 10.5539/ass.v11n26p112]
29. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. Journal of Consulting and Clinical Psychology 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
30. Nordin L, Rorsman I. Cognitive behavioural therapy in multiple sclerosis: a randomized controlled pilot study of acceptance and commitment therapy. Journal of Rehabilitation Medicine 2012 Jan;44(1):87-90 [FREE Full text] [doi: 10.2340/16501977-0898] [Medline: 22234322]
31. Smout M, Hayes L, Atkins P. The empirically supported status of acceptance and commitment therapy: An update. Clinical Psychologist 2012;16(3):97-109. [doi: 10.1111/j.1742-9552.2012.00051.x]
32. Eilenberg T, Fink P, Jensen JS, Rief W, Frostholm L. Acceptance and commitment group therapy (ACT-G) for health anxiety: a randomized controlled trial. Psychol Medicine 2016 Jan;46(1):103-115. [doi: 10.1017/S0033291715001579] [Medline: 26281857]
33. Flaxman PE, Bond FW. A randomised worksite comparison of acceptance and commitment therapy and stress inoculation training. Behaviour Research and Therapy 2010 Aug;48(8):816-820. [doi: 10.1016/j.brat.2010.05.004] [Medline: 20627269]
34. McConachie DA, McKenzie K, Morris PG, Walley RM. Acceptance and mindfulness-based stress management for support staff caring for individuals with intellectual disabilities. Research in Developmental Disabilities 2014 Jun;35(6):1216-1227. [doi: 10.1016/j.ridd.2014.03.005] [Medline: 24685937]
35. Lappalainen P, Kaipainen K, Lappalainen R, Hoffrén H, Myllymäki T, Kinnunen M, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013;2(1):e1 [FREE Full text] [doi: 10.2196/resprot.2389] [Medline: 23611946]
36. Jones HA, Heffner JL, Mercer L, Wyszynski CM, Vilardaga R, Bricker JB. Web-based acceptance and commitment therapy smoking cessation treatment for smokers with depressive symptoms. Journal of Dual Diagnosis 2015;11(1):56-62 [FREE Full text] [doi: 10.1080/15504263.2014.992588] [Medline: 25671683]
37. Gifford EV, Kohlenberg BS, Hayes SC, Antonuccio DO, Piasecki MM, Rasmussen-Hall ML, et al. Acceptance-based treatment for smoking cessation. Behavior Therapy 2004;35(4):689-705. [doi: 10.1016/j.brat.2010.05.004] [Medline: 20627269]
38. Bricker JB, Bush T, Zbikowski SM, Mercer LD, Heffner JL. Randomized trial of telephone-delivered acceptance and commitment therapy versus cognitive behavioral therapy for smoking cessation: a pilot study. Nicotine & Tobacco Research 2014 Nov;16(11):1446-1454 [FREE Full text] [doi: 10.1093/ntr/ntu102] [Medline: 24935757]
39. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]
40. Juarascio A, Shaw J, Forman E, Timko CA, Herbert J, Butryn M, et al. Acceptance and commitment therapy as a novel treatment for eating disorders: an initial test of efficacy and mediation. Behavior Modification 2013 Jul;37(4):459-489. [doi: 10.1177/0145445513478633] [Medline: 23475153]
41. Higgins JPT, Greene S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. In: The Cochrane Collaboration. UK: Wiley; 2011:154.
42. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemporary Clinical Trials 2007 Feb;28(2):105-114. [doi: 10.1016/j.cct.2006.04.004] [Medline: 16807131]
43. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
44. Borenstein M, Hedges L, Higgins J, Rothstein HR. Introduction to Meta-Analysis. 1st edition. England: John Wiley & Sons; 2009.
45. Lappalainen P, Granlund A, Siltanen S, Ahonen S, Vitikainen M, Tolvanen A, et al. ACT Internet-based vs face-to-face? A randomized controlled trial of two ways to deliver Acceptance and Commitment Therapy for depressive symptoms: an 18-month follow-up. Behaviour research and therapy 2014 Oct;61:43-54. [doi: 10.1016/j.brat.2014.07.006] [Medline: 25127179]
46. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-Based Acceptance and Commitment Therapy for Depressive Symptoms With Minimal Support: A Randomized Controlled Trial. Behavior Modification 2015 Aug 6:805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
47. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-Based Self-Help for Preventing Mental Health Problems in Universities: Comparing Acceptance and Commitment Training to Mental Health Education. Journal of Clinical Psychology 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]

48. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. Journal of American College Health 2014;62(1):20-30 [FREE Full text] [doi: 10.1080/07448481.2013.843533] [Medline: 24313693]
49. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behaviour Research and Therapy 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
50. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. Journal of Behavioral Medicine 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
51. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki A, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
52. Pots WT, Fledderus M, Meulenbeek PA, ten Klooster PM, Schreurs KM, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]
53. 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]
54. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
55. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
56. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
57. Wangberg S, Bergmo TS, Johnsen J. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65.
58. 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]
59. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
60. 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]
61. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems 2009;24(1):28.
62. 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]
63. Strecher VJ. The internet: Just another smoking cessation tool? Addiction 2008;103(3):485-486. [doi: 10.1111/j.1360-0443.2008.02144.x]
64. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
65. Ost LG. Det empiriska stödet för KBT vid psykiska störningar [The empiri-cal support for CBT in psychiatric disorders]. In: KBT: Kognitiv beteendeterapi inom psykiatrin. Stockholm, Sweden: Natur och Kultur; 2006.
66. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
67. Kivi M, Eriksson MC, Hange D, Petersson EL, Vernmark K, Johansson B, et al. Internet-based therapy for mild to moderate depression in Swedish primary care: short term results from the PRIM-NET randomized controlled trial. Cogn Behav Ther 2014 Jun;43(4):289-298 [FREE Full text] [doi: 10.1080/16506073.2014.921834] [Medline: 24911260]

Abbreviations

ACBS: Association for Contextual Behavioral Science
ACT: Acceptance and commitment therapy
CMD: common mental disorder
CBT: cognitive behavior therapy

PTSD: Post-Traumatic Stress Disorder
OCD: obsessive compulsive disorder
WLC: wait list control
RCT: Randomized controlled trial
HADS: Hospital and depression scales
BDI: Beck Depression Inventory
BAI: Beck Anxiety Inventory
DASS: Depression Anxiety Stress Scales
CES-D: Center for Epidemiologic Studies Depression Scale
QOLI: Quality of Life Inventory
GHQ-12: General health Questionnaire 12 items
SCL-90: Symptom checklist 90 items
MHC-SF: Mental health Continuum Short Form
AIS: Avoidance and Inflexibility Scale
AAQ-II: Acceptance and Action Scale
FFMQ: Five Facet Mindfulness Questionnaire
PIPS: Psychological Inflexibility Scale
ELS: Engaged Living Scale

Edited by G Eysenbach; submitted 14.06.16; peer-reviewed by S Langrial, K Vernmark, E Kaiser; comments to author 13.07.16; revised version received 20.07.16; accepted 20.07.16; published 24.08.16.

Please cite as:
Brown M, Glendenning A, Hoon AE, John A
Effectiveness of Web-Delivered Acceptance and Commitment Therapy in Relation to Mental Health and Well-Being: A Systematic Review and Meta-Analysis
J Med Internet Res 2016;18(8):e221
URL: http://www.jmir.org/2016/8/e221/
doi:10.2196/jmir.6200
PMID:27558740

Original Paper

An Internet-Based Intervention for Depression in Primary Care in Spain: A Randomized Controlled Trial

Jesús Montero-Marín[1,2,3], PhD; Ricardo Araya[4], MD, PhD; María C Pérez-Yus[2,3,5], PhD; Fermín Mayoral[2,6], MD, PhD; Margalida Gili[2,7], MD, PhD; Cristina Botella[8,9], PhD; Rosa Baños[9,10], PhD; Adoración Castro[2,7], PhD; Pablo Romero-Sanchiz[6], PhD; Yolanda López-Del-Hoyo[2,3,11], PhD; Raquel Nogueira-Arjona[12], PhD; Margarita Vives[2,7], MD, PhD; Antoni Riera[8], PhD; Javier García-Campayo[2,3,13], MD, PhD

[1]Faculty of Health Sciences and Sports, University of Zaragoza, Zaragoza, Spain

[2]Primary Care Prevention and Health Promotion Research Network, RedIAPP, ISCIII, Madrid, Spain

[3]Aragon Institute for Health Research (IIS Aragon), Zaragoza, Spain

[4]London School of Hygiene and Tropical Medicine, London, United Kingdom

[5]Aragon Health Sciences Institute, Zaragoza, Spain

[6]Mental Health Clinical Management Unit, Institute of Biomedical Research of Malaga (IBIMA), Regional Universitary Hospital Carlos Haya, University of Malaga, Málaga, Spain

[7]University of Balearic Islands, Palma de Mallorca, Spain

[8]Universitat Jaume I, Castellón de la Plana, Spain

[9]Ciber Fisiopatología Obesidad y Nutrición (CB06/03) Instituto Salud Carlos III, Madrid, Spain

[10]University of Valencia, Valencia, Spain

[11]University of Zaragoza, Zaragoza, Spain

[12]University of Malaga, Málaga, Spain

[13]University Hospital Miguel Servet, Zaragoza, Spain

Corresponding Author:
María C Pérez-Yus, PhD
Aragon Health Sciences Institute
Avda. San Juan Bosco, 13. 50009. Zaragoza.
Zaragoza,
Spain
Phone: 34 (+34) 976715895
Fax: 34 (+34) 976714670
Email: mcperezy@gmail.com

Abstract

Background: Depression is the most prevalent cause of illness-induced disability worldwide. Face-to-face psychotherapeutic interventions for depression can be challenging, so there is a need for other alternatives that allow these interventions to be offered. One feasible alternative is Internet-based psychological interventions. This is the first randomized controlled trial (RCT) on the effectiveness of an Internet-based intervention on depression in primary health care in Spain.

Objective: Our aim was to compare the effectiveness of a low-intensity therapist-guided (LITG) Internet-based program and a completely self-guided (CSG) Internet-based program with improved treatment as usual (iTAU) care for depression.

Methods: Multicenter, three-arm, parallel, RCT design, carried out between November 2012 and January 2014, with a follow-up of 15 months. In total, 296 adults from primary care settings in four Spanish regions, with mild or moderate major depression, were randomized to LITG (n=96), CSG (n=98), or iTAU (n=102). Research completers at follow-up were 63.5%. The intervention was Smiling is Fun, an Internet program based on cognitive behavioral therapy. All patients received iTAU by their general practitioners. Moreover, LITG received Smiling is Fun and the possibility of psychotherapeutic support on request by email, whereas CSG received only Smiling is Fun. The main outcome was the Beck Depression Inventory-II at 3 months from baseline. Mixed-effects multilevel analysis for repeated measures were undertaken.

Results: There was no benefit for either CSG [(B coefficient=-1.15; *P*=.444)] or LITG [(B=-0.71; *P*=.634)] compared to iTAU, at 3 months. There were differences at 6 months [iTAU vs CSG (B=-4.22; *P*=.007); iTAU vs LITG (B=-4.34; *P*=.005)] and 15

months [iTAU vs CSG (B=-5.10; *P*=.001); iTAU vs LITG (B=-4.62; *P*=.002)]. There were no differences between CSG and LITG at any time. Adjusted and intention-to-treat models confirmed these findings.

Conclusions: An Internet-based intervention for depression combined with iTAU conferred a benefit over iTAU alone in the Spanish primary health care system.

Trial Registration: Clinicaltrials.gov NCT01611818; https://register.clinicaltrials.gov/prs/app/action/SelectProtocol?selectaction=Edit&uid=U0001NPQ&ts=2&cx=gctdh2&sid=S0003KJ6 (Archived by WebCite at http://www.webcitation.org/6jbsUvUDz)

(J Med Internet Res 2016;18(8):e231) doi:10.2196/jmir.5695

Introduction

Depression is the most prevalent cause of illness-induced disability worldwide [1]. It is among the most common reasons for consulting a general practitioner (GP), and it carries considerable personal and economic burden [2]. Antidepressants are a common form of treatment for depressive patients in primary care (PC) [3], but many patients would also like to receive psychotherapy [4]. Psychological treatments for depression are shown to be effective in PC, especially when GPs refer patients for treatment [5]. There is evidence that psychological treatments achieve results as effective as those achieved by antidepressant medication, reducing the number of physician consultations and hospital days, and obtaining better results in adherence, relapse prevention, and reducing chronicity [6].

Nevertheless, delivering face-to-face psychotherapeutic interventions to a population is challenging given the lack of specialized resources [7], so there is a need for other alternatives that allow these interventions to be offered. One feasible alternative is Internet-based psychological interventions [8]. The Internet offers a way of providing psychological treatments for depression [9] that may even attract people who are reluctant to use traditional mental health services [10] because of barriers such as possible stigmatization processes [11]. In general, Internet-based psychological treatments seem to be effective for the treatment of depression. Although the effects seem to be more favorable for guided or assisted interventions [12-14], stand-alone Internet-based treatments for depression have also shown to be effective [15].

Until now, there have been no studies on the effectiveness of Internet-based treatments for depression in the context of PC in Spain. Therefore, the main objective of this study was to compare the effectiveness of a low-intensity therapist-guided (LITG) Internet-based program and a completely self-guided (CSG) Internet-based program with improved treatment as usual (iTAU) care for the treatment of major depression in PC in Spain.

Methods

Hypotheses

The main hypothesis was that both Internet-based interventions, CSG and LITG, would be more effective in reducing depressive symptoms than iTAU, in the context of PC at 3 months after baseline. A secondary hypothesis was that, in the context of PC, where there is a more frequent and closer contact with the GPs, the offer of additional help is unlikely to improve outcomes when using Internet-based interventions.

Design

This study was a multicenter, three-arm, parallel, randomized controlled trial (RCT). Adults presenting with depressive symptoms in PC were randomized to receive either iTAU from their GP or an Internet-based intervention program (Smiling is Fun) for depression, in this case, either with psychotherapist support (LITG) or without it (CSG). The trial protocol of the study [16], the manual used to implement the program [17], and a study on expectations of depressed PC patients have already been published [18,19].

Recruitment of Participants and Baseline Assessment

We recruited patients with major depression, aged 18-65 years, able to understand and read Spanish, with mild or moderate severity symptoms according to the Spanish Beck Depression Inventory-II (BDI-II) (14-19: mild depression; 20-28: moderate depression) [20], with symptoms lasting longer than 2 weeks, with access to Internet at home, and having an email account. Major depression was identified using the MINI International Neuropsychiatric Interview 5.0, which can establish major depression diagnoses according to the Diagnostic and Statistical Manual, version IV (DSM-IV) and International Classification of Diseases [21,22]. We excluded patients who had been receiving any psychological treatment during the previous year, those with severe psychiatric disorder in Axis I (eg, alcohol/substances abuse or dependence, psychotic disorders, dementia), and patients with severe depression (score ≥29 on the BDI-II), who were referred by their GPs for treatment.

Participants were recruited in PC settings, between November 2012 and January 2014, in the Spanish regions of Aragon, Andalusia, the Balearic Islands, and Valencia. GPs identified potential participants through using a case-finding questionnaire. Eligible individuals were then interviewed in the clinic within the following 3 days by an independent researcher, who assessed inclusion and exclusion criteria, using the MINI psychiatric interview and other questionnaires. Informed consent to enter the trial was sought from patients who fulfilled study criteria, followed by randomization, carried out by an independent researcher. Patient safety was systematically monitored. The Ethical Review Board of the regional health authority approved the study on April 7, 2010 (ref: PI10/01083).

Randomization, Concealment, and Blinding

Participants were individually randomized using blocked randomization to one of the three groups. Blocks were administered in each of the regions, using a computer-generated

random number sequence. A person who had no other involvement in the study managed the random allocation to groups. This procedure was implemented through a remote central telephone line. The sequence was concealed until all individuals had been randomized. Although patients were not informed of the group allocation, the nature of the intervention meant that it was virtually impossible to keep this completely blind. Study personnel conducting the outcome assessments were blind to the participants' allocation.

Follow-Up

Collection of follow-up data took place between March 2013 and June 2015. Participants were assessed online at 3 (time-1), 6 (time-2), and 15 (time-3) months post-baseline assessment. These moments deviated from the registered protocol. Post-treatment evaluation was stated at 3 months after the beginning of the intervention in order not to favor participants in the intervention conditions who could take longer than the estimated intervention duration comparing to control group. By setting the same evaluation moments for all participants, we ensured that measurements would be comparable. Participants were sent an email with a link to an online platform that hosted the questionnaires. No other protocols were used to increase compliance with the research data collection, but a phone call was made before each wave assessment to increase response rates.

Control Group

All the patients included in the study (whether in the control or intervention arms) received iTAU. This treatment was provided by their GPs, who had previously received a 3-hour training program to update their knowledge on how to diagnose and treat depression in primary care, based on the National Institute for Health and Care (NICE) guidelines [23]. The training mostly dealt with the appropriate use of antidepressants. In case of suicide risk or severe social dysfunction, or if worsening of symptoms was detected, patients were referred to mental health facilities.

Intervention Groups

Smiling is Fun is an Internet-delivered, self-help program for the treatment of depression, based on similar programs that have proven effective in other countries [24]. The program consists of 10 cognitive behavioral therapy modules, covering different psychological techniques for coping with depression. These modules need to be completed in a sequential way. The program recommends working on every module for at least a week, with the following modules: (1) Medication management (psychoeducation I), (2) Sleep hygiene (psychoeducation II), (3) Motivation for change (motivation), (4) Understanding emotional problems (psychoeducation III), (5) Learning to move on (behavioural activation), (6) Learning to be flexible (cognitive therapy), (7) Learning to enjoy (positive psychology I), (8) Learning to live (positive psychology II), (9) Living and learning (positive psychology III), and (10) From now on, what else? (relapse prevention). A more specific and detailed description of the module contents can be found elsewhere [16,17].

Patients in the intervention groups were allocated to LITG or to CSG Internet-based programs. In LITG, 4 trained psychotherapists randomly contacted the patients by email to offer help with any difficulties or problems encountered when using the program. Patients could ask the psychotherapists questions or advice via email messages with a maximum of three contacts over the treatment period. They could also ask a technician for help to resolve problems of a technical nature. In CSG, there was no contact with any therapist, and only technical questions could be asked regarding the computer program.

To maximize adherence, if participants did not access the program for a week, they received an automated email encouraging them to use the program and to complete the tasks for each module. In addition, the program offered continuous feedback to the users on their progress via (1) a self-monitored activity report, providing feedback on how their mood was related to the activities performed, (2) the calendar, providing feedback about homework and tasks already completed, and (3) graphs and other feedback about activity levels, emotional distress, and negative and positive emotionality.

Among those patients on medication, GPs and patients were advised not to increase dosages in any of the three groups (iTAU, CSG, LITG), but decreasing medication was permitted.

Instruments

Demographic Variables

We gathered sociodemographic data such as age, sex, living with family or alone, level of studies (university vs secondary or less), employment (employed vs unemployed), and income according to national minimum wage, as well as clinical variables such as taking antidepressant medication (yes vs no) and the number of GP visits in the previous 12 months.

Outcomes

The Spanish version of BDI-II [19], as a continuous variable, was used as the primary outcome measure at time 1 (3 months after baseline). The BDI-II is one of the most widely used instruments to evaluate presence and severity of depressive symptoms. It is a self-reported measure, which includes the cardinal cognitive, emotional, and somatic symptoms of depression, and it can be linked to diagnosis from the DSM-IV. The studies published show good agreement between BDI-II and the clinical diagnosis of depression, and good psychometric properties for the scale [25,26]. Scores can range from 0-63.

Secondary outcomes included the visual analogue scale (VAS) of the EuroQol (EQ-5D) [27], in its Spanish version [28], and the Short Form Health Survey (SF-12v1) [29], in its Spanish version [30], as measures of health-related quality of life and functioning. The VAS is a vertical line on which the best and worst possible health states are scored 100 or zero respectively. The SF-12 scoring algorithm yields a physical component scale and a mental component scale, and both were used as continuous variables applying Spanish norms, with a mean of 50 (SD 10) [30].

Sample Size

Estimated sample size in protocol was 450 participants [16], but there were recruitment problems in one of the four participant regions, Valencia, where there were multiple stakeholders, so the recruitment was delayed for more than one year. In view of this, the steering committee of the project decided to rule out recruitment in that region, repeating the process of sample size calculation, according to the new situation. This new power calculation was based on testing differences between LITG Internet-based program and iTAU care alone, and we based our sample size estimation on an expected difference in the primary outcome of at least 0.5 standard deviation (SD) [14]. This size has been considered as a clinically relevant criterion [24,31]. Previous PC studies of depressed patients in England and Spain have found BDI-II means and SDs of 22 and 12, respectively, but as our inclusion cut-off points attenuate variability, we used an SD of 6. Thus, a difference of 3 points across these two groups was our target (around 15%). In order to detect this difference between LITG and iTAU, assuming a common SD of 6 points, a 5% significance level and a statistical power of 80%, we needed 63 subjects in each group. We expected a dropout rate of around 30% [12,32], so we inflated the numbers to reach a total sample size of around 300 patients (100 per arm). This change in the number of participants was more realistic for the new situation of recruitment.

Data Analysis

First, descriptive data were compared to assess the balance of a number of variables across arms at baseline. All analyses followed a pre-specified plan [16], based on the Consolidated Standards of Reporting Trial (CONSORT) guidelines [33] (Multimedia Appendix 1). The primary between-group analysis was carried out on an intention-to-treat basis for BDI-II total scores, using a multilevel mixed-effects analysis for repeated measures, and calculating regression coefficients (B), unadjusted and adjusted for baseline scores, sex, and age [34]. Sensitivity analyses were conducted to assess the effects of missing data. Missing values were replaced by multiple imputations based on chained equations, after ensuring that data were missing at random [35]. Secondary analysis comprised comparisons of SF-12 Mental and Physical subscale scores, as well as EuroQol VAS scores using the same analytical strategy. Effect sizes between groups were calculated by means of Hedge's *g*, and group by time interactions (3 groups and 4 time points) through chi-square tests, unadjusted and adjusted for baseline, sex, and age, with the associated degrees of freedom of (r-1) x (c-1), where r is the number of groups and c is the number of time points. We also performed a Complier Average Causal Effect (CACE) analysis to assess the impact of the number of sessions on the outcome. We theoretically defined compliance as attendance at >6 sessions, but we also performed a parallel assessment of the impact for each session added separately.

We used two-sided tests at the 5% significance level, taking into account Bonferroni's criterion whenever there were multiple comparisons. All the analyses were performed with Stata 12.

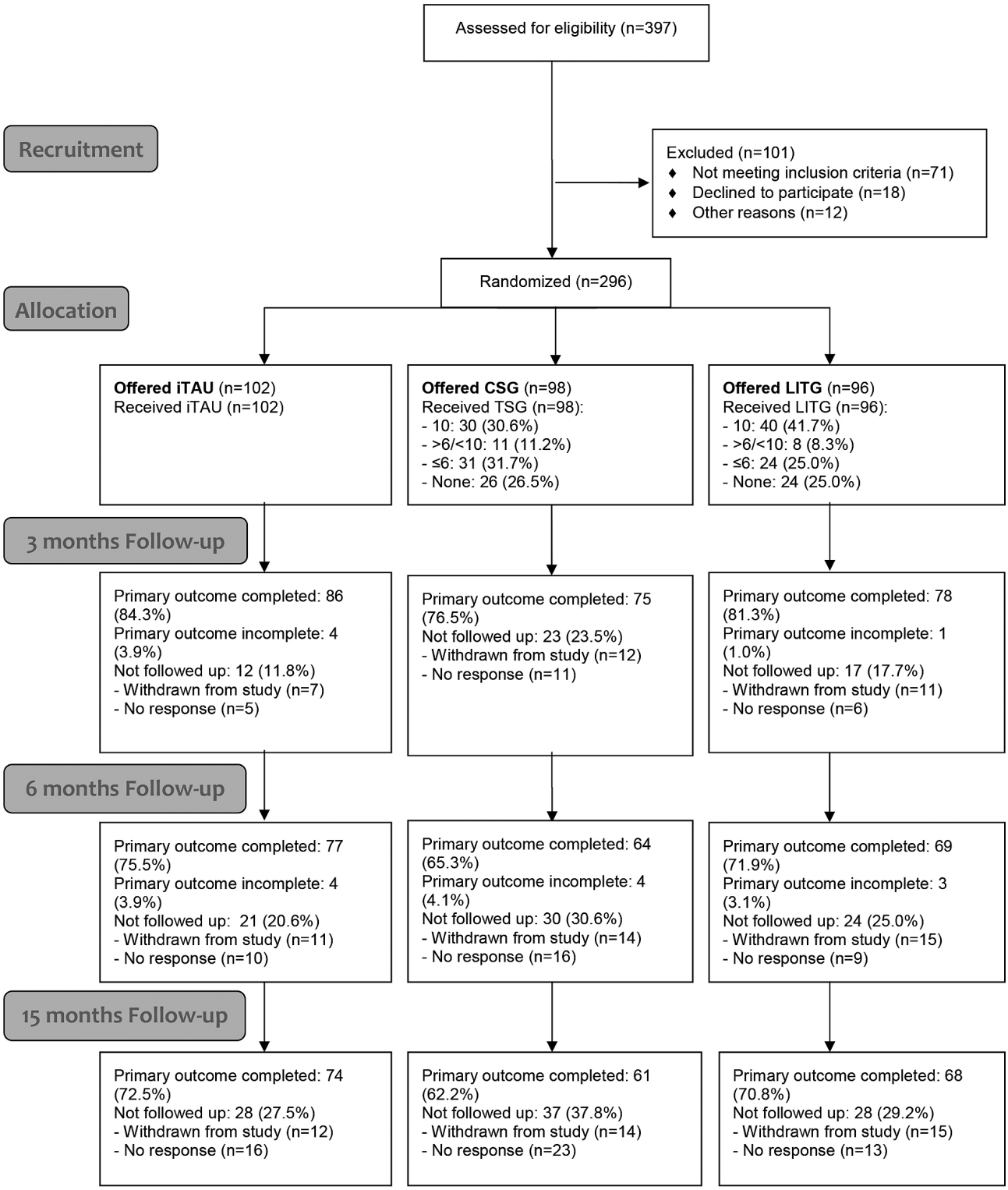
Results

A total of 46 GPs took part in the study. Of 397 potential participants, 296 were randomized (see Figure 1), with 102 allocated to iTAU, 98 to the CSG, and 96 to the LITG. The randomized groups were well balanced in all variables at baseline (Table 1). Recruitment varied across regions: 126 participants from Aragon, 106 from Andalusia, 44 from the Balearic Islands, and 20 from Valencia. Follow-up primary outcome data were obtained for 239 (80.7%) of the participants at Time 1, 210 (70.9%) at Time 2, and 203 (68.6) at Time 3. There were no significant differences among groups in terms of attrition rate (iTAU=34.3%; CSG=41.8%; LITG=33.3%; χ^2_2=0.42; *P*=.812). Only age was significantly related to attrition (at Time 3) [completers (n=203): mean 44.28 years (SD 10.22) vs missing (n=93): mean 39.99 years (SD 10.01); *P*=.001]. No baseline-level differences in other sociodemographic or primary or secondary outcomes were observed between completers and non-completers at different waves, so dropouts were considered as to be random [36]. The median depression severity across all groups in BDI-II at baseline was 23, which broadly equated with a depression of moderate severity [20].

Table 1. Baseline characteristics of participants across groups.

Characteristics at baseline	iTAU, n=102	CSG, n=98	LITG, n=96
Sociodemographics			
Age, mean (SD)	43.04 (9.66)	42.57 (11.94)	43.19 (9.30)
Sex female, n (%)	76 (74.5)	72 (73.5)	76 (79.2)
Living with family, n (%)	92 (90.2)	90 (91.8)	82 (85.4)
University education, n (%)	30 (29.4)	29 (29.6)	32 (33.3)
Employed, n (%)	54 (52.9)	51 (52.0)	52 (54.2)
Income, n (%)			
<1 national minimum wage	27 (26.5)	34 (34.7)	22 (22.9)
1-2 national minimum wage	42 (41.2)	33 (33.7)	40 (41.7)
≥3 national minimum wage	33 (32.4)	31 (31.6)	34 (35.4)
On medication, n (%)	91 (89.2)	84 (85.7)	88 (91.7)
Number of GP visits, median (Q_1-Q_3)	5 (2-8)	5 (3-10)	5 (3-8)
Clinical measures			
Depression severity			
BDI-II, mean (SD); median (Min to Max)	22.18 (5.25); 23 (14-28)	22.33 (4.85); 23 (14-28)	22.36 (4.91); 23 (14-28)
Perceived health			
EuroQol VAS, mean (SD); median (Min to Max)	57.04 (15.77); 57 (20-90)	55.45 (19.23); 50 (10-100)	56.04 (18.34); 60 (0-100)
Physical Health SF-12, mean (SD); median (Min to Max)	48.87 (11.26); 49.47 (23.52-66.60)	48.52 (11.61); 49.87 (16.76-65.59)	48.60 (11.16); 50.99 (26.12-66.43)
Mental Health SF-12, mean (SD); median (Min to Max)	28.71 (10.43); 26.68 (10.51-55.69)	28.03 (9.33); 26.16 (5.53-56.02)	28.60 (8.91); 26.83 (14.09-56.09)

Figure 1. Flow diagram.



Primary Analyses

There were no clear differences between either CSG or LITG compared with iTAU at Time 1 (Table 2). However, there were differences between iTAU versus CSG, and iTAU versus LITG at Time 2 and at Time 3, with both computerized interventions performing better than usual care. There were no significant differences between CSG and LITG at any time. The adjusted models confirmed these findings (Table 2). Models with missing values replaced through multiple imputations showed small reductions in regression coefficients, but the main differences remained unaltered: Time 2, iTAU vs CSG (B=-2.91; *P*<.001); iTAU vs LITG (B=-3.97; *P*<.001); Time 3, iTAU vs CSG (B=-3.69; *P*<.001); iTAU vs LITG (B=-4.26; *P*<.001). In keeping with the above results, there was a significant group x time interaction (unadjusted: χ^2_6=19.23; *P*=.003; adjusted: χ^2_6=21.27; *P*=.001; imputed: χ^2_6=343.16; *P*<.001). The CACE analysis and the estimated additional intervention effect per

sessions attended also showed there was an additional benefit with more sessions attended (Table 4).

Table 2. Primary outcome analysis with observed data[a].

	iTAU (a) mean (SD)	CSG (b) mean (SD)	LITG (c) mean (SD)	*g* (a-b)	*P* (a-b)	B (95% CI) (a-b)	*g* (a-c)	*P* (a-c)	B (95% CI) (a-c)	*g* (b-c)	*P* (b-c)	B (95% CI) (b-c)
BDI-II	n=67	n=57	n=64									
Time 0	21.76 (5.39)	22.59 (4.78)	21.73 (4.83)									
Time 1	17.91 (11.06)	16.59 (10.60)	17.08 (10.24)	0.12	.444	-1.15 (-4.08 to 1.79)	0.08	.634	-0.71 (-3.61 to 2.20)	-0.05	.764	0.44 (-2.45 to 3.34)
Adjusted					.359	-1.35 (-4.23 to 1.54)		.613	-0.74 (-3.60 to 2.12)		.674	0.61 (-2.23 to 3.45)
Time 2	18.12 (12.15)	14.27 (10.00)	13.56 (11.56)	0.34	.007	-4.22 (-7.28 to -1.16)	0.38	.005	-4.34 (-7.36 to -1.33)	0.07	.938	-0.12 (-3.14 to 2.90)
Adjusted					.003	-4.55 (-7.56 to -1.55)		.004	-4.31 (-7.27 to -1.35)		.862	0.26 (-2.70 to 3.22)
Time 3	16.72 (10.97)	11.53 (10.72)	11.39 (10.96)	0.48	.001	-5.10 (-8.20 to -1.99)	0.48	.003	-4.62 (-7.66 to -1.58)	0.01	.758	0.48 (-2.57 to 3.54)
Adjusted					<.001	-5.47 (-8.51 to -2.42)		.002	-4.62 (-7.61 to -1.63)		.574	0.86 (-2.14 to 3.85)

[a]*g*: Hedge's g as an effect size measure; B: regression coefficients; adjusted: adjusted analysis controlling baseline, sex, and age; a-b: iTAU vs CSG comparison; a-c: iTAU vs LITG comparison; b-c: CSG vs LITG comparison.

Secondary Analyses

Results with the Mental SF-12 showed a similar pattern of differences as in the primary outcome (Table 3). Analysis with imputed values attenuated coefficients but the main differences were maintained: Time 2, iTAU vs CSG (B=4.79; *P*<.001); iTAU vs LITG (B=4.76; *P*<.001); Time 3, ITAU vs CSG (B=4.63; *P*<.001); iTAU vs LITG (B=5.54; *P*<.001)]. Similarly, group x time interactions were observed (unadjusted: χ^2_6=27.05; *P*<.001; adjusted: χ^2_6=31.01; *P*<.001; imputed: χ^2_6=422.56; *P*<.001), and there were also dose-response effects associated with the number of sessions (Table 4). There were no differences across arms for Physical SF-12 at any time (Table 3). Some differences were found between iTAU vs LITG at Time 2 and Time 3 when using EUROQOL (Table 3), which were confirmed with adjusted and imputed models: imputed Time 1: B=2.49; *P*<.001; Time 2: B=7.26; *P*<.001; Time 3: B=7.66; *P*<.001. At Time 3, there were also EUROQOL differences between iTAU vs CSG in the adjusted model. Using imputed data, these differences were also seen at earlier times (Time 2: B=3.78; *P*<.001; Time 3: B=5.30; *P*<.001). When imputing values, we also found differences between CSG vs LITG at Time 1 (B=2.75; *P*<.001), Time 2 (B=3.48; *P*<.001), and Time 3 (B=2.35; *P*<.001). Group x time effects were not found with unadjusted data (χ^2_6=11.10; *P*=.085), but there were interactions with the adjusted (χ^2_6=12.88; *P*=.045) and imputed (χ^2_6=219.54; *P*<.001) models. Dose-response effects were also found (Table 4).

Table 3. Secondary outcome analyses with observed data[a].

	iTAU (a) mean (SD)	CSG (b) mean (SD)	LITG (c) mean (SD)	*g* (a-b)	*P* (a-b)	B (95% CI) (a-b)	*g* (a-c)	*P* (a-c)	B (95% CI) (a-c)	*g* (b-c)	*P* (b-c)	B (95% CI) (b-c)
Mental SF-12	n=64	n=55	n=64									
Time 0	29.13 (11.18)	28.59 (8.90)	27.95 (8.78)									
Time 1	35.41 (12.19)	34.72 (12.46)	36.97 (12.57)	0.06	.380	-1.83 (-5.91 to 2.25)	-0.13	.682	0.85 (-3.20 to 4.89)	-0.18	.216	2.63 (-1.54 to 6.80)
Adjusted					.341	-1.94 (-5.94 to 2.06)		.589	1.10 (-2.87 to 5.06)		.187	3.01 (-1.06 to 7.07)
Time 2	36.05 (12.38)	42.35 (11.03)	42.22 (13.24)	-0.53	.003	6.49 (2.26 to 10.72)	-0.48	.005	5.97 (1.80 to 10.13)	0.01	.801	-0.56 (-4.89 to 3.78)
Adjusted					.002	6.67 (2.53 to 10.81)		.002	6.41 (2.33- 10.49)		.863	-0.37 (-4.59 to 3.85)
Time 3	36.35 (12.12)	43.44 (11.66)	43.65 (13.41)	-0.59	.002	6.78 (2.46 to 11.11)	-0.57	.003	6.53 (2.29 to 10.76)	-0.02	.901	-0.28 (-4.69 to 4.12)
Adjusted					.001	7.03 (2.80 to 11.26)		.001	7.09 (2.95 to 11.24)		.989	-0.03 (-4.31 to 4.25)
Physical SF-12	n=64	n=55	n=64									
Time 0	48.74 (11.67)	47.83 (12.29)	47.98 (10.87)									
Time 1	47.91 (10.31)	48.84 (11.89)	49.20 (10.58)	-0.08	.443	1.13 (-1.76 – 4.02)	-0.12	.638	0.69 (-2.17 – 3.54)	-0.03	.760	-0.44 (-3.26 – 2.38)
Adjusted					.382	1.23 (-1.53 – 4.00)		.542	0.85 (-1.89 – 3.60)		.799	-0.35 (-3.02 – 2.33)
Time 2	47.56 (10.74)	47.42 (12.18)	46.87 (10.79)	0.01	.820	0.35 (-2.66 to 3.35)	0.06	.906	-0.18 (-3.13 to 2.77)	0.05	.727	-0.52 (-3.46 – 2.41)
Adjusted					.677	0.61 (-2.26 to 3.47)		.946	0.10 (-2.72 to 2.92)		.727	-0.49 (-3.27 – 2.28)
Time 3	47.53 (11.78)	47.65 (11.89)	48.05 (9.85)	-0.01	.864	0.27 (-2.81 to 3.34)	-0.05	.655	0.68 (-2.32 to 3.68)	-0.04	.781	0.42 (-2.56 to 3.41)
Adjusted					.830	0.32 (-2.61 to 3.25)		.620	0.73 (-2.14 to 3.59)		.755	0.45 (-2.37 to 3.27)
EuroQol VAS	n=66	n=57	n=65									
Time 0	57.80 (15.81)	53.56 (20.05)	57.46 (18.23)									

	iTAU (a) mean (SD)	CSG (b) mean (SD)	LITG (c) mean (SD)	*g* (a-b)	*P* (a-b)	B (95% CI) (a-b)	*g* (a-c)	*P* (a-c)	B (95% CI) (a-c)	*g* (b-c)	*P* (b-c)	B (95% CI) (b-c)
Time 1	62.12 (18.12)	63.11 (21.61)	65.85 (21.34)	-0.05	.924	0.30 (-5.91 to 6.52)	-0.19	.418	2.55 (-3.61 to 8.70)	-0.13	.495	2.24 (-4.19 to 8.66)
Adjusted					.816	0.71 (-5.32 to 6.75)		.385	2.66 (-3.33 to 8.64)		.541	1.94 (-4.28 to 8.17)
Time 2	62.33 (20.83)	65.81 (21.44)	69.83 (20.21)	-0.16	.152	4.73 (-1.73 to 11.19)	-0.36	.016	7.73 (1.38 to 14.09)	-0.19	.381	2.99 (-3.70 to 9.68)
Adjusted					.088	5.45 (-0.81 to 11.71)		.013	7.84 (1.68 to 14.01)		.478	2.34 (-4.13 to 8.81)
Time 3	62.59 (20.37)	68.89 (22.79)	72.45 (15.93)	-0.29	.028	7.39 (0.82 to 13.97)	-0.54	.014	8.10 (1.65 to 14.54)	-0.18	.842	0.69 (-6.09 to 7.47)
Adjusted					.011	8.31 (1.94 to 14.44)		.010	8.19 (1.95 to 14.44)		.970	-0.13 (-6.68 to 6.43)

[a]*g*: Hedge's g as an effect size measure; B: regression coefficients; adjusted: adjusted analysis controlling baseline, sex, and age; a-b: iTAU vs CSG comparison; a-c: iTAU vs LITG comparison; b-c: CSG vs LITG comparison.

Table 4. Dose-response in primary and secondary outcomes.

Variables	Raw[a]		Adjusted[b]		Imputed[c]	
	B	*P*	B	*P*	B	*P*
BDI-II						
CACE analysis[d]	-6.93 (-12.52 to -1.33)	.016	-7.24 (-12.33 to -2.15)	.006	-8.07 (-9.42 to -6.72)	<.001
Effect per session	-0.65 (-1.18 to -0.13)	.015	-0.68 (-1.16 to -0.20)	.006	-0.73 (-0.85 to -0.61)	<.001
Mental SF-12						
CACE analysis[d]	9.42 (3.08 to 15.75)	.004	10.27 (4.01 to 16.52)	.001	10.59 (9.05 to 12.12)	<.001
Effect per session	0.89 (0.29 to 1.48)	.004	0.97 (0.38 to 1.56)	.001	0.95 (0.82 to 1.09)	<.001
Physical SF-12						
CACE analysis[d]	0.01 (-5.71 to 5.74)	.996	0.65 (-3.85 to 5.16)	.776	0.29 (-1.07 to 1.64)	.676
Effect per session	<1.01 (-0.54 to 0.54)	.996	0.06 (-0.36 to 0.49)	.775	0.03 (-0.10 to 0.15)	.676
EuroQol VAS						
CACE analysis[d]	11.44 (0.76 to 22.12)	.036	12.18 (2.22 to 22.13)	.017	13.12 (10.58 to 15.66)	<.001
Effect per session	1.08 (0.08 to 2.07)	.034	1.15 (0.22 to 2.07)	.016	1.18(0.96 to 1.41)	<.001

[a]Using raw models.

[b]Adjusting baseline, sex and age.

[c]Using imputed data.

[d]Compliance as attendance at >6 sessions. B: regression coefficients. Controls were considered as receiving 0 sessions.

Internet-Based Program Usage

At Time 3, 72.4% (n=71) of CSG participants and 75.0% (n=72) of LITG participants had accessed the Internet-based programs (χ^2_1=0.35; *P*=.556). In CSG, the median number of modules completed was 4 (interquartile range: 0-10), with 41.8% attending >6 modules; 30 participants in this group completed sessions. In LITG, the median number of modules completed was 6 (interquartile range: 0-10), with 50.0% attending >6 modules and 40 participants completing 10 modules. There was

no significant difference between CSG and LTIG in terms of modules completed (Z=-1.20; *P*=.228). A total of 17 email contacts were made with 13 participants of the LITG program, which represented 11.9% of the patients in this group. Support requests were monitored, and a content analysis of them showed that the topics were related to (1) requesting information on depressive symptoms, (2) counseling with regard to events and difficulties in life, and (3) additional support to follow the program recommendations. As per protocol, CSG received no email contact from therapists.

Other Treatments Received

Most of the participants received drug treatment (Table 1). There were no major differences in the use of medication (as a dichotomous criterion of drug intake) across groups (iTAU: 80.6%; CSG: 68.3%; LITG: 67.6%; χ^2_2=3.67; *P*=.160). The same pattern was found for the use of mental health services, psychiatrist and/or psychologist (iTAU: 29.9%; CSG: 19.7%; LITG: 18.8%; χ^2_2=3.09; *P*=.214), and number of visits to the GP [iTAU: median=2 (Q_1=0; Q_3=4); CSG: median=1 (Q_1=0; Q_3=3); LITG: median=1 (Q_1=0; Q_3=4); χ^2_2=0.22; *P*=.643].

Discussion

Principal Findings

As far as we are aware, this is the first RCT on the effectiveness of an Internet-based intervention for the treatment of depression in PC health services in Spain. Other Web-based interventions for treating depressive symptoms in the Spanish language have been developed in Mexico [37] and Chile [38], but the efficacy of these interventions in reducing depressive symptoms has not yet been evaluated in an RCT. Therefore, this trial is novel in that it allowed the testing of a new Spanish program (Smiling is Fun), which adapted some of the techniques used in other available programs [24,39]. This trial compared two Internet-based interventions, with and without psychotherapeutic support, with usual care. We observed differences in the medium and long term in favor of psychotherapy, but not in the short term. This was somewhat consistent with previous studies on depression when comparing face-to-face psychotherapy (alone or plus usual care), with usual care based on medication treatments [40]. We found a larger effect size than in previous works, particularly in the group without additional support, and we also found that the offer of support did not yield additional benefits in terms of better adherence or outcomes, contrary to other previously published studies [41,42].

The way patients were recruited in PC settings, with face-to-face contacts in a confidential context with the involvement of the GP, might explain why our results were better than those of other trials [14]. Studies in which patients suffering from depression were referred by their GP for treatment had generally shown higher effect sizes, with values similar to our study [5]. This aspect of the trial may have also contributed to the low number of email contacts requesting assistance to the psychotherapists, and therefore, to the similarity in the results for the supported and the unsupported groups. This same result has also been found in other works [15], although effect sizes in Internet-based interventions with therapist support typically seem larger than those without therapist support [14]. In our study, the type of support received by email on request was less intense than in other efficacy trials, which additionally include a group with some kind of weekly contact [14,15]. Patients in our study did not make much use of the support offered. We do not know the real reasons for this, but the lack of initial face-to-face contact with the psychotherapist might have hampered the establishment of an adequate alliance—something that might merit interest for further studies [19].

The possibility of receiving attention and care from a GP in a PC setting could also help minimize the stigma associated with referral to mental health services [11] and could improve adherence to the program. In fact, our procedure was as close as possible to the usual practice, and thus, it might increase the likelihood that the program might be used in the PC health services in Spain, where most depressed patients are treated [43]. Nevertheless, GPs may experience difficulties in recruiting patients because of their overloaded schedule, as has been described in other countries [44].

Limitations

This study encountered several limitations. First, attrition at follow-up was significant, but replacing missing values through imputations confirmed the main findings. As shown in other similar trials of computerized interventions, attrition rates are often large [12,32]. If anything, our retention rate was probably as good, if not better, than that achieved in other similar trials. Second, the potential effect of GPs or psychotherapist was not taken into account and may have been a source of variability. Finally, this trial was not powered to detect small differences between the two computerized interventions. This part of the study needs to be tested properly in a separate trial also controlling for the possible effects of psychotherapists and GPs.

Conclusions

A Spanish-language Internet-based intervention for the treatment of depression (Smiling is Fun) added to usual care proved to be more effective than treatment as usual alone at follow-up assessments. Pending cost-effectiveness analysis, these results suggest that it might be worth investing in this program for PC clinics in Spain, and possibly in other Spanish-speaking settings. The kind of low-intensity support offered in the program did not show additional improvement on the effectiveness of the computerized intervention. It remains to be seen whether or not any other forms of online/telephone support might yield further gains.

Acknowledgments

This study was financed by the Instituto de Salud Carlos III of the Spanish Ministry of Economy and Competitiveness with the PI10/01083 grant (Eficacia y coste-efectividad de un programa de psicoterapia asistida por ordenador para el tratamiento de la

depresión mayor en atención primaria: estudio controlado, randomizado y cualitativo). The project also received funding from the Network for Prevention and Health Promotion in primary Care (RD12/0005) grant from the Instituto de Salud Carlos III of the Ministry of Economy and Competitiveness (Spain), co-financed with European Union ERDF funds.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.1 [33].

[PDF File (Adobe PDF File), 7MB - jmir_v18i8e231_app1.pdf]

References

1. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]
2. Ustün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive disorders in the year 2000. Br J Psychiatry 2004 May;184:386-392 [FREE Full text] [Medline: 15123501]
3. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
4. Lester H, Howe A. Depression in primary care: three key challenges. Postgrad Med J 2008 Oct;84(996):545-548. [doi: 10.1136/pgmj.2008.068387] [Medline: 19017840]
5. Cuijpers P, van SA, van SA, Andersson G. Psychological treatment of depression in primary care: a meta-analysis. Br J Gen Pract 2009 Feb;59(559):e51-e60 [FREE Full text] [doi: 10.3399/bjgp09X395139] [Medline: 19192368]
6. Huhn M, Tardy M, Spineli LM, Kissling W, Förstl H, Pitschel-Walz G, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014 Jun;71(6):706-715. [doi: 10.1001/jamapsychiatry.2014.112] [Medline: 24789675]
7. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17 [FREE Full text] [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
8. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014 May;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
9. van BW, Cuijpers P, van SA, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014 Jul;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
10. Cuijpers P, van SA, Warmerdam L, van Rooy MJ. Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. BMC Health Serv Res 2010;10:181 [FREE Full text] [doi: 10.1186/1472-6963-10-181] [Medline: 20579332]
11. 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]
12. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
13. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
14. Kleiboer A, Donker T, Seekles W, van SA, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
15. 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 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
16. López-del-Hoyo Y, Olivan B, Luciano JV, Mayoral F, Roca M, Gili M, et al. Low intensity vs. self-guided internet-delivered psychotherapy for major depression: a multicenter, controlled, randomized study. BMC Psychiatry 2013;13:21 [FREE Full text] [doi: 10.1186/1471-244X-13-21] [Medline: 23312003]
17. Botella C, Garcia-Palacios A, Banos R, Quero S. Tratamiento psicológico de la depresión: aplicación presencial y online. Editorial Síntesis 2016:-.

18. Montero-Marín J, Carrasco JM, Roca M, Serrano-Blanco A, Gili M, Mayoral F, et al. Expectations, experiences and attitudes of patients and primary care health professionals regarding online psychotherapeutic interventions for depression: protocol for a qualitative study. BMC Psychiatry 2013;13:64 [FREE Full text] [doi: 10.1186/1471-244X-13-64] [Medline: 23425435]
19. Montero-Marín J, Prado-Abril J, Botella C, Mayoral-Cleries F, Baños R, Herrera-Mercadal P, et al. Expectations among patients and health professionals regarding Web-based interventions for depression in primary care: a qualitative study. J Med Internet Res 2015 Mar;17(3):e67 [FREE Full text] [doi: 10.2196/jmir.3985] [Medline: 25757358]
20. Sanz J, García-Vera M, Espinosa R, Fortún M, Vázquez C. Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. Clínica y salud 2005;16(2):3.
21. Ferrando L, Bobes J, Gibert J, Soto M, Soto O. Mini-International Neuropsychiatric Interview (Versión en Español 5.0). Madrid: IAP; 1998.
22. Lecrubier Y, Sheehan D, Weiller E. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 1997;12(5):224-231.
23. García-Herrera J, Nogueras-Morillas V, Muñoz-Cobos F, Morales-Asensio J. Guía de Práctica Clínica para el tratamiento de la depresión en Atención Primaria Clinical Practice Guide on the Treatment of Depression in Primary Care. Málaga: GAEDAP; 2011. URL: http://www.guiasalud.es/GPC/GPC_488_Depresion_AP_resum.pdf [accessed 2016-08-12] [WebCite Cache ID 6jhxGQ4kx]
24. Purves DG, Bennett M, Wellman N. An open trial in the NHS of Blues Begone: a new home based computerized CBT program. Behav Cogn Psychother 2009 Oct;37(5):541-551. [doi: 10.1017/S1352465809990282] [Medline: 19703330]
25. Beck A, Steer R. BDI-II - Beck Depression Inventory Manual. Second Volume. San Antonio: The Psychological Corporation; 1996.
26. Schotte CK, Maes M, Cluydts R, De DD, Cosyns P. Construct validity of the Beck Depression Inventory in a depressive population. J Affect Disord 1997 Nov;46(2):115-125. [Medline: 9479615]
27. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199-208.
28. Badia X, DeCharro F. EuroQol, un instrumento para valorar la salud. EQ 5-D guía del usuario (Spanish version). Med Clin 1999;114:6-14.
29. Ware JJ, Kosinski M, Keller S. 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]
30. Vilagut G, Ferrer M, Rajmil L. El cuestionario de salud SF-36 español: una década de experiencia y nuevos desarrollos. Gaceta sanitaria 2005;19(2):135-150.
31. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. Br J Clin Psychol 2006 Nov;45(Pt 4):499-514. [doi: 10.1348/014466505X84782] [Medline: 17076960]
32. 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]
33. 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]
34. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. BMC Med Res Methodol 2014;14:49 [FREE Full text] [doi: 10.1186/1471-2288-14-49] [Medline: 24712304]
35. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
36. National Research Council (US) Panel on Handling Missing Data in Clinical Trials. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: National Academies Press; 2010.
37. Lara M, Tiburcio M, Aguilar A, Sánchez-Solís A. A four-year experience with a Web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica 2014;35:398-406.
38. Espinosa H, Carrasco A, Moessner M, Cáceres C, Gloger S, Rojas G, et al. Acceptability Study of “Ascenso”: An Online Program for Monitoring and Supporting Patients with Depression in Chile. Telemed J E Health 2016 Jul;22(7):577-583. [doi: 10.1089/tmj.2015.0124] [Medline: 26741190]
39. Marks IM, Cavanagh K, Gega L. Computer-aided psychotherapy: revolution or bubble? Br J Psychiatry 2007 Dec;191:471-473 [FREE Full text] [doi: 10.1192/bjp.bp.107.041152] [Medline: 18055948]
40. Spielmans GI, Berman MI, Usitalo AN. Psychotherapy versus second-generation antidepressants in the treatment of depression: a meta-analysis. J Nerv Ment Dis 2011 Mar;199(3):142-149. [doi: 10.1097/NMD.0b013e31820caefb] [Medline: 21346483]
41. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
42. 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]

43. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013 Aug;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
44. Fernández A, Saameño JAB, Pinto-Meza A, Luciano JV, Autonell J, Palao D, et al. Burden of chronic physical conditions and mental disorders in primary care. Br J Psychiatry 2010 Apr;196(4):302-309 [FREE Full text] [doi: 10.1192/bjp.bp.109.074211] [Medline: 20357307]

Abbreviations

BDI-II: Beck Depression Inventory-II
CACE: Complier Average Causal Effect
CSG: completely self-guided Internet-based program
DSM-IV: Diagnostic and Statistical Manual, version IV
GP: general practitioner
iTAU: improved treatment as usual care
LITG: low-intensity therapist-guided Internet-based program
PC: primary care
RCT: randomized controlled trial
SF-12: Short Form Health Survey

Edited by G Eysenbach; submitted 26.02.16; peer-reviewed by N Görg, D Richards, D Meyer; comments to author 24.03.16; revised version received 26.04.16; accepted 23.06.16; published 26.08.16.

Please cite as:
Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, Baños R, Castro A, Romero-Sanchiz P, López-Del-Hoyo Y, Nogueira-Arjona R, Vives M, Riera A, García-Campayo J
An Internet-Based Intervention for Depression in Primary Care in Spain: A Randomized Controlled Trial
J Med Internet Res 2016;18(8):e231
URL: http://www.jmir.org/2016/8/e231/
doi:10.2196/jmir.5695
PMID:27565118

Original Paper

Depression Screening Using Daily Mental-Health Ratings from a Smartphone Application for Breast Cancer Patients

Junetae Kim[1], BS; Sanghee Lim[2], PhD; Yul Ha Min[3], RN, PhD; Yong-Wook Shin[4], MD, PhD; Byungtae Lee[1], PhD; Guiyun Sohn[5], MD; Kyung Hae Jung[6], MD, PhD; Jae-Ho Lee[7,8,9], MD, PhD; Byung Ho Son[5], MD, PhD; Sei Hyun Ahn[5], MD, PhD; Soo-Yong Shin[8,9], PhD; Jong Won Lee[5], MD, PhD

[1]College of Business, KAIST, Seoul, Republic Of Korea

[2]Carey Business School, The Johns Hopkins University, Baltimore, MD, United States

[3]College of Nursing, Gachon University, Seongnam, Republic Of Korea

[4]Department of Psychiatry, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic Of Korea

[5]Department of Surgery, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic Of Korea

[6]Department of Oncology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic Of Korea

[7]Department of Emergency Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic Of Korea

[8]Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic Of Korea

[9]Ubiquitous Health Center, Asan Medical Center, Seoul, Republic Of Korea

Corresponding Author:
Jong Won Lee, MD, PhD
Department of Surgery
University of Ulsan College of Medicine
Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu
Seoul, 138-736
Republic Of Korea
Phone: 82 2 3010 5603
Fax: 82 2474 9027
Email: jongwonlee116@gmail.com

Abstract

Background: Mobile mental-health trackers are mobile phone apps that gather self-reported mental-health ratings from users. They have received great attention from clinicians as tools to screen for depression in individual patients. While several apps that ask simple questions using face emoticons have been developed, there has been no study examining the validity of their screening performance.

Objective: In this study, we (1) evaluate the potential of a mobile mental-health tracker that uses three daily mental-health ratings (sleep satisfaction, mood, and anxiety) as indicators for depression, (2) discuss three approaches to data processing (ratio, average, and frequency) for generating indicator variables, and (3) examine the impact of adherence on reporting using a mobile mental-health tracker and accuracy in depression screening.

Methods: We analyzed 5792 sets of daily mental-health ratings collected from 78 breast cancer patients over a 48-week period. Using the Patient Health Questionnaire-9 (PHQ-9) as the measure of true depression status, we conducted a random-effect logistic panel regression and receiver operating characteristic (ROC) analysis to evaluate the screening performance of the mobile mental-health tracker. In addition, we classified patients into two subgroups based on their adherence level (higher adherence and lower adherence) using a k-means clustering algorithm and compared the screening accuracy between the two groups.

Results: With the ratio approach, the area under the ROC curve (AUC) is 0.8012, indicating that the performance of depression screening using daily mental-health ratings gathered via mobile mental-health trackers is comparable to the results of PHQ-9 tests. Also, the AUC is significantly higher (P=.002) for the higher adherence group (AUC=0.8524) than for the lower adherence group (AUC=0.7234). This result shows that adherence to self-reporting is associated with a higher accuracy of depression screening.

Conclusions: Our results support the potential of a mobile mental-health tracker as a tool for screening for depression in practice. Also, this study provides clinicians with a guideline for generating indicator variables from daily mental-health ratings. Furthermore,

our results provide empirical evidence for the critical role of adherence to self-reporting, which represents crucial information for both doctors and patients.

(J Med Internet Res 2016;18(8):e216) doi:10.2196/jmir.5598

KEYWORDS

depression; smartphone applications; mental health; breast cancer (neoplasms)

Introduction

Mental distress can impair treatment processes and outcomes, such as adherence to treatment recommendations, satisfaction with care, and quality of life [1-3]. However, when present, depression is detected less than 30% of the time in cancer patients, mainly due to the time constraints of both patients and clinicians as well as patients' reluctance to undergo depression screening tests [4,5]. To reduce the burden on patients, researchers have developed simpler screening tools that use only one or two questions, such as the Distress Thermometer and the Patient Health Questionnaire-2 (PHQ-2) [5,6]. However, these screening methods are still problematic when dealing with patients who rarely visit a doctor, because these patients do not have a chance to be tested. To alleviate this problem, doctors recommend that patients track their Patient Reported Outcomes (PROs) on paper as a form of mental-status diary [7,8]. However, the inconvenience of keeping daily mental-health ratings on paper leads to a low rate of using such diaries [7,9].

The rapid increase in the use of mobile phones, and specifically smartphones, has prompted health care providers to consider mobile phone apps as a way to collect mental PROs. Such apps are known as mental-health trackers [10,11]. Despite the potential benefits of mental-health trackers in the setting of oncology treatment, prior studies have focused on evaluating the feasibility of data collection and overall response rates [7,10,12], with only a few studies evaluating the validity of the data in depression screening [11,13]. These validity studies show whether the mood ratings sent via text message can be used as a proxy for depression assessment [13] and whether the scores reported via mobile phones are consistent with the ones reported via a paper-based test using traditional depression screening questionnaires [11]. Therefore, it is still unclear whether the daily mental-health ratings, which are gathered using simple instruments and facial emoticon scales via mobile mental-health trackers, can be used to screen for depression for clinical purposes.

The contribution of our study is mainly threefold. First, we provide a performance evaluation of mobile mental-health trackers. Some researchers have raised concerns about using a simpler and shorter depression screening survey designed for a mobile phone, fearing that it may increase measurement errors. However, we argue that the shorter recall period involved in using a mobile mental-health tracker can compensate for potential measurement errors. Prior studies have shown that the accuracy of human memory substantially decreases as the recall period increases [14-17]. The use of mobile mental-health trackers can reduce the patient recall period since it is possible for patients to easily report mental-health ratings on a daily basis. Our results show that daily mental-health ratings collected via mobile mental-health trackers provide results comparable to those of traditional depression screening tools.

Second, we propose three data-processing approaches (average, ratio, and frequency) for generating indicator variables from daily mental-health ratings and evaluate the performance of screening accuracy among these three approaches. From a practical perspective, there has been no discussion on the approach for transforming daily mental-health ratings reported via mobile phone apps to create indicator variables used for depression screening. Although several studies have conducted analyses using cross-sectional time-series data of patients' mood ratings, these studies focus on finding factors associated with mood variation [18-20]. Thus, our proposed approaches help clinicians transform daily mental-health ratings reported via a mobile phone app to generate indicator variables for depression screening.

Last, we show the effects of adherence on screening accuracy. The vast amount of daily data collected from patients through mobile mental-health trackers can be a burden for clinicians. Therefore, it is crucial to devise a systematic approach that can automatically distinguish useful data from data that may only increase noise, bias, and variability, which are common pitfalls of mobile data [21].

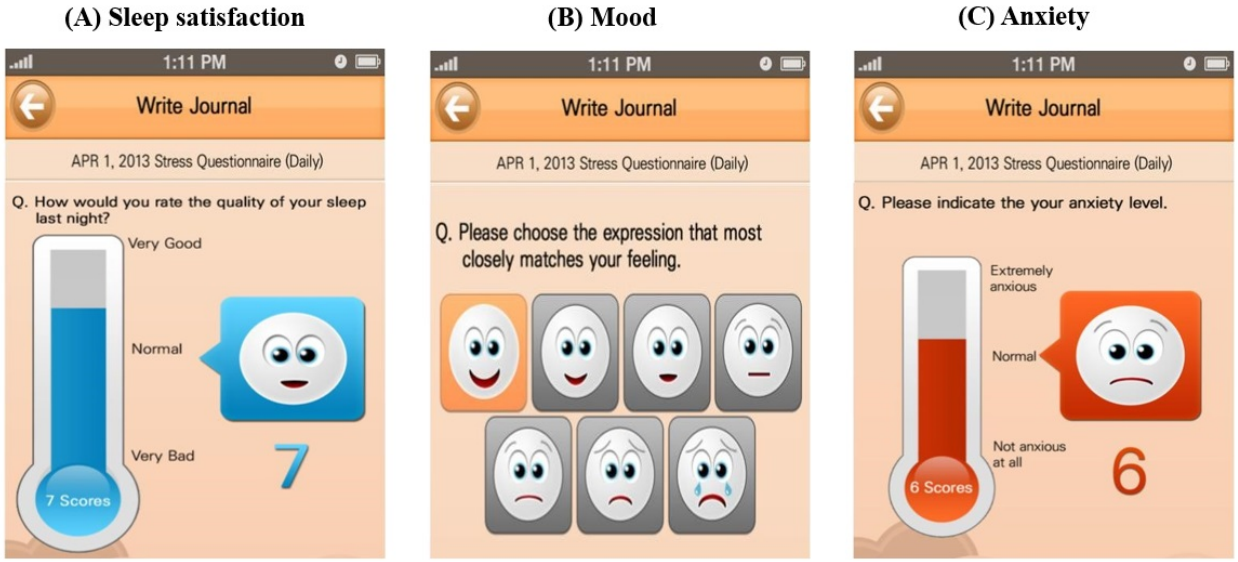
We note that adherent patients tend to make an extra effort to complete a suggested treatment plan [22], as observed in cancer patients reporting mental PROs in a supportive mental-health care setting. We expect that the patient adherence to self-reporting PROs is positively associated with the quantity and quality of data and thus increases the statistical accuracy of the model [23]. We tested this argument by categorizing patients into higher adherence and lower adherence groups and comparing their screening accuracy.

Research Setting

In early 2013, the largest hospital in South Korea developed a mobile phone app called Pit-a-Pat [10] (Figure 1), which was designed to collect PROs of breast cancer patients. The patients kept ratings on three mental-health items on a daily basis via the app: anxiety, mood, and sleep satisfaction, which have been shown to be associated with depression [24,25]. These ratings were reported using facial emoticon scales (Figure 1). Patients reported their sleep satisfaction level on a scale ranging from 0 (very bad) to 10 (very good), but this scale was reversed so that it ranged from 0 (very good) to 10 (very bad) in order to make it consistent with other measures, whose values increased with the severity of depression. Patients recorded their mood level on a scale from 0 (none) to 7 (very severe), and anxiety level on a scale from 0 (none) to 10 (very severe).

The app also administered the PHQ-9 test on a biweekly basis. This test is one of the most widely used depression screening tools in primary care settings [24,26-29].

Figure 1. Three mental logs in the Pit-a-Pat app: (A) Sleep satisfaction, (B) Mood, and (C) Anxiety.



Methods

Our analysis broadly consists of two parts. One is the evaluation of the accuracy of the mobile mental-health tracker for depression screening. The other is the examination of the effects of adherence on screening accuracy.

Evaluation of the Empirical Model for Depression Screening

Dependent Variable

Depression as a dependent variable was measured based on the PHQ-9 test results. The PHQ-9 test consists of nine items, each scored from 0-3 points, with the final score calculated as the sum of the scores for the nine items. We used a cut-off of 5 points for a depression diagnosis based on prior literature [26,27,29,30]. This relatively low cut-off value reduces the possibility that cancer patients who have depressive tendencies are classified as "normal." Several studies have reported that depression severity tends to be underestimated in the cancer treatment setting [31-33], despite the high cost of failing to detect depression due to its negative impact on health outcomes [2,3]. For this reason, in the depression treatment setting, researchers put more emphasis on improving a true-positive rate rather than a true-negative rate [34] because it is far more important to correctly identify depressed people than to correctly identify people without depression.

Three Data-Processing Approaches for Generating Indicator Variables

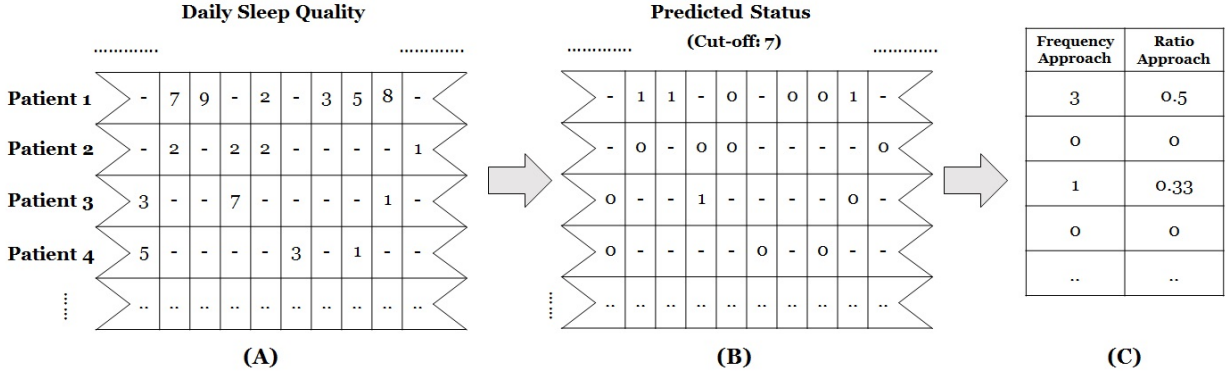
We used three approaches to construct the biweekly indicator variables from daily mental-health ratings: (1) average, (2) frequency, and (3) ratio. We generated the indicator variables by making daily mental-health ratings line up with the time interval of PHQ-9 questionnaires, which was biweekly. For instance, using the daily mental-health ratings reported from April 1-14, 2015, we generated the indicator variables and matched them with the PHQ-9 score measured on April 14, 2015, which recorded a patient's depressive tendencies during the same period.

The average approach calculates the average of each mental-health rating of a patient in a biweekly period. The average approach is easy for doctors to implement because it generates a continuous variable and does not require the doctor to calculate the optimal cut-off values [35]. However, practical guidelines suggest measuring the severity of depression by counting the number of days that people have depressive tendencies during specified periods [24,34].

For this reason, we tested two additional approaches: the frequency approach, which counts the number of depressed days during a 2-week period, and the ratio approach, which calculates the ratio of the number of days when the score indicates depression to the total number of days that the ratings are reported during the 2-week period. To construct these indicator variables, we followed two steps to transform the daily mental-health ratings into discrete scales (Figure 2). We first assigned a score of 1 to the days when the reported score on a given day was above a certain cut-off value. For example, if the score for sleep satisfaction on a particular day was higher than the cut-off value (eg, 7 points), we considered the patient to be depressed on that day and assigned the day a value of 1 (Figure 2, Step 2). Thus, each day was assigned a binary value (1=depressed, 0=normal) to indicate whether or not a patient was depressed on that day. Second, the biweekly depressed status was calculated based on each approach (Figure 2, Step 3). For instance, if there are three values of 1 and three values of 0 during a given period (6 days), the indicator variables calculated by the frequency and ratio approaches are 3 and 0.5, respectively. We applied this procedure to all types of daily mental-health ratings. Then, we conducted an analysis on each

approach to evaluate the screening performance of the three approaches.

Figure 2. Illustration of data conversion from daily mental-health logs into biweekly indicators with frequency and ratio approaches: (A) Daily scores of sleep quality during 2 weeks, (B) assigned scores of 1 to the days when the reported score is higher than the cut-off value, (C) calculated scores in a biweekly format.



Model Specification

Our model is designed to identify depression using indicator variables generated from three types of mental-health ratings:

$$\text{Depressed}_{i,t} = \text{Sleep}_{i,t} + \text{Mood}_{i,t} + \text{Anxiety}_{i,t} + e_{i,t}$$

where subscripts *i* and *t* indicate each patient and each biweekly period, respectively. The dependent variable, Depressed, takes a binary value (0=normal, 1=depressed). Because we are primarily interested in the extent to which daily mental-health ratings can identify depression, we did not include control variables in the main model. The model parameters were estimated using a random-effect logistic regression model [36-38] (see Multimedia Appendix 1). Logistic regression is one of the classifier models used for estimating the probability of a binary dependent variable based on indicator (ie, independent) variables. We used a random-effect model instead of a fixed-effect model due to the estimation efficiency of the former. Our dataset is a short-panel set, meaning that the number of patients is far greater than the number of time stamps of the observations. Estimation efficiency can be an issue with a fixed-effect model because a fixed-effect model should estimate the parameters of the dummy variables, whose number is the same as the number of patients in our sample. Moreover, our dataset contains some patients who reported a PHQ-9 test result just once during the study period. These patients would be excluded in the analysis of a fixed-effect model. Therefore, a random-effect model is preferred for our situation.

Receiver Operating Characteristic Analysis

Receiver operating characteristic (ROC) analysis is used to evaluate the screening accuracy of our model. ROC is a graphic plot, which is widely used to demonstrate the prediction accuracy of a classifier model. It plots the true positive rate (ie, sensitivity) against the false positive rate (ie, 1-specificity) at various threshold values (0< values <1) of predicted probability calculated based on logistic regression models. The area under the ROC curve, referred to as the area under the curve (AUC), represents the probability that a classifier model ranks a positive case higher than a negative case. Therefore, a higher AUC implies a better prediction performance of a classifier model. Rough criteria for assessing the performance of ROCs note that having an AUC higher than 0.7 is considered to be clinically acceptable [39].

Procedures for Calculating Cut-Off Points of Each Mental-Health Rating

As discussed earlier, we constructed dichotomy variables for the ratio and the frequency approaches. Dichotomy variables need optimal cut-off values [35] because patients' mental states should be dichotomized into two groups. We calculated the optimal cut-off value by simulating models. First, we predicted a patient's mental status by using each daily rating item with an arbitrary cut-off value. Then, we compared the AUCs of all possible cut-off values and selected the one that gives the highest AUC as the optimal cut-off. For example, to determine the cut-off for the Anxiety variable, we (1) selected an arbitrary cut-off value, (2) calculated Anxiety based on the ratio or the frequency approaches (Step 2 in Figure 2), (3) estimated a simplified model, $\text{Depressed}_{it} = \text{Anxiety}_{it} + e_{it}$, and (4) calculated the AUC through ROC analysis. As Anxiety can take a value from 0 to 10, we repeated this procedure 10 times and then selected the cut-off value that produced the highest AUC.

Robustness Analyses of Mental-Health Ratings in Detecting Depression

The primary purpose of our empirical analysis is to test the performance of depression screening. Thus, the consistency of the screening accuracy of our model is important. We conducted two additional analyses to ensure the robustness of our results.

First, we conducted a robustness analysis to validate our model by employing the five-fold cross-validation procedure. We (1) randomly partitioned the data into five subsets where the sample size is approximately 100, (2) calculated each cut-off value of indicator variables using four of the subsets as a training set, (3) generated indicator variables for the ratio and frequency approaches, (4) ran a random effect logistic regression using a training set, (5) calculated the predicted probability for the remaining subset as a test dataset, and (6) employed ROC

analysis and calculated the AUC. Steps 2-5 were repeated five times by alternating training and test datasets.

Second, one may be concerned with potential bias if patients submitted mental-health ratings on a day they took a PHQ-9 test. Thus, we conducted the analyses using a subsample that excluded the daily ratings reported on the days when a PHQ-9 test was taken.

Evaluation of Adherence Impact on Screening Accuracy

Conceptualization of Adherence to PROs as Composite Construct

In prior studies on adherence to self-reporting in the mobile and Internet health care setting, researchers tended to measure adherence with only one dimension—the response rate to technology during a given period [10,40]. This practice may be too simple because of the multidimensional characteristics of adherence [22,41-44]. We classified the patients into higher and lower adherence groups based on three dimensions: (1) activeness, (2) timeliness, and (3) persistence. Activeness refers to the degree to which the activities of a patient adhere to particular guidelines [22,44]. We calculated Activeness as the total number of days when daily mental-health ratings were reported. For Timeliness, which captures the noncompliance of a patient with treatment plans in the literature [41,42], we counted the total number of days when ratings were reported without delay because the app allowed users to submit ratings for the past few days. Persistence, defined as continuous involvement with clinical treatment during the prescribed period [43], was measured with two variables: (1) the number of biweekly periods between the first and last days in which the patient reported daily ratings (ie, the total duration), and (2) the total number of biweekly periods with reported ratings. The total duration is an important dimension of Persistence because it captures a discontinuity effect if patients stop using the app after a few weeks. However, there could be cases where a patient submits only two ratings, one very early and the other much later in the study period. Therefore, we also considered the total number of biweekly periods with reported ratings. This measure is still different from Activeness because it captures the lower adherence of patients who report ratings very actively during the first few weeks and then subsequently use the app rarely.

We considered a patient's adherence to using a mobile mental-health tracker as a composite construct of these three factors. These factors (activeness, timeliness, and persistence) address different aspects of adherence, and the relative importance of the three is unclear. Moreover, the way patients adhere to using mobile mental-health trackers can vary depending on their personalities.

K-Means Clustering Analysis and Receiver Operating Characteristic Comparison Test

To classify patients based on their adherence levels, we used a *k*-means clustering algorithm (see Multimedia Appendix 1) [45,46]. The *k*-means clustering classifies subjects into homogeneous subgroups, where each observation belongs to the cluster with the smallest intracluster distance and the largest intercluster distance. The number of clusters can be determined based on statistical criteria, such as the Akaike Information Criterion (AIC) [47]. However, a statistical approach often returns so many clusters that it becomes complicated to interpret the characteristics of clusters. For this reason, a researcher's judgment is also often used [46]. We classified patients into two clusters (ie, higher- and lower-adherent patients) for easier interpretation of results. We also examined the results with three clusters as a robustness check, as we describe below.

After patients were classified into higher- and lower-adherence groups, we compared the AUCs of each group in an ROC analysis. In addition, to support our approach to measure adherence as a multidimensional variable, we compared the AUCs of high- and low-adherence groups when the groups are classified based on prior studies [10,40], using the response rate only when the groups are classified based on our approach using activeness, timeliness, and persistence.

Robustness Analyses of Impact of Adherence on Screening Accuracy

To test the robustness of our finding that the screening accuracy is higher for patients with a higher level of adherence, we examined two potential sources of bias that may affect our ROC comparison tests—the length of data collection periods by patients and the number of clusters.

First, we examined whether the difference in the length of data collection periods by patients influences the results. Because each patient started using the app at a different time during the study period, the measure of persistence can be biased for patients who started using the app very early or very late in the study period. For example, persistence can be underestimated for patients who started using the app later in the study period. Likewise, persistence can be overestimated for patients who started using it earlier. Therefore, we examined whether our results are robust if we consider only the rating data collected during the first 24-week period (ie, half of the total study period) for each patient. We also analyzed the subsample excluding patients who joined the study during the last 12 weeks, which was the average usage period of the patients in our sample.

Second, we examined whether the results are maintained when patients are classified into three groups instead of two. This analysis also allows us to address the possibility of outliers in each group (high and low) who may have skewed the results.

Results

Sample Description

A total of 85 breast cancer patients provided informed consent to participate in this study (Institutional Review Board No. 2012-0709). These patients submitted 5817 daily mental-health ratings from early April 2013 to late March 2014. We excluded 25 ratings reported by 7 patients who did not complete a PHQ-9 test. As a result, 5792 daily mental-health ratings reported by 78 patients during 24 biweekly periods were used for our analysis. The 78 patients in our sample provided 497 PHQ-9 test results, which consisted of 270 normal statuses and 227 depressed statuses when using a cut-off score of 5 [26,27,29,30].

On average, there were 6.4 observations per patient, with the number of observations ranging from 1 (n=11) to 24 (n=1). The cumulative percentage of the number of days in which patients report ratings at least 11 days during 2 weeks (14 days) is 65.59%. Table 1 shows the demographic information for the 78 patients in our sample.

Table 1. Participant characteristics in the two study groups.

Characteristic		Total, n (%) or mean (SD) (n=78)	Lower adherence, n (n=58)	Higher adherence, n (n=20)	*P* [a]
Age, years					
	Mean (SD)	44.35 (7.01)	44.24 (7.07)	44.65 (7.02)	.83 (*t*)
	≤39	18 (23.1%)	14	4	
	40-49	40 (51.3%)	28	12	.66 (χ^2)
	≥50	20 (25.6%)	16	4	
Cohabitation [b]					
	No	17 (21.8%)	11	6	.31 (χ^2)
	Yes	61 (78.2%)	47	14	
Children					
	None	12 (15.4%)	10	2	
	1	17 (21.8%)	12	5	.45 (χ^2)
	2	43 (55.1%)	33	10	
	3 or 4	6 (7.7%)	3	3	
Marital status					
	Divorced	3 (3.8%)	3	0	
	Single	6 (7.7%)	5	1	.49 (χ^2)
	Married	69 (88.5%)	50	19	
Education level					
	Up to high school	37 (47.4%)	29	8	.44 (χ^2)
	College degree or higher	41 (52.6%)	29	12	
Employed					
	Yes	46 (59.0%)	32	14	.25 (χ^2)
	No	32 (41.0%)	26	6	
Adherence dimension [c]					
	Activeness	68.55 (60.06)	37.66	158.15	<.001 (*F*)
	Timeliness	51.31 (44.08)	29.81	113.65	<.001 (*F*)
	Duration	6.45 (5.67)	3.79	14.7	<.001 (*F*)
	Persistence	6.58 (5.76)	3.67	14.5	<.001 (*F*)

[a]Tested null hypotheses: *t* test: lower and higher adherence groups have the same mean; χ^2 test: characteristic categories and adherence groups are independent; *F* test: lower and higher adherence groups have the same mean.

[b]Cohabitation refers to patients living with family members.

[c]Key variables by groups are classified using *k*-means clustering (see Results section).

Table 2 shows the summary statistics of the daily mental-health ratings and indicator variables, which are obtained based on the ratio approach. Also, the calculated optimal cut-off values for each indicator variable of the ratio and the frequency approaches are listed in Table 2. For both the ratio and frequency approaches, the cut-off scores for sleep, anxiety, and mood were identified as 7, 6, and 4 points, respectively. Table 3 shows the correlation matrix of these variables.

Table 2. Summary statistics of daily mental-health ratings and indicator variables based on the ratio approach.

	n	Mean	SD	Min.	Med.	Max.	Skew.	Kurt.	Cut-off[d]
Sleep rating[a,b]	5792	4.99	2.03	1.00	5	10	–0.10	2.685	
Mood rating[a]	5792	3.19	1.29	1.00	3	7	0.372	2.832	
Anxiety rating[a]	5792	4.21	2.07	0.00	5	10	–0.075	2.295	
Depressed	497	0.46	0.50	0.00	0.00	1.00	0.174	1.030	5
Sleep[b,c]	497	0.21	0.23	0.00	0.14	1.00	1.167	3.951	7
Mood[c]	497	0.40	0.38	0.00	0.33	1.00	0.388	1.590	4
Anxiety[c]	497	0.30	0.32	0.00	0.2	1.00	0.801	2.474	6

[a]Daily mental-health ratings.

[b]Sleep rating and Sleep indicate sleep dissatisfaction (the reversed scale of sleep satisfaction).

[c]The indicator variables based on the ratio approach.

[d]The cut-off value of Depressed is selected based on prior literature. The cut-off values for Sleep, Mood, and Anxiety are calculated based on the simulation analysis described in the Methods section. The cut-off values obtained by using the frequency and average approaches are the same.

Table 3. Correlation matrix of daily mental-health ratings and indicator variables by the ratio approach.

	Depressed	Sleep rating	Mood rating	Anxiety rating
Sleep rating[a,b]		1		
Mood rating[a]		0.62	1	
Anxiety rating[a]		0.48	0.61	1
	Depressed	Sleep	Mood	Anxiety
Depressed	1.00			
Sleep[b,c]	0.36	1.00		
Mood[c]	0.42	0.38	1.00	
Anxiety[c]	0.40	0.22	0.47	1.00

[a]Correlation matrix of daily mental-health ratings.

[b]Correlation matrix of indicator variables by the ratio approach.

[c]Sleep rating and Sleep indicate sleep dissatisfaction (the reversed scale of sleep satisfaction).

Evaluation of the Empirical Model for Depression Screening

Performance of Each Approach to Data Processing in Detecting Depression

Table 4 presents the results of our model with three different approaches to constructing indicator variables. With the ratio approach, all three types of mental-health ratings are statistically significant ($P \leq .001$) in predicting the mental status of patients. This result indicates that each type of mental-health rating addresses different dimensions of patient mental status. For example, consider the case in which two patients reported the same level of anxiety and mood condition but a differing sleep condition. Our result suggests that holding other variables fixed, a one-tenth-unit (0.1) increase in Sleep (ie, an increase in the ratio of depressed days to the total number of reported days in a given biweekly period by 0.1) is associated with a 31.3% increase in the likelihood of the patient being depressed, since exp(0.272)=1.313. Similarly, all other things being equal, a one-tenth-unit increase in Mood and Anxiety is associated with about a 19% increase in the likelihood of the patient being depressed. Likewise, all three types of mental-health ratings with the average approach are statistically significant ($P<.05$), and a one-tenth-unit increase in Sleep , Mood , and Anxiety increases the odds of the patient being depressed by 4%, 7%, and 4%, respectively. The Sleep and Mood ratings with the frequency approach are statistically significant ($P<.05$). However, the Anxiety rating is not significant. A one-tenth-unit increase in Sleep , Mood, and Anxiety in the frequency model is related to increases of about 1%, 2%, and 1%, respectively, in the likelihood of the patient being depressed.

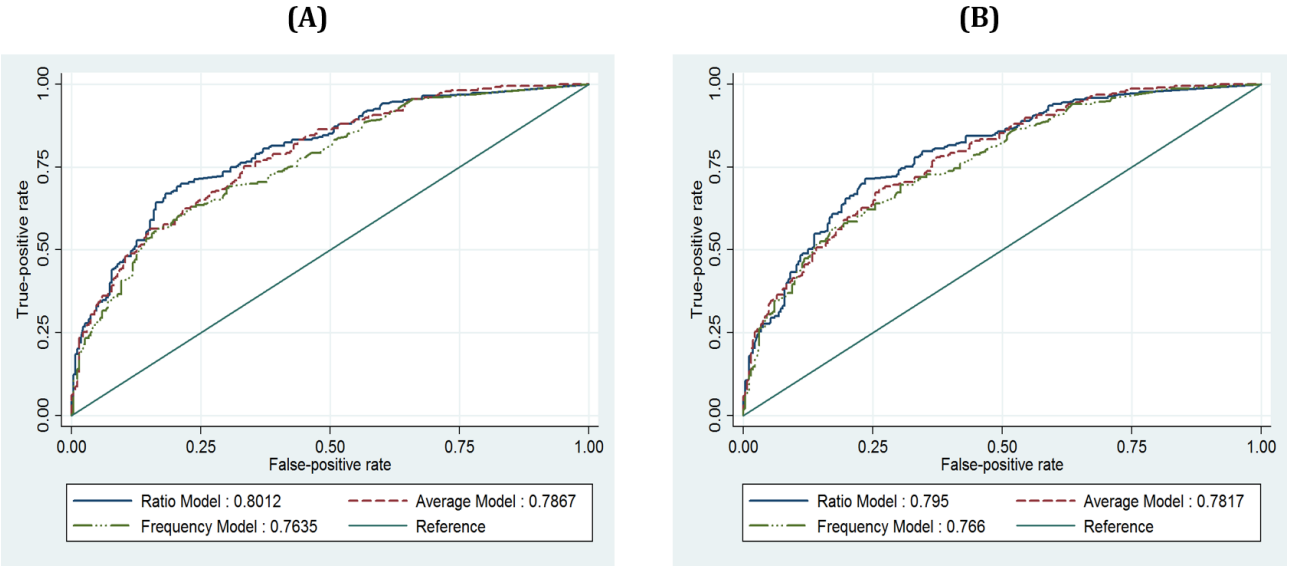
Table 4. Results of random effect logistic panel regression[a](the 497 observations were constructed from the 5792 daily mental-health ratings reported via the mental-health tracker).

	Ratio (*P*)	Average (*P*)	Frequency (*P*)
Sleep	2.722 (<.001)	0.348 (.036)	0.139 (.046)
Mood	1.783 (.001)	0.728 (.004)	0.177 (.001)
Anxiety	1.782 (.001)	0.396 (.005)	0.080 (.133)
Constant	–1.965 (<.001)	–6.002 (<.001)	–1.404 (<.001)
Observations, n	497	497	497
Patients, n	78	78	78

[a]Dependent variable: Mental status, which is 0 if normal (PHQ-9 score <5) and 1 if depressed (PHQ-9 score ≥5).

Chart A in Figure 3 shows the ROC curves of the predicted results of three models and the corresponding AUCs. The AUCs calculated from the ROCs of the ratio, the average, and the frequency approaches are 0.8012, 0.7867, and 0.7635, respectively. The AUC of the ratio approach is not statistically different (*P*=.150) from that of the average approach, but it differs significantly (*P*=.001) from that of the frequency approach. This result shows that the accuracy of depression screening by using the ratio and the average approaches is statistically indifferent, while the accuracy by using the frequency approach is slightly lower in our empirical result.

Figure 3. Results of ROC analysis: (A) ROC curves calculated from three models (full samples), (B) ROC curves calculated from three models (subsample excluding the daily logs reported on the day the PHQ-9 is administered).



Robustness Analyses of Mental-Health Ratings in Detecting Depression

First, we conducted a five-fold cross-validation test. With the ratio approach, the AUCs of the five subsets range from 0.7228 to 0.8568. The aggregated result of five subsets yields an AUC of 0.7836. With the average approach, the AUCs range from 0.7234 to 0.8488, and the AUC of the aggregated result is 0.7755. The AUCs of the frequency approach range from 0.7107 to 0.8188, and the AUC of the aggregated result is 0.7385. These results suggest that the risk of overfitting is low for our model according to rough criteria that AUCs higher than 0.7 are clinically acceptable [39].

Second, we conducted the analyses using a subsample excluding the daily ratings reported on the day the PHQ-9 is administered. The subsample consists of 480 observations taken from 5022 daily ratings, which still leaves a sufficient number of daily ratings for our analysis. All coefficients of the new analysis results using the subsample with the three approaches are statistically significant, with similar magnitude (*P*<.05), thus confirming the results of the main analysis. Chart B in Figure 3 shows the ROC curves and corresponding AUCs. The resulting AUC is 0.795 with the ratio approach, 0.7817 with the average approach, and 0.766 with the frequency approach. The difference between the AUCs of the ratio and of the average approach is not statistically significant (*P*=.197), but the one between the ratio and the frequency approach is statistically significant (*P*=.02). This result is consistent with the main analysis, indicating that the accuracy of depression screening of all three approaches is acceptable (AUC >0.7), while the results using the ratio and the average approaches are statistically higher than that of the frequency approach.

Evaluation of Adherence Impact on Screening Accuracy

Effect of Adherence to Self-Reported PROs on Screening Accuracy Based on Composite Construct

Table 1 provides descriptive statistics of four variables (ie, Activeness, Timeliness, Duration, and Persistence) used to determine a patient's adherence level. Among the 78 study patients, 58 and 20 were classified into the lower and higher adherence groups, respectively. The 497 observations in the biweekly panel dataset comprised 208 observations in the lower adherence group and 289 observations in the higher adherence group. The analysis of variance (ANOVA) test results show that the differences in the means of the four variables between the two are statistically significant (P<.001). Also, we conducted a t test and a Pearson's chi-square test to examine whether the classification result was associated with other latent factors (see Table 1), such as a patient's baseline severity of depression and demographics. The results show that the demographic variables are not significantly related to the adherence level (P>.05).

Figure 4 and Table 5 (see Composite construct) show the results of the ROC comparison test by adherence level measured according to our suggestion. These results show the AUCs calculated by the ratio (higher: 0.8524, lower: 0.7234), the average (higher: 0.8425, lower: 0.7016), and the frequency (higher: 0.8529, lower: 0.6664) approaches, respectively. All AUCs of the higher adherence group are statistically higher (P<.01) than those of the lower adherence group. These results support our argument that adherence to self-reporting is associated with an increase in the accuracy of depression screening.

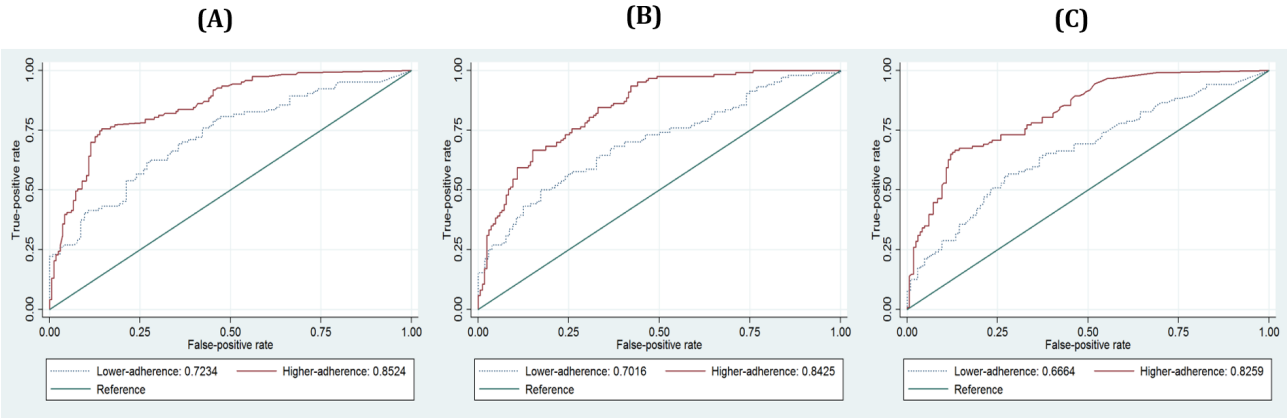
Table 5. Result for ROC comparisons of subsamples by adherence level (null hypothesis: χ^2test, AUCs of higher and lower adherence groups are the same).

Layers	Adherence (n=number of observations)	Ratio approach		Average approach		Frequency approach	
		AUC	*P*	AUC	*P*	AUC	*P*
Composite construct[a]	Lower (n=208)	0.7234	.002	0.7016	.001	0.6664	<.001
	Higher (n=289)	0.8524		0.8425		0.8259	
Prior method[b]	Lower (n=138)	0.7594	.269	0.7290	.104	0.6588	.002
	Higher (n=359)	0.8113		0.8076		0.8198	

[a]Adherence is clustered based on composite constructs of three factors: Activeness, timeliness, and persistence.

[b]Adherence is clustered based on the response rate during 2 weeks.

Figure 4. Graphs for ROC comparisons of subsamples by adherence level: (A) ROCs by adherence levels with the ratio approach, (B) ROCs by adherence levels with the average approach, (C) ROCs by adherence levels with the frequency model.



Effect of Adherence to Self-Reported PROs on Screening Accuracy Based on Prior Method

Table 5 (see Prior Method section) presents the result of the comparison of the AUCs when the adherence level is measured based on the response rate during a 2-week period (ie, the number of days when ratings are reported, to 14 days) following prior studies [10,40]. When we use the response rate only, the results show the AUCs calculated by the ratio (higher: 0.8113, lower: 0.7594), the average (higher: 0.8076, lower: 0.7290), and the frequency (higher: 0.8198, lower: 0.6588) approaches. While the comparison of AUCs with the frequency approach is statistically different between high- and low-adherence groups (P<.01), those with the ratio and the average approaches are not significantly different (P>0.1). These results show that adherence measured based on only one dimension, the response rate, is not sufficient to distinguish two groups that produce different qualities of PROs in terms of screening accuracy. On the other hand, our approach to measure adherence using three dimensions classifies patients into two distinct groups, supporting our suggestion.

Robustness Analyses of Impact of Adherence on Screening Accuracy

First, we examined whether our results are robust if we consider only the PROs collected during the first 24-week period (ie,

half of the total study period) for each patient (see Table 6). This subsample analysis shows that screening accuracy with all three approaches is statistically higher (P<.05) for patients in the higher adherence group than for the ones in the lower adherence group. The analysis results with the subsample excluding patients who joined the study during the last 12 weeks are also consistent with our main results (P<.05).

Second, we examined whether our results are maintained when patients are clustered into three groups. The ANOVA test results support that the differences between three groups are statistically significant (P<.001). These results show that the screening accuracy is higher for patients with a higher level of adherence (P<.05) (see Table 6).

Table 6. Results obtained in the robustness analysis.

Layers	Adherence (n=number of observations)	Ratio approach		Average approach		Frequency approach	
		AUC	*P*	AUC	*P*	AUC	*P*
6 months[a]	Lower (n=161)	0.728	.016	0.7239	.033	0.6774	.007
	Higher (n=336)	0.8364		0.8205		0.8088	
Without late starters[a]	Lower (n=171)	0.7405	.006	0.7015	.001	0.6796	.002
	Higher (n=273)	0.8599		0.8540		0.8283	
3 groups[b]	Lower (n=113)	0.6767	.006	0.7134	.020	0.6003	.002
	Middle (n=159)	0.7893		0.7542		0.7986	
	Higher (n=225)	0.8512		0.8446		0.8114	

[a]Null hypothesis: χ^2test. The AUCs of higher and lower adherence groups are the same.

[b]Null hypothesis: χ^2test. The AUCs of higher, middle, and lower adherence groups are the same.

Discussion

Principal Findings

This study provides several academic implications as well as important practical implications for health care providers and patients. First, this study is the first attempt to examine the depression screening performance of mobile mental-health trackers, which collect daily mental-health ratings from patients. Our results show that the depression screening performance of mobile mental-health trackers is comparable to the traditional method, administration of a PHQ-9 test, in the clinical setting.

There may be a concern about collecting data with a small number of questions related to depressive tendencies. However, based on our findings, we argue that the portability of mobile phones, which enable patients to report their mental-health ratings on a daily basis, compensates for this disadvantage. The memory recall issue is particularly critical for cancer patients because their mental status is often unstable due to the side effects of cancer treatment [48,49]. A shorter recall period may reduce the potential for measurement errors and offset the limitation of a shorter survey. Another concern may involve the inconvenience for patients to report mental-health ratings every day. However, the use of a facial emoticon scale should reduce this inconvenience. Prior studies in psychology suggest that using a face emoticon scale demands less cognitive effort and is less of a burden to patients in interpreting questionnaire items [50,51]. Also, a face emoticon scale can actually make participation in the survey more enjoyable [52]. Therefore, the use of a face emoticon scale adapted to the small phone screen may facilitate user participation, potentially making the data more useful.

Second, this study provides empirical evidence that patient adherence to self-reporting via mobile mental-health trackers has a positive effect on the depression screening accuracy. For clinicians, it might be inefficient to analyze a large number of PROs obtained from various sources such as mobile or wearable devices because these data are susceptible to common instrumentation pitfalls such as noise, bias, and variability [21]. To design a systematic approach to distinguish meaningful PROs from noises, we employed the concept of adherence because it is known that adherent patients tend to make more efforts to successfully comply with suggested treatment guidelines [22]. The PROs reported by patients with higher adherence tend to be of high quality and quantity, and our results show that the accuracy of depression screening is higher for those patients [23].

Third, we provide a new perspective on measuring adherence to self-reporting as a multidimensional construct, consisting of activeness, timeliness, and persistence. Prior empirical studies of adherence to mobile PROs have tended to focus on the total number of PROs only (activeness) without considering that the overall adherence level can decrease over time [10,40]. By incorporating the degree of patient autonomy in timely reporting (timeliness) over the entire treatment period (persistence), this new measurement allows us to capture the time effects over both short- and long-term horizons. Our empirical analysis shows that the difference in screening accuracy between high- and low-adherence groups is clearer when the groups are classified based on three dimensions (activeness, timeliness, persistence) than when they are classified based on activeness only, supporting our argument that time dimensions are also important aspects of patient adherence.

Fourth, our results also have important implications for patients. Reporting daily mental-health ratings can be a significant burden for patients and can have an adverse effect on their mental status [53,54]. However, these burdens may be reduced if patients recognize the clinical benefits of reporting their outcomes (ie,

PRO) [55]. Our results can help patients understand the positive effects of adherence and provide motivation for them to adhere to self-reporting and thereby improve the quality of treatment.

Practical Implications

Our study provides clinicians with a practical guideline for transforming daily mental-health ratings reported via a mobile mental-health tracker to generate indicator variables for depression screening. As new technologies generate new types of data, doctors are challenged to deal with them. For instance, they may have questions such as "Should a variable be dichotomized? How should we determine the cut-off values?" and "How should we transform daily PROs into a biweekly format? Is a ratio approach better than a frequency approach?".

These questions are not easy, and the answers vary depending on the situation. The average approach is easy for doctors to implement because it does not require clinicians to calculate the cut-off value [35]. Calculating the optimal cut-off value could be burdensome for them because cut-off values may be different according to demographics or the scale of the questionnaire [35]. Also, the optimal cut-off value cannot be calculated a priori without sufficient data. Thus, clinicians must wait for a certain amount of time until they obtain a sufficient amount of data to get the optimal cut-off value. This implies that the ratio and the frequency approaches cannot be used during early periods in which doctors are just starting to implement screening depression using daily mental-health ratings. Our empirical results show that the accuracy of depression screening by the average approach is clinically acceptable [39]. Thus, during the early period, using the average approach may be more appropriate.

As more data are accumulated, clinicians may choose the ratio or the frequency approach. In some cases, doctors may want to see a simple count of depressed days of a patient during a certain period in order to compare it with the results of traditional depression screening tests. In this case, although the scores of daily mental-health ratings as continuous variables, as well as the values obtained based on the average approach, may provide clinicians with much detailed information, doctors still need to determine whether the scores are high enough to consider a patient to be depressed [35]. Therefore, a systematic way to construct reasonable cut-offs is still needed, and we believe our proposed approaches (the ratio and the frequency approaches) and empirical results of their performance are valuable for clinicians.

It should be noted that the results of data processing depend on the nature of data, such as missing values and outliers, and each approach has its own limitations when dealing with these issues. For instance, the average approach is susceptible to outliers. The frequency approach considers depressed days only, ignoring the difference between normal days and days when ratings are not reported. The ratio approach considers the days when ratings are reported, ignoring the presence of omitted days. Thus, it is important to note that the relative superiority of data-processing approaches varies by situations. We recommend doctors choose an appropriate approach based on their clinical purposes.

Limitations

Our study is a derivation study, and we still need future validation studies using different patient samples before this measure is more broadly adopted. First, we used three variables to evaluate mental health—sleep, mood, and anxiety—to gather information on patients' mental status. Although we selected these three variables based on prior studies [24,25], there may be other important dimensions to assess daily mental status. However, to the best of our knowledge, there has been no study investigating which types of mental-health PROs should be considered for the mobile mental-health trackers. Therefore, a natural extension of our study would be to investigate the optimal choice of the mental dimensions to be used in mobile mental-health trackers. Second, we did not account for the methods of dealing with missing ratings. Further studies on this issue may be useful for improving the accuracy of depression screening in the mobile phone setting. For instance, future research may examine how the quantity of missing values affects the screening performance or how missing values can be effectively imputed to improve depression screening using ratings via mobile phone apps.

Third, our study was conducted in a breast cancer treatment setting in South Korea. Therefore, our results may not be generalizable to other types of mental illness or to patients with different diseases, especially patients with more severe malignancies, such as pancreatic and rectal cancers. Furthermore, South Korea is known for a high percentage of mobile phone users compared with other countries. Mobile app development technology and data management skills are considered to be of high quality. Therefore, the implications from our study may not be applicable in an environment where complementary infrastructures are not adequately supported. In this regard, our study warrants further research on the assessment of the use of mobile mental-health trackers in other settings.

Conclusion

Self-reported daily mental-health ratings obtained via a mobile phone app can be used for screening for depression in breast cancer patients. Adherence to self-reporting can improve the efficacy of mobile phone based approaches for managing distress in this population.

Acknowledgments

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry for Health and Welfare, Republic of Korea (No. 1220170), and the Gachon University research fund of 2015 (GCU-2015-0049).

Authors' Contributions

All authors have made substantial contributions to the conception and design of the research, the acquisition of data, or the analysis and the interpretation of data. JWL and S-YS as co–corresponding authors, and JK and SL as co–first authors, conceived the study and directly participated in the whole process throughout the study.
Y-WS participated in the questionnaire design in the app and the validity of the statistical analysis for depression screening. S-YS, J-HL, and JWL designed the app and data structure. JK, SL, BL, and JWL directly participated in the research design and coordination. JK, SL, and S-YS conducted the statistical analysis and the interpretation. YHM, GS, BHS, SHA, and JWL contributed to the acquisition of clinical data by soliciting patients' participation in the study. YHM, KHJ, and JWL participated in the test subject group design. JK, SL, and JWL drafted the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information for (1) number of days patients report logs in a 2-week period, (2) random-effects logistic panel regression models, and (3) k-means clustering algorithm.

[PDF File (Adobe PDF File), 819KB - jmir_v18i8e216_app1.pdf]

References

1. Jacobsen PB. Screening for psychological distress in cancer patients: challenges and opportunities. J Clin Oncol 2007 Oct 10;25(29):4526-4527 [FREE Full text] [doi: 10.1200/JCO.2007.13.1367] [Medline: 17846452]
2. Stommel M, Given BA, Given CW. Depression and functional status as predictors of death among cancer patients. Cancer 2002 May 15;94(10):2719-2727 [FREE Full text] [Medline: 12173342]
3. Trask PC, Paterson AG, Hayasaka S, Dunn RL, Riba M, Johnson T. Psychosocial characteristics of individuals with non-stage IV melanoma. J Clin Oncol 2001 Jun 1;19(11):2844-2850. [Medline: 11387356]
4. Gessler S, Low J, Daniells E, Williams R, Brough V, Tookman A, et al. Screening for distress in cancer patients: is the distress thermometer a valid measure in the UK and does it measure change over time? A prospective validation study. Psychooncology 2008 Jun;17(6):538-547. [doi: 10.1002/pon.1273] [Medline: 17973237]
5. Gil F, Grassi L, Travado L, Tomamichel M, Gonzalez J, Southern European Psycho-Oncology Study Group. Use of distress and depression thermometers to measure psychosocial morbidity among southern European cancer patients. Support Care Cancer 2005 Aug;13(8):600-606. [doi: 10.1007/s00520-005-0780-0] [Medline: 15761700]
6. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007 Feb;57(535):144-151 [FREE Full text] [Medline: 17263931]
7. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
8. Turnbull Macdonald GC, Baldassarre F, Brown P, Hatton-Bauer J, Li M, Green E, et al. Psychosocial care for cancer: a framework to guide practice, and actionable recommendations for Ontario. Curr Oncol 2012 Aug;19(4):209-216 [FREE Full text] [doi: 10.3747/co.19.981] [Medline: 22876147]
9. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002 May 18;324(7347):1193-1194 [FREE Full text] [Medline: 12016186]
10. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
11. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
12. Reid SC, Kauer SD, Dudgeon P, Sanci LA, Shrier LA, Patton GC. A mobile phone program to track young people's experiences of mood, stress and coping. Development and testing of the mobiletype program. Soc Psychiatry Psychiatr Epidemiol 2009 Jun;44(6):501-507. [doi: 10.1007/s00127-008-0455-5] [Medline: 19011723]
13. Aguilera A, Schueller SM, Leykin Y. Daily mood ratings via text message as a proxy for clinic based depression assessment. J Affect Disord 2015 Apr 1;175:471-474 [FREE Full text] [doi: 10.1016/j.jad.2015.01.033] [Medline: 25679202]
14. Ayers M, Reder L. A theoretical review of the misinformation effect: Predictions from an activation-based memory model. Psycho Bull Rev 1998 Mar;5(1):1-21. [doi: 10.3758/BF03209454]
15. Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol Bull 1993 Jul;114(1):3-28. [Medline: 8346328]
16. Odinot G, Wolters G. Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. Appl. Cognit. Psychol 2006 Nov;20(7):973-985. [doi: 10.1002/acp.1263]

17. Windschitl PD. Memory for faces: evidence of retrieval-based impairment. J Exp Psychol Learn Mem Cogn 1996 Sep;22(5):1101-1122. [Medline: 8805817]
18. Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. Modeling mood variation associated with smoking: an application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. Addiction 2009 Feb;104(2):297-307 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02435.x] [Medline: 19149827]
19. Hedeker D, Demirtas H, Mermelstein RJ. A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. Stat Interface 2009;2(4):391-401 [FREE Full text] [Medline: 20357914]
20. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
21. 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]
22. Sandman L, Granger BB, Ekman I, Munthe C. Adherence, shared decision-making and patient autonomy. Med Health Care Philos 2012 May;15(2):115-127. [doi: 10.1007/s11019-011-9336-x] [Medline: 21678125]
23. Guisasola A, Baeza J, Carrera J, Sin G, Vanrolleghem P, Lafuente J. The Influence of Experimental Data Quality and Quantity on Parameter Estimation Accuracy: Andrews Inhibition Model as a Case Study. Education for Chemical Engineers 2006 Jan;1(1):139-145. [doi: 10.1205/ece06016]
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, (DSM-5). Washington, DC: American Psychiatric Association; 2013.
25. Mayers AG, Grabau EAS, Campbell C, Baldwin DS. Subjective sleep, depression and anxiety: inter-relationships in a non-clinical sample. Hum Psychopharmacol 2009 Aug;24(6):495-501. [doi: 10.1002/hup.1041] [Medline: 19582759]
26. Baldacci F, Vedovello M, Ulivi M, Vergallo A, Poletti M, Borelli P, et al. Triggers in allodynic and non-allodynic migraineurs. A clinic setting study. Headache 2013 Jan;53(1):152-160. [doi: 10.1111/head.12012] [Medline: 23278553]
27. Kravitz RL, Franks P, Feldman MD, Tancredi DJ, Slee CA, Epstein RM, et al. Patient engagement programs for recognition and initial treatment of depression in primary care: a randomized trial. JAMA 2013 Nov 6;310(17):1818-1828 [FREE Full text] [doi: 10.1001/jama.2013.280038] [Medline: 24193079]
28. Kroenke K, Spitzer R. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
29. Lazenby M, Dixon J, Bai M, McCorkle R. Comparing the distress thermometer (DT) with the patient health questionnaire (PHQ)-2 for screening for possible cases of depression among patients newly diagnosed with advanced cancer. Palliat Support Care 2014 Feb;12(1):63-68. [doi: 10.1017/S1478951513000394] [Medline: 24468423]
30. McLennon SM, Bakas T, Jessup NM, Habermann B, Weaver MT. Task difficulty and life changes among stroke family caregivers: relationship to depressive symptoms. Arch Phys Med Rehabil 2014 Dec;95(12):2484-2490 [FREE Full text] [doi: 10.1016/j.apmr.2014.04.028] [Medline: 24858447]
31. Fann JR, Thomas-Rich AM, Katon WJ, Cowley D, Pepping M, McGregor BA, et al. Major depression after breast cancer: a review of epidemiology and treatment. Gen Hosp Psychiatry 2008;30(2):112-126. [doi: 10.1016/j.genhosppsych.2007.10.008] [Medline: 18291293]
32. Hardman A, Maguire P, Crowther D. The recognition of psychiatric morbidity on a medical oncology ward. J Psychosom Res 1989;33(2):235-239. [Medline: 2724199]
33. Hegel MT, Moore CP, Collins ED, Kearing S, Gillock KL, Riggs RL, et al. Distress, psychiatric syndromes, and impairment of function in women with newly diagnosed breast cancer. Cancer 2006 Dec 15;107(12):2924-2931 [FREE Full text] [doi: 10.1002/cncr.22335] [Medline: 17103381]
34. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
35. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006 May 6;332(7549):1080 [FREE Full text] [doi: 10.1136/bmj.332.7549.1080] [Medline: 16675816]
36. Conaway M. A Random Effects Model for Binary Data. Biometrics 1990 Jun;46(2):317-328. [doi: 10.2307/2531437]
37. Hill RC, Griffiths WE, Lim GC. Principles of Econometrics. 4th edition. Hoboken, NJ: Wiley; 2011.
38. Li B, Lingsma HF, Steyerberg EW, Lesaffre E. Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. BMC Med Res Methodol 2011;11:77 [FREE Full text] [doi: 10.1186/1471-2288-11-77] [Medline: 21605357]
39. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 2013 Apr;9(2):280-299 [FREE Full text] [doi: 10.1007/s11306-012-0482-9] [Medline: 23543913]
40. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol 2013 Jul 10;31(20):2580-2585 [FREE Full text] [doi: 10.1200/JCO.2012.47.6804] [Medline: 23733753]
41. Bastard M, Pinoges L, Balkan S, Szumilin E, Ferreyra C, Pujades-Rodriguez M. Timeliness of clinic attendance is a good predictor of virological response and resistance to antiretroviral drugs in HIV-infected patients. PLoS One 2012;7(11):e49091 [FREE Full text] [doi: 10.1371/journal.pone.0049091] [Medline: 23145079]

42. Blacher RJ, Muiruri P, Njobvu L, Mutsotso W, Potter D, Ong'ech J, et al. How late is too late? Timeliness to scheduled visits as an antiretroviral therapy adherence measure in Nairobi, Kenya and Lusaka, Zambia. AIDS Care 2010 Nov;22(11):1323-1331. [doi: 10.1080/09540121003692235] [Medline: 20711886]
43. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]
44. Lutfey KE, Wishner WJ. Beyond "compliance" is "adherence". Improving the prospect of diabetes care. Diabetes Care 1999 Apr;22(4):635-639 [FREE Full text] [Medline: 10189544]
45. Hartigan J, Wong M. Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics 1979;28(1):100-108. [doi: 10.2307/2346830]
46. Shmueli G, Patel N, Bruce P. Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner. 2nd edition. Hoboken, NJ: Wiley; 2010.
47. Steinley D. K-means clustering: A half-century synthesis. Br J Math Stat Psychol 2006;59(1):1-34. [doi: 10.1348/000711005X48266]
48. Badger TA, Braden CJ, Mishel MH. Depression burden, self-help interventions, and side effect experience in women receiving treatment for breast cancer. Oncol Nurs Forum 2001 Apr;28(3):567-574. [Medline: 11338763]
49. Vahdaninia M, Omidvari S, Montazeri A. What do predict anxiety and depression in breast cancer patients? A follow-up study. Soc Psychiatry Psychiatr Epidemiol 2010 Mar;45(3):355-361. [doi: 10.1007/s00127-009-0068-7] [Medline: 19458878]
50. Bieri D, Reeve RA, Champion GD, Addicoat L, Ziegler JB. The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. Pain 1990 May;41(2):139-150. [Medline: 2367140]
51. McKinley S, Coote K, Stein-Parbury J. Development and testing of a Faces Scale for the assessment of anxiety in critically ill patients. J Adv Nurs 2003 Jan;41(1):73-79. [Medline: 12519290]
52. Derham P. Using Preferred, Understood or Effective Scales? How Scale Presentations Effect Online Survey Data Collection. Australas J Mark Soc Res 2011;19(2):13-26.
53. Donaldson MS. Taking stock of health-related quality-of-life measurement in oncology practice in the United States. J Natl Cancer Inst Monogr 2004(33):155-167. [doi: 10.1093/jncimonographs/lgh017] [Medline: 15504926]
54. Snyder CF, Jensen R, Courtin SO, Wu AW, Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res 2009 Sep;18(7):793-800 [FREE Full text] [doi: 10.1007/s11136-009-9497-8] [Medline: 19544089]
55. Locklear T, Miriovsky B, Willig J, Staman K, Bhavsar N, Weinfurt K, et al. National Institutes Of Health. 2014. Strategies for Overcoming Barriers to the Implementation of Patient-Reported Outcomes Measures URL: https://www.nihcollaboratory.org/Products/Strategies-for-Overcoming-Barriers-to-PROs.pdf [accessed 2016-08-01] [WebCite Cache ID 6jRJd5G9h]

Abbreviations

AUC: area under the ROC curve
PHQ-9: Patient Health Questionnaire-9
PRO: patient reported outcome
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 13.02.16; peer-reviewed by J Kim, A Landman, K Singh, S Schueller, J Pollak; comments to author 16.03.16; revised version received 17.05.16; accepted 20.07.16; published 04.08.16.

Please cite as:
Kim J, Lim S, Min YH, Shin YW, Lee B, Sohn G, Jung KH, Lee JH, Son BH, Ahn SH, Shin SY, Lee JW
Depression Screening Using Daily Mental-Health Ratings from a Smartphone Application for Breast Cancer Patients
J Med Internet Res 2016;18(8):e216
URL: http://www.jmir.org/2016/8/e216/
doi:10.2196/jmir.5598
PMID:27492880

Original Paper

mHealth Intervention to Improve Diabetes Risk Behaviors in India: A Prospective, Parallel Group Cohort Study

Angela Pfammatter[1], PhD; Bonnie Spring[1], PhD; Nalini Saligram[2], PhD; Raj Davé[3], MBA; Arun Gowda[4], MD; Linelle Blais[5], PhD; Monika Arora[6], PhD; Harish Ranjani[7], PhD; Om Ganda[8], MD; Donald Hedeker[9], PhD; Sethu Reddy[8], MD; Sandhya Ramalingam[2], PhD

[1]Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Arogya World, Naperville, IL, United States

[3]Health ARC, Pennington, NJ, United States

[4]Focus Scientific Research Center, Bengaluru, India

[5]Emory Centers for Training and Technical Assistance, Behavioral Sciences and Health Education, Emory University, Atlanta, GA, United States

[6]Public Health Foundation of India, Haryana, India

[7]Madras Diabetes Research Foundation, Chennai, India

[8]Joslin Diabetes Center, Boston, MA, United States

[9]University of Chicago, Chicago, IL, United States

Corresponding Author:
Angela Pfammatter, PhD
Northwestern University Feinberg School of Medicine
680 N Lake Shore Dr, Suite 1400
Chicago, IL, 60611
United States
Phone: 1 312 503 1574
Fax: 1 312 503 0982
Email: angela.pfammatter@northwestern.edu

Abstract

Background: In low/middle income countries like India, diabetes is prevalent and health care access limited. Most adults have a mobile phone, creating potential for mHealth interventions to improve public health. To examine the feasibility and initial evidence of effectiveness of mDiabetes, a text messaging program to improve diabetes risk behaviors, a global nonprofit organization (Arogya World) implemented mDiabetes among one million Indian adults.

Objective: A prospective, parallel cohort design was applied to examine whether mDiabetes improved fruit, vegetable, and fat intakes and exercise.

Methods: Intervention participants were randomly selected from the one million Nokia subscribers who elected to opt in to mDiabetes. Control group participants were randomly selected from non-Nokia mobile phone subscribers. mDiabetes participants received 56 text messages in their choice of 12 languages over 6 months; control participants received no contact. Messages were designed to motivate improvement in diabetes risk behaviors and increase awareness about the causes and complications of diabetes. Participant health behaviors (exercise and fruit, vegetable, and fat intake) were assessed between 2012 and 2013 via telephone surveys by blinded assessors at baseline and 6 months later. Data were cleaned and analyzed in 2014 and 2015.

Results: 982 participants in the intervention group and 943 in the control group consented to take the phone survey at baselne. At the end of the 6-month period, 611 (62.22%) in the intervention and 632 (67.02%) in the control group completed the follow-up telephone survey. Participants receiving texts demonstrated greater improvement in a health behavior composite score over 6 months, compared with those who received no messages $F(1, 1238) = 30.181$, $P<.001$, 95% CI, 0.251-0.531. Fewer intervention participants demonstrated health behavior decline compared with controls. Improved fruit, vegetable, and fat consumption ($P<.01$) but not exercise were observed in those receiving messages, as compared with controls.

Conclusions: A text messaging intervention was feasible and showed initial evidence of effectiveness in improving diabetes-related health behaviors, demonstrating the potential to facilitate population-level behavior change in a low/middle income country.

Trial Registration: Australian New Zealand Clinical Trials Registry (ACTRN): 12615000423516; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=367946&isReview=true (Archived by WebCite at http://www.webcitation.org/6j5ptaJgF)

(J Med Internet Res 2016;18(8):e207) doi:10.2196/jmir.5712

KEYWORDS

mHealth; diabetes; health promotion

Introduction

Diabetes is estimated to affect 387 million globally, with disease prevalence expected to increase to 592 million by the year 2035 [1]. Approximately 80% of people with diabetes live in low- and middle-income countries (LMICs) creating a pressing need for prevention and treatment efforts focused on these regions. With 3.6 billion mobile subscriptions worldwide in 2014, mobile phones hold great potential as an intervention delivery channel [2]. Mobile phone uptake is high, even in parts of the globe that lack basic electricity and sanitation infrastructure. Especially since the mobile phone is carried by the user throughout much of the day, mobile health intervention holds the potential to reach and help remote target populations cost-effectively [2,3].

Text messaging interventions could offer a particularly useful health promotion intervention delivery channel for LMICs. Texts can be delivered to the most basic mobile phones and do not require consistent connectivity or Internet capability. Hence, messaging can be delivered inexpensively and automatically to widely geographically dispersed people from different socioeconomic strata [4]. Recent systematic reviews of text messaging health promotion interventions indicate that text messaging has not only proved most efficacious for smoking cessation but also shows some promise for improving other health behaviors, yielding small-to-medium effects comparable with those of print- or computer-delivered interventions [5,6]. However, the vast majority of studies to date have been conducted in high-income countries and in a single language [7]. Since LMICs are disproportionally affected by diabetes as well as underresourced compared with higher income countries, research specifically examining text messaging intervention as part of a public health initiative in LMICs is needed.

India is illustrative of other LMICs, in having a high prevalence of diabetes, poor health care access, and yet a high penetration of mobile device use. Despite its relatively low rates of overweight and obesity, more than 62 million of India's 1.2 billion residents are diagnosed with diabetes [8,9]. The size and heterogeneity of India's population, limited access to health services, and scarcity of quality-controlled clinical laboratory facilities create barriers to intervening preventively on diabetes. Access challenges are particularly acute in India's rural areas, where almost 70% of the population resides [10]. Growing evidence indicates that Indians have heightened genetic risk and a lowered disease threshold in response to diabetes risk factors including age, obesity, abdominal adiposity, and high body fat percentage, resulting in increased risk of diabetes at younger age and lower body mass than other ethnic groups [8]. The elevated and increasing risk of diabetes in the Indian population creates an urgent need for effective interventions that can be scalable to all regions. With its large number of mobile phone subscribers (900 million) [11], India offers a test bed to examine whether an mHealth intervention has the potential to reduce diabetes risk in a LMIC population.

To date, few chronic disease prevention interventions have been tested in the Indian population. Two studies using in-person or telephone counseling showed feasibility, acceptability, and preliminary evidence of efficacious screening and treatment for prediabetic and diabetic patients in rural India. However, these interventions used either individual telephone counseling or a fully equipped mobile van [12,13], requiring extensive personnel and equipment resources that preclude scalability and national implementation in an LMIC. To date, the sole test of a diabetes-related text messaging intervention in India was conducted in men with impaired glucose tolerance, limiting generalizability to the context of prevention in those at risk, rather than population-level prevention [14]. To slow the epidemic of diabetes in India, scalable prevention interventions are needed that can address cultural, geographical, and language barriers across the entire population.

mDiabetes was a text message, public health program developed to address awareness of diabetes and the corresponding risk behaviors. The program was planned to be disseminated through Nokia's mobile platform. The investigators leveraged an opportunity to employ a pre-post evaluation of effectiveness of the messages disseminated. Thus, this study aimed to evaluate whether this 6-month text messaging intervention alone, unaccompanied by costly, burdensome in-person visits or telephone coaching, was acceptable to end users and could improve behavioral risk factors for diabetes in all segments of the Indian population. We hypothesized that those receiving the text messaging intervention, as compared with an untreated comparison group, would show positive changes in 4 health behaviors that lessen diabetes risk: engagement in exercise, avoidance of fat foods, fruit intake of 2 servings a day or more, and vegetable intake of 2 servings a day or more. Furthermore, we explored whether the intervention increased the prevalence of those able to change more than a single health behavior, an important outcome since risk behaviors cluster, and diabetes prevention requires engaging in multiple healthy behaviors [15,16].

Methods

Study Design and Participants

Text messages for the mDiabetes program were developed by Emory University and reviewed by a Behavior Change Task Force assembled as part of a 2011 Clinton Global Initiative Commitment by Arogya World. The 56 messages were designed

to motivate improvement in diabetes risk behaviors and increase awareness about the causes and complications of diabetes. Based on feedback from Indian consumers, messages were culturally tailored to be more acceptable and actionable by the population. In 2012, Nokia held a 22% marketshare in India, [17] which enabled them to invite one million individuals from all over India to opt in to receive health messages; enrollment was closed once that number was reached. Texts (see Multimedia Appendix 1) were available in one of 12 languages based on participant preference. Texts arrived in a predetermined order and frequency (twice a week) to a dedicated inbox on the Nokia phone. A comparison group was drawn from a database that included all mobile phone users in India, after excluding subscribers to Nokia service.

Ipsos, a global market research company, implemented the evaluation by randomly selecting samples to be interviewed from the Nokia and non-Nokia cohorts and conducting the phone surveys in multiple languages. Intervention participants and controls were recruited in India for a 6-month prospective study. The intervention group of 982 was randomly selected from the one million Nokia phone customers who opted in to receive mDiabetes messages. The control group (n=943) was randomly selected from the database including all non-Nokia mobile phone users in India. The sole eligibility criteria were that participants in both groups be adults aged 18 years and older. The research protocol was approved by an independent ethics review committee of the Centre for Chronic Disease Control of India, New Delhi. The protocol was registered in the Australian New Zealand Clinical Trials Registry, #ACTRN12615000423516.

Procedures

Between November and December of 2012, Ipsos staff assessed baseline levels of behaviors by interviewing all study participants by mobile telephone in the participant's choice of language. A similar follow-up interview was repeated 6 months later in mid 2013. Interviews were scripted, conducted by personnel who were kept blind to the interviewee's treatment assignment, and lasted approximately 20 minutes. The baseline and 6-month interviews comprised 19 questions asking participants to self-report demographic information, including residential location, age, and health behaviors. To assess physical activity, participants were asked, "Do you exercise currently?" with response options "yes" or "no." Number of fruit and of vegetable servings was assessed with response options of "0 to 1 servings, 2-3 servings, or 4 or more servings." High fat food intake was assessed with the question, "Do you consistently avoid eating high fat food/fried food such as samosas, vadai, bajji, bondas, etc.?" with response options " yes" or "no." Diabetes preventive behaviors were coded as: endorsing exercise, endorsing avoidance of fatty food, endorsing consuming 2 or more servings of fruit, endorsing consuming 2 or more servings of vegetables.

Over 6 months, participants in the active intervention condition were sent 56 unique messages related to diabetes: one message per day for the first 6 days, then 2 messages per week. Participants in the control group received no study contact until the end of the 6-month period.

Statistical Analysis

Data checking, cleaning, and analyses were conducted in 2014 and 2015. The study was designed to detect a 10% difference between groups on change in a composite health behavior score with 80% power and a 2-sided significance level of α=0.05. Power analysis produced a sample size estimate of 384 for each group. Assuming that rural versus urban location would be a meaningful covariate, the objective was to recruit equal numbers in each area. Therefore, with an estimated 15% attrition, target recruitment was 450 participants for intervention and control, stratified by geographic region, resulting in a total sample size of at least 1800.

The match between each individual's postintervention and preintervention survey was confirmed on the basis of phone number and demographic data. Differences between the intervention and control groups in baseline measures and retention were tested by chi-square. All available surveys were included in the baseline analyses, but only matched pairs of surveys were included in the longitudinal analysis. To compare the lifestyle behavior changes over time, a composite healthy behavior change score was constructed. Each instance of pre/postbehavior change from unhealthy status to healthy status was assigned a numeric score of 1; no change was assigned a score of 0; change from healthy to unhealthy was assigned a score of −1. The scores for each of the 4 behaviors were then summed, generating a composite change score that ranged from 4 (when all 4 behaviors moved from unhealthy status in the prephase to healthy status after treatment) to −4 (when all 4 behaviors changed from healthy to unhealthy). Healthful behavior change was operationalized by the composite healthy lifestyle improvement score [18].

Normality of the distribution of the composite score was evaluated before conducting the primary analysis comparing the 2 treatment groups on healthful behavior change. An analysis of covariance was conducted with baseline level of behavior, gender, and urban/rural location included as covariates. Secondary analyses using logistic regression with the same covariates compared the groups on the presence of each of the 4 preventive behaviors at postintervention. All analyses were performed using in IBM SPSS Statistics (version 22) [19].

Results

During the recruitment phase of the study, 982 participants in the intervention group and 943 in the control group consented to take the phone survey (Table 1). A majority of the sample were male (88.52%), lived in an urban location (68.78%), and resided in the North of India (67.06%). At both baseline and 6-month follow-up, the intervention group had fewer males (P<.001) and fewer individuals living in urban locations (P<.001) than the control group. At the end of the 6-month period, 611 (62.22%) in the intervention and 632 (67.02%) in the control group completed the follow-up telephone survey (P=.028).

Table 1. Participant baseline characteristics.

	Overall; n	Control; n	Experimental; n	*P* value
Baseline	1925	943 (48.99%)	982 (51.01%)	
Male	1704 (88.52%)	881 (93.43%)	823 (83.81%)	<.001[a]
Urban	1324 (68.78%)	867 (91.94%)	457 (46.54%)	<.001[a]
North India	1291 (67.06%)	653 (69.25%)	638 (64.97%)	.047[a]
Mean age (SD)	32.2 (10.6)	32.83 (9.39)	31.66 (11.64)	.016[a]
Consumes fruit	46 (2.39%)	31 (3.29%)	15 (1.53%)	.033[a]
Consumes vegetables	75 (3.90%)	53 (5.62%)	22 (2.24%)	<.001[a]
Consumes fat	569 (29.56%)	267 (28.31%)	302 (30.75%)	.241
Exercises	1094 (56.83%)	601 (63.73%)	493 (50.20%)	<.001[a]

[a]It indicates statistical significance ($P<.05$).

After controlling for covariates (baseline behavior, gender, urban/rural location), the treatment groups differed significantly in their composite healthy change scores $F(1, 1238) = 30.181$, $P<.001$, Bonferroni adjusted 95% CI (0.251-0.531), such that participants in the intervention group reported greater improvement in their aggregated diabetes risk behaviors over time than participants in the control group. Figure 1, demonstrating the distribution of change scores for each group, indicates that 36.55% of participants in the control condition reported a decline in their number of healthy behaviors, as compared with 24.71% of those in the intervention group.

Analyses shown in Table 2 revealed that, as compared with controls, more participants in the experimental group improved their fruit and vegetable intake and reduced their fat intake after intervention. No differential change in exercise was observed between the groups (Figure 2). In addition, 128 (20.95%) of the intervention group compared with 73 (11.55%) of the control group improved 2 or more health behaviors.

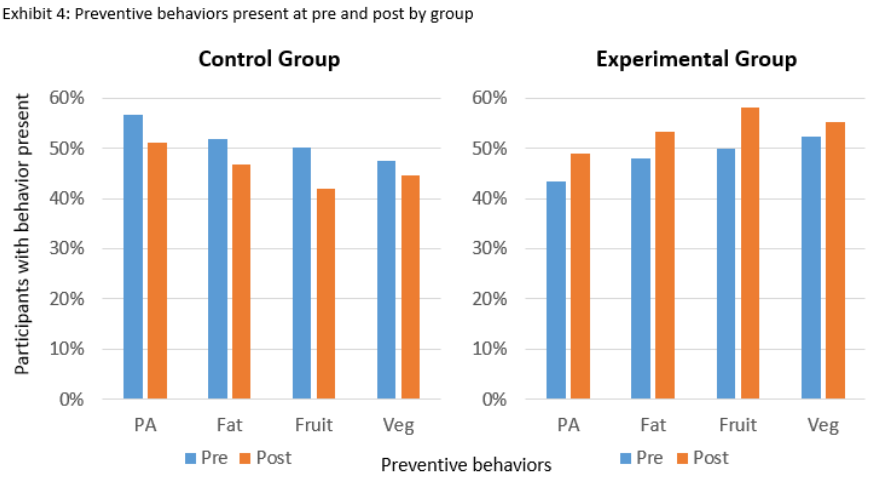
Table 2. Logistic regression of behaviors at the end of intervention by group with covariates.

Health Behavior	Predictor	β	SE β	Wald's χ^2	d*f*	p	e^{β}
Fruit Consumption (n=1243)	Constant	−0.836	0.098	72.611	1	<.001[a]	0.434
	Baseline	0.543	0.122	19.797	1	<.001[a]	1.721
	Location	0.230	0.141	2.657	1	.103	1.259
	Gender	−0.132	0.199	0.441	1	.507	0.876
	Group	0.549	0.133	17.023	1	<.001[a]	1.731
	Test			χ^2	d*f*	P	
	Overall model evaluation			51.919	4	<.001[a]	
Vegetable Consumption (n=1243)	Constant	−0.357	0.110	10.452	1	<.001[a]	0.700
	Baseline	0.984	0.124	62.628	1	<.001[a]	2.674
	Location	0.135	0.153	0.774	1	.379	1.144
	Gender	0.217	0.220	0.972	1	.324	1.242
	Group	0.561	0.140	16.025	1	<.001[a]	1.753
	Test			χ^2	d*f*	P	
	Overall model evaluation			98.424	4	<.001[a]	
Fat Consumption (n=1243)	Constant	0.465	0.134	12.055	1	.001[a]	1.593
	Baseline	0.463	0.144	10.332	1	.001[a]	1.589
	Location	0.446	0.179	6.215	1	.013[a]	1.562
	Gender	0.462	0.271	2.902	1	.088	1.587
	Group	0.510	0.156	10.739	1	.001[a]	1.665
	Test			χ^2	d*f*	P	
	Overall model evaluation			46.373	4	<.001[a]	
Physical activity (n=1243)	Constant	−0.163	0.113	2.094	1	.148	0.849
	Baseline	0.963	0.121	63.347	1	<.001[a]	2.620
	Location	0.009	0.145	0.004	1	.951	1.009
	Gender	0.048	0.203	0.055	1	.815	1.049
	Group	0.094	0.137	0.469	1	.494	1.098
	Test			χ^2	d*f*	P	
	Overall model evaluation			65.295	4	<.001[a]	

[a] indicates statistical significance (*P*<.05).

Figure 1. Composite behavior change scores by group.

Figure 2. Preventive behaviors present at pre and postintervention by group.



Discussion

Principal Findings

This study demonstrated the feasibility and initial evidence of effectiveness of a diabetes prevention text messaging intervention in 12 languages delivered to a sample of nearly 1000 Indian adults, drawn from both rural and urban regions. As compared with those who were not sent diabetes-related messages, those receiving text messages designed to enhance awareness of diabetes risk factors displayed greater improvement in diabetes preventive behaviors over a 6-month period. In addition, the text messaging intervention prevented health behaviors from deteriorating. Importantly, the minimalist and scalable text messaging intervention improved more than one risk behavior simultaneously, an important feature since multiple co-occurring risk behaviors are implicated in the development of diabetes and other chronic diseases. Fat and fruit and vegetable intakes all improved at 6 months in the group receiving messaging, as compared with controls after controlling for baseline differences between the groups. Even though the group receiving text messages reported a small increase in exercise relative to controls, the difference was not significant. Consistent with findings from other studies, a more robust intervention may be needed to produce improvement in physical activity [18]. Overall, results support the feasibility, acceptability, and preliminary effectiveness of a low-cost, low-burden text messaging intervention to prevent degradation of health behaviors over time and to promote the acquisition of diabetes preventive behaviors.

Diabetes imposes a devastating socioeconomic burden globally, especially in the developing world. Although effective diabetes

preventive interventions exist, including those evaluated in the Diabetes Prevention Program, the Finnish Diabetes Prevention Study, the Malmo Study of Sweden, and the Da Qinq IGT and Diabetes Study of China, most successful treatments are very burdensome and costly [20-23]. An urgent need exists to develop effective, scalable interventions that are adaptable for different cultures at low per-patient cost. As mobile phone use continues to increase, the use of mHealth interventions to address public health problems grows increasingly feasible and attractive [23]. Results of this study adds to the emerging evidence that population level healthy lifestyle change may be attainable through low-cost mobile health interventions [24].

Limitations

This study had several limitations that should be considered when interpreting the results. First, the absence of random assignment to treatment groups resulted in some imbalances in the baseline characteristics of the groups. We addressed these by controlling for covariates in the analyses; however, it is possible that inherent differences between the intervention and control populations remained. Second, the selection of participants may have been biased toward those having greater than average interest in health, given that all participants agreed to answer questions about their health. Measurement error and social desirability bias may have been operative since participants' behaviors were self-reported rather than objectively measured. Although these selection biases are of concern, they apply similarly and nondifferentially to participants in both groups. This study was constrained by time and cost of asking multiple questions in multiple languages and as such was unable to use longer, more rigorous assessments of diet quality. Thus, to assess initial feasibility and potential for public health benefit, brief questions were used. Another measurement challenge that affected both groups was that, to avoid appearing rude or insulting, interviewers inferred the participant's gender by listening to his/her voice on the telephone, rather than directly asking. The reliability of interviewers' reports of gender is unknown. Generalizability of these findings beyond the Indian Nokia user context cannot be assumed. However, given high rates of cellphone penetration in other LMICs, we expect that our findings are likely to generalize, particularly if messages are culturally tailored as they were in this study.

The results from this study demonstrate that health promotion text messaging interventions are feasible to implement in 12 languages, across a large LMIC population, and can be effective in improving behaviors that heighten the risk of diabetes and other chronic diseases. With its low cost and burden, text messaging intervention holds the potential to represent sound and effective public health investment. The cost of Arogya World's one million-person program was $0.65 per person, including program development, transmission, and measurement. The current global epidemic of noncommunicable disease creates an urgent need for proven, simple, transportable, readily executable strategies that offer hope for chronic disease prevention at the population level particularly in LMICs. Population-level mobile health promotion interventions like the one described, show promise to meet that need. Additional evaluation using randomized trial design with measurement of objective outcomes will add to the evidence base for the use of mobile technology in chronic disease prevention.

Acknowledgments

The authors acknowledge study partners Nokia and Emory University and Mallory Stasko and H. Gene McFadden for their expertise, hard work, and support. This study was directed by Arogya World, which received funding from Aetna International, Lifescan Inc. (a Johnson & Johnson company), and MSD India. The funding sources had no involvement in the conduct of the study, writing of the manuscript, or decision to submit for publication. The research protocol was approved by the Centre for Chronic Disease Control of India, New Delhi and registered in the Australian New Zealand Clinical Trials Registry #ACTRN12615000423516. Linelle Blais received compensation to write the text messages used in the current study. Om Ganda has a grant with Amarin Pharma for an ongoing multi-center RCT, receives consulting fees from Amgen and Sanofi, and receives speaker honoraria from Merck and Janssen. S. Sethu Reddy was vice-president of medical affairs at Merck, Sharp, and Dohme until December of 2013 and receives speaker honoraria from Merck, Sharp, and Dohme. No financial disclosures were reported by other authors.

Authors' Contributions

Arogya World (NS, RD, AG, and SR) conceptualized and designed the mDiabetes initiative. NS, BS, and SR, designed the research study. AP, BS, NS, SSR, OG, DH, and SR analyzed and interpreted the data. LB developed the text messages. MA led the Ethics Review. NS, LB, MA, HR, and SR were members of Arogya World's Behavior Change Task Force and commented on the messages and the behavior change data. AP prepared and coordinated the submission of the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

This study was directed by Arogya World, which received funding support from Aetna International, Lifescan Inc. (a Johnson & Johnson company), and MSD India. The funding sources had no involvement in the conduct of the study, writing of the manuscript, or decision to submit for publication. The corresponding author constructed the manuscript, had final responsibility for the decision to submit, and had full access to all the data in the study.

Linelle Blais received compensation as part of a contract with Arogya World to write the text messages used in this study.

Multimedia Appendix 1

mDiabetes text messages sent.

[PDF File (Adobe PDF File), 126KB - jmir_v18i8e207_app1.pdf]

References

1. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]
2. Kaplan WA. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? Global Health 2006;2:9 [FREE Full text] [doi: 10.1186/1744-8603-2-9] [Medline: 16719925]
3. Wei J, Hollin I, Kachnowski S. A review of the use of mobile phone text messaging in clinical and healthy behaviour interventions. J Telemed Telecare 2011;17(1):41-48. [doi: 10.1258/jtt.2010.100322] [Medline: 21097565]
4. Pellegrini CA, Pfammatter AF, Conroy DE, Spring B. Smartphone applications to support weight loss: current perspectives. Adv Health Care Technol 2015 Jul;1:13-22. [doi: 10.2147/AHCT.S57844] [Medline: 26236766]
5. 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]
6. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on EpidemiologyPrevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on CardiovascularStroke Nursing, Council on Functional GenomicsTranslational Biology, Council on Quality of CareOutcomes Research,Stroke Council. 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]
7. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
8. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 May;27(5):1047-1053. [Medline: 15111519]
9. Ramachandran A, Snehalatha C, Vijay V. Low risk threshold for acquired diabetogenic factors in Asian Indians. Diabetes Res Clin Pract 2004 Sep;65(3):189-195. [doi: 10.1016/j.diabres.2004.03.012] [Medline: 15331198]
10. Censusindia. 2011. Rural urban distribution of population URL: http://censusindia.gov.in/2011-prov-results/paper2/prov_results_paper2_india.html [accessed 2016-07-15] [WebCite Cache ID 6j1ceVRx0]
11. TRAI. 2014. Highlights on Telecom Subscription Data as on 31st March, 2014 URL: http://www.trai.gov.in/WriteReadData/WhatsNew/Documents/Press%20Release-TSD-Mar,14.pdf [accessed 2016-07-15] [WebCite Cache ID 6j1db306Q]
12. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V, Indian Diabetes Prevention Programme (IDPP). The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). Diabetologia 2006 Feb;49(2):289-297. [doi: 10.1007/s00125-005-0097-z] [Medline: 16391903]
13. Mohan V, Deepa M, Pradeepa R, Prathiba V, Datta M, Sethuraman R, et al. Prevention of diabetes in rural India with a telemedicine intervention. J Diabetes Sci Technol 2012;6(6):1355-1364 [FREE Full text] [Medline: 23294780]
14. 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]
15. Spring B, King AC, Pagoto SL, Van Horn L, Fisher JD. Fostering multiple healthy lifestyle behaviors for primary prevention of cancer. Am Psychol 2015;70(2):75-90 [FREE Full text] [doi: 10.1037/a0038806] [Medline: 25730716]
16. Spring B, Moller A, Coons M. Multiple health behaviors: Overview and implications. Journal of Public Health 2012;34:i3-i10. [Medline: 22363028]
17. CMR. India; 2013. India Mobile Handsets Market Review, CY 2012 URL: http://cmrindia.com/more-than-221-million-mobile-handsets-shipped-in-india-during-cy-2012-a-y-o-y-growth-of-20-8-nokia-retains-overall-leadership/ [accessed 2016-06-02] [WebCite Cache ID 6hyFHK3UR]
18. Spring B, Moller AC, Colangelo LA, Siddique J, Roehrig M, Daviglus ML, et al. Healthy lifestyle change and subclinical atherosclerosis in young adults: Coronary Artery Risk Development in Young Adults (CARDIA) study. Circulation 2014 Jul 1;130(1):10-17. [doi: 10.1161/CIRCULATIONAHA.113.005445] [Medline: 24982115]
19. IBM Corp. 2013. IBM SPSS Statistics for Windows, Version 22 URL: http://www.ibm.com/analytics/us/en/technology/spss/ [accessed 2016-07-15] [WebCite Cache ID 6j1kTfqsF]
20. Hoskin MA, Bray GA, Hattaway K, Khare-Ranade PA, Pomeroy J, Semler LN, for the Diabetes Prevention Program Research Group. Prevention of Diabetes Through the Lifestyle Intervention: Lessons Learned from the Diabetes Prevention Program and Outcomes Study and its Translation to Practice. Curr Nutr Rep 2014 Dec 1;3(4):364-378 [FREE Full text] [doi: 10.1007/s13668-014-0094-2] [Medline: 25383256]

21. Herman WH, Edelstein SL, Ratner RE, Montez MG, Ackermann RT, Orchard TJ, Diabetes Prevention Program Research Group. Effectiveness and cost-effectiveness of diabetes prevention among adherent participants. Am J Manag Care 2013;19(3):194-202 [FREE Full text] [Medline: 23544761]
22. Ali MK, Echouffo-Tcheugui J, Williamson DF. How effective were lifestyle interventions in real-world settings that were modeled on the Diabetes Prevention Program? Health Aff (Millwood) 2012 Jan;31(1):67-75 [FREE Full text] [doi: 10.1377/hlthaff.2011.1009] [Medline: 22232096]
23. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796 [FREE Full text] [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
24. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile Health Devices as Tools for Worldwide Cardiovascular Risk Reduction and Disease Management. Circulation 2015 Nov 24;132(21):2012-2027. [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]

Abbreviations

LMIC: low/middle income country

Edited by G Eysenbach; submitted 22.03.16; peer-reviewed by S Coughlin, S Bahanshal; comments to author 12.05.16; revised version received 15.06.16; accepted 04.07.16; published 05.08.16.

Please cite as:
Pfammatter A, Spring B, Saligram N, Davé R, Gowda A, Blais L, Arora M, Ranjani H, Ganda O, Hedeker D, Reddy S, Ramalingam S
mHealth Intervention to Improve Diabetes Risk Behaviors in India: A Prospective, Parallel Group Cohort Study
J Med Internet Res 2016;18(8):e207
URL: http://www.jmir.org/2016/8/e207/
doi:10.2196/jmir.5712
PMID:27496271

Original Paper

The Top Chinese Mobile Health Apps: A Systematic Investigation

Jeffrey Hsu[1], MD, MA; Di Liu[2], BMed; Ya Min Yu[2], BMed; Hui Tong Zhao[2], BMed; Zhi Rou Chen[2], BS; Jiao Li[2], PhD; Wei Chen[1], MD

[1]Division of Cardiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Institute of Medical Information and Library, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Corresponding Author:
Wei Chen, MD
Division of Cardiology
Peking Union Medical College Hospital
Chinese Academy of Medical Sciences and Peking Union Medical College
No 1. Shuaifuyuan
Dongcheng District
Beijing, 100730
China
Phone: 86 10 6915 5068
Fax: 86 10 6915 5068
Email: chenwei6@medmail.com.cn

Abstract

Background: China's mHealth market is on track to become a global leader by industry size. The Chinese mobile app market and health care system have peculiarities that distinguish them from other app markets. To date, Chinese mHealth apps have not been systematically investigated.

Objective: The objective of this study was to provide an overview of Chinese mHealth apps as of December 2015.

Methods: We identified and investigated the most downloaded apps from the iOS and Android platforms. For each app, we analyzed and recorded its main service offered, mHealth initiative, disease and specialty focus, app cost, target user, Web app availability, and emphasis on information security. Standard descriptive statistics were used.

Results: A total of 234 apps met the inclusion criteria and were investigated. The apps targeting nonhealth care professionals focused on providing telemedicine and appointment-making services. The apps targeting health care professionals focused on education and peer reviewed articles. The most common disease-specific apps focused primarily on diabetes, hypertension, and hepatitis management. Most apps were free and available on both iOS and Android platforms.

Conclusions: The primary mHealth initiatives targeted by the apps reflect Chinese patients' demand for access to medical care. Disease-specific apps are also representative of disease prevalence in China. Government press releases suggest that new policies on the horizon may shift the industry.

(J Med Internet Res 2016;18(8):e222) doi:10.2196/jmir.5955

KEYWORDS

mobile health applications; mHealth; medical informatics; China

Introduction

mHealth, an area of electronic health, is the provision of health services and information via mobile technologies such as mobile phones [1,2]. mHealth apps have epitomized the typical mHealth service. These apps have the potential to cut costs, promote patient engagement, and improve health outcomes [3]. Much has been reported about the services in developed countries [3,4]. To our knowledge, mHealth apps in Mainland China have not been systematically evaluated.

China's mHealth industry is a rapidly growing sector with a year-on-year growth rate of 29% in 2014 and a forecasted growth rate of 49% in 2015 [5]. The market size is expected to reach ¥7.18 billion (approximately US $1.08 billion) in 2016 and ¥12.5 billion (approximately US $1.90 billion) in 2017 [6]. In China, patients often have difficulty gaining access to appropriate medical care [7-9]. mHealth has the potential to

provide widely accessible services that can be individually tailored and easily adopted.

An understanding of the Chinese health care system and smartphone usage provides a framework for understanding mHealth apps in China. The Chinese health care system has peculiarities that make mHealth a viable option. On the surface, the overall Chinese medical system can be comparable with those in advanced countries. In 2011, the doctor to patient ratio was not profoundly different between China and other developed countries: 1.5 per 1000 patients in China versus 2.5 in the United States and 2.7 in the United Kingdom [10]. The imbalances within the Chinese health care system become apparent when comparing urban versus rural areas. Health care expenditure varied by nearly 4-fold in 2009 between urban and rural areas [11]. As the overall health care system grew between 1980 and 2006, the number of beds in rural areas actually decreased. Also, the quality of care was far inferior in rural areas: the infant mortality rate was 16.1% in rural areas versus 5.8% in urban areas [11]. Due to the imbalance of medical resources, patients flock to urban areas seeking medical resources. The displacement of rural patients to urban areas causes difficulties in obtaining access to high-quality care, since rural and urban patients all compete for access to the same medical resources.

The smartphone usage rate is also unique in China. In 2014, 62% of the Chinese population between 16 and 59 years old owned a smartphone. By city tiers, smartphone ownership among the same age range was 94% in tier 1 cities and 75% to 88% in tier 2 cities, while the rate in rural areas was 32% [12]. Also, the Android platform holds a strong foothold with approximately 70% share of the smartphone market [13]. The main Android app stores in China are operated by Baidu, 360, and Tencent; Google Play has been absent since a 2010 censorship dispute [14]. The app markets in China, whether iOS or Android, all have multistep quality and content screening before apps are showcased in the store [15-18]. However, the specific rejection criteria differ across each store and are summarized in Multimedia Appendix 1.

The purpose of this study was to provide an overview of the leading mHealth apps in Mainland China as of December 2015. This study investigated each app with regard to availability, service, and data security to understand the current state of the Chinese mHealth market. In this study, we focused on medical-related apps instead of general health care. Hereafter, mHealth refers to mobile apps as they pertain to medicine.

Methods

Selection of Apps

We sampled apps from both the Android (Google, Mountain View, CA, USA) and iOS (Apple Inc, Cupertino, CA, USA) mobile phone app stores. For Android, we sampled apps from the 3 largest Android app stores in China operated by Tencent (Tencent Holdings Limited, Shenzhen, China), Baidu (Baidu, Inc, Beijing, China), and 360 (Qihoo 360 Technology Co. Ltd, Beijing, China). The 3 stores make up nearly 60% of the Chinese Android market share [19]. For iOS, we used App Annie (App Annie, San Francisco, CA, USA) to gather the list on China's iOS market. We obtained the sample of apps on December 5, 2015.

We collected the top 100 apps according to each app store. The Android stores listed free and paid apps together. The iOS apps separately listed free and paid apps; thus, we collected both the top 100 free and top 100 paid apps. We systematically evaluated the free apps and but did only a cursory assessment of the paid ones.

We selected apps from the medical category for further evaluation. We then reviewed the apps for potential inclusion into the study. The inclusion criteria were as follows: simplified Chinese language, service tailored toward Mainland China (excluding Taiwan, Hong Kong, and Macau regions), and services pertaining to health care and medicine, not general health. For example, we omitted weight loss, exercise, smoking cessation, and menstrual cycle management apps (Figure 1).

The initial screening was completed by 4 authors: LD, YYM, ZHT, and CZR. The apps included for further evaluation were randomly assigned to these 4 authors. Each app was reviewed using information from the app store description, the app's website (if available), and the app itself. We then reviewed each app collectively to ensure accuracy. If discrepancies existed, the particular app was discussed and a consensus was reached. For each app, we identified the main service offered, mHealth initiative, disease and specialty focus, app cost, target user, Web app availability, and emphasis on information security.

Categorization of Apps

We recorded what service was offered based on the app store description and assigned the app to a corresponding mHealth initiative. mHealth initiative describes where an app's service lies along the continuum of medical service delivery. In this study, we adopted the categorization of mHealth initiatives from 2 previous reports [1,6] and tailored it for the Chinese market. The 10 health initiatives used in this study were as follows: (1) appointment making, (2) reminders, (3) telemedicine, (4) records and patient monitoring, (5) pharmacy, (6) disease awareness, (7) clinical decision support, (8) discussion forums, (9) medical education and scholarly articles, and (10) other (Table 1). Each app could have one or more identified mHealth factors.

Figure 1. Flowchart of the selection process for Chinese mHealth apps.

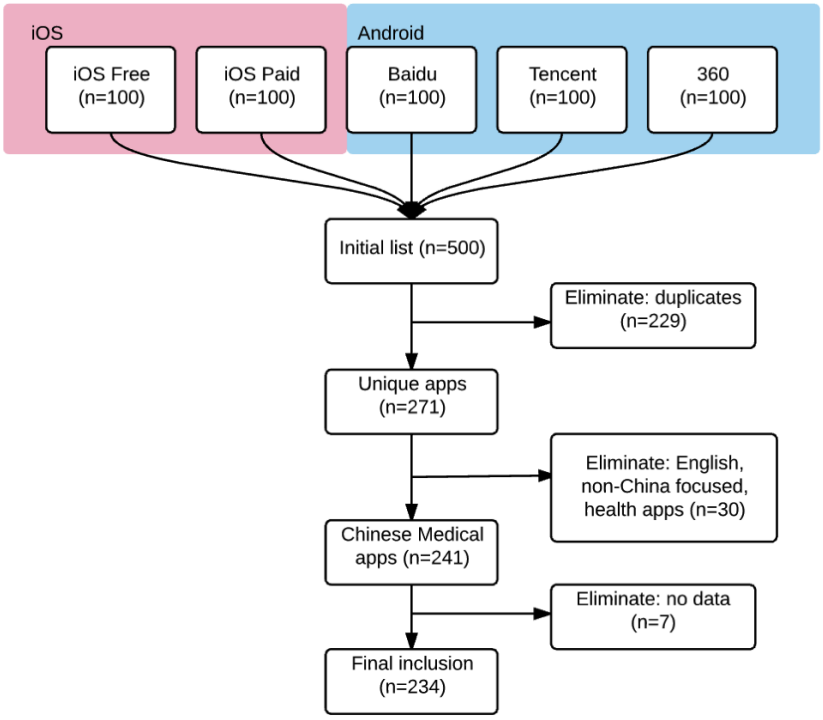


Table 1. mHealth initiatives previously reported and used in the categorization of Chinese mHealth apps in this study.

Previously published categories[a]	Categories adapted for this study
Call center	Appointment making
Reminders	Reminders
Telemedicine	Telemedicine
Records	Records and patient monitoring[b]
Treatment	Pharmacy
Awareness	Disease awareness
Patient monitoring[b]	Records and patient monitoring
Decision support	Clinical decision support
Discussion forum	Discussion forum
Scholarly articles	Medical education and scholarly articles
Other	Other

[a]Sources: Ryu [1] and Xiaohui et al [6].

[b]In previous reports, patient monitoring was an individual factor. In this study, patient monitoring was categorized together with records.

We assessed the disease and medical specialty of an app by the app's name, app store description, and website. We matched the diseases we identified to its closest *International Classification of Diseases, Tenth Revision* counterpart.

App cost referred only to the cost at download. Additional fees incurred after usage were difficult to systematically evaluate and thus were not considered. We considered cost to be a binomial variable.

For target users, we classified apps as being focused on health care professionals (HCPs), non-HCPs, or both. HCPs referred to those included in the World Health Organization's health professional categorization [20]. Since clinicians in public Chinese hospitals are typically required to conduct research [4,21], we also included clinical research-focused apps, even though the WHO categorizes this profession under life sciences. Apps identified as "both" had an HCP and a non-HCP version available.

For Web app availability, we examined the app store and website to determine whether a Web-based version of the app was available.

Finally, we examined whether an app emphasized information security. China lacks an industry standard or legislation regarding medical information safety or privacy similar to the Health Insurance Portability and Accountability Act in the United States [22]. Thus, for information security, we evaluated whether each app self-reported relevant information security measures. We labeled information security for each app as absent, present, or complete. For apps to have complete information security, they had to present documentation or a link referencing a third-party auditor.

Categorizations were made based on the description in the corresponding app store. For apps that were present in multiple app stores, we ensured the consistency of the descriptions across stores and eliminated duplicates. Some apps had a patient and a clinician version. We evaluated these as one entry. Descriptive statistics were used to describe the characteristics of the apps. Heat maps were used to identify the areas of the market receiving the most traction.

Results

There were 241 unique apps that met the inclusion criteria. When we analyzed the apps, 7 were not available or had been taken offline, and thus we eliminated them from the final list. We analyzed a total of 234 apps. Of these, 195 were available in both iOS and Android app stores. However, 22 (9.4%) and 17 (7.3%) of the apps were available exclusively in the iOS and Android app stores, respectively.

The most common medical initiatives were telemedicine, disease awareness, appointment making, and records and patient monitoring (Figure 2). The least common service factors were reminders and clinical decision support. Apps classified under other services included pharmaceutical drug information, drug delivery, insurance plans, and online-to-offline (O2O) health checkup services. We subdivided each medical initiative into the corresponding target user. Apart from clinical decision support and medical education, which primarily focused on HCPs, all other health initiatives were mainly aimed at non-HCP users.

Of the apps, 185 targeted non-HCPs, while 34 targeted HCPs, and 15 had both versions available. A total of 210 (89.7%) of the mHealth apps in China were free. All the paid apps were from the iOS app store. Nearly one-third (154/234, 65.8%) of the apps had both a mobile and a Web-based version. 227 of 234 apps (97.0%) of the apps did not mention information security (Figure 3). Of the 7 apps (3.0%) that mentioned information security, none had undergone external auditing.

We created heat maps to examine the distribution of apps along the medical initiative, and the disease and medical specialty spectrum. The most common diseases were diabetes, hypertension, liver disease (general), and infertility (Figure 4, top). For diabetes, the apps were focused on record keeping and patient education. The most common specialties were general medicine, obstetrics and gynecology, endocrinology, pharmacology, and traditional Chinese medicine (Figure 4, bottom). Apps classified under general medicine covered an assortment of specialties without an emphasis on any particular one.

Figure 2. Distribution of Chinese mHealth apps according to medical initiative and user type. HCP: health care professional.

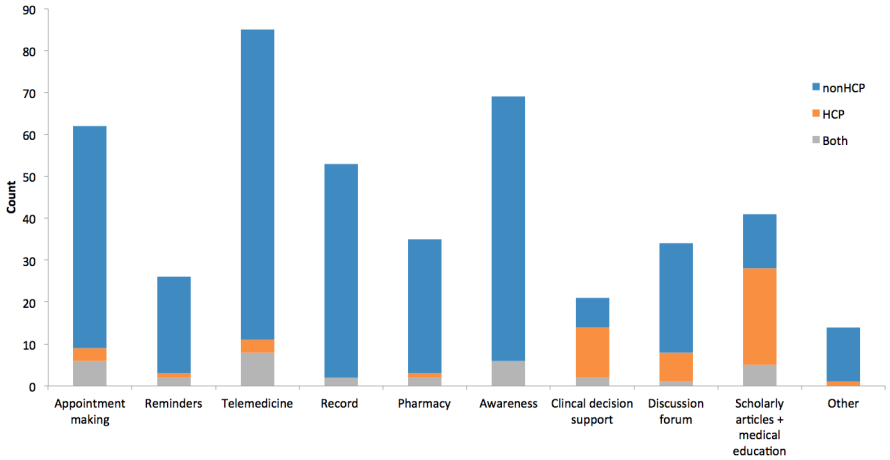


Figure 3. Features of the Chinese mHealth app landscape. Top left: Venn diagram of Chinese mHealth apps' target users; top right: distribution of paid and free apps; bottom left: portion of apps with and without a Web-based version; bottom right: portion of apps with medical information security measures. HCP: health care professional.

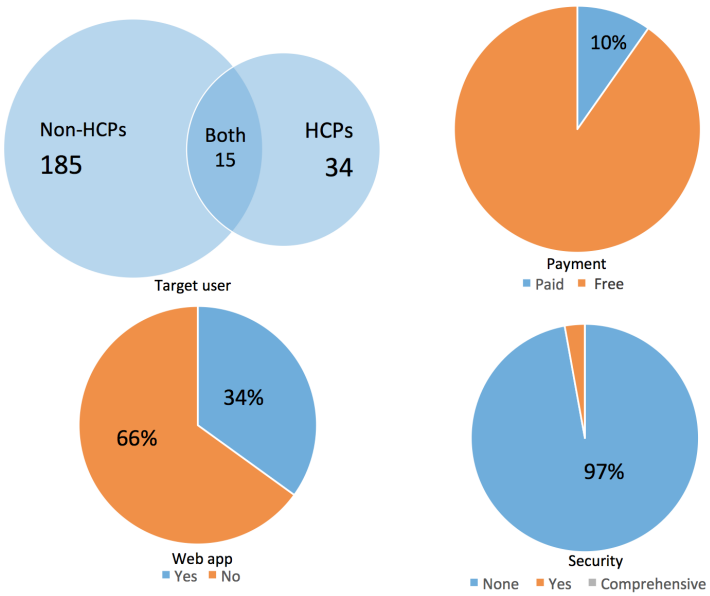
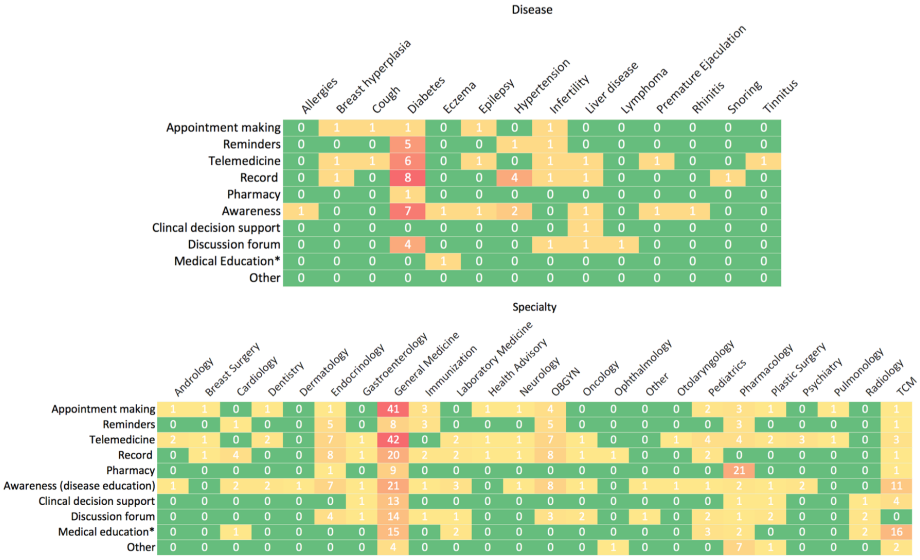


Figure 4. Heat map of medical initiative targeted by applications from (top) each disease and (bottom) medical specialty. Note white numbers refer to the count of apps within each specialty. Warmer colors refer to higher count. *Medical education includes dissemination of scholarly articles. OBGYN: obstetrics and gynecology; TCM: traditional Chinese medicine.



Discussion

Principal Findings

This study provided a snapshot of the Chinese mHealth industry in December 2015. The primary mHealth initiatives targeted by the apps reflected Chinese patients' demand for access to medical care. The primary initiatives were telemedicine, disease awareness, appointment making, and record keeping, followed by medical education and scholarly information. The overwhelming majority of the apps were not specific to a medical specialty. However, the apps that targeted a disease reflected prevalent conditions in China such as diabetes, hypertension, and hepatitis. The target users among apps were mainly non-HCPs. An overwhelming majority did not mention information security. We discuss each aspect further below.

mHealth Initiatives

Through the heat map and mHealth initiative analyses (Figure 4), we found that among the most common mHealth initiatives were telemedicine, appointment making, and medical education. Telemedicine can broadly be defined as the use of telecommunications technologies to provide medical information and services [23]. The use of telemedicine to diagnose and prescribe medication has yet to be legalized in China. This implies that the actual service delivered in telemedicine was limited to general medical inquiries, with limited medical actions

involved. Appointment making reflects a phenomenon across China of patients lacking access to clinicians [7,9].

Here, we discuss the role and background of hospital appointment making. In 2009, the Chinese Ministry of Health required all public hospitals of level 3 [24] or higher to allow appointment-making services and prohibited partnering with third-party agencies to profit from these services [25]. In recent years, for a given public hospital, available appointment dates have been provided directly on the hospital's appointment-making system, public (municipal or provincial) third-party systems, or private third parties such as apps. (Of note, in the Peking Union Medical College Hospital in particular, the ratio of self-distributed to public third-party distribution is 7 to 3.) These apps obtain a number of appointments by signing agreements with hospitals or with individual doctors. Apps solve the problem of difficult access by (1) aggregating hospital appointment availabilities on their platform, thus preventing patients from waiting at fully booked hospitals or (2) partnering with individual doctors to provide consultations on personal free time.

Apps focused on medical education and scholarly articles targeted two fundamental demands in the Chinese health care market: access to reliable information, and clinicians' need to publish peer reviewed articles. The proliferation of discussion forums and unregulated "medical" articles on the Internet and apps makes reliable information a rare commodity. The abundance of medical education apps appears to target this need.

On the other hand, a clinician's professional advancement is dependent on many factors, of which publishing peer reviewed articles is crucial [21,26]. The availability of apps to provide articles and assist with the writing process can be an attempt to supply this demand. These apps provide writing resources such as editing and actual writing, and provide access to scholarly articles. This access is often through a Chinese translation or a summary of the original article (most often in English).

Disease- and Specialty-Specific Apps

We examined the distribution of diseases and medical specialties in the mHealth industry through a heat map plotted against medical initiative. When examining the frequency of apps on a disease basis, we determined diabetes to be the most common, followed by hypertension and hepatitis. This is consistent with epidemiological surveys pointing to these diseases as the most prevalent in Mainland China. According to a report from the Chinese Center for Disease Control, the prevalence of hypertension among Chinese adults was 33.5% (amounting to roughly 330 million hypertension patients) [27]. A recent study showed the prevalence of type 2 diabetes at 11.6% of the population (amounting to about 90 million diabetes patients) [28]. Since mHealth apps for these diseases focused almost exclusively on disease monitoring and recording, market forces driven by pharmaceutical drug sales may not be the key factor for app providers. Rather, it may be the distribution of monitoring devices. Diabetes is monitored by glucometers and single-use disposable test strips, versus repeated-use blood pressure machines for hypertension. The market opportunity for diabetes in mHealth is believed to be greater than that for hypertension due to the nature of disease management and monitoring, despite hypertension having nearly triple the patient volume.

Unlike many developed countries in the world, in China hepatitis B is endemic. Approximately one-third of worldwide cases are in Mainland China [29]. A report noted 120 million carriers of hepatitis B virus in China and 30 million patients who are chronically infected [30]. The existence of apps targeting hepatitis is consistent with the disease's endemic nature in China.

The heat map of medical specialty by health initiative provided more information about the industry as a whole. By far the most heavily targeted specialty, or lack thereof, is general medicine. The apps here mainly focused on appointment making and telemedicine. The telemedicine services offered were often a "lite" version of medical history taking, since patients were provided with general answers and then encouraged to seek specific guidance through in-app appointments. Thus, despite telemedicine and appointment making having lexical and implicit differences, Chinese mHealth apps providing telemedicine services were more synonymous than they appeared based on the above analysis. This finding reinforces the difficulty of accessing care.

Another popular area was pharmacology. The apps in pharmacology provided patients with access to online or offline pharmacies and provided clinicians with pocket drug references. Pharmacology-related apps aimed to provide convenience for patients who require drug refills and cannot repeatedly travel to community pharmacies or hospitals. Community pharmacies often do not carry a full repertoire of prescription drugs, and the current state insurance policy covers in-hospital prescriptions for 14 days for able-bodied patients and up to 30 days for disabled patients. On the other hand, pharmacology apps targeting clinicians, pharmacists, and other HCPs provided references for dosing, interactions, and alternative drugs. The incorporation of pharmacology apps into HCPs' daily practice appears to be a mainstay in China and many other countries [31,32].

Web App

Accessing an app from different media can allow users to experience the app differently and can serve different purposes. The apps in this study focused on mobile phones. However, about one-third of the apps also had a Web-based version. The main difference between the Web app and mobile app was the ability for users to view historical data organized in reports or graphs in the Web-based version. It has been reported that data visualization and context awareness could enhance an app's utility [33].

Target User and Payment

An app's target user can provide information about market opportunities and underlying market forces facing app developers. Our analysis showed that app developers preferred targeting the non-HCP user, perhaps due to easier user acquisition and a larger potential user base. The HCP-oriented apps mainly provided services through medical education or scholarly articles. Few focused on medical care delivery or integration into hospital care. This can be explained by financial

compensation related to publication requirements [21,26]. A physician's compensation is directly tied to his or her position in a hospital hierarchy. Among many factors, publication quotas are mandatory evaluation metrics in many hospitals. We believe that this demand on physicians compliments an app developers' need to grow and retain a steady user base. There were 10 apps with both an HCP and a non-HCP version available. These apps all allowed communication between the two parties.

Most apps in China were free for users to download and use. All apps surveyed from the Android app stores were free. Chinese users are likely accustomed to using free apps.

Security

We examined the lack of information security from two perspectives: market forces and government regulations. Nearly all apps failed to mention on their website or user agreement form about securing the users' information. This can imply gross negligence from the developer or use of the data for other purposes. The latter is likely the underlying motive because, as noted by notable industry agencies, information can be knowingly or unknowingly sold to marketers for financial gains [34].

We also considered information security from a policy perspective. We examined policies pertaining to medical apps by reviewing documents from the People's Republic of China State Council's official website [35] dated between 2013 and 2015. The search was limited to "Internet," "Internet plus health care," and "Internet plus medicine." All results were read and screened by a Chinese Health Care Policy analyst (author LD) for relevance. Among the more than three thousand policies and documents available, no direct laws governing medical apps, let alone mHealth security, were found, and only 11 official releases were identified that can directly or indirectly affect the operation of medical apps (Table 2).

Table 2. Documents[a] by the Chinese State Council or General office of the State Council related to the governance of mHealth.

Document number	Document content	Targeted industry
1	Internet plus action plan	Manufacturing, transportation, information technology
2	Action outline on promoting the development of big data	Manufacturing, transportation, information technology
3	Medical services and health care system planning guideline (2015–2020)	Health
4	Circular on boosting the development of e-commerce	Comprehensive government affairs
5	Guideline on strengthening support for consumer services to upgrade consumption	Commerce
6	Guideline on further boosting consumption as a key component in driving economic development	Commerce
7	Guideline on pushing integrated medical and nursing care for the elderly	Health
8	Circular on strengthening the Patriotic Hygiene Campaign in the new era	Health
9	Opinions on using big data technology to improve the government's supervisory responsibilities and services for market entities.	Manufacturing, transportation, information technology
10	Opinions on cracking down on infringement of intellectual property rights and the production of fake and inferior commodities in cyberspace to safeguard the healthy development of e-commerce	Science, education, intellectual property
11	The legislation working plan of 2015	General

[a]Source: State Council of the People's Republic of China [35].

The scope of mHealth apps can cross into many industries. Documents mentioning mHealth targeted industries such as information technology, manufacturing, health care, commerce, and intellectual property. The fact that many policies, across various industries, mention mHealth suggests that the Chinese government recognizes the future potential of the industry but has yet to dictate a clear stance on how to regulate this relatively new industry. The lack of specific policies toward mHealth may be due to the difficulty in pinpointing which industry it lies in.

Platform

Most apps were available on both app platforms. Android dominates the Chinese smartphone market at over 70% market share [13]. The iOS app store is known to have more apps available than the Android store. However, many are not specific to the Chinese market. It is possible that mHealth app companies are developing the apps domestically and targeting consumers on both platforms.

Limitations

There are limitations to this study. First, the number of apps sampled is small and cannot fully explain the market. This study was meant to provide a snapshot of the industry as a whole rather than details along each medical specialty. To fully understand the availability and characteristics of the apps for each specialty, the analysis should be done in disease or specialty verticals. Second, this study focused on medical apps while excluding general health apps. This allowed for a more homogeneous analysis, since apps targeting healthy users and

sick users are likely different in nature. However, we excluded a large portion of apps from the analysis.

Prospective

In the future, a shift in mHealth apps from delivering purely online services to an O2O approach is expected. O2O, a concept common in Chinese e-commerce, refers to an integration of offline businesses into online commerce [36,37]. Online services in health care can refer to mobile apps, websites, or other digital tools. Offline services include services delivered in physical sites such as hospitals, clinics, pharmacies, and health centers. The idea of O2O in mHealth refers to the integration of online services delivered via apps with "traditional" health service providers. Possible applications include prescription apps that allow patients with existing prescriptions to have their drugs delivered by a local pharmacy, and third-party apps that provide electronic health records services linked to the electronic medical records of a regional clinic.

As of July 2016, the government released legislation banning mHealth apps from providing appointments for patients. The legislation requires appointments to be made directly through the hospital, not a third-party provider. The reasons behind the ban were to ensure the integrity of the hospitals and protect patients' interests. Most Chinese hospitals are nonprofit public institutions. Providing public resources to partnering companies for financial gain violates their nonprofit nature. In addition, third-party appointment-making apps commonly tack on a cost premium for each service-seeking patient. Finally, allowing third-party appointment-making apps to serve individuals who can afford a premium is a detriment to nonpremium-paying patients. Thus, it behooves the government to ban this service to protect the general public and prevent private-public partnerships from hoarding publicly available resources.

A growing trend, internationally and domestically, is allowing clinicians to send prescriptions via apps. Services offered in this space have obvious financial incentives and safety risks. The development of the prescription app industry is expected to catch the government's attention. There is precedence for the government to interject and control specific health markets. How or when that will occur is to be determined. As policy shifts loom on the horizon, mHealth providers must react. One common theme found in China and in the United States is that the "grassroots entrepreneurial nature of the market" [3] appears to be the main driver of the mHealth industry.

Conclusion

mHealth in China is a large and continuously growing market. The potential to disrupt the traditional health care market exists. At the end of 2015, the Chinese mHealth market targeted the nonprofessional user. The services offered heavily focused on the demands of HCPs and non-HCPs, such as publishing peer reviewed papers and gaining access to clinicians, respectively. Also, a unifying policy or standard from the Chinese central government or the China Food and Drug Administration to govern this industry is lacking, but evidence shows that the government is cognizant of the potential this industry and regulations may have in the near future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81250010 and 81470426), the Beijing National Science Foundation (7132221), the Peking Union Medical College Hospital Young and Middle-aged Investigation Fund (PUMCH-2013-016), the Peking Union Medical College Youth Fund, the Fundamental Research Funds for the Central Universities (20122012J01 and 3332013102), 2012 Janssen Research Council China (JRCC), and the project sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (SRF for ROCS, and SEM) to Wei Chen.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Common reasons for app rejection across iOS and Chinese Android app stores. Android app stores include 360, Baidu, and Tencent.

[PNG File, 94KB - jmir_v18i8e222_app1.png]

References

1. Ryu S. mHealth: new horizons for health through mobile technologies: based on the findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthcare Inform Res. 2012. (3) p. 231-233 URL: http://synapse.koreamed.org/DOIx.php?id=10.4258/hir.2012.18.3.231 [WebCite Cache ID 6jdcuxJr8]
2. Arie S. Can mobile phones transform healthcare in low and middle income countries? BMJ 2015;350:h1975.
3. Sama P, Eapen Z, Weinfurt K, Shah B, Schulman K. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
4. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR mHealth uHealth 2015;3(1):e28.

5. Market Share of Mobile Health Market in China. Hamburg: Statista; 2016. URL: http://www.statista.com/statistics/328927/mobile-health-market-scale-in-china/ [accessed 2016-03-26] [WebCite Cache ID 6jXNecYvz]
6. Yu X, Han H, Du J, Wei L, Li C, Zhang X, et al. mHealth in China and the United States: How Mobile Technology Is Transforming Healthcare in the World's Two Largest Economies. Washington, DC: Center for Technology Innovation at Brookings; 2014. URL: https://www.brookings.edu/wp-content/uploads/2016/06/mHealth_finalx.pdf [accessed 2016-08-08] [WebCite Cache ID 6jbpFNTRd]
7. Gao J, Tang S, Tolhurst R, Rao K. Changing access to health services in urban China: implications for equity. Health Policy Plann 2001;16(3):302-312.
8. Knowledge@Wharton. "Ticking Time Bombs": China's Health Care System Faces Issues of Access, Quality and Cost. Philadelphia, PA: Wharton School, Uniiversity of Pennsylvania; 2013 Jun 26. URL: http://knowledge.wharton.upenn.edu/article/ticking-time-bombs-chinas-health-care-system-faces-issues-of-access-quality-and-cost/ [accessed 2016-08-08] [WebCite Cache ID 6jbpTtAn5]
9. Süssmuth-Dyckerhoff C, Wang J. China's health care reforms. Health Int. 2010. p. 54-67 URL: http://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/healthcare%20systems%20and%20services/health%20international/hi10_china_healthcare_reform.ashx [WebCite Cache ID 6jddojVvz]
10. The World Bank. World Development Indicators. Washington, DC: The World Bank Group; 2016. URL: http://data.worldbank.org/data-catalog/world-development-indicators [accessed 2016-08-08] [WebCite Cache ID 6jbq2M4HC]
11. Chen Y, Yin Z, Xie Q. Suggestions to ameliorate the inequity in urban/rural allocation of healthcare resources in China. Int J Equity Health 2014;13:34 [FREE Full text] [doi: 10.1186/1475-9276-13-34] [Medline: 24884614]
12. eMarketer. In China, Tier 1 Cities Rule for Smartphone Usage. New York, NY: eMarketer, Inc; 2015 Aug 18. URL: http://www.emarketer.com/Article/China-Tier-1-Cities-Rule-Smartphone-Usage/1012873 [accessed 2016-08-08] [WebCite Cache ID 6jbqVUDE8]
13. Reed B. IOS, Windows Phone Seen Clawing Market Share Back from Android in China.: BGR Media, LLC; 2013 Sep 26. URL: http://bgr.com/2013/09/25/china-smartphone-market-share/ [accessed 2016-08-08] [WebCite Cache ID 6jbqjwTJn]
14. Matthews C. Time Magazine. 2015 Sep 04. Google is Heading Back to China URL: http://time.com/4023367/google-china/ [accessed 2016-08-08] [WebCite Cache ID 6jbr4aHll]
15. App Review [in Chinese].: Apple Inc; 2016. URL: https://developer.apple.com/app-store/review/cn/ [accessed 2016-08-08] [WebCite Cache ID 6jbrEurRB]
16. App Release Guidelines [in Chinese]. Beijing: Qihoo 360 Technology Co. Ltd; 2016. URL: http://dev.360.cn/wiki/index/id/18 [accessed 2016-08-08] [WebCite Cache ID 6jddfUJse]
17. App Review Specifications [in Chinese]. Beijing: Baidu, Inc; 2016. URL: http://app.baidu.com/dev-doc/verify.html [accessed 2016-08-08] [WebCite Cache ID 6jbxgpcIm]
18. App Access Review Principles [in Chinese]. Shenzhen: Tencent Holdings Limited; 2016. URL: http://wiki.open.qq.com/wiki/%E5%BA%94%E7%94%A8%E6%8E%A5%E5%85%A5%E5%AE%A1%E6%A0%B8%E5%8E%9F%E5%88%99 [accessed 2016-08-08] [WebCite Cache ID 6jbxyDtFN]
19. Yoo E. The Top Ten Android App Stores in China 2015.: Technode; 2016. URL: http://technode.com/2015/09/22/ten-best-android-app-stores-china/ [accessed 2016-08-08] [WebCite Cache ID 6jbyEdiWG]
20. World Health Organization. Classifying Health Workers: Mapping Occupations to the International Standard Classification. Geneva: WHO; 2010. URL: http://www.who.int/hrh/statistics/Health_workers_classification.pdf [accessed 2016-08-08] [WebCite Cache ID 6jbyYoSPM]
21. Yuan HF, Xu WD, Hu HY. Young Chinese doctors and the pressure of publication. Lancet 2013 Feb;381(9864):e4. [doi: 10.1016/S0140-6736(13)60174-9] [Medline: 23374483]
22. Public Law 104-191: Health Insurance PortabilityAccountability Act of 1996. Washington, DC: US Government Publishing Office; 1996 Aug 21. URL: https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf [accessed 2016-08-08] [WebCite Cache ID 6jbrfyp9B]
23. Perednia D, Allen A. Telemedicine technology and clinical applications. JAMA 1995 Feb 08;273(6):483. [doi: 10.1001/jama.1995.03520300057037]
24. The Measures for the Administration of the Hospital Grade [in Chinese].: Ministry of Health P.R. China; 1989. URL: http://baike.baidu.com/link?url=n0LhGo1rvNF9NGab8lH6zuTS1L6v997KHooqEg9u-61bSoGBOUdBtW1bRlK1oOzb-lL0B1jo41irgPKIeD8eYa [accessed 2016-08-08] [WebCite Cache ID 6jbsNM5wU]
25. Bulletin of the Ministry of Health.: Ministry of Health P.R.China; 2009. Opinions of the Ministry of Health on the appointment and treatment of medical services in public hospitals [in Chinese] URL: http://www.moh.gov.cn/mohbgt/s9514/200910/43105.shtml [accessed 2016-08-08] [WebCite Cache ID 6jbwlb2S6]
26. Chen S, Pan Y, Yao Q, Yao L, Liu Z, Xiang L. Publication pressure on Chinese doctors--another view. Lancet 2014 Sep 13;384(9947):956. [doi: 10.1016/S0140-6736(14)61630-5] [Medline: 25220974]
27. Li Y, Wang L, Jiang Y, Li X, Zhang M, Hu N. Prevalence of hypertension among Chinese adults in 2010. Zhonghua Yu Fang Yi Xue Za Zhi 2012 May;46(5):409-413. [Medline: 22883726]

28. Kao P, Han Q, Liu S, Li X, Inman K, Chia N. Letter to the Editor: The surge of type 2 diabetes mellitus in China: an international alert: physical exercise and low-caloric diet may reduce the risks of type 2 diabetes mellitus and dementia. Ann Clin Lab Sci 2016;46(1):114-118. [Medline: 26927355]
29. Custer B, Sullivan S, Hazlet T, Iloeje U, Veenstra D, Kowdley K. Global epidemiology of hepatitis B virus. J Clin Gastroenterol 2004;38(10 Suppl 3):S158-S168. [Medline: 15602165]
30. Liu J, Fan D. Hepatitis B in China. Lancet 2007 May 12;369(9573):1582-1583. [doi: 10.1016/S0140-6736(07)60723-5] [Medline: 17499584]
31. Aungst T. Medical applications for pharmacists using mobile devices. Ann Pharmacother 2013;47(7-8):1088-1095. [doi: 10.1345/aph.1S035] [Medline: 23821609]
32. Davies MJ, Collings M, Fletcher W, Mujtaba H. Pharmacy apps: a new frontier on the digital landscape? Pharm Pract. 2014. (3) p. 453 URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4161411/ [WebCite Cache ID 6jddVDwG2]
33. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: a developer's perspective. J Syst Software 2011 Nov;84(11):2022-2033 [FREE Full text] [doi: 10.1016/j.jss.2011.06.049]
34. PwC Health Research Institute. Top Health Industry Issues of 2015: Outlines of a Market Emerge. New York, NY: PriceWaterhouseCooper LLP; 2014 Dec. URL: https://www.pwc.com/us/en/health-industries/top-health-industry-issues/assets/pwc-hri-top-healthcare-issues-2015.pdf [accessed 2016-08-08] [WebCite Cache ID 6jbua23ly]
35. The State Council of the People's Republic of China. Beijing; 2016. URL: http://english.gov.cn/ [accessed 2016-08-08] [WebCite Cache ID 6jbuFGds4]
36. Zhang J. Customer loyalty forming mechanism of O2O e-commerce. Int J Bus Soc Sci 2014;5(1):164-169.
37. Zhu X, Song B, Ni Y, Ren Y, Li R. The O2O Model—From Online/Offline to the O2O Model. In: Business Trends in the Digital Era. Singapore: Springer; Jun 01, 2016:191-212.

Abbreviations

HCP: health care professional
O2O: online-to-offline

Edited by G Eysenbach; submitted 15.05.16; peer-reviewed by L Lim, YK Lee; comments to author 09.06.16; revised version received 14.07.16; accepted 29.07.16; published 29.08.16.

Please cite as:
Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, Chen W
The Top Chinese Mobile Health Apps: A Systematic Investigation
J Med Internet Res 2016;18(8):e222
URL: http://www.jmir.org/2016/8/e222/
doi:10.2196/jmir.5955
PMID:27573724

Original Paper

Social Network Behavior and Engagement Within a Smoking Cessation Facebook Page

Heather Cole-Lewis[1], MA, MPH, PhD; Adler Perotte[2], MA, MD; Kasia Galica[1], MA; Lindy Dreyer[3], BMA; Christopher Griffith[4], BA; Mary Schwarz[4], BA; Christopher Yun[4], BCS; Heather Patrick[5], MA, PhD; Kisha Coa[1], MPH, PhD; Erik Augustson[6], MPH, PhD

[1]ICF International, Rockville, MD, United States

[2]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, United States

[3]ICF International, Washington, DC, United States

[4]ICF International, Fairfax, VA, United States

[5]Applied Behavior Change Science, Science and Product, Envolve PeopleCare, Bethesda, MD, United States

[6]Tobacco Control Research Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

Corresponding Author:
Erik Augustson, MPH, PhD
Tobacco Control Research Branch
Division of Cancer Control and Population Sciences
National Cancer Institute
9609 Medical Center Drive
Rockville, MD, 20850
United States
Phone: 1 240 276 6774
Fax: 1 240 276 7907
Email: augustse@mail.nih.gov

Abstract

Background: Social media platforms are increasingly being used to support individuals in behavior change attempts, including smoking cessation. Examining the interactions of participants in health-related social media groups can help inform our understanding of how these groups can best be leveraged to facilitate behavior change.

Objective: The aim of this study was to analyze patterns of participation, self-reported smoking cessation length, and interactions within the National Cancer Institutes' Facebook community for smoking cessation support.

Methods: Our sample consisted of approximately 4243 individuals who interacted (eg, posted, commented) on the public Smokefree Women Facebook page during the time of data collection. In Phase 1, social network visualizations and centrality measures were used to evaluate network structure and engagement. In Phase 2, an inductive, thematic qualitative content analysis was conducted with a subsample of 500 individuals, and correlational analysis was used to determine how participant engagement was associated with self-reported session length.

Results: Between February 2013 and March 2014, there were 875 posts and 4088 comments from approximately 4243 participants. Social network visualizations revealed the moderator's role in keeping the community together and distributing the most active participants. Correlation analyses suggest that engagement in the network was significantly inversely associated with cessation status (Spearman correlation coefficient = −0.14, P=.03, N=243). The content analysis of 1698 posts from 500 randomly selected participants identified the most frequent interactions in the community as providing support (43%, n=721) and announcing number of days smoke free (41%, n=689).

Conclusions: These findings highlight the importance of the moderator for network engagement and provide helpful insights into the patterns and types of interactions participants are engaging in. This study adds knowledge of how the social network of a smoking cessation community behaves within the confines of a Facebook group.

(J Med Internet Res 2016;18(8):e205) doi:10.2196/jmir.5574

KEYWORDS

social network analysis; smoking cessation; Facebook; social support; Web-based communities; social media; communication

Introduction

Although the prevalence of smoking has been steadily declining, an estimated 42.1 million adults in the United States still currently smoke cigarettes [1,2]. Tobacco use continues to be the single largest preventable cause of death and disease in the United States, accounting for 1 of every 5 deaths [1,2]. Reducing the number of smoking individuals remains crucial for improving public health in the United States. For this reason, the National Cancer Institutes' (NCIs) Smokefree.gov program exists as an open access Web-based smoking cessation intervention designed to provide resources and support to current smokers as well as family and friends looking to support someone who wants to quit smoking.

The Smokefree.gov program evolved from a single website to a suite of resources and supporting platforms as technology and literature for Web-based health interventions has advanced. There has been an emphasis on providing personalized resources for specific subgroups of smokers, and this analysis focuses on the Smokefree Women Facebook page, a resource developed specifically for women. Approximately 15% of women in the United States smoke [3], and female smokers face unique challenges in quitting, including weight concerns, quitting while pregnant, stress, and depression and thus can potentially benefit from gender-specific smoking cessation resources [4]. To this end, the Smokefree Women Facebook page was launched in 2009 to engage women and their social networks in the conversation on quitting smoking. The Smokefree Women Facebook page is meant to serve as a virtual support community and provides quit smoking and health information from the NCI to encourage members to lead healthier, smoke-free, lives and to engage women and their social networks in the conversation on quitting smoking. The purpose of the current analysis is to better understand the social network behavior and interactions of participants within the Smokefree Women Facebook group.

Systematic reviews of the literature on Web-based social networks for health indicate the need for further use and evaluation of tools such as Facebook for health-related information dissemination, surveillance, observation, prediction, and behavior change [5-7]. Maher et al [7] call for novel applications of computational methodologies to allow for nuanced understanding of social network sites without manipulation of the network. One such approach, social network analysis, is a collection of computational methods that can provide insight into the structure of a social networking site and interactions and behaviors among participants.

Existing research using social network analysis uncovers phenomena that may contribute to the efficacy of interventions. For example, in a controlled social network experiment, it was found that social reinforcement signals facilitated by clustering of social ties improved individual behavior adoption—adoption in this study being the act of registering for an Internet-based health forum [8]. Beyond benefits for individual action, clustered networks served to diffuse the behavior more quickly than random networks [6]. In addition, a social network analysis study on the smoking cessation website "QuitNet" concluded that characteristics for a sustainable social network include persistence of members over time, heterogeneity of cessation status, and bidirectional communications [9]. In a content-based network analysis of "QuitNet," theme-based thresholds (eg, support theme, advice theme) for identification of meaningful theme-based social subnetworks and identification of opinion leaders and subcommunity clusters within the theme-based networks [10]. Further application of social network analysis in existing behavior change support communities may uncover underlying psychosocial mechanisms useful for driving innovation and strategy in public health interventions.

To our knowledge, few if any studies have performed social network analysis within Facebook groups. Given the nature of the platform, network behavior in a Facebook group is inherently confounded by the Facebook algorithm [11]. In other words, all content, interactions, and users within and outside the group are subject to various manipulations that serve the goal of the host. Rather than deter the study of networks on Facebook, these conflicts mandate their study. Many health interventions, resources, programs, and naturally spawned social support groups live on Facebook due to its ease of usage and large existing audience. To support and improve the functionality of these communities, visualizations of network behavior within the confines of the platform are necessary.

Launched in 2009, the Smokefree Women Facebook group saw improved growth and engagement after a significant strategy shift documented by Post et al [12]. The strategy modification focused on repurposing user-generated content to encourage engagement in lieu of primarily promoting Smokefree resources [12]. With over 22,000 fans ("likes" on the page) at the time of data collection averaging 1700 user comments, 110 shares, and 6300 likes monthly, the community was ripe for an exploration of user interactions.

The goal of this study was to analyze patterns of participation, self-reported smoking cessation length, and interactions within the Smokefree Women Facebook group. In addition to visualizing the network structure, this study sought to analyze user content for potential themes and explore correlates of self-reported cessation length with placement within the network. The study combined quantitative and qualitative methods to answer the following exploratory research questions: (1) What are the characteristics of the network structure? and (2) How do people interact in the social networking site?

Methods

Sample

The Smokefree Women Facebook page is an open-access smoking cessation community. Built as an extension of the NCIs' Web-assisted tobacco cessation intervention, Smokefree.gov, the Smokefree Women Facebook page was specifically created for the purpose of helping individuals

achieve sustained abstinence from smoking. On the Smokefree Women Facebook page, participants communicate and interact with one another and the moderator by likes, comments, comment likes, and shares (see Table 1 for description). The page moderator ("Page Admin" or "Smokefree Women"), a trained public health professional, frequently posts unique and participant-generated content to the page to facilitate engagement, share information, and support women in smoking cessation. At the time of data collection, the Smokefree Women Facebook page had over 22,000 fans, or participants who had "liked" the page, with up to 2500 participants actively interacting on the Smokefree Women Facebook page monthly. Our sample consisted of individuals who interacted (eg, posted, commented) on the Smokefree Women Facebook page between February 2013 and March 2014 (n=4243).

Table 1. Facebook interactions and definitions.

Interaction	Definition
Moderator posts	Content posted by the moderator on the Facebook page wall (visible to all participants who visit the page and visible on "Home" newsfeed of participants based on Facebook algorithm)
Participant posts	Content posted by participants to the Smokefree Women Facebook page wall (visible on the left side of the page to all participants who visit the page but not necessarily visible on "Home" newsfeed of participants)
Shares	When a participant shares content from the Smokefree Women page on their own wall
Comments	When a participant comments on a post
Likes	When a participant clicks the "thumbs up" button on a post to indicate "liking"
Comment likes	When a participant "likes" the comment of another participant

Phase I

In Phase I of this study, we examined the network structure by conducting social network analysis, which provides a visual and descriptive analysis of the network, including a metric for participant engagement—centrality. We also explore user interactions through automated text analysis, which provides insight on how participant engagement in the network is related to content shared.

Data Collection: Full Sample

Data were collected retrospectively in March 2014 from Simply Measured, a social media management marketing platform for all interactions on the Smokefree Women Facebook page, between February 2013 and March 2014. Table 1 provides an overview of Facebook interactions used for this study. All data published on Facebook are publicly available. Personal identifiers were masked to all except members of the research team directly involved in data analysis.

Analysis Strategies: Social Network Analysis

Networks consist of nodes and edges where an edge connects 2 nodes, and network structure is determined by the pattern of connectivity between all nodes [13]. In the Smokefree Women social network, page participants are considered to be nodes (dots), and an edge (line) between 2 participants represents that they have interacted during the study period. Interactions consisted of comments on posts unless otherwise noted.

Centrality measures are used to evaluate which of the nodes in a network are most important to the network [13]. As network structure is determined using comments and posts, for the purposes of this study, centrality is also a measure of participant engagement in the network. Eigenvector centrality is the measure of centrality used in this study due to the exploratory nature of the research questions. Eigenvector centrality was calculated for each participant in the study.

Eigenvector centrality is a measure of centrality that is based on a recursive definition where a node's importance is determined by the importance of adjacent nodes. This is equivalent to evaluating the leading eigenvector of the adjacency matrix. For instance, consider 2 nodes—A and B—that are each connected to 2 additional nodes. Suppose node A is connected to 2 nodes that are connected to several other nodes and thus link 2 communities, whereas node B is connected to 2 nodes that are at the periphery of a network and not connected to any other nodes; node A will have high eigenvector centrality, whereas node B will have low eigenvector centrality.

Visualization of the Network

For the sake of simplicity, for all visualizations, the network was treated as an undirected graph where edges (connections between nodes) of the network do not have a direction and only indicate that 2 individuals have interacted. Moreover, the network was treated as unweighted (unless otherwise noted), meaning that an edge can represent 1 interaction with a person or 10 interactions with the same person. This unweighted approach focuses on interpretation of the breadth of interactions between people in the network, as opposed to depth of interactions between any 2 people.

The network was visualized with the Smokefree Women Facebook moderator, both included and excluded. Therefore, when the moderator was not present, the visualization represented participant-to-participant interactions specifically. Moreover, there is a magnification of the network without the presence of the moderator and highlighted centrality with a blue-to-red increasing color scale, where blue nodes are the least central, and red nodes are the most central.

A force-directed layout algorithm was used to position the nodes in the network for the visualizations. Specifically, Fruchterman-Reingold's [14] algorithm, based on physical forces, was used. Nodes are attracted or repelled based on the connectivity of the network in a way that produces a visually

appealing representation [14]. For the purpose of this visualization only, the edges of the graph were weighted to increase the interpretability of the graph. The weights between the edges of the network were set to be the number of interactions between the 2 participants represented by the nodes.

All statistical analyses were conducted with the open-source computing tool, Python version 2.7.5, specifically using SciPy, package version 0.12.0 for plotting and positioning of the force-directed algorithm and NetWorx version 1.7 package for centrality calculations.

Analysis Strategies: Automated Text Analysis

Automated text analysis was used to see if a difference existed in topics discussed by participants who are highly engaged and those who are not highly engaged in the network. To identify hubs in the network (ie, participants who have the highest centrality) and characterize the content they contribute compared with other participants in the network, ordered centrality values were plotted, and a threshold for determining the 2 groups (central and peripheral) was chosen to be near the elbow of this curve by visual inspection. To conduct automated analysis of content posted by participants, all participants in the network were divided into 2 groups, split according to the threshold. Subsequently, the top 30 terms preferentially used by those in each group were identified.

To determine the propensity of a term to be used by one group versus the other, a ratio of smoothed relative frequencies was used. Although the moderator was included in the network for calculating the centrality scores, the moderator was excluded from the text analysis to focus on the language that participants use themselves.

Phase II

In Phase II of this study, mixed methods were used to further explore participant interactions on the site. For a randomly sampled subset of the population, qualitative content analysis was used to identify salient themes being discussed and self-reported cessation length. Furthermore, correlational analysis was conducted to determine how participant engagement, as measured by centrality, was associated with self-reported cessation length.

Data Collection: Subsample

A random sequence generator was used to identify a uniform random sample of 500 participants who interacted on the Smokefree Women Facebook page during the study period. Participants included in this subsample are also included in Phase I of the study. However, in Phase II of the study, qualitative content analysis is conducted to gather more detailed information about information shared by these participants. The sample size of 500 participants was chosen because the size of the dataset was feasible for manual coding, yet likely robust enough to provide a representation of participants in the network.

Analysis Strategies: Qualitative Content Analysis

Applied thematic analysis was conducted using an inductive methodology described by Guest et al [15]. Researchers first independently reviewed a subset of the data for familiarization and then reviewed a second time to inductively identify salient themes. These themes were then cross-referenced with previous content analyses of similar topics in an effort to use consistent terms (eg, Burri et al [16]). Once themes were finalized, 2 coders independently analyzed data for 125 of 500 participants (25%) in the subsample to assess inter-rater reliability. An inter-rater reliability of at least .8 agreement using Cohen's kappa is considered 'good'. Once the inter-rater reliability threshold of .8 was reached, the remaining sample (n=375) was split, and participant data were coded by 1 of the 2 coders. One coder conducted an additional analysis assessing each post from all 500 participants to extract any self-report of cessation length (measured in days).

Analysis Strategies: Correlation Analysis

Correlation between self-reported cessation length and the centrality of participants to the network was analyzed to determine the relationship between Facebook interactions and cessation behaviors. During the qualitative data analysis in Phase II, the longest self-reported cessation length for each of the 500 participants in the subsample was identified. One coder went through all the posts of the subsample of 500 participants and documented any posts reporting cessation length. If a participant reported cessation length in more than 1 post, the longest self-reported cessation length was used in this analysis.

Spearman rank correlation was calculated between centrality in the network and longest self-reported cessation length. Spearman rank correlation was used due to the nonlinear nature of the data. Statistical significance of the correlation was evaluated with no correlation as the null hypothesis.

To evaluate how those actively in the process of quitting compare with those who have been smoke free for some time, a subgroup analysis was performed to evaluate the aforementioned correlation for participants who report longest cessation length of less than 1 year and those who report longest cessation length of more than 1 year.

Results

Between February 2013 and March 2014, there were 875 posts and 4088 comments from participants on the Smokefree Women Facebook Page and 1166 posts from the moderator. Roughly 4243 people interacted on the page through posts and comments during the 13-month period of data collection. It is of note that participants who interacted on the page did not have to be fans of the Smokefree Women Facebook Page, and thus interactions observed may have been drawn from outside of the 22,000 fans of the page. Additional information about total and average participant interactions (posts and comments) for the entire time period is provided in Table 2.

Table 2. Summary of participant Facebook interactions (n=4243).

Action	Participants taking action at least once	Average per participant
Posts	875	2.23
Comments	4088	3.59

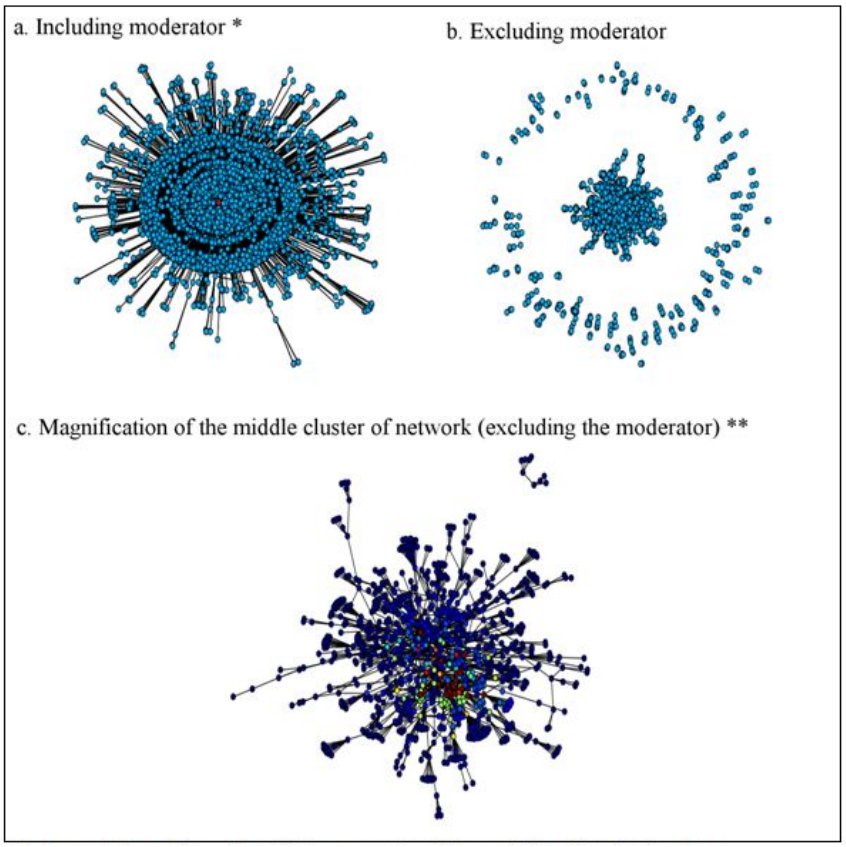
Phase I

Social Network Analysis

Figure 1 displays the network structure with and without the moderator. Each participant is indicated by a node, and Facebook interactions between participants are indicated by edges (lines). The moderator is indicated by the red dot in the center of the visualization of Figure 1 A because the moderator is the most engaged participant in the network. The moderator is visibly the connector of every person in the network. In Figure 1 B, absence of the moderator indicates that there is a large, distinct cluster of participants who interact with many other participants in the middle of the network visualization. A separate ring of participants who interact with few others forms around the middle cluster.

Figure 1 C is a magnified visualization of the cluster of participants in the center of the network without the presence of the moderator. Centrality is illustrated with a blue-to-red increasing color scale. Blue nodes are the least central, and red nodes are the most central. There are several highly engaged people who serve as hubs or large connectors, even without the presence of the moderator, as indicated by the bright colors near the center of the network visualization.

Figure 1. Visualizations of social network. *Moderator indicated by red dot. **Most engaged participants indicated by bright colors.



Automated Text Analysis

A plot of the ordered centralities revealed a threshold for centrality of 0.025 (centrality range: 1×10−7 to 0.7; centrality mean: 0.004). Approximately 100 participants were above this threshold, and 4129 fell below; thus, these participants were labeled as high- and low-engagement participants. Automated text analysis of topics discussed by hubs in the network, or participants with highest centrality compared with other less-engaged participants revealed that hubs, who are most connected to other participants, used more terms of encouragement and congratulations, whereas less-engaged participants discussed more issues related to seeking help, smoking status, and strategies for cessation. The top 30 ranked terms for each group are presented in Table 3.

Table 3. Terms used by more-engaged participants and less-engaged participants.

Rank	More-engaged participant terms	Less-engaged participant terms
1	smokefree	Years
2	good	Quit
3	congrats	smoke
4	go	smoking
5	far	Two
6	saved	Year
7	stay	Help
8	keep	I'm
9	great	cigarette
10	strong	since
11	positive	Need
12	well	Cold
13	hang	turkey
14	wtg[a]	Ago
15	yes	Free
16	today	proud
17	get	days
18	don't	stop
19	water	cigarettes
20	way	smoked
21	think	nicotine
22	better	God
23	awesome	trying
24	come	started
25	job	chantix
26	take	pack
27	cravings	electronic
28	weeks	th[b]
29	wow	Like
30	deep	months

[a]wtg is an acronym for "way to go."

[b]th likely indicates an ordinal number suffix (eg, 5th, 6th, 7th).

Phase II

Content Analysis

Qualitative analysis of content posted by a subset of 500 randomly sampled participants yielded 1698 unique posts or comments from those participants. Table 4 includes the full list of codes. The most frequently occurring themes of posts and comments were providing support (42.52%, 721 of 1698), announcing number of days smoke free (40.58%,689 of 1698), and giving detailed advice (14.61%,248 of 1698; Table 4). The overall tone of conversation was positive (85.32%, n=1447 of 1698).

Table 4. Codes, definitions, and frequencies.

Category	Code	Definition/example	n (%) of Messages (of 1,698)
Message type	Post	Post directly on SFW[a] Facebook wall	165 (9.72)
	Comment	Comment on an existing post on SFW Facebook wall	1533 (90.28)
Tone toward smoking cessation	Positive	The emotional tone or sentiment of the message	1447 (85.32)
	Neutral		229 (13.50)
	Negative		20 (1.18)
Core content	Providing Support	For example, “You can do it!”	721 (42.52)
	Giving Advice	For example, “Try this…”	248 (14.61)
	Seeking Help	For example, “I can't quit, I need ideas.”	83 (4.89)
	Declaration of Days Smokefree (announce quit date)	For example, “I have been smokefree xx days!”	689 (40.58)
	Relapse	For example, “I broke down yesterday and had a cigarette”	81 (4.77)
	Return from Relapse	For example, “I went back to smoking but I am back and ready to quit again”	67 (3.95)

[a]SFW: Smokefree Women.

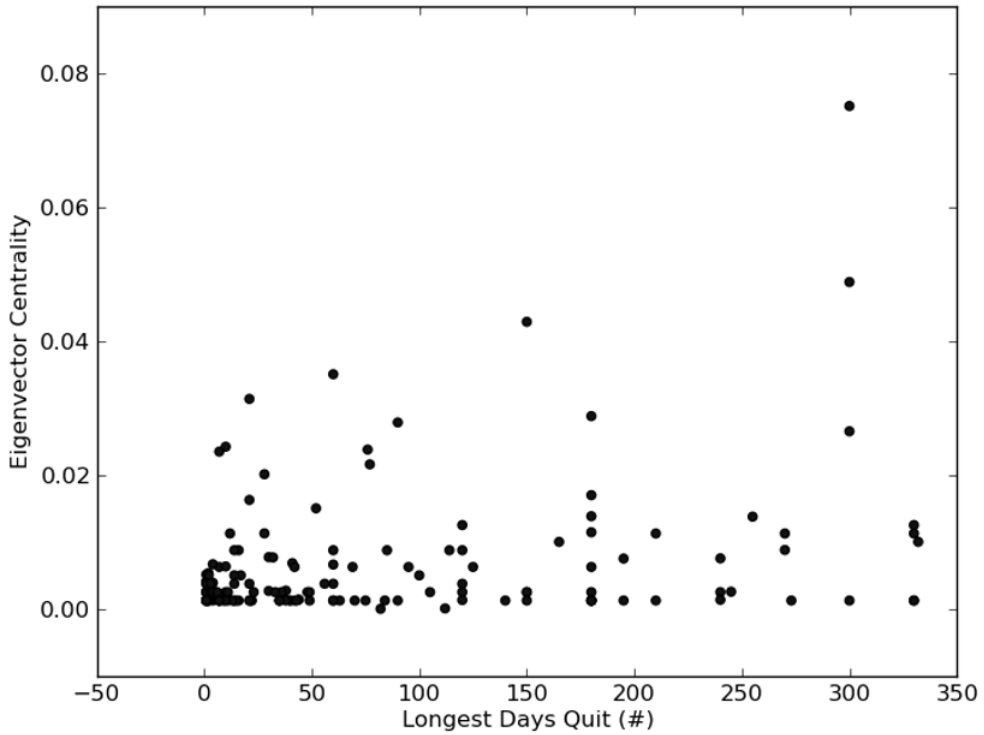
Correlation Analysis

Of the random subsample of 500 participants, 243 people reported how long they had stopped smoking. The longest reported period of cessation was 35 years, and the shortest period was 1 day. Seven participants reported smoking cessation of exactly 1 year, and median time of smoking cessation was 5 months. There is a significant inverse correlation between cessation length and centrality at the 0.05 level (Spearman correlation coefficient = −0.14, P=.03, N=243), meaning that participants who reported longer cessation lengths were less engaged.

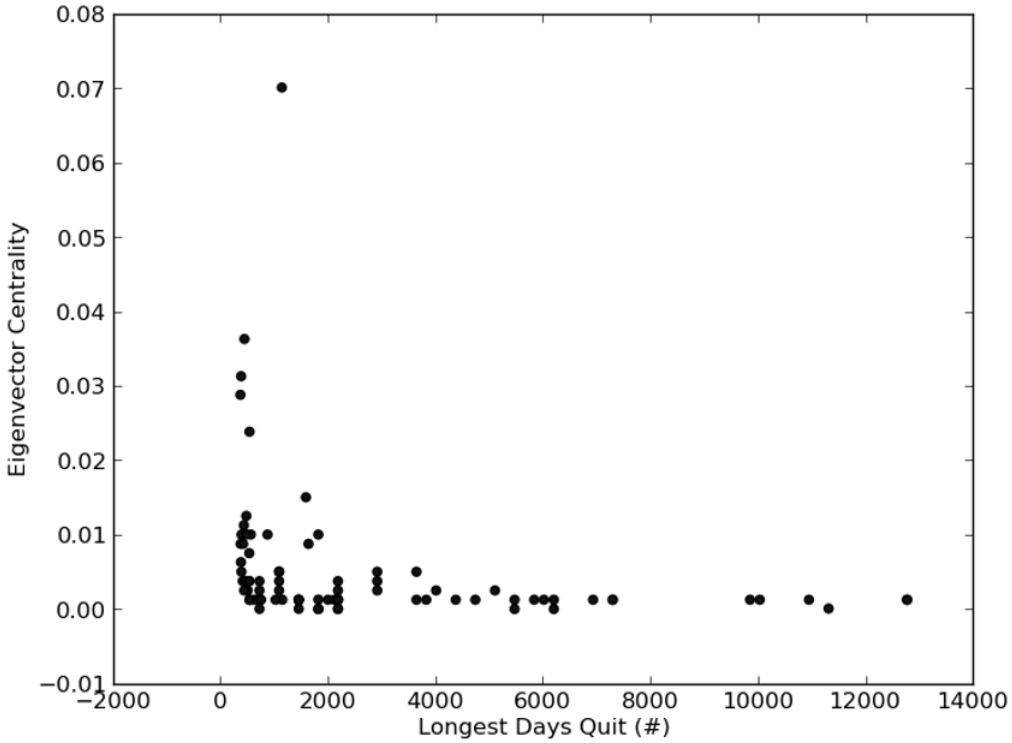
However, splitting the population into those who have been smoke free for less than 1 year versus those who have been smoke free for more than 1 year demonstrates a positive correlation for those who have been smoke free for less than 1 year (Spearman correlation coefficient = 0.20, P=.01, N=155) and a strong inverse correlation for those who have been smoke free for more than 1 year (Spearman correlation coefficient = −0.59, P<.001, N=81; Figure 2).

Figure 2. Cessation Length Versus Eigenvector Centrality.

a. Quit times less than one year (spearman correlation coefficient = 0.20, p=0.01, N=155)



b. Quit times greater than one year (spearman correlation coefficient = -0.59, p<0.001, N=81)



Discussion

Principal Findings

Network visualization with and without the moderator indicates that participants interact with each other in many small clusters linked by a series of hubs or highly engaged participants (see Figure 1). This suggests that the network is robust to random attack (ie, loss of a participant without regard to their position in the network) but sensitive to selective attack (ie, loss of specific members who are hubs of the network). For example, it is clear that the network is affected quite severely by the loss of the moderator, a particularly important hub of the network [16].

This does not come as a surprise, as participant interaction on a Facebook page is driven, in part, by moderator posts. It is of note that the Facebook platform uses a proprietary algorithm to serve page moderator posts to Facebook participants who have previously interacted with the page. These posts appear in a participant's home newsfeed, where most interactions on Facebook take place. The more interaction (eg, comments, likes, shares) a moderator post receives, the more Facebook participants are likely to see the post in their newsfeed [11].

Although the moderator serves as a connector of each person in the network, there are several hubs in the network that serve as large connectors of other participants that are less engaged. This finding supports existing evidence for "super participants" in social network sites [17,18]. Having these highly connected super participants is an advantage for the Smokefree Women Facebook page because they help maintain participant-to-participant interaction, rather than only moderator-to-participant interaction. Furthermore, because there are many hubs, when certain hubs leave the network, there are others who continue interacting and connecting participants.

Automated text assessment of content in the network supports findings from the qualitative content analysis in this study, indicating that participants primarily come to the network to provide and receive support and advice, as well as to mark milestones in their smoke-free journey. When stratified by centrality, automated text analysis suggests that highly engaged hub participants use language that is more congratulatory and supportive, whereas other less central participants seek support, discuss strategies for cessation, and announce their smoke-free status.

These findings are further enhanced by findings from the correlational analysis of self-reported smoke-free status, which suggest that participants become more central to the network as they maintain their smoke-free status and use the network for social support but become less involved in the network as maintaining their smoke-free status becomes less difficult and requires less community support. This is consistent with previous research, which found that social network site participants who had recently stopped smoking were more likely to be the first to respond to posts [19]. It is possible that those who have quit more recently feel connected to challenges of other community members and hence assume a more central role.

Taken together, findings from this study suggest that participants who are less central, or are not hubs, are a combination of people at the beginning of their smoke-free journey and people who have been smoke-free for an extended time and only come back to the network to announce their sustained smoke-free status. On contrary, participants who are more central and connect many people (ie, hubs) appear to be those who are further along in their cessation journey and come to the network to provide support and perhaps in the process gain support.

Limitations

Limitations of this research include lack of analysis of participant demographics; lack of analysis of shares and likes for the study period; and use of self-report for cessation status. This study did not analyze participant demographic information due to privacy restrictions of Facebook that prevented this information from being publicly available. However, as this analysis was focused on properties of the network, demographic details were not central to the goals of this study.

In addition, the study did not include data on shares and likes, additional actions that could be taken by participants in the network, because these data were not available in an automated fashion from Facebook. In a separate analysis, share and like data were manually collected for a period that spanned 3 months. Network visualizations conducted in Phase I were replicated for these 3 months, using all possible Facebook interactions (ie, posts, comments, shares, likes, comment likes). No observable differences were identified when comparing network visualizations for the full sample using only posts and comments as participant interactions, with the subset using all possible Facebook interactions (data not shown).

Moreover, smoke-free status was self-reported by participants on their own volition, sometimes in response to comments on moderator posts asking how long they have been smoke free and at other times in general conversation. Given that the status of everyone in the network is not known, the possibility exists that the number of people who have made a cessation attempt while engaged with the network is underestimated in this study. Furthermore, the correlation analysis may be biased because it did not take into account potential relapse. Nonetheless, it is possible to gain insight on how many participants have made at least 1 cessation attempt through the subset analysis. Future research should explore opportunities to obtain this information from the entire study population and examine the association between cessation length and engagement prospectively.

Implications for Research and Practice

Ultimately, researchers of social network sites for health seek to understand whether participation in a social network site such as the Smokefree Women Facebook page can lead to better health outcomes (eg, increased quit rates). Although this study was not designed to answer that question, through observation of naturally occurring interactions of the social network and use of methods such as social network analysis, study findings provide insight into the network structure of the social networking site that stand to inform research and practice. Future research may seek to integrate social analysis data with survey data on use of smoking cessation–related social media sites and

smoking status or cessation length data to determine the mechanisms by which social media sites can facilitate quit attempts and sustained cessation.

Practitioners may use findings from this study to improve design, implementation, and program evaluation of social network sites focused on health behavior. On the basis of the findings from this study, practitioners may consider the following: (1) developing personas that mimic participants at various points of the cessation trajectory and tailoring the experience with the social network site to fit characteristics of each persona; (2) assessing the best strategies for moderation of the network to determine whether moderator posts that attract comments such as questions, requests for advice, and direct quotes are more effective in providing social support to participants than moderator posts that solicit likes and shares, given the large role the moderator plays in the network; (3) determining whether participants benefit from interaction with only other participants on the Facebook page, or if they also receive benefit from participants in their personal network that may not be participants in that particular social network site for health behavior; and (4) exploring the use of paid advertising on Facebook to boost posts to ensure that users who are new to the community and most in need of support see the posts that will benefit them most.

Conflicts of Interest

None declared.

References

1. Jamal A, Agaku I, O'Conner E, King B, Kenemer J, Neff L. Morbidity and Mortality Weekly REport. Atlanta, GA: Centers for Disease Control and Prevention; 2014 Nov 28. Current Cigarette Smoking Among Adults — United States, 2005–2013 URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6347a4.htm?s_cid=mm6347a4_w [accessed 2016-07-07] [WebCite Cache ID 6ipZPZKrI]
2. U.S. Department of HealthHuman Services. A Report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services; 2014. The Health Consequences of Smoking - 50 Years of Progress URL: http://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf [accessed 2015-08-01] [WebCite Cache ID 6fZOeg2td]
3. Agaku I, King B, Dube S. Morbidity and Mortality Weekly Report. Atlanta, GA: Centers for Disease Control and Prevention; 2014 Jan 17. Current Cigarette Smoking Among Adults — United States, 2005–2012 URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6302a2.htm [accessed 2014-08-01] [WebCite Cache ID 6fZQaz3gB]
4. Bottorff JL, Haines-Saah R, Oliffe JL, Sarbit G. Gender influences in tobacco use and cessation interventions. Nurs Clin North Am 2012 Mar;47(1):55-70. [doi: 10.1016/j.cnur.2011.10.010] [Medline: 22289398]
5. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
6. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, Lau Annie Y S. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
7. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
8. Centola D. The spread of behavior in an online social network experiment. Science 2010 Sep 3;329(5996):1194-1197 [FREE Full text] [doi: 10.1126/science.1185231] [Medline: 20813952]
9. Cobb N, Graham A, Abrams D. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
10. Myneni S, Cobb N, Cohen T. Finding meaning in social media: content-based social network analysis of QuitNet to identify new opportunities for health promotion. Stud Health Technol Inform 2013;192:807-811. [Medline: 23920669]
11. Facebook. 2014. How news feed works URL: https://www.facebook.com/help/327131014036297/ [accessed 2016-07-07] [WebCite Cache ID 6ipaUqj69]
12. Post S, Taylor S, Sanders A, Goldfarb J, Hunt Y, Augustson E. If you build (and moderate) it, they will come: the Smokefree Women Facebook page. J Natl Cancer Inst Monogr 2013 Dec;2013(47):206-208 [FREE Full text] [doi: 10.1093/jncimonographs/lgt019] [Medline: 24395993]
13. Goldenberg A, Zheng A, Fienberg S, Airoldi E. Foundations and Trends in Machine Learning. 2009 Dec. A survey of statistical network models URL: https://arxiv.org/abs/0912.5410 [accessed 2016-07-07] [WebCite Cache ID 6ipariz4M]
14. Fruchterman T, Reingold E. Graph drawing by force-directed placement. Software—Practice & Experience 1991 Nov;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
15. MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications, Inc; 2011.
16. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. Nature 2000 Jul 27;406(6794):378-382. [doi: 10.1038/35019019] [Medline: 10935628]

17. Burri M, Baujard V, Etter J. A qualitative analysis of an internet discussion forum for recent ex-smokers. Nicotine Tob Res 2006 Dec;8 Suppl 1:S13-S19. [Medline: 17491166]
18. Chou W, Prestin A, Lyons C, Wen. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
19. Selby P, van Mierlo T, Voci S, Parent D, Cunningham J. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]

Edited by G Eysenbach; submitted 29.02.16; peer-reviewed by L Laranjo, R Grainger; comments to author 17.03.16; revised version received 23.06.16; accepted 26.06.16; published 02.08.16.

Please cite as:
Cole-Lewis H, Perotte A, Galica K, Dreyer L, Griffith C, Schwarz M, Yun C, Patrick H, Coa K, Augustson E
Social Network Behavior and Engagement Within a Smoking Cessation Facebook Page
J Med Internet Res 2016;18(8):e205
URL: http://www.jmir.org/2016/8/e205/
doi:10.2196/jmir.5574
PMID:27485315

Original Paper

Reaching Adolescent Gay, Bisexual, and Queer Men Online: Development and Refinement of a National Recruitment Strategy

Tonya L Prescott[1], BA; Gregory Phillips II[2], PhD; L. Zachary DuBois[2,3], PhD; Sheana S Bull[4], MPH, PhD; Brian Mustanski[2], PhD; Michele L Ybarra[1], MPH, PhD

[1]Center for Innovative Public Health Research, San Clemente, CA, United States

[2]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Department of Anthropology, California State University Long Beach, Long Beach, CA, United States

[4]Community and Behavioral Health, Colorado School of Public Health, University of Colorado, Denver, CO, United States

Corresponding Author:
Michele L Ybarra, MPH, PhD
Center for Innovative Public Health Research
555 N. El Camino Real #A347
San Clemente, CA, 92672
United States
Phone: 1 877 302 6858 ext 801
Fax: 1 877 302 6858
Email: michele@innovativepublichealth.org

Abstract

Background: Using social networking websites to recruit research participants is increasingly documented in the literature, although few studies have leveraged these sites to reach those younger than 18 years.

Objective: To discuss the development and refinement of a recruitment protocol to reach and engage adolescent gay, bisexual, and other teenaged men who have sex with men (AGBM). Participants were recruited for development and evaluation activities related to Guy2Guy, a text messaging–based human immunodeficiency virus infection prevention program.

Methods: Eligibility criteria included being between 14 to 18 years old; being a cisgender male; self-identifying as gay, bisexual, and/or queer; being literate in English, exclusively owning a cell phone, enrolled in an unlimited text messaging plan, intending to keep their current phone number over the next 6 months, and having used text messaging for at least the past 6 months. Recruitment experiences and subsequent steps to refine the Internet-based recruitment strategy are discussed for 4 research activities: online focus groups, content advisory team, beta test, and randomized controlled trial (RCT). Recruitment relied primarily on Facebook advertising. To a lesser extent, Google AdWords and promotion through partner organizations working with AGBM youth were also utilized.

Results: Facebook advertising strategies were regularly adjusted based on preidentified recruitment targets for race, ethnicity, urban-rural residence, and sexual experience. The result was a diverse sample of participants, of whom 30% belonged to a racial minority and 20% were Hispanic. Facebook advertising was the most cost-effective method, and it was also able to reach diverse recruitment goals: recruitment for the first focus group cost an average of US $2.50 per enrolled participant, and it took 9 days to enroll 40 participants; the second focus group cost an average of US $6.96 per enrolled participant, and it took 11 days to enroll 40 participants. Recruitment for the first content advisory team cost an average of US $32.52 per enrolled participant; the second cost US $29.52 per participant. Both recruitment drives required 10 days to enroll 24 participants. For the beta test, recruitment cost an average of US $17.19 per enrolled participant, and it took 16 days to complete enrollment of 20 participants. For the RCT, recruitment cost an average of US $12.54 per enrolled participant, and it took 148 days to enroll 302 participants. Google AdWords campaigns did not result in any enrolled participants of whom the research staff members were aware.

Conclusions: Internet-based strategies can be a cost-efficient means to recruit and retain hard-to-reach populations from across the country. With real-time monitoring of participant demographic characteristics, diverse samples can be achieved. Although Facebook advertising was particularly successful in this study, alternative social media strategies can be explored in future research as these media are ever-changing.

(J Med Internet Res 2016;18(8):e200) doi:10.2196/jmir.5602

KEYWORDS

Facebook; mHealth; recruitment methods; intervention development; HIV; adolescent; AGBM; sexual minority

Introduction

Among the youth in the United States, 9 in 10 new human immunodeficiency virus (HIV) infections occur among adolescent gay, bisexual, and other teen men who have sex with men (AGBM) between the ages of 13-19 years [1]. Clearly, to affect HIV incidence, efforts to reach and engage AGBM are critical. In the face of this obvious need, it is perhaps surprising that few validated prevention programs are available [2]. In part, this lack of health programming reflects the challenge of successfully recruiting and retaining these hard-to-reach youth. More specifically, these challenges to recruitment and retention include the need to obtain parental permission or an institutional review board (IRB) waiver for those younger than 18 years [3]; gaining the trust of youth who may face social stigma for their sexual minority status; and recruiting a sufficient sample size, particularly if the sampling frame is local rather than national. Internet-based recruitment has the potential to address some of these challenges, but it can also introduce challenges in terms of validating the respondent and ensuring a diverse sample. Certainly with the popularity of social networking sites, particularly Facebook (FB), among adolescents [4], researchers are beginning to use Web-based recruitment strategies to engage hard-to-reach populations [5-11]. Even with the increasing use of other social networking sites (eg, Twitter, Instagram, Vine) FB remains the most popular site with youth, including 72% of adolescent men [4]. Furthermore, among youth who use social networking sites, 1 in 3 report using FB exclusively. Importantly, too, racial and ethnic minority youth are well represented among FB users. Usage is relatively similar by urbanicity (ie, 77% of urban, 75% of rural, 67% of suburban teenagers) and socioeconomic status (eg, 77% of youth from households earning less than US $50,000 annually compared with 68% of those from households earning more than US $50,000 annually) as well [4]. Given this diversity and breadth of reach, particularly among hard-to-reach youth, social networking sites may be an ideal place for researchers to recruit study participants.

Although the use of social networking websites to recruit research participants is increasingly documented in the literature [12], few studies have leveraged social networking websites to specifically reach *children* (ie, 18 years of age and younger). We also were unable to find any studies that documented the recruitment experiences of sexual minority youth in particular. To address this research gap, we will discuss our experiences developing and refining a national Internet-based recruitment protocol targeting AGBM 18 years of age and younger, primarily through FB advertising and also with Google AdWords and assistance from lesbian, gay, bisexual, and transgender (LGBT)–focused organizations. Participants were recruited for development and evaluation activities related to Guy2Guy, a text messaging–based HIV prevention and healthy sexuality program for AGBM. Lessons learned here have the potential to inform future HIV prevention programs aiming to recruit hard-to-reach populations, particularly sexual minority children.

Methods

The research protocol was reviewed and approved by Chesapeake Institutional Review Board, the Center for Innovative Public Health Research's IRB of record, and the Northwestern University IRB. Youth provided informed assent or consent, depending on their age, and completed a capacity to consent assessment [13-15]. A waiver of parental permission for participants younger than 18 years was obtained, primarily because requiring parental consent could increase risk to participants who may be victimized by their parents as a result of disclosing their sexual minority status [16]. The waiver also avoided the potential for fatal sampling bias that might occur if only youth who were out to their parents chose to enroll [17]. In addition to a waiver of parental permission, a Certificate of Confidentiality [18] was obtained from the National Institutes of Health to protect participants' data from subpoenas and other law enforcement efforts.

Given the wide reach of both social media and text messaging, we developed a recruitment plan that facilitated enrollment of youth across the United States. The recruitment protocol was refined across 4 different national research activities that were delivered either online or via text messaging. These research activities included online focus groups to better understand AGBM sexual decision making, online content advisory teams to vet the proposed program messages, a beta test of the Guy2Guy text messaging program, and the randomized controlled trial (RCT) of Guy2Guy. Here, we describe the recruitment experiences and resulting recruitments of our strategy across time and research activities.

Eligibility Criteria

Eligibility criteria for all research activities included being 14-18 years of age; being cisgender men (ie, those whose current gender identity and sex assigned at birth are both male); self-identifying as gay, bisexual, and/or queer; and being English-speaking. To promote the likelihood of the sample reflecting those who might use the text messaging–based program if it were publicly available, participants were also required to be exclusive owners of a cell phone, be enrolled in an unlimited text messaging plan, intend to keep their current phone number for the next 6 months, and have used text messaging for at least the past 6 months. Exclusion criteria included knowing another person already enrolled in the program and participating in a previous study activity (eg, those who took part in the focus groups were ineligible to take part in the RCT). The same eligibility criteria were applied for each study development activity with one exception: male gender identity was directly queried after the focus groups.

Recruitment

Participants were recruited primarily using FB ads, which must adhere to character limits (ie, 25 characters for the headline, 90

characters for the body text) and include an image. These ads must also be approved by FB. Our ads were targeted based on the user's location (United States), age (14-18 years), and sex (note "gender" is the term used by FB; male). We also included a list of 65 keywords (referred to as "interests" in the FB ad manager) relevant to AGBM including pop culture interests (eg, LGBT community, Katy Perry). Two types of ads were displayed depending on how the user accessed FB: News Feed ads (ie, ads are embedded in the dynamic news field central column) and right column ads (ie, ads are displayed in the static column on the right side of the webpage). People accessing FB on their desktop computer received both types of ads; those accessing the social networking site on a mobile device were only shown News Feed ads.

During the online focus groups, we asked partner organizations working with AGBM youth to promote the study. In subsequent research activities, we also used Google AdWords, which grants free advertising to the Center for Innovative Public Health Research for being a nonprofit organization. These ads included a title line of up to 25 characters, 2 description lines with up to 35 characters each, and a URL field. Ads linked to the project website, which included a brief project description and a Web-based contact form for those who wanted to be contacted for potential participation. The contact form was purposefully brief to reduce burden and promote form completion. As such, it consisted of 15 questions that queried study eligibility questions as well as demographic characteristics that were used to facilitate the recruitment of a diverse sample.

Enrollment

Web-based eligibility screeners received were emailed to the project coordinator. Ineligible candidates were sent a link to a sexual health website [19] relevant to sexual minority youth via email. Eligible candidates and those whose eligibility was unclear (ie, because of "do not want to answer" responses) were sent a text message by research staff to schedule an appointment via telephone. On the call, research staff confirmed the candidate's eligibility, explained the study, conducted a decisional capacity assessment [13-15], and obtained verbal assent or consent to participate. Those in the beta test and RCT also completed a self-safety assessment to determine whether it was safe for them to receive text messages about sensitive topics (eg, anal sex, being gay or bisexual) on their cell phones. The assent form, which included contact information for the IRB and principal investigator, was emailed to participants for later reference. All study recruitment and enrollment documents (eg, Web-based screener, assent or consent forms, self-safety assessment) can be found online [20]. For the focus groups and content advisory teams, enrollment occurred at the time of verbal assent or consent. Participants in the beta test and RCT were enrolled after they completed the baseline survey and confirmed receipt of a text message from the study program to ensure that their phone was compatible with the study software.

To ensure a diverse sample for the beta test and RCT, recruitment goals were preidentified by creating allocation "bins" (ie, recruitment goals) for race, ethnicity, urban-rural residence, and sexual experience (ie, ever versus never had anal or vaginal sex). Race and ethnicity targets were based on the 2010 US population demographics, which were the most current data available at the time of recruitment (ie, 72.4% white, 12.6% black, 16.3% Hispanic) [21]. Our targets were revised slightly so that the sample had a greater percentage of minority race (35%) and Hispanic ethnicity (20%). We also targeted 50% of the sample to be sexually inexperienced (ie, never had anal or vaginal sex) and 20% living in a rural residence. Urban-rural residence targets were identified to ensure sufficient representation among youth living in outer communities that often have fewer available LGBT resources. A Web-based study interface was programmed to automatically track the demographic characteristics of enrolled participants, allowing staff to monitor in real time the characteristics of the sample. Youth were enrolled sequentially until their bin was filled.

Protecting Against Deception by Participants When Recruiting Online

Several measures were taken to limit the potential for deception by participants during the recruitment and enrollment process (ie, the same participant enrolling in the study multiple times, an ineligible candidate providing fake responses to become eligible), as this is often a concern when conducting Web-based research in which researchers are unable to see the candidate face-to-face [22]. The project description provided in the Web-based screener did not include study eligibility criteria or mention an incentive. Furthermore, to reduce a candidate's ability to identify and provide the "right" answers for eligibility, the eligibility form did not exclusively include questions necessary to determine study eligibility, but also included questions to target enrollment (eg, race, ZIP code, sexual experience) and additional questions (eg, how they found the website). The eligibility screener also captured the candidate's Internet protocol address, which allowed us to identify possible duplicate entries. Additionally, eligibility questions were queried again over the phone and compared with the answers provided previously on the Web-based screener. Discrepant responses for questions expected to be consistent (eg, age) were questioned further, whereas discrepant responses for more fluid questions (eg, sexual identity) were not further explored.

Measures

The outcome measures for this enrollment case study included the following.

Advertisement Metrics

Quantification of study interest was based on measures provided by the FB and Google AdWords analytics reporting. "Clicks" referred to the number of total clicks on the ad. "Reach" referred to the number of people the ad was shown to. "Unique clicks" referred to the number of unique people who clicked on the ad. "Click-through rate" (CTR) referred to the number of clicks received divided by the number of impressions (ie, number of times the ad was shown). "Unique click-through rate" (uCTR) referred to the number of unique clicks received divided by the number of unique people the ad reached. The "average cost per click" (CPC) was calculated as the amount spent advertising divided by the CTR.

Recruitment Efficiency

The length of time to recruit the target sample and the number of youth who completed a screener needed to successfully enroll an eligible participant were indicators of recruitment efficiency. For the RCT, we also reported the number of phone calls required to enroll 1 person to quantify recruitment efficiency.

Sample Diversity

Because health disparities are apparent by race and ethnicity [1] and by rural versus urban community residence (as youth in rural communities often lack access to LGBT resources compared with those in urban settings), we believed it critically important to ensure the enrollment of a diverse sample. As such, we had a complex sampling scheme based on race, ethnicity, urban versus rural residence (determined by ZIP code), age, and sexual identity. For the sake of parsimony, we report the sample characteristics for the RCT only.

Results

Online Focus Groups

A total of 4 focus groups were conducted to inform the development of program content and the protocol: 2 with sexually experienced youth and 2 with sexually inexperienced youth. A more detailed description of the focus group methods and results are described elsewhere [23-25].

Advertisement Metrics

Facebook ads for the first round of focus groups (ie, 1 with sexually experienced youth, 1 with sexually inexperienced youth) were submitted to FB for approval on November 9, 2012 and approved the same day. The ads ran for 4 days (ie, November 9-12). On the basis of FB recommendations, pricing was set with a maximum bid per click of US $0.90 and maximum daily budget of US $25. Additionally, our LGBT-focused partner organization posted an announcement regarding the study on its website forum from November 3 to 7, 2012.

At a total cost of US $100, with an average of US $0.68 per click, the FB ad campaign resulted in 148 clicks and a CTR of 0.04 (Table 1). We spent an average cost of US $2.50 per enrolled participant (n=40).

In January 2013, another recruitment effort was implemented to enroll youth for the second set of focus groups. Facebook ads were submitted 5 days before the intended recruitment start date to ensure we had approval; ads were approved the same day of submission. The FB ads ran for 9 days (ie, January 14-22, 2013). Pricing was set again with a maximum bid per click of US $0.90 with a maximum daily budget of US $25. Another LGBT-focused partner organization also posted an announcement of the study on its FB page on January 12, 2013.

A total of 13 screeners were received during the first 3 days of recruitment. Of these, 6 appeared to be eligible. To invigorate enrollment, our previous LGBT-focused organization partner emailed the recipients in its mailing list on January 16, 2013. We also modified the FB recruitment ads by removing the targeted interests or keywords (eg, Katy Perry), adding targeting criteria to include teenaged men who were “interested in men,” updating the image to one thought to be more relevant (Figure 1), increasing the daily ad budget to US $100, and changing the maximum bid per click to US $0.74. These changes doubled our selected audience (ie, the number of people the ad had the potential to reach based on ad targeting criteria) from 21,000 to 46,000 and resulted in 143 newly completed screeners in a 24-hour period.

The FB ads were active for a total of 7 days, posted nonconsecutively between January 14 and 22, 2013 (ie, ads were paused or stopped during the ad campaign depending on recruitment needs). We spent an average of US $6.96 per enrolled participant in this second effort (Table 1; n=40).

Table 1. Facebook advertisement placement metrics by study activity.

Facebook ad placement metrics by study activity	Placement	Impression device	Reach	Clicks	Unique clicks	CTR[a]	uCTR[b]	Cost[c]	CPC[c,d]	Cost[c] per unique click	Cost[c] per each enrolled participant
Focus group 1[e] (n=80)											
	N/A[f]	N/A	93,823	148	148	0.04	N/A	$100.00	$0.68	N/A	$2.50
Focus group 2[g] (n=80)											
	N/A	N/A	78,235	846	774	0.11	1.11	$278.35	$0.39	$0.42	$6.96
Content advisory team 1[h] (n=24)											$32.52
	Right column ads on home page	Desktop or laptop	220,332	1667	1525	0.06	0.27	$483.99	$0.16	$0.18	
	Right column ads	Desktop or laptop	161,051	839	767	0.04	0.22	$296.54	$0.19	$0.21	
	Unknown placement	Other	335	0	0	0	0	$0.00	$0.00	$0.00	
Content advisory team 2[i] (n=24)											$29.52
	Right column ads on home page	Desktop or laptop	52,184	418	381	0.07	0.51	$101.12	$0.21	$0.23	
	Right column ads	Desktop or laptop	37,327	168	157	0.05	0.35	$46.29	$0.22	$0.23	
	News Feed	Desktop or laptop	17,175	271	221	1.45	1.23	$95.44	$0.34	$0.42	
	News Feed	Mobile phone or tablet	133,202	1432	1256	1.82	1.51	$465.65	$0.33	$0.39	
	Unknown placement	Other	3	0	0	0	0	$0.00	$0.00	$0.00	
Beta test[j] (n=20)											$17.19
	Right column ads on home page	Desktop or laptop	13,276	71	67	0.11	0.4	$24.36	$0.27	$0.28	
	Right column ads	Desktop or laptop	12,071	84	79	0.1	0.53	$35.05	$0.28	$0.28	
	News Feed	Desktop or laptop	7006	144	115	1.56	1.36	$40.90	$0.26	$0.30	
	News Feed	Mobile phone or tablet	34,886	1430	1031	3.13	2.35	$243.58	$0.23	$0.31	
RCT[k] (n=302)											$12.54
	Right column ads on home page	Desktop or laptop	54,304	607	548	0.17	0.79	$455.27	$0.72	$0.77	
	Right column ads	Desktop or laptop	43,736	411	386	0.12	0.81	$346.01	$0.68	$0.72	
	News Feed	Desktop or laptop	34,234	728	607	1.36	1.28	$349.81	$0.57	$0.64	
	News Feed	Mobile phone or tablet	194,084	10,535	7333	3.54	3.12	$2635.36	$0.35	$0.42	

[a] CTR: click-through rate.

[b] uCTR: unique click-through rate.

[c] All costs are in US dollars.

[d] CPC: cost per click.

[e] Facebook (FB) ad pricing structure: maximum bid per click of US $0.90. Daily budget: US $25.

[f] N/A: not applicable (FB did not have the information available at the time of recruitment).

[g] Initial FB ad pricing structure: maximum bid per click of US $0.90. Daily budget: US $25. Modifications to ad: 3 days into recruitment, the interest targeting was removed (eg, Katy Perry), criteria to target teenaged men "interested in men" was added, and the ad image to be more relevant to the population. Daily ad budget was also increased to US $100 and the maximum bid per click changed to US $0.74.

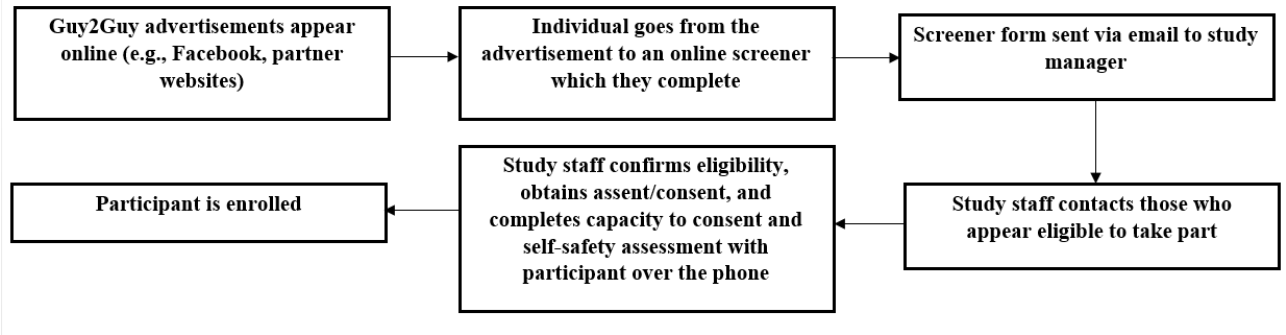
[h] FB ad pricing structure: maximum bid per click of US $0.69. Daily budget: US $150. Modifications to ad: before launch of content advisory team 1 ads, multiple variations of the same ad, in which the headline and tagline were different, were created.

[i] Daily budget: US $150. Modifications to ad: before launch of content advisory team 2 ads, pricing was modified to be optimized for clicks as opposed to preidentifying a maximum bid per click.

[j] Daily budget: US $100. The same FB ad settings as described in the content advisory team 2 were used in the beta test effort.

[k] RCT: randomized controlled trial. Daily budget ranged from US $50 to $100. Modifications to ad: during the RCT, FB added the option to select if one was attracted to "men or women," which was integrated into the ad campaign (eg, allowing us to better target those that may identify as bisexual). Previously just the "men" or "women" options were available. New ad images were introduced toward the end of field that were thought to be more salient to specific populations (ie, younger participants, nonwhite race). Given the time span spent on recruitment ads were regularly adjusted to target based on age (eg, if we wanted to reach more 14-year-olds, ads were modified to specifically target those who are 14 years of age based on their FB profile).

Figure 1. Guy2Guy flowchart.



Recruitment Efficiency

We received 209 completed screeners from the first round of focus groups, of which 94/209, 45.0% did not meet the study eligibility criteria. The most common reasons for ineligibility were age 41/94, 44%; not being cisgender male 20/94, 21%; and not being enrolled in an unlimited text messaging plan 12/94, 13%. For the second round of focus groups, we received 251 completed screeners, of which 99/251, 39.4% did not meet the study eligibility criteria. The most common reasons for ineligibility were not being a cisgender male 30/99, 30%; not being the exclusive owner of a cell phone 20/99, 20%; and not being enrolled in an unlimited text messaging plan 17/99, 17%.

Among the 115 candidates who appeared eligible from their Web-based screener responses during the first round of focus groups, 77/115, 67.0% candidates were contacted during the enrollment period. The remaining eligible candidates were not contacted because either the target sample size was met or their particular allocation bin was full; this applies for all study activities. It took 9 days to enroll 40 participants (ie, 20 sexually experienced, 20 sexually inexperienced) for the first round of focus groups. For the second round of focus groups, among the 152 candidates who appeared eligible given their screener responses, 75/152, 49.3% were contacted. Candidates were contacted until our target sample size of 20 participants was met for each focus group. The second round of focus groups required 11 days to enroll 40 participants.

Content Advisory Teams

Advertisement Metrics

Next, we recruited a content advisory team, comprising 10 young participants from our target population, to review the draft of our program content. To do so, we made several changes to the FB campaign originally launched for the focus groups: we created multiple versions of the same ad in which the headline and tagline varied, increased the daily budget to US $150, and changed the maximum bid per click to be between US $0.25 and $0.69. The FB ads were live for 10 days between July 5 and 14, 2013. These changes resulted in an average cost of US $32.52 per enrolled participant (Table 1; n=24).

We also created a Google AdWords campaign of 187 keywords (eg, gay guys, gay teen, gay chat rooms). Ads were targeted by: network (all), device (all), location (United States), and language (English). The bid strategy was focused on clicks, manual maximum CPC bidding, and a set maximum daily budget of US $329 per day. A total of 3 different Google AdWords campaigns ran for 7 days (July 6-12, 2013) and received 962 clicks, 79,315 impressions, a CTR of 1.2%, and CPC of US $1.61. The total cost of the campaign, which was included in the advertising budget grant covered by Google AdWords, was US $1547.07.

After integrating findings from the first content advisory team into the content, we recruited a second content advisory team to confirm the modifications to the messaging. The FB recruitment campaign was further modified to optimize the pricing to get more clicks (ie, FB automatically bids to maximize the clicks to a website that can be achieved with the campaign budget) as opposed to preidentifying a maximum bid per click

[26]. The FB ads were active for 7 days between August 28 and September 5, 2013, resulting in an average cost of US $29.52 per enrolled participant (n=24).

We used the same targeting criteria as described above in our Google AdWords campaign. As an exception, we set the network specifically to Google search networks (ie, a group of search-related websites, such as Google search sites and non-Google search sites—like AOL—that partner with Google to show search ads, where the ad may appear), rather than both search and display networks (ie, display networks consist of a group of more than a million websites, videos, and apps where the ad may appear) [27]. A total of 4 Google AdWords campaigns ran for 10 days (August 28 to September 6, 2013) using the same list of keywords noted above. Google AdWords received 2055 clicks, 209,257 impressions, a CTR of 0.98%, a CPC of US $1.59, and a total cost of US $3267.79.

Recruitment Efficiency

We received 271 Web-based screeners from the first content advisory team, of which 132/271, 48.7% did not meet the study eligibility criteria. The most common reasons for ineligibility were age 45/132, 34.1%, not being the exclusive owner of a cell phone 42/132, 31.8%, and not being enrolled in an unlimited text messaging plan 10/132, 7.6%. For the second content advisory team, 246 completed screeners were received, of which 84/246, 34.1% did not meet the study eligibility criteria. The most common reasons for ineligibility were not being enrolled in an unlimited text messaging plan 30/84, 36%; not being the exclusive owner of a cell phone 21/84, 25%; and having not used text messaging for 6 months or more 17/84, 20%.

Among the 139 eligible candidates based on the Web-based screener responses received during the first content advisory team, 79/139, 56.8% candidates were contacted during the enrollment period. For the second content advisory team, among the 162 eligible candidates, 69/162, 42.6% candidates were contacted. A total of 48 participants were enrolled in both content advisory teams. None of the participants were enrolled through the Google AdWords campaign; all were from FB outreach efforts.

It took 10 days to enroll 24 participants in the first content advisory team as well as the second content advisory team, including the time to schedule and complete the enrollment telephone calls.

Beta Test

Advertisement Metrics

The same FB ad settings as described above in the second content advisory team were used in the beta test effort, with the exception of reducing the daily budget to US $100. The FB ads were active for 7 nonconsecutive days between March 9 and 29, 2014, and we spent an average cost of US $17.19 per each enrolled participant (Table 1; n=20).

In an attempt to increase the reach of our Google AdWords, 6 campaigns were created for the beta test recruitment effort. Additional headlines hypothesized to be more attention-grabbing (ie, by asking a question: “Are you a gay or bi teen?”) were added. Google AdWords were active for 6 days (March 12-17, 2014) and generated 599 clicks, 42,092 impressions, a CTR of 1.4%, CPC of US $1.43, and a total cost of US $854.46.

Recruitment Efficiency

We received 236 completed screeners during the beta test recruitment, of which 68/236 28.8% did not meet the study eligibility criteria at the initial screening. The most common reasons for ineligibility were not being the exclusive owner of a cell phone 21/68, 31%; did not intend to keep the current cell phone for 6 months or more 21/68, 31%; and not being cisgender male 15/68, 22%.

Among the 168 eligible candidates based on the Web-based screener responses received, 51/168, 30.4% candidates were contacted during the enrollment period; 40/51, 78.4% candidates responded; and 20/40, 50.0% participants were enrolled in the beta test. Again, no participants were enrolled through the Google AdWords campaign. It should be noted that the question querying how candidates heard about the research study was removed from the beta test Web-based screener. That said, no participants reported hearing about the study through Google AdWords.

It took 16 days to enroll 20 participants into the beta test. The length of time to complete enrollment exceeded that for previous recruitment activities because we newly implemented our targeting strategy to ensure a diverse sample of participants. This protocol was added at this step to ensure feasibility during the RCT enrollment effort.

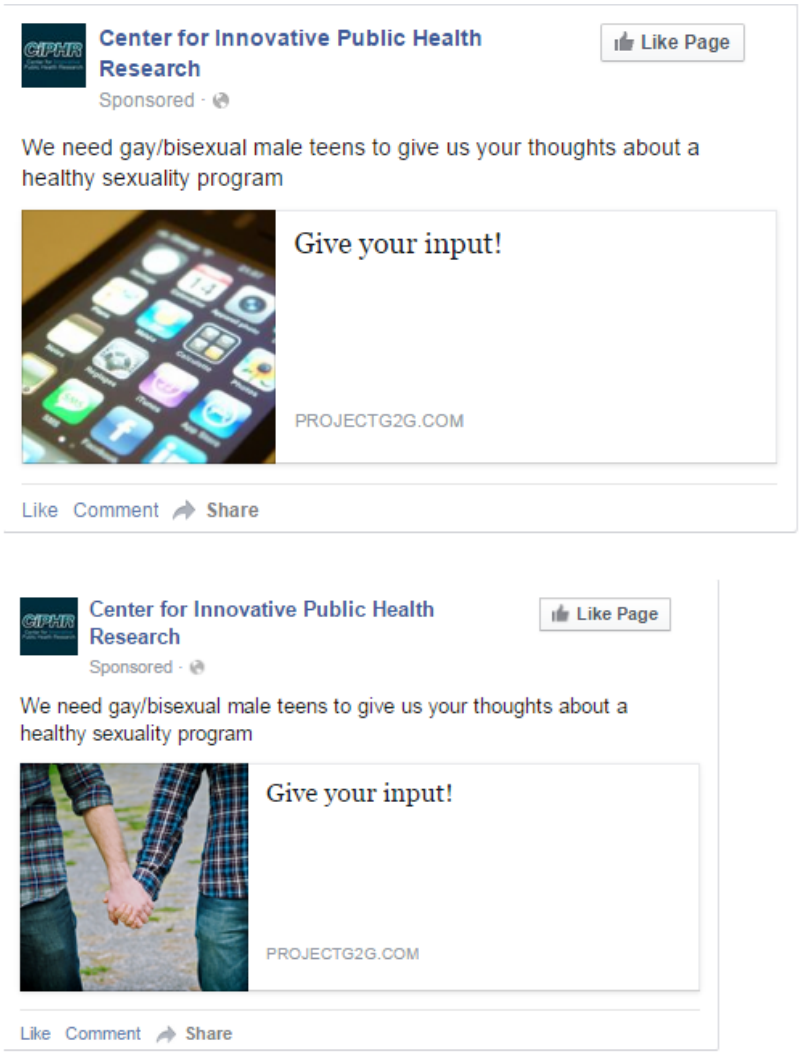
Randomized Controlled Trial

Advertisement Metrics

The same FB ad settings used in the beta test were implemented in the RCT. We were able to take advantage of a newly added FB targeting category that allowed us to advertise to users who were “interested in men and women”; previously, we were only able to target those who were “interested in men.” New ad images were used toward the end of field that were thought to be more salient to specific populations (ie, younger participants, nonwhite race; Figure 2, numbers 2-5). The FB ads were active for 52 nonconsecutive days between June 20 and October 31, 2014, and cost an average of US $12.54 to enroll each of the 302 RCT participants (Table 1). Notably, 32.3% (US $1221.55/$3787.08) of the recruitment money was spent to reach and enroll the last 10% of the sample, as these represented the youth who were particularly difficult to reach (eg, 14- to 15-year-old black youth; 14-year-old youth). Excluding these participants, the average cost per enrolled participant was US $9.47.

In total, 4 Google AdWords campaigns using the same targeting criteria as above were active for 15 days (June 16-30, 2014) before they were discontinued. The ads generated 1324 clicks, 132,843 impressions, a CTR of 1.00%, a CPC of US $1.59, and a total cost of US $2105.29.

Figure 2. Comparison of the original versus modified focus group Facebook recruitment advertisements.



Recruitment Efficiency

We received 1522 completed screeners during the RCT recruitment, of which 411/1522, 27.0% were ineligible based on their responses to the Web-based screener. The most common reasons for ineligibility were not being cisgender male 129/411, 31.4%; did not intend to keep current cell phone for 6 months or more 108/411, 26.3%; and not being enrolled in an unlimited text messaging plan 77/411; 18.7%. Among the 1111 eligible or potentially eligible screeners received, 600/1111, 54.0% candidates were contacted sequentially based on their allocation bin during the enrollment period. Those who were eligible but their allocation bin was filled were not contacted. Among those contacted, 494/600, 82.3% candidates responded, and 342/494, 69.2% candidates spoke to study staff, ultimately resulting in the enrollment of 328/342, 95.9% participants, of whom 302/328, 92.1% were randomized into the RCT.

It took 148 days to enroll and randomize 302 participants at an average rate of 13.7 participants per week (range 1-28 participants). Research staff aimed to enroll between 15 and 20 participants per week, so that a manageable number of participants were actively receiving the intervention at any given time. An average of 1.54 contact attempts were required to enroll study participants (range 1-6).

Sample Diversity

Screeners were more commonly completed by sexually experienced 865/1522, 56.8% than inexperienced 628/1522, 41.3% young men. Inexperienced youth were also slightly more likely to respond to staff outreach about the program, 258/494, 52.2% versus 236/494, 47.8%. In total, 1036/1522, 68.1% of the screeners were from white youth, compared with 91/1522, 5.98% from black or African American youth and 39/1522, 2.6% from Asian youth.

Facebook advertising was regularly adjusted based on the allocation bins (Figure 3). For example, if all the bins for 18-year-olds were filled, we adjusted the targeting on FB ads to include only 14- to 17-year-olds. Also, when noticeably fewer screeners were received from 14- to 15-year-olds, in part because of fewer FB profiles of young people in this age range, we created FB ads that specifically targeted these younger teenagers. This purposeful sampling strategy (Table 2) resulted in a sample that was 14.2% black or African American and 23.2% living in rural settings.

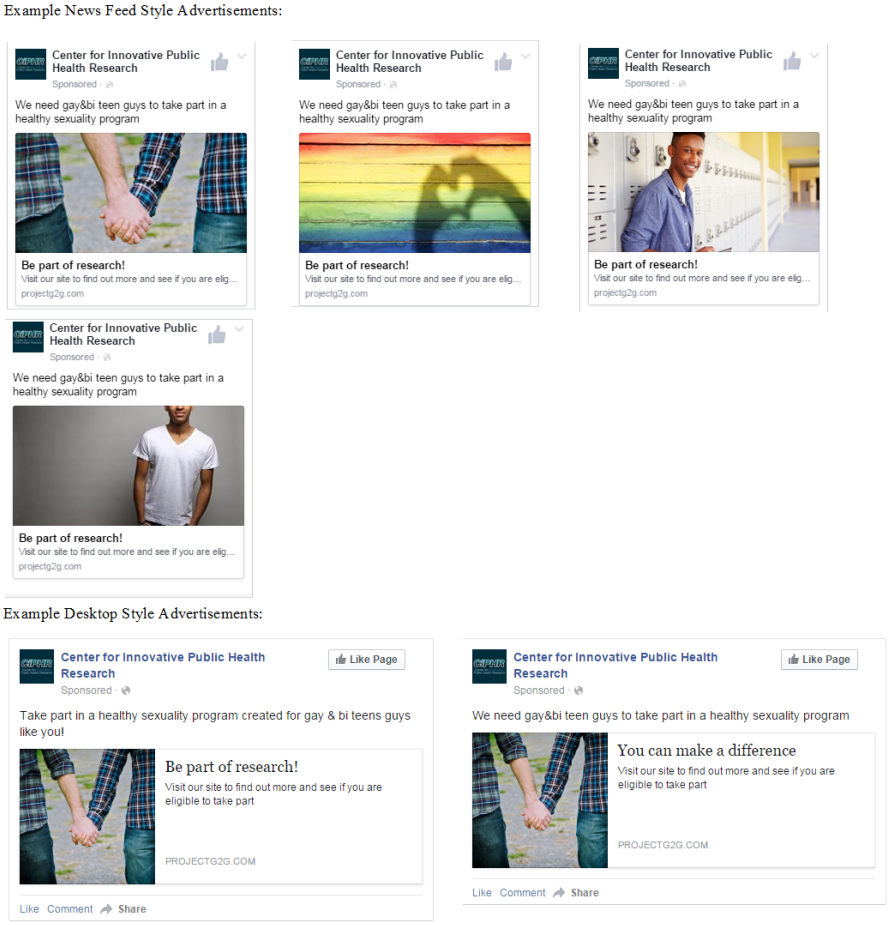
Table 2. Randomized controlled trial randomized participant demographic characteristics based on allocation bins (N=302).

Participant characteristics[a]	n (%)
Sexually experienced	
Yes	153 (50.7)
No	149 (49.3)
Age, years	
14-15	116 (38.4)
16-18	186 (61.6)
Race	
White	204 (67.5)
Black	43 (14.2)
All other	55 (18.2)
Hispanic	
Yes	67 (22.2)
No	235 (77.8)
Sexual identity	
Gay and/or queer	195 (64.6)
Bisexual	107 (35.4)
Type of community	
Rural	70 (23.2)
Urban	232 (76.8)

[a]Demographic characteristics reflect participant responses during the phone screener, as most of these questions were not again queried in the baseline survey. With exception, sexual experience was assigned based on response to the baseline survey as a more comprehensive battery of questions were asked that would subsequently lead to participants being assigned to the sexually experienced or sexually inexperienced program.

Challenges in balancing the sample were encountered when participants provided different responses on the phone or baseline survey than what they had reported in the Web-based screener, resulting in the allocation bin being overfilled (Figure 4). For example, although the same behaviorally based items to query sexual experience were used in the Web-based screener and baseline survey, 29 participants indicated a different sexual experience history than they had first reported in the Web-based screener. Similar issues were also observed with other demographic characteristics used for allocation (eg, race, n=38; ethnicity, n=22).

Figure 3. Example of Facebook recruitment advertisements used during the randomized controlled trial.



One Facebook advertisement image used is excluded because of copyright permissions. The image depicted 2 African American teens—one with his arm around the other.

Figure 4. Example of real-time enrollment monitoring tool on the Guy2Guy study administrative interface.

Allocation data

Demographic Allocation (Experienced)								
	14-15 yrs.				16-18 yrs.			
	Not Hispanic		Hispanic		Not Hispanic		Hispanic	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
White	13 / 7	23 / 23	2 / 2	8 / 7	10 / 10	33 / 35	5 / 3	12 / 10
Black	0 / 2	2 / 9	0 / 1	2 / 2	5 / 3	13 / 13	1 / 1	1 / 3
All Other	0 / 2	2 / 6	1 / 1	2 / 2	3 / 2	11 / 10	0 / 1	4 / 2

Demographic Allocation (Inexperienced)								
	14-15 yrs.				16-18 yrs.			
	Not Hispanic		Hispanic		Not Hispanic		Hispanic	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
White	11 / 7	22 / 23	1 / 2	7 / 7	10 / 10	35 / 35	2 / 3	10 / 10
Black	0 / 2	5 / 9	0 / 1	0 / 2	1 / 3	12 / 13	0 / 1	1 / 3
All Other	1 / 2	10 / 6	1 / 1	3 / 2	3 / 2	10 / 10	0 / 1	4 / 2

Randomization Allocation				
	Experienced		Inexperienced	
	Intervention	Control	Intervention	Control
Gay / Queer	45	49	53	48
Bisexual	32	27	20	28

Within each bin, the number on the right is the target sample size, the number on the left is the number currently enrolled. Black font reflects a bin that has not yet reached its preset target sample size. Orange font reflects a bin that has reached or gone over the target sample size. For example, 13/7 signifies that 13 white, rural, non-Hispanic, sexually experienced 14- to 15-year-olds were enrolled, but the target based on the allocation bin was 7.

Discussion

On the basis of our iterative development of an Internet-based recruitment strategy aimed at 14- to 18-year-old AGBM, our findings suggest that protocols designed to use FB to reach a national sample of diverse youth will be most effective if they use images salient to the target population, are broad (eg, targeted based on minimum eligibility criteria, such as age, sex) rather than too specific (eg, not targeted on interests or keywords), are regularly monitored for performance, and the ad targets are regularly modified to meet recruitment needs.

Our findings suggest that not all Web-based advertising has the same effect when trying to reach AGBM adolescent men. Indeed, our Google AdWords outreach effort did not result in a single participant being enrolled. This is likely because our reach—based on the keyword targeting we utilized—was too diffuse. Beyond keywords, we could only target Google AdWords campaigns by location (eg, United States). Therefore, the people who saw our ads went well beyond our very specific population of interest. On the other hand, FB advertising was much more successful, resulting in the enrollment of 405/450, 90.0% of our study sample across the 4 research activities. The FB ad manager allowed the ad to be targeted on multiple criteria beyond location, such as gender, age, and the relationship interests that FB users indicate in their profile (ie, interested in men, interested in men and women). In sexual health research, this ability to target ads based on relationship interest can narrow the ad audience down to the target audience, in our case AGBM, thereby increasing recruitment cost-effectiveness and efficiency. It is worth noting that our experiences may not generalize to other populations and research topics, however [28,29].

Although our specific recruitment goals were identified before recruitment (eg, race, ethnicity, age), the ability to reach recruitment goals is nonetheless dependent on having those from diverse backgrounds actually see the recruitment advertisement and subsequently complete a screener. Although FB ad manager did not allow for tailoring based on diversity targets (eg, race, ethnicity) at the time, resulting screeners reflected a sufficiently diverse group of youth that allowed us to follow-up with youth who met our diversity goals. It should be noted that updating the ad image so that it better resonates with a specific population or targeting the ad to specific cities or states that have demographic profiles of interest can also invigorate sample diversity.

Limitations

These findings should be considered within the context of the study's limitations. For example, by targeting FB users through the "interested in" targeting category, we targeted those youth who are more likely out because they indicated their dating preferences on a social networking profile. However, FB users can choose to make specific parts of their profile private (eg, only visible to them or their friends), so that they can still complete their profile while limiting who is able to see their information (eg, parents, nonfriends). Additionally, by recruiting on FB, this inherently excludes those youth who are not on FB. Given that FB continues to be the most popular social networking site for teenagers, however [30], this recruitment method still provides a unique opportunity to reach a national sample of adolescents. Indeed, some may wonder why we did not use other Web-based recruitment strategies such as the Turk. This is because our goal was to test the interest of our intervention among AGBM teenagers in places where they "hang out" online rather than places where people go to earn money.

Conclusions

This research has elucidated important and effective strategies for utilizing social media for the recruitment of youth to research. The utilization of social media for recruitment of hard-to-reach youth, particularly sexual minority teenaged males for whom privacy may be an issue, is increasingly important as targeting these youth for health, support, and sexual health programs is a public health imperative. Although our research indicates the cost-effectiveness and overall efficiency of using Facebook in particular for recruiting AGBM into research, alternative social media strategies should be explored as they become more popular with youth. It will be essential to continue documenting successful outlets for recruitment of hard-to-reach youth. Given that the research project we described herein was sensitive in nature and we were able to demonstrate success in Web-based recruitment, we are confident that these strategies can be utilized to recruit AGBM for research related to sexual health topics and perhaps other topics. Additional research is needed to demonstrate the success of these strategies with other populations.

Acknowledgments

The authors would like to acknowledge all study participants for their vital role in completing this study. This study was supported by a grant from the National Institute of Mental Health (R01MH096660). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. HIV Surveillance in Adolescents and Young Adults URL: http://www.cdc.gov/hiv/pdf/statistics_surveillance_Adolescents.pdf [accessed 2016-06-12] [WebCite Cache ID 6iCn1qZ9f]

2. Mustanski BS, Newcomb ME, Du Bois SN, Garcia SC, Grov C. HIV in young men who have sex with men: a review of epidemiology, risk and protective factors, and interventions. J Sex Res 2011 Mar;48(2-3):218-253 [FREE Full text] [doi: 10.1080/00224499.2011.558645] [Medline: 21409715]
3. Fisher CB, Mustanski B. Reducing health disparities and enhancing the responsible conduct of research involving LGBT youth. Hastings Cent Rep 2014 Sep;44 Suppl 4:S28-S31 [FREE Full text] [doi: 10.1002/hast.367] [Medline: 25231783]
4. Lenhart A. Pew Research Center. Washington, DC: Pew Internet & Life Project; 2015. Teens, social media, & technology overview URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2015-11-12] [WebCite Cache ID 6czUnl1N1]
5. Gilligan C, Kypri K, Bourke J. Social networking versus facebook advertising to recruit survey respondents: a quasi-experimental study. JMIR Res Protoc 2014;3(3):e48 [FREE Full text] [doi: 10.2196/resprot.3317] [Medline: 25230740]
6. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
7. Ramo DE, Rodriguez TM, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
8. Bull SS, Levine D, Schmiege S, Santelli J. Recruitment and retention of youth for research using social media: Experiences from the Just/Us study. Vulnerable Children and Youth Studies 2013 Jun;8(2):171-181. [doi: 10.1080/17450128.2012.748238]
9. Ezell JM, Saltzgaber J, Peterson E, Joseph Christine L M. Reconnecting with urban youth enrolled in a randomized controlled trial and overdue for a 12-month follow-up survey. Clin Trials 2013 Oct;10(5):775-782 [FREE Full text] [doi: 10.1177/1740774513498320] [Medline: 23983157]
10. 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]
11. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
12. Lane TS, Armin J, Gordon JS. Online Recruitment Methods for Web-Based and Mobile Health Studies: A Review of the Literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
13. Dunn LB, Jeste DV. Enhancing informed consent for research and treatment. Neuropsychopharmacology 2001 Jun;24(6):595-607 [FREE Full text] [doi: 10.1016/S0893-133X(00)00218-9] [Medline: 11331139]
14. The UCSD Task Force on Decisional Capacity. UCSD. San Diego, CA: The University of California; 2003. Procedures for Determination of Decisional Capacity in Persons Participating in Research Protocols URL: https://irb.ucsd.edu/DecisionalCapacityAssessment.pdf [accessed 2016-06-12] [WebCite Cache ID 6iCod2Rhh]
15. Moser DJ, Schultz SK, Arndt S, Benjamin ML, Fleming FW, Brems CS, et al. Capacity to provide informed consent for participation in schizophrenia and HIV research. Am J Psychiatry 2002;159(7):1201-1207. [doi: 10.1176/appi.ajp.159.7.1201] [Medline: 12091200]
16. Mustanski B. Ethical and regulatory issues with conducting sexuality research with LGBT adolescents: a call to action for a scientifically informed approach. Arch Sex Behav 2011 Aug;40(4):673-686. [doi: 10.1007/s10508-011-9745-1] [Medline: 21528402]
17. Mustanski B. Future directions in research on sexual minority adolescent mental, behavioral, and sexual health. J Clin Child Adolesc Psychol 2015;44(1):204-219 [FREE Full text] [doi: 10.1080/15374416.2014.982756] [Medline: 25575125]
18. National Institutes of Health. NIH. 2016. Certificates of Confidentiality (CoC) URL: https://humansubjects.nih.gov/coc/index [WebCite Cache ID 6iCnutQZL]
19. Centers for Disease Control and Prevention. Atlanta, GA: Centers for Disease Control and Prevention; 2014. Lesbian, Gay, Bisexual, and Transgender Health URL: http://www.cdc.gov/lgbthealth/ [accessed 2016-07-13] [WebCite Cache ID 6iwrSJywL]
20. Center for Innovative Public Health Research. CiPHR. 2015. Guy to Guy (G2G) URL: https://innovativepublichealth.org/projects/guy-to-guy/ [accessed 2016-06-12] [WebCite Cache ID 6iCnrKc3U]
21. United States Census Bureau. Profile of general population and housing characteristics: 2010 demographic data URL: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_DP_DPDP1&prodType=table [WebCite Cache ID 6iCnxvgy5]
22. Konstan J, Simon RB, Ross M, Stanton J, Edwards W. The story of subject naught: A cautionary but optimistic tale of internet survey research. J Comput Mediat Commun 2005;10(2). [doi: 10.1111/j.1083-6101.2005.tb00248.x]
23. DuBois LZ, Macapagal KR, Rivera Z, Prescott TL, Ybarra ML, Mustanski B. To have sex or not to have sex? An online focus group study of sexual decision making among sexually experienced and inexperienced gay and bisexual adolescent men. Arch Sex Behav 2015 Oct;44(7):2027-2040. [doi: 10.1007/s10508-015-0521-5] [Medline: 25925896]
24. Mustanski B, DuBois LZ, Prescott TL, Ybarra ML. A mixed-methods study of condom use and decision making among adolescent gay and bisexual males. AIDS Behav 2014 Oct;18(10):1955-1969. [doi: 10.1007/s10461-014-0810-3] [Medline: 24906532]

25. Ybarra ML, DuBois LZ, Parsons JT, Prescott TL, Mustanski B. Online focus groups as an HIV prevention program for gay, bisexual, and queer adolescent males. AIDS Educ Prev 2014 Dec;26(6):554-564 [FREE Full text] [doi: 10.1521/aeap.2014.26.6.554] [Medline: 25490735]
26. Facebook. Bidding and optimization. Menlo Park, CA: Facebook; 2016. URL: https://www.facebook.com/unsupportedbrowser [accessed 2016-07-07] [WebCite Cache ID 6iphsqvgx]
27. Google. Mountain View, CA: Google; 2016. Choosing the campaign type that's right for you URL: https://support.google.com/adwords/answer/2567043 [accessed 2016-06-12] [WebCite Cache ID 6iCoC8M0H]
28. Ybarra ML, Prescott TL, Holtrop JS. Steps in tailoring a text messaging-based smoking cessation program for young adults. J Health Commun 2014 Dec;19(12):1393-1407. [doi: 10.1080/10810730.2014.901441] [Medline: 24766267]
29. Ybarra ML, Prescott TL, Espelage DL. Stepwise Development a Text Messaging-Based Bullying Prevention Program for Middle School Students (BullyDown). JMIR Mhealth Uhealth 2016;4(2):e60 [FREE Full text] [doi: 10.2196/mhealth.4936] [Medline: 27296471]
30. Madden M. Pew Research Center. Washington, DC: Pew Research & Internet Project; 2013. Teens Haven't Abandoned Facebook (Yet) URL: http://www.pewinternet.org/2013/08/15/teens-havent-abandoned-facebook-yet/ [accessed 2016-06-12] [WebCite Cache ID 6iCoII9IU]

Abbreviations

AGBM: adolescent gay, bisexual, and other teenaged men who have sex with men
CPC: cost per click
CTR: click-through rate
FB: Facebook
HIV: human immunodeficiency virus
IRB: institutional review board
LGBT: lesbian, gay, bisexual, and transgender
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 04.02.16; peer-reviewed by M Prietula, K Horvath; comments to author 09.03.16; revised version received 12.06.16; accepted 13.06.16; published 04.08.16.

Please cite as:
Prescott TL, Phillips II G, DuBois LZ, Bull SS, Mustanski B, Ybarra ML
Reaching Adolescent Gay, Bisexual, and Queer Men Online: Development and Refinement of a National Recruitment Strategy
J Med Internet Res 2016;18(8):e200
URL: http://www.jmir.org/2016/8/e200/
doi:10.2196/jmir.5602
PMID:27492781

Original Paper

Frequent Surfing on Social Health Networks is Associated With Increased Knowledge and Patient Health Activation

Dafna Grosberg[1], RN, MHA; Haya Grinvald[2], MPH; Haim Reuveni[3], MD; Racheli Magnezi[2], MHA, MBA, PhD

[1]Bar Ilan University, Interdisciplinary Department for Social Sciences, Ramat Gan, Israel

[2]Bar Ilan University, Public Health and Health Systems Management Program, Ramat Gan, Israel

[3]Ben Gurion University of the Negev, Department of Health Policy and Management, Beer Sheva, Israel

Corresponding Author:
Racheli Magnezi, MHA, MBA, PhD
Bar Ilan University
Public Health and Health Systems Management Program
Bar Ilan
Ramat Gan, 52900
Israel
Phone: 972 35317128
Fax: 972 39013221
Email: magnezir@biu.ac.il

Abstract

Background: The advent of the Internet has driven a technological revolution that has changed our lives. As part of this phenomenon, social networks have attained a prominent role in health care. A variety of medical services is provided over the Internet, including home monitoring, interactive communications between the patient and service providers, and social support, among others. This study emphasizes some of the practical implications of Web-based health social networks for patients and for health care systems.

Objective: The objective of this study was to assess how participation in a social network among individuals with a chronic condition contributed to patient activation, based on the Patient Activation Measure (PAM).

Methods: A prospective, cross-sectional survey with a retrospective component was conducted. Data were collected from Camoni, a Hebrew-language Web-based social health network, participants in the diabetes mellitus, pain, hypertension, and depression/anxiety forums, during November 2012 to 2013. Experienced users (enrolled at least 6 months) and newly enrolled received similar versions of the same questionnaire including sociodemographics and PAM.

Results: Among 686 participants, 154 of 337 experienced and 123 of 349 newly enrolled completed the questionnaire. Positive correlations (P<.05) were found between frequency and duration of site visits and patient activation, social relationships, and chronic disease knowledge. Men surfed longer than women (χ^2_3=10.104, P<.05). Experienced users with diabetes surfed more than those with other illnesses and had significantly higher PAM scores (mean, M=69.3, standard deviation, SD=19.1, PAM level 4; Z=−4.197, P<.001) than new users (M=62.8, SD=18.7, PAM level 3). Disease knowledge directly predicted PAM for all users (β=.26 and .21, respectively). Frequency and duration of social health network use were correlated with increased knowledge about a chronic disease. Experienced surfers had higher PAM than newly enrolled, suggesting that continued site use may contribute to increased activation.

Conclusions: Web-based social health networks offer an opportunity to expand patient knowledge and increase involvement in personal health, thereby increasing patient activation. Further studies are needed to examine these changes on other aspects of chronic illnesses such as quality of life and costs.

(J Med Internet Res 2016;18(8):e212) doi:10.2196/jmir.5832

KEYWORDS

Internet; social networks; social media; pain; Patient Activation Measure (PAM); chronic disease; diabetes mellitus

Introduction

According to Pew [1] and Harris [2] polls, more than 70% of Internet users looked on the Internet for health information. Across all age groups, Internet use has increased from 2005 through 2013 [3]. Individuals 18 to 29 years of age are most likely to use social networking sites (89%), and men and women surf equally among them [1].

Patients with chronic illness are less likely to use the Internet, although use is still high at 62% [4]. Moreover, research indicates that Internet users with a chronic illness are more likely to blog or contribute to a Web-based discussion, a listserv, or other Web-based group forum that helps people with personal issues or health problems [5,6].

Patient activation describes how much a patient is involved in his or her health care. It includes self-efficacy, behavior, and knowledge and can predict healthy behaviors, preventive care measures, disease-specific self-care behavior, and information seeking [7]. Activated patients are better at self-management [8-10], which affects health outcomes [11], decreases health care costs, and increases quality of life [12].

The Internet can impact self-management and activate patients. Patient activation varies by socioeconomic status [6,13], and the Internet offers a potential mechanism to reduce these inequalities. Studies include Web-based self-management [14], patient portals [15,16], Web-based patient activation interventions [17], and Web-based interventions for a variety of health-related issues ranging from chronic conditions, health promotion, and mental health. Overall, there have been small, but statistically significant effects [18]. A review of studies of chronically ill samples and the impact of participation in Internet programs that combine health information with Web-based peer support, decision support, or help with behavior, found improvements in users' knowledge, social support, health behaviors, and clinical outcomes. Participants' self-efficacy was enhanced as well [19,20]. In addition, eHealth, Web-based interventions provide cost-effective patient empowerment or at least suggest promising evidence thereof [21].

Little is known about Web-based social networks. A review on the impact of Web-based social networks on health behavior changes concluded that there is very modest evidence that interventions incorporating Web-based social networks may be effective [22,23].

We previously found that a social health network (without interventions) led to increased patient activation, as active participants serve as role models for other participants for health management [24]. They learn self-efficacy, gain knowledge, and learn to behave in ways consistent with self-management. In this study, we examined the association of continued participation in a social health network with patient activation after 3 months of enrollment.

Methods

The Platform

The research was conducted using the Camoni (which means "Like me" in Hebrew) site, which was established in 2008. Camoni is the first Hebrew-language social network. It is targeted to individuals with chronic conditions and it helps them find others facing similar health issues.

The Camoni site is comprised of 16 communities (diabetes mellitus, chronic pain, heart disease, hypertension, obesity, eating disorders, multiple sclerosis, spinal injury, lung disease, kidney disease, stroke, osteoporosis, Crohn's disease, cancer, obesity, and depression), defined according to chronic health conditions. A medical expert (physician, nurse, or psychologist) heads each community. Camoni offers advice, the opportunity to consult with experts, and the chance to converse with other patients who face the same health condition. The site includes blogs, forums, support groups, internal mail, and chats. It also explains each health condition, its diagnosis, and offers practical advice on how to maintain one's health and cope with the disease. Registration is required for active participation on the site, which is open to all. More than 21,000 patients have registered, and 400 to 500 join monthly.

Recruitment

This was a cross-sectional survey with a retrospective element. Data were collected from November 1, 2012, to October 31, 2013. The study focused on the 4 largest Camoni communities: diabetes mellitus (n=5015), pain (n=4255), depression (n=3877), and hypertension (n=3821). Eating disorders were not included despite its size (n=4500) due to community sensitivities. The respondents were divided into 3 groups. The first group, defined as "newly enrolled," was comprised of those who had just joined the site. The second group defined as "experienced respondents," was comprised of those who had used the site for at least 6 months before the beginning of the study. The third group included a subgroup defined as "experienced and new respondents" who answered the initial questionnaire and a follow-up approximately 3 months after the registration. The study was approved by the Ethics Committee of Bar Ilan University.

Instruments

Two similar versions of the same questionnaire were used. The first was for experienced users and included 49 questions, written in the present tense. For example, "How much support are you getting from the site" The second questionnaire was for the newly enrolled respondents and included 48 questions, written in the future tense. For example, "How much support do you expect to get from the site?" The questionnaire was sent to everyone who registered at the site during the study period. Both questionnaires queried sociodemographics, the self-reported effect of the use of the site, and personal involvement in the chronic illness.

The questionnaire was accessible through Google Docs. An invitation to participate and a link were available on the Camoni home page of each of the 4 disease communities. During the recruitment period, reminders were placed in the monthly

newsletter sent to all Camoni respondents who had not declined the option.

Demographic and Health Characteristics

Respondents were asked to provide information about basic demographic variables such as age, gender, education, family status, income, and diagnosis, as well as information regarding their use of the website.

Personal Involvement in Health Care Related to Site Use

This section was based on the work of Lemire et al [25]. It measures the site's influence on respondents' health behaviors. Two similar versions were used. The questionnaire for veteran users included 20 items, and the questionnaire for the new respondents included 17 items. The number of questions for each group differed because some questions were not relevant to new users. For example: "How frequently do you visit the site on average?" Possible responses ranged from "several times a day" to "less than once a month"; "About how long do you spend on the site each time?" Possible responses ranged from "less than 10 minutes" to "more than an hour"; and "Are you an active user of the site (write or post notices, tips, etc.)?" with a "yes" or "no" response. Both versions used a 4-point scale (1=strongly disagree to 4=strongly agree). There were no reversed items. The items were divided into 3 dimensions: (1) Knowledge about the disease—for example, "The site allows me to develop a better understanding of my personal health" (4 items, α=.846); (2) Social relationships—for example, "The site helps me feel less lonely" (4 items, α=.811); and (3) Involvement in personal health—for example, "The site makes me feel more confident about the choices I plan on making" (10 items, α=.916).

Patient Activation Measure (PAM)

This section was based on the work of Hibbard et al [26]. It captures the degree to which patients have the beliefs, knowledge, and skills to manage their condition(s), collaborate with their providers, maintain their health, and access appropriate and high-quality care [27]. The PAM includes 13 items on a 4-point scale (1=strongly disagree to 4=strongly agree). An interval scale of 0 to 100 was calculated (Cronbach α=.899). The interval score can be divided into 4 levels. Level 1: May not yet believe that the patient role is important (PAM score of 45.2 or lower); Level 2: Lacks confidence and knowledge to take action (PAM score of 47.4-52.9); Level 3: Beginning to take action (PAM score of 56.4-66.0); and Level 4: Has no difficulty maintaining behaviors over time (PAM score of 68.5 or above). The PAM questionnaire was translated into Hebrew and was validated [28].

Statistical Analysis

Statistical analyses were performed using the SPSS statistical software, version 21. Chi-square tests for independence were used to examine the dependency between the duration of the site visits and sociodemographic attributes of the experienced users. Spearman correlations were used to examine the correlation between site visits and personal empowerment. Mann–Whitney and Wilcoxon nonparametric tests were used to examine the differences in PAM interval score. A Mann–Whitney test was used to examine the differences in the PAM interval scores between new and experienced users.

Predictors of PAM were analyzed through structural equation modeling with AMOS, version 21, creating a path analysis with the maximum likelihood method. Structural equation modeling was carried out to examine predictors of the PAM interval score. A multiple-group analysis yielded a good model fit (χ^2_4=2.4; *P*=.664; comparative fit index =1.000; normed fit index =.999; root mean square error of approximation =.000). As can be seen in Figure 1, for the new respondents, PAM was predicted directly by knowledge about the disease (β=.26) and only by that factor. Social relationship and involvement in personal health had a strong correlation with knowledge about the disease and among themselves. In Figure 2, for the experienced users, PAM level was also predicted directly by knowledge about the disease (β=.21). The correlations between the 3 factors of empowerment were also high and significant among the experienced users.

Figure 1. Model 1: New respondents.

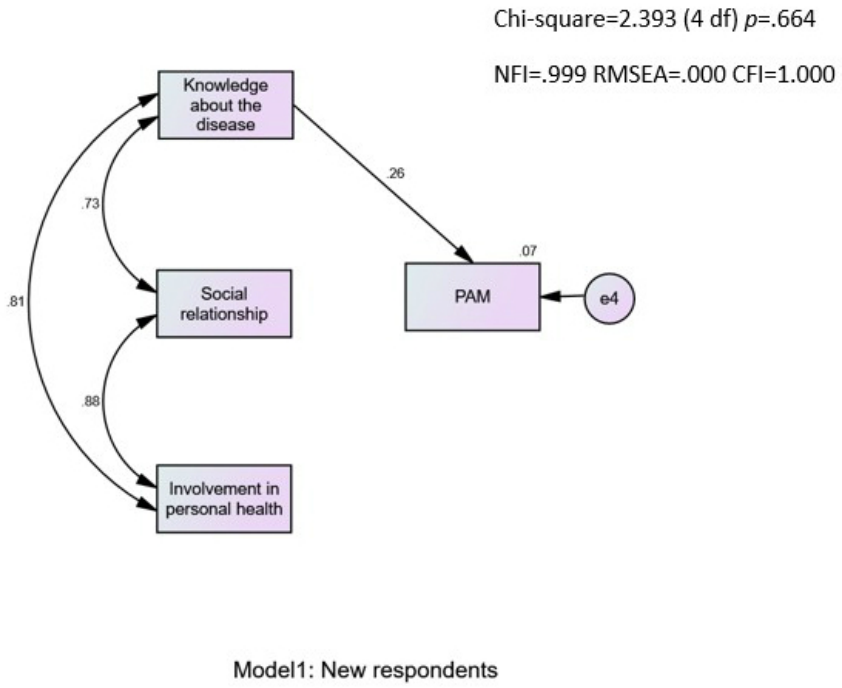
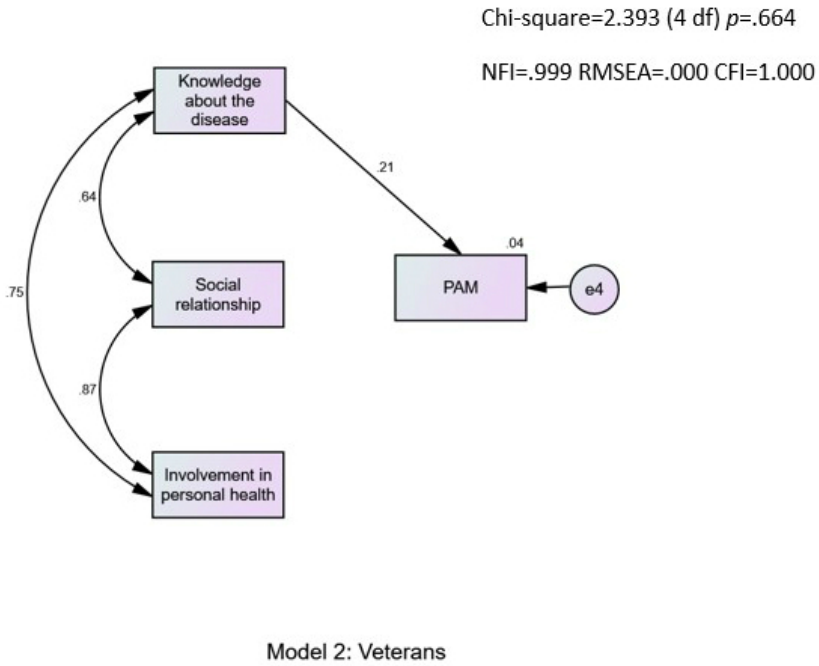


Figure 2. Model 2: Veteran users.



Results

Significant Findings

Table 1 describes the demographics of the new respondents and the experienced users.

There were several differences between the groups. New registrants were younger, included more singles, were less educated, and had lower incomes than the experienced users.

Table 2 describes the relationship between frequency and duration of visits to the Camoni site and knowledge about the disease, social relationships, and involvement in personal health among experienced users.

Table 1. Descriptive statistics of study variables.

Variable	New users N (%)	Experienced users N (%)	Total N (%)
Gender[a]			
Male	123 (35.2)	154 (45.7)	277 (40.4)
Female	226 (64.8)	183 (54.3)	409 (59.6)
Family status[a]			
Never married	63 (18)	46 (14)	109 (16.1)
Married	178 (52.0)	203 (60.4)	381 (56.2)
Divorced/separated	86 (25)	62 (19)	148 (21.8)
Widower	15 (4)	25 (7)	40 (6)
Age[a], years			
15-19	4 (1)	2 (0.6)	6 (0.9)
20-39	74 (23)	28 (9)	102 (15.8)
40-59	162 (49.2)	126 (39.9)	288 (44.7)
60+	89 (27.1)	160 (50.6)	249 (38.6)
Education[a]			
Elementary or high school	202 (58.6)	149 (44.2)	351 (51.5)
Academic	143 (41.4)	188 (55.8)	331 (48.5)
Income[a]			
Lower than average	189 (55.8)	138 (42.5)	327 (49.2)
Average	72 (21)	85 (26)	157 (23.6)
Above average	78 (23)	102 (31.4)	180 (27.1)
Chronic illness			
Diabetes mellitus	96 (27)	164 (48.1)	260 (37.6)
Pain	167 (47.7)	112 (32.8)	279 (40.4)
Depression	129 (36.9)	68 (20)	197 (28.5)
Hypertension	137 (39.1)	150 (44.0)	287 (41.5)
Number of illnesses			
1	152 (53.8)	186 (58.3)	338 (49.2)
2	106 (32.2)	96 (30)	202 (34.3)
3+	51 (14)	37 (12)	88 (17)

[a] The chi-square test for independence significant at $P<.05$

Table 2. Spearman correlations between the frequency and duration of visits and the 3 factors of personal empowerment in health, among experienced users.

Correlations	Knowledge about the disease (N=362)	Social relationships (N=374)	Involvement in personal health (N=374)
Frequency of visiting the site (1=less than once a month, 7=a few times a day)	0.251[a]	0.271[a]	0.292[a]
Duration of the visits (1=less than 10 minutes, 4=more than an hour)	0.136[a]	0.152[a]	0.190[a]

[a] $P<.01$

We found significant positive correlations between both the frequency and duration of visits to the site and personal empowerment in health. Experienced users (mean, M=69.3, standard deviation, SD=19.1, PAM level 4) had significantly

higher PAM scores (P<.001) than did new enrollees (M=62.8, SD=18.7, PAM level 3).

Men tended to surf longer than women, as 15.8% of the men surfed more than an hour compared with 8% of women; 52% of the women surfed 10 to 30 minutes, (P<.05) compared with 38% of the men. Among the experienced users, 75.7% with a high school education surfed 10 minutes or more compared with only 61.2% of those with an academic education (P<.05).

No significant dependency was found between the experienced users' socioeconomic level and duration of visits (P>.05). Experienced users with diabetes mellitus tended to surf more than experienced users with other illnesses. We found that 16.9% of the respondents with diabetes surfed for more than an hour. Only 7.1% of the respondents with other illnesses surfed for more than an hour (P<.05).

Follow-Up Questionnaire

A third group was comprised of 55 respondents who answered a follow-up questionnaire approximately 3 months after the first one. They included 31 new users and 24 experienced users. Their demographics were similar to those of the new and experienced users combined. The Wilcoxon test was used to examine the interval change in the PAM score. The PAM scores increased significantly between the initial score (M=61.89, SD=20.92, PAM level 3) and the score 3 months later (M=71.08, SD=19.70, PAM level 4; Z=−3.625, P<.001).

Experienced respondents were asked 2 questions: what is the frequency of use of the site and the average amount of time spent surfing. A (P<.01) positive correlation was found between frequency and time spent and the 3 indices of health empowerment (confidence from knowledge acquired about the disease, a sense of shared support, and personal involvement in treatment).

We found significantly higher PAM scores among experienced users compared with the newly enrolled (mean 62.8 and 69.3, respectively Z=−4.197, P<.001). More of the experienced users (56.6%) had the highest PAM rating (level 4) across all 4 diseases (diabetes mellitus, pain, depression, and high blood pressure) compared with the newly enrolled (42.3%). Continued involvement in the site seems to be associated with increased PAM scores.

Discussion

Surfer Demographics

This study is unique in that we examined the association of participating in a social health network (Camoni) with patient activation, among patients with depression, diabetes mellitus, pain, or high blood pressure. The main findings among Camoni surfers indicate that those with at least 6-month experience on the site had the highest PAM (level 4) compared with new visitors (level 3; Z=−3.938, P<.00). Among those who participated an additional 3 months, 54.5% reached level 4 PAM scores.

We found that more patients with chronic diseases surfing on Camoni were from lower socioeconomic groups (less educated, lower income) and were both passively and actively using the social health network site compared with those with higher education and income levels. Other studies have reported inconsistent results regarding factors such as income and age with Web-based health-related information seeking [29-31]. As computer access increases over time, lower income groups are increasingly using the Internet [32]. The greatest growth is seen among those who have low incomes and are less educated. This is particularly important because lower socioeconomic groups often have lower patient activation [4,14,33-34] and poorer health outcomes. In the present study, we also saw that the social health network seemed to benefit lower income and less-educated participants.

In our sample, more men used the site than did women. Although earlier studies reported that women were more likely to use health and general social network sites [4], more recent data indicate that this trend is changing. A 2014 Pew survey found that men and women use social networking sites equally [35].

We also found that individuals ages 40 to 59 years were more likely to use the social network site (both new and experienced users) than their younger counterparts did. Prior studies reported that older people use the Internet less and that older generations prefer to wait for a physician consultation [17,34,36]. However, older adults are increasingly likely to turn to the Internet and social networking sites [37]. Increasing computer literacy has the potential to increase patient well-being greatly. We know that 75% of adults ages 65 years and older are living with a chronic condition [33]. As the population continues to age, more and more elderly will likely turn to social networks for support and information, thus improving patient activation and well-being.

Duration of Visit

We found that increased frequency and duration of visits to the social network were associated with increased PAM scores. Those who surf more frequently and for longer times had higher scores on 3 empowerment measurements: patient feels confidence in acquired knowledge about the disease; feelings of support and the ability to share their experience; and level of involvement in treatment.

Disease and Health Knowledge

Specific health groups turn to the social health network for several reasons [4-10], and most respondents want to obtain information and are not seeking social relationships [12-15]. They expect the website to be specifically directed to their chronic health condition, and they seek knowledge about the disease and about treatment options. Likewise, we found that knowledge had the greatest influence on patient activation. This is consistent with the literature on patient activation and the assumption that consumers will make more prudent health and health care choices and will increase their "activation" when they are given relevant and quality information [9]. Several recent studies considered the role of Internet use and patient activation in the chronically ill [38-40] as well as in the general population [41]. Those who access health information over the Internet are more likely to have higher PAM.

No difference was found in PAM levels between those who post on the site and lurkers (those who merely observe but do not post). We suggest that this is further evidence of the importance of knowledge acquisition over the social element. This finding is similar to that of other studies that reported that those who posted more frequently were no more likely to be activated or empowered than lurkers were [42]. It is most likely that simply reading and searching for information within the social network site was sufficient to increase activation levels [43] and that obtaining information is a key factor.

We suggest that the finding that respondents with diabetes mellitus particularly benefited from the website is related to the increased importance of self-care and knowledge for patients with diabetes. Most of the day-to-day care for diabetic patients is carried out by the patients and their families. People with diabetes in particular require active participation in their medical care [44]. Patients need to monitor blood glucose levels, adjust insulin doses exercise, and monitor their diet with consequences that are often immediate. They therefore may be more inclined to seek information and peer support and become more involved in their health care.

Limitations of this Study

This is a cross-sectional study, and an association is not necessarily causation. Longitudinal studies are needed to determine the arrow of causation. It may be that of continued participation in a social health network leads to increased patient activation, or that increased patient activation facilitates participation in a social health network.

Another limitation of the study arises from the sampling method. The study population was self-selected, and therefore, the results may be biased. It included only Israelis, was conducted in the Hebrew language, and it probably did not include new immigrants or the non-Jewish population. It is possible that people with certain characteristics tend to respond to Web-based surveys. We cannot know who chose not to participate in the Web-based questionnaire and whether the results would have been the same had they participated. Also, it is possible that those who actively participated and those who were passive participants (lurkers) differed. Another limitation was that we did not analyze the different diseases separately, according to the type of knowledge and support provided, although all participants had similar exposure to the contents of the site. We did not have information about other sources of knowledge or support that the participants might have accessed, such as additional medical consultations, family members who provided support, other forums, or Internet sites where they received information, or avenues such as medical books or pamphlets. Also, only 4 of the 16 communities were selected. Perhaps other diseases would have led to different results. In addition, we resurveyed the users 3 months after registration. We cannot know how many times they actually entered the site compared with what they reported. Had we conducted the second survey after a longer period, the results might have been different.

Finally, response to the follow-up questionnaire was lower than expected. This finding is similar to that of others who described [45] similar poor follow-up on Internet questionnaires.

Such limitations are expected in all studies among anonymous, surfing patients. As a result, these types of studies describe trends and attitudes and probably represent an iceberg phenomenon of social networks.

Conclusions

Among patients with a chronic illness, surfing on an Internet-based, health-related social network was associated with improved knowledge and health activation. This was stronger among those with less education, lower income, men, older age, and patients with diabetes.

Knowledge about one’s disease had a direct effect on patient activation in a structural equation model. Social health networks seem to provide an opportunity for people from different strata to increase activation by increasing knowledge about their disease and about their personal health.

Acknowledgments

This study was supported by a grant from the National Institute for Health Policy Research, Israel.

The authors thank Lisa C. Bergman, PhD, and Faye Schreiber, MS, for their assistance.

Conflicts of Interest

None declared.

References

1. Pew Research Center. 2014. Pew Research Center's Internet Project January Omnibus Survey, January 23-26, 2014 URL: http://www.pewinternet.org/2014/01/14/january-2014-omnibus-week-4/ [accessed 2016-03-30] [WebCite Cache ID 6gNQo2TAI]
2. Taylor H. Harris Interactive. Harris Poll Shows Number of Cyberchondriacs - Adults Who Have Ever Gone Online for Health Information - Increases to an Estimated 160 Million Nationwide URL: http://media.theharrispoll.com/documents/Harris-Interactive-Poll-Research-Cyberchondriacs-2007-07.pdf [accessed 2016-03-29] [WebCite Cache ID 6gNQxc4Cz]
3. Fox S. Pew Research Center. 2013 Sep 30. The new environment for better health care decisions URL: http://www.pewinternet.org/2013/09/30/the-new-environment-for-better-health-care-decisions/ [accessed 2016-03-29] [WebCite Cache ID 6gNR7x85z]

4. Fox S, Purcell K. Pew Research Center. 2010 Mar 24. Chronic disease and the Internet URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2016-03-29] [WebCite Cache ID 6gNRENRCy]
5. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
6. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008 Oct(8):1-9. [Medline: 18946947]
7. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is patient activation associated with outcomes of care for adults with chronic conditions? J Ambul Care Manage 2007 Mar;30(1):21-29. [Medline: 17170635]
8. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
9. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
10. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008 Oct(8):1-9. [Medline: 18946947]
11. Remmers C, Hibbard J, Mosen DM, Wagenfield M, Hoye RE, Jones C. Is patient activation associated with future health outcomes and healthcare utilization among patients with diabetes? J Ambul Care Manage 2009;32(4):320-327. [doi: 10.1097/JAC.0b013e3181ba6e77] [Medline: 19888008]
12. Marshall R, Beach MC, Saha S, Mori T, Loveless MO, Hibbard JH, et al. Patient activation and improved outcomes in HIV-infected patients. J Gen Intern Med 2013 May;28(5):668-674 [FREE Full text] [doi: 10.1007/s11606-012-2307-y] [Medline: 23288378]
13. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [Medline: 10413387]
14. Fowles JB, Terry P, Xi M, Hibbard J, Bloom CT, Harvey L. Measuring self-management of patients' and employees' health: further validation of the Patient Activation Measure (PAM) based on its relation to employee characteristics. Patient Educ Couns 2009 Oct;77(1):116-122. [doi: 10.1016/j.pec.2009.02.018] [Medline: 19356881]
15. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun;33(6):1275-1281 [FREE Full text] [doi: 10.2337/dc09-2153] [Medline: 20299481]
16. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
17. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
18. Thiboutot J, Sciamanna CN, Falkner B, Kephart DK, Stuckey HL, Adelman AM, et al. Effects of a web-based patient activation intervention to overcome clinical inertia on blood pressure control: cluster randomized controlled trial. J Med Internet Res 2013;15(9):e158 [FREE Full text] [doi: 10.2196/jmir.2298] [Medline: 24004475]
19. 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]
20. 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]
21. Elbert NJ, van OH, van RW, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
22. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
23. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
24. Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res 2014;16(1):e12 [FREE Full text] [doi: 10.2196/jmir.2630] [Medline: 24413148]
25. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008 Oct;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
26. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]

27. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
28. Magnezi R, Glasser S. Psychometric properties of the hebrew translation of the patient activation measure (PAM-13). PLoS One 2014;9(11):e113391 [FREE Full text] [doi: 10.1371/journal.pone.0113391] [Medline: 25411841]
29. Dickinson A, Gregor P. Computer use has no demonstrated impact on the well-being of older adults. International Journal of Human-Computer Studies 2006 Aug;64(8):744-753. [doi: 10.1016/j.ijhcs.2006.03.001]
30. Huang C. Internet use and psychological well-being: a meta-analysis. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):241-249. [Medline: 20557242]
31. Weaver JB, Mays D, Lindner G, Eroglu D, Fridinger F, Bernhardt JM. Profiling characteristics of internet medical information users. J Am Med Inform Assoc 2009;16(5):714-722 [FREE Full text] [doi: 10.1197/jamia.M3150] [Medline: 19567794]
32. Perrin A, Duggan M. Pew Research Center. 2015 Jun 26. Americans' Internet Access: 2000-2015 URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [accessed 2016-03-30] [WebCite Cache ID 6gNRMiZsm]
33. Fox S, Duggan M. Pew Research Center. 2013 Nov 26. Part one: Who lives with chronic conditions URL: http://www.pewinternet.org/2013/11/26/part-one-who-lives-with-chronic-conditions/ [accessed 2016-03-29] [WebCite Cache ID 6gNRXIcos]
34. Taylor K. Paternalism, participation and partnership - the evolution of patient centeredness in the consultation. Patient Educ Couns 2009 Feb;74(2):150-155. [doi: 10.1016/j.pec.2008.08.017] [Medline: 18930624]
35. Pew Research Center.: Pew Research Center; 2016 Mar 29. Social Networking Fact Sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2016-03-29] [WebCite Cache ID 6gNRcq6o0]
36. Falkum E, Førde R. Paternalism, patient autonomy, and moral deliberation in the physician-patient relationship. Attitudes among Norwegian physicians. Soc Sci Med 2001 Jan;52(2):239-248. [Medline: 11144780]
37. Smith A. Pew Research Center. 2014 Apr 03. Older Adults and Technology Use URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/Accessed: [accessed 2016-03-29] [WebCite Cache ID 6gNRhGzVk]
38. Lui C, Col JR, Donald M, Dower J, Boyle FM. Health and social correlates of Internet use for diabetes information: findings from Australia's Living with Diabetes Study. Aust J Prim Health 2015;21(3):327-333. [doi: 10.1071/PY14021] [Medline: 25007788]
39. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
40. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
41. Smith S, Pandit A, Rush S. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015;18:3262-3273.
42. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
43. Petrovcic A, Petric G. Differences in intrapersonal and interactional empowerment between lurkers and posters in health-related online support communities. Computers in Human Behavior 2014;34:39-48 [FREE Full text]
44. Etzwiler DD. Diabetes translation. A blueprint for the future. Diabetes Care 1994 Jun;17 Suppl 1:1-4 [FREE Full text] [Medline: 8088217]
45. Bälter KA, Bälter O, Fondell E, Lagerros YT. Web-based and mailed questionnaires: a comparison of response rates and compliance. Epidemiology 2005 Jul;16(4):577-579. [Medline: 15951679]

Abbreviations

M: mean
PAM: Patient Activation Measure
SD: standard deviation

Edited by G Eysenbach; submitted 07.04.16; peer-reviewed by E Black, D Lee; comments to author 11.05.16; revised version received 19.06.16; accepted 29.06.16; published 10.08.16.

Please cite as:
Grosberg D, Grinvald H, Reuveni H, Magnezi R
Frequent Surfing on Social Health Networks is Associated With Increased Knowledge and Patient Health Activation
J Med Internet Res 2016;18(8):e212
URL: http://www.jmir.org/2016/8/e212/
doi:10.2196/jmir.5832
PMID:27511272

Original Paper

Mechanisms of Communicating Health Information Through Facebook: Implications for Consumer Health Information Technology Design

Hannah K Menefee[1], MPH; Morgan J Thompson[2], BA; Thomas M Guterbock[3], PhD; Ishan C Williams[4], PhD; Rupa S Valdez[5], PhD

[1]Department of Population Health Sciences, Virginia Tech, Blacksburg, VA, United States

[2]Department of Psychology, The College of William and Mary, Williamsburg, VA, United States

[3]Department of Public Health Sciences; Center for Survey Research; Department of Sociology, University of Virginia, Charlottesville, VA, United States

[4]School of Nursing, University of Virginia, Charlottesville, VA, United States

[5]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Rupa S Valdez, PhD
Department of Public Health Sciences
University of Virginia
PO Box 800717
Charlottesville, VA, 22908
United States
Phone: 1 4349822510
Fax: 1 43432433418
Email: rsv9d@virginia.edu

Abstract

Background: Consumer health information technology (IT) solutions are designed to support patient health management and have the ability to facilitate patients' health information communication with their social networks. However, there is a need for consumer health IT solutions to align with patients' health management preferences for increased adoption of the technology. It may be possible to gain an understanding of patients' needs for consumer health IT supporting their health information communication with social networks by explicating how they have adopted and adapted social networking sites, such as Facebook, for this purpose.

Objective: Our aim was to characterize patients' use of all communication mechanisms within Facebook for health information communication to provide insight into how consumer health IT solutions may be better designed to meet patients' communication needs and preferences.

Methods: This study analyzed data about Facebook communication mechanisms use from a larger, three-phase, sequential, mixed-methods study. We report here on the results of the study's first phase: qualitative interviews (N=25). Participants were over 18, used Facebook, were residents or citizens of the United States, spoke English, and had a diagnosis consistent with type 2 diabetes. Participants were recruited through Facebook groups and pages. Participant interviews were conducted via Skype or telephone between July and September 2014. Data analysis was grounded in qualitative content analysis and the initial coding framework was informed by the findings of a previous study.

Results: Participants' rationales for the use or disuse of a particular Facebook mechanism to communicate health information reflected six broad themes: (1) characteristics and circumstances of the person, (2) characteristics and circumstances of the relationship, (3) structure and composition of the social network, (4) content of the information, (5) communication purpose, and (6) attributes of the technology.

Conclusions: The results of this study showed that participants consider multiple factors when choosing a Facebook mechanism for health information communication. Factors included what information they intended to share, what they were trying to accomplish, attributes of technology, and attributes and communication practices of their social networks. There is a need for consumer health IT that allows for a range of choices to suit the intersectionality of participants' rationales. Technology that

better meets patients' needs may lead to better self-management of health conditions, and therefore, improve overall health outcomes.

(J Med Internet Res 2016;18(8):e218) doi:10.2196/jmir.5949

KEYWORDS

consumer health information; medical informatics; Facebook; social networks; health communication

Introduction

Recent government initiatives encouraging the development and implementation of health information technology (IT) have expanded the definition of health care delivery locations to include home and community environments [1]. With easier and increasing access to health IT solutions, patients' interactions with their health information and health systems can now occur outside of the clinical encounter [1]. Patients' engagement in their health care, or patients having the knowledge, skills, confidence, and tools to manage their health, has proven beneficial for improving long-term health outcomes [2,3]. Consumer health IT may be one means of supporting patient engagement [4,5]. Although there is no consensus definition of consumer health IT [6], the term generally refers to electronic technology designed specifically to support laypeople with health and health care management. These technologies can position patients to actively participate in their care by facilitating self-management of their conditions and improve coordination and communication among their care team [1,3].

It is imperative that consumer health IT is designed to align with patients' health management needs [7-11]. One challenge that patients with chronic conditions face is effectively engaging others in their health management [12-15]. Successful health management often involves communicating health information with a larger social network (eg, friends, family members, others with similar diagnoses) [12-14,16]. Several consumer health IT solutions have the ability to support patients' health information communication with members of their social network. Examples of these technologies include CaringBridge, Microsoft HealthVault, MyChart, and online health communities such as PatientsLikeMe, QuitNet, and TuDiabetes. However, these existing solutions are currently limited in relation to what, to whom, and how health information is communicated [14].

Designing consumer health IT that effectively supports health information communication between patients and their social network members requires gaining a deep understanding of existing communication practices. Such an effort may generate both descriptive (ie, narrative explication of communication approaches) and prescriptive (ie, specific design actions) design guidance [14]. Several studies have sought to understand the ways that patients communicate health information face to face, over the phone, through video chat, and using other common non social media based communication mechanisms for the purpose of drawing design guidance for consumer health IT [12-14,16,17]. Others have sought to draw such design guidance from patients' communication practices within online health communities [18-21]. However, previous studies have only partially drawn consumer health IT design guidance from the ways that patients have adopted and adapted readily available information technology for health management, which are not specifically designed for this purpose (ie, from technologies that are not consumer health IT). These technologies include social networking sites (SNS), which have been adopted and adapted by patients for health management [22-26], including communicating health information with social network members [27-31].

In this study, we use Facebook as a case study from which to draw design guidance for consumer health IT that supports health information communication between patients and their social networks. Facebook was selected as an object of study because it is a general information technology that takes the form of an SNS rather than consumer health IT. Furthermore, its popularity, with 70% of site users engaged daily [32], and usage as a site for health communication [27,33] contributed to our choice. Facebook was also chosen as a case for this study because of the diverse communication mechanisms offered to users, which span those that may be used in ways that are public versus private, synchronous versus asynchronous, and active versus passive. This SNS enables communication with social networks defined by personal relationship (eg, friends) and shared health condition (eg, other group members). A list and brief discussion of these mechanisms is found in Multimedia Appendix 1. By understanding why patients choose one Facebook mechanism over another for health information communication, designers (ie, individuals creating consumer health IT solutions) can gain insight into what communication mechanisms should be replicated, abandoned, or modified for consumer health IT supporting health information communication between patients and their social networks. Such insight can then be synthesized with other assessments of participant needs and incorporated within a participatory design process [34] to create an acceptable consumer health IT solution for a particular population.

Previous research on health information communication on Facebook has primarily sought to understand how patients use the group mechanism. Patients' interactions on health-related Facebook groups are structured around informational and emotional support and community building [23,35-39]. Medical and lifestyle information, including users' personal experiences, opinions, and advice are frequently communicated and highly valued by users of health-related Facebook groups [35,36,38,39]. In addition to informational support, emotional support in the form of encouragement and affirmation to help motivate other users is frequently found in posts to health-related Facebook groups [36,38-40]. Informational and emotional support cultivates health-related Facebook groups as spaces for companionship and community [36,38]. Some health-related Facebook groups also develop group identity by rejecting

common stereotypes to reframe their condition in a positive light [35]. Moreover, health-related Facebook groups often become spaces for the marketing of health services or products [23,41], enabling expression of experiences with health care centers, providers, and insurance companies [39], raising general awareness, and supporting fundraising efforts [22,37,41]. Users of health-related Facebook groups also include more than just patients. Family members, advertisers, and researchers participate and often communicate health information within these groups as well [23,39].

While these previous studies make meaningful contributions to our understanding of how patients engage with Facebook groups for health information communication, they lack insight into how and why patients engage with other Facebook mechanisms to communicate information about their health. Expanding analysis efforts to encompass other Facebook mechanisms enables understanding of how patients communicate not only with others with similar diagnoses (ie, within Facebook groups), but how they communicate health information with their larger social networks (ie, with Facebook friends and friends of friends). Accordingly, the purpose of this study was to characterize patients' rationales for using diverse communication mechanisms within Facebook for health information communication, each of which presents the Facebook user with a choice regarding who is the target audience, whether the communication is synchronous or asynchronous, and whether the communication is relatively active or passive. From this analysis, we hope to provide insight into how consumer health IT may be better designed to meet patients' health information communication needs and preferences.

Methods

Study Details

This study analyzed data from a larger, three-phase, sequential, mixed-methods study. The larger study's first phase used in-depth participant interviews (N=25) to gain a broad understanding of patients' health information communication practices on Facebook. These interviews are the focus of this paper.

Sample

Eligible study participants self-reported in a pre-study survey that they were over 18, used Facebook, were residents or citizens of the United States, spoke English, and had a diagnosis consistent with type 2 diabetes. Type 2 diabetes was the focus of this study because it is a chronic disease that is highly prevalent, requires intense personal engagement, and disproportionately affects underserved populations [14]. The University of Virginia's (UVa) Social and Behavioral Sciences Institutional Review Board approved this study, and all participants provided informed consent.

Recruitment

The recruitment process has previously been described in detail [42]. In short, we began by identifying diabetes-related Facebook groups and pages as well as those specific to racial and ethnic minorities. The latter were targeted to promote sample diversity. Consenting moderators posted a link to, and information about, our study's Facebook group. Each member of our Facebook group was asked to complete an online pre-study survey, deployed via Qualtrics [43] by UVa's Center for Survey Research. The pre-study survey focused on demographic characteristics, including race, ethnicity, gender, socioeconomic status, geographic location, and frequency of Facebook use. Maximum variance sampling [44] based on these demographic characteristics was used to target individuals for interview participation. Participants were provided US$25 in compensation.

Data Collection

Semistructured interviews were conducted via Skype or telephone between July and September 2014 and ranged from approximately 21 minutes to 68 minutes in length (average=43 minutes). Participants were asked about their general use of Facebook (eg, "How large is your social network on Facebook?"), how they used specific Facebook mechanisms (eg, "Do you chat or send private messages? Why or why not?"), and their Facebook-related privacy concerns and settings (eg, "What concerns, if any, do you have regarding your privacy on Facebook?"). Participants were next asked systematically about components of their health information communication work system, including to whom and why health information is communicated (ie, social subsystem), what and how health information is communicated (ie, technical subsystem), and how factors such as their economic, political, cultural, and health statuses impacted their health information communication practices (ie, external environment) [45,46]. All interviews were audio recorded, professionally transcribed, and reviewed by the research team prior to analysis.

Data Analysis

Data analysis was grounded in qualitative content analysis [47] and facilitated by the qualitative analysis software program QSR NVivo 10 [48]. The unit of analysis was the full response provided by the participant for each question. Simultaneous coding was used when units of analysis contained multiple meanings [49]. The initial coding framework was informed by the findings of a previous study [50] and was revised based on participant responses in this study. The original framework included the six broad themes (ie, person, relationship, third party, message, goals, context) as well as categories under each theme. Data from this study were deductively coded into these existing themes and categories. When a unit of analysis did not fit into a preexisting theme and/or category, a new theme and/or category was created. Two investigators independently coded the interviews (HKM & MJT), and discrepancies were referred to the corresponding author (RSV) for resolution. The three authors discussed any coding disagreements. Outcomes of the discussions were used to revise the coding framework.

Results

Sample Characteristics

A total of 25 individuals participated in the interviews. Most were female and between the ages of 30 and 64. All major racial and ethnicity categories were represented. The majority of participants were white and non-Hispanic. All participants had

received at least a high school diploma or equivalent and most participants worked full time (35 hours/week or more). Individuals from all marital statuses, household incomes, and geographic regions were represented. The majority of participants were married, had a yearly household income between US$75,000 and $149,000 and were from suburban areas. Individuals represented all brackets of self-rated health status and years with type 2 diabetes. Over half of the participants rated their health as good and had been diagnosed with diabetes for 3-19 years. Most participants used Facebook more than once a day. Sample characteristics are shown in Table 1.

Participants' Rationales for Health Information Communication Mechanism Use

Participants' rationales for the use of a particular mechanism to communicate health information reflected six broad themes. Each theme represents the types of rationales provided for communicating or not communicating health information through a specific Facebook mechanism: (1) characteristics and circumstances of the person, (2) characteristics and circumstances of the relationship, (3) structure and composition of the social network, (4) content of the information, (5) communication purpose, and (6) attributes of the technology. Participants' responses to open-ended questions are presented verbatim to provide evidence for analytical categories. The quotes selected are representative rather than exhaustive of the analytic category. It is important to note that the multidimensionality of participants' rationales means that an illustrative quote may contain more than one meaning (and that, as stated above, quotes were often coded under multiple rationales).

Theme 1: Characteristics and Circumstances of the Person

Participants' choice of Facebook communication mechanism reflected their own and their social network members' characteristics and/or circumstances.

Individuals' Time Available to Talk About Health

Participants reflected on the time they had available to use specific Facebook mechanisms for health information communication. Some participants noted that posting health information on one's timeline would result in the post being available to view by most of their social network and that the publicity of such messages would result in undesired feedback and consume their time for response:

> *"I would post, 'Hey, I have some kind of issue with my kidneys.' All that's gonna do is open up, 'What kind of issue? Oh my gosh, are you okay? What are they gonna do?' I have no room to answer a million questions."* [173, F, Black or African American, 30-49]

Individuals' Degree of Health-Related Knowledge and Experience

Participants discussed the role of their and their social network members' health-related knowledge and experience. They stressed the importance of not posting health information about which they had limited knowledge to their own or friends' timelines on Facebook:

> *"[T]here's a lot of stuff that comes in on a feed on your Facebook...I am not going to share something that I don't know anything about."* [69, F, Hispanic or Latino, Other race, 50-64]

Participants also considered the health-related knowledge and similarities in experiences of individuals that could be reached through a specific Facebook mechanism:

> *"If I have an issue going on and maybe I want to know if someone else has experienced the same issues, I want to know if they tried out a certain product before I buy it. I may post that in the discussion board for that diabetes group."* [90, F, Declined to answer race, 30-49]

Moreover, participants stressed the ability to reach individuals with diverse knowledge and experience using the group mechanism:

> *"Groups has a wide variety of people, and they've got lots of information, different attitudes about it, different answers. I can get a wide variety from those people."* [120, F, White, 50-64]

Individuals' Temperament and Mood

Participants noted that their overall dispositions played major roles in determining how they communicated health information on Facebook. Shyness, depression, and feeling tired all deterred participants from using public mechanisms of communication such as their own timeline or groups. Participants also recognized that specific moods altered their approaches to using these mechanisms for health information communication:

> *"Well, in terms of images of myself, personally, I'm a little insecure about my weight and that kind of stuff, so unless I'm really in a good mood or something...when I was relaxed or whatever, I wasn't too shy about posting those kinda pictures—me sitting on the island or something like that. That's the primary reason. Appearance."* [11, M, Black or African American, 30-49]

Individuals' Patterns of Technology Use

Social network members' patterns of communication mechanism use also influenced participants. Specifically, others' degree of engagement with specific mechanisms impacted participants' frequency and likelihood of use. Participants engaged with a communication mechanism when they knew that the social network member that they were trying to reach would be available through that mechanism:

> *"[Messenger is] a fast form of communication with some people like my Weight Watchers leader. She's on messenger all the time."* [185, F, White, 65+]

Participants' decision of whether or not to engage with communication mechanisms involving multiple individuals similarly depended in part on the level of group activity. Thus, one participant noted that even though there are many people

in his Facebook groups, there are not many that are active members, and expressed:

> *"...if people were more active, or more communicative, I would be the same way."* [11, M, Black or African American, 30-49]

Individuals' States of Well-Being

Participants' states of well-being determined how they communicated health information through Facebook. Certain Facebook mechanisms deterred participants with physical limitations because of the normative length of the message exchanged via that mechanism:

> *"I can't see well, and a lot of times my fingers hurt bad, and I don't like to private message because of that."* [120, F, White, 50-64]

However, participants also expressed reaching out to others in their social network broadly through groups and posting on timelines when facing poor health:

> *"Primarily I've been well, since I signed onto Facebook. There's always a chance that yeah, if I was feeling under the weather I would put a post on there, just so everyone would know."* [130, M, White, 65+]

Theme 2: Characteristics and Circumstances of the Relationship

Participants' choice of Facebook communication mechanism reflected characteristics and circumstances of the relationships between them and their social network members.

Definition of Relationship

The roles that social network members played in participants' everyday lives (ie, specific social or biological ties) influenced the choice of communication mechanism on Facebook. For example, participants posted on their own and friends' timelines simply because of the nature of the relationship:

> *"[T]he choice of my friends are all people that have done the same thing I did, are working hard, making an income, taking care of themselves. The fact that they're my friends is why I'm posting to them, and we have a lot in common."* [130, M, White, 65+]

Established Patterns of Relationship

The established communication patterns between participants and their social network members influenced the choice of Facebook mechanism for health information communication. Participants noted that they used certain Facebook mechanisms habitually to communicate to specific social network members and that their communication of health information to the same individuals followed this pre-existing routine:

> *"...once they know that there's something, that you've got this problem, then talking back and forth on Facebook [messages]—because, as I said, that's the way we do a lot of our conversing."* [114, F, Black or African American, 50-64]

Likewise, established communication patterns hindered the communication of health information via the same mechanisms:

> *"Pretty much the people that I message are not ones that I share the diabetes information with."* [139, F, White, 30-49]

Theme 3: Structure and Composition of the Social Network

Participants' choice of Facebook communication mechanism reflected the structure and composition of participants' and their social network members' social networks.

Mutual Friends Within Social Networks

The degree of interconnectedness between participants and members of their social networks influenced the use of some Facebook mechanisms for health information communication. Often participants were concerned that the information posted on Facebook timelines would be brought into offline conversations with mutual friends:

> *"I would be very careful about what I'd say about someone else because so many of my friends are interconnected."* [55, F, White, 65+]

Friends of Friends Within Social Networks

Participants acknowledged that the ability of certain Facebook mechanisms to reach friends of friends was useful when wanting to provide support to other type 2 diabetes patients. A few mentioned sharing pages related to diabetes on their timelines for the benefit of people beyond their immediate Facebook social network:

> *"I share some pages for diabetes, cuz I do know some people that have it or maybe their spouses. If I see anything that I like, I like somebody else to see it. Cuz not everybody gets the same feed. I might share a page."* [69, F, Hispanic or Latino, Other race, 50-64]

Social Network Insufficiency

Participants stressed that certain Facebook mechanisms gave them access to information and support not available to them through existing social network members. This was particularly true of the group mechanism:

> *"I think the reason why I'm in those groups, though, is because I was kind of pissed off that there was really no help that was worth anything from my doctor. I mean she gave me some general stuff, but nothing that was of any substance. That's kinda why I participate in the groups."* [92, M, White, 30-49]

Theme 4: Content of the Information

Participants' choice of Facebook communication mechanism reflected the substance of the health information.

Gravity of Information

Participants considered the seriousness of their messages. For instance, one participant noted that she did not consider one piece of information to be major or serious and therefore felt the information was appropriate to communicate using Facebook messages:

> *"[Y]ou know, at that time, it didn't make sense for me to sit and try to call 10 different people, I could do it*

> *in like two or three minutes on Facebook. It wasn't something that was major, serious, or deathly, it was something that [my mother] had been through a number of times before."* [114, F, Black or African American, 50-64]

Personal Nature of Information

The degree to which participants considered the health information to be private influenced their use of Facebook mechanisms. For example, one participant felt that general health information unrelated to her diabetes was appropriate for her Facebook timeline, but that private health information should not be discussed using this mechanism:

> *"I wouldn't talk about, like I said, anything that was, that I consider very, very, very private—but general stuff, I mean, if you look and you see my leg is broke, I don't see why I can't talk about a broken leg on Facebook."* [114, F, Black or African American, 50-64]

Sensitivity of Information

The emotional weight of the health information also influenced the use of Facebook mechanism. Participants acknowledged that some health information would invoke strong reactions from their social network members. The intensity of these messages determined which Facebook communication mechanism, if any, was used:

> *"[S]ay I had a diagnosis of cancer or anything, a diagnosis that I knew would really be shocking to someone, I would tell them face to face. I wouldn't send my sister or brother a Facebook message and say, "Oh, by the way, I've got cancer." I wouldn't do that. I would personally call them."* [114, F, Black or African American, 50-64]

Specialized Nature of Information

A majority of participants described the use of Facebook groups when discussing a particular diabetes topic, expressing that this mechanism was more appropriate than discussing this type of detailed information on their timelines. Furthermore, participants explained the relationship between level of specialization and communication via Facebook groups:

> *"In the groups, it's specifically for something special. I might be a member of the Rheumatoid Arthritis page, and we'll be talking about methotrexate or something. If I posted something about methotrexate on my newsfeed, 80 percent of the people are not going to have a clue what I'm talking about…I can be more specific on the group pages."* [120, F, White, 50-64]

Relevance of Information

The Facebook mechanism used was also determined by the perceived novelty of the health information. When such information was perceived as relevant to many individuals, participants used the sharing mechanism or posted to their timelines:

> *"I may see something in a diabetes group on Facebook, go click on it, "Oh, that's really interesting. I'm gonna share it," and go back to Facebook and share it and put it on my page. That's typically how I do it. On occasion, I'll see something through another source, read it, "Oh that's interesting," and click on it and share it on Facebook that way."* [139, F, White, 30-49]

Participants also considered how relevant the health information was to what was typically communicated through the Facebook mechanism:

> *"Because it's relevant to the group that I'm part of. Mostly because it's relevant or it's relevant to a conversation that's occurring in the group or within an individual conversation."* [33, F, White, 50-64]

Even if participants believed in the relevance of the information to a specific individual, they sometimes used mechanisms that had a wider reach instead of imposing the information on a specific person:

> *"They're just things [posted to own timeline] that I'm interested in and may be able to help other people. I don't really—putting it on their timeline is kind of like 'pushy'."* [33, F, White, 50-64]

Scarcity of Information

The existence and availability of health information to others in their Facebook network influenced participants' use of mechanisms for communicating health information. Participants used mechanisms targeted to a wider social network when they believed the information they were sharing was not well known to this audience. For example, one participant focused on the relative lack of information regarding type 2 diabetes:

> *"I would say just because I feel like there is not enough people talking about type 2 diabetes, it shapes my posts in the sense that I post a lot about diabetes health. It has made me—my health status has made me an advocate for informing other people about type 2 diabetes."* [173, F, Black or African American, 30-49]

Theme 5: Communication Purpose

Participants' choice of Facebook communication mechanism reflected the goals they sought to achieve in relation to their health information.

Ensure Personal Communication

Participants sought to use the Facebook mechanism they felt would ensure a more personal connection when communicating health information. For the majority of participants, preserving this connection was achieved by using the private messaging mechanism:

> *"It's a private thing, and it's personal; so people are more likely to respond to personal messages."* [173, F, Black or African American, 30-49]

Reduce Burden

Participants' rationales for using specific Facebook mechanisms to communicate health information included reflection on the degree of convenience versus hardship on themselves and others.

The ability to use certain Facebook mechanisms like messages to address a specific group was mentioned by participants as more convenient than other mechanisms requiring one-on-one interaction:

> *"Sometimes it's easier to get a blanket statement out there, especially to the friends."* [146, M, White, 30-49]

Similarly, mechanisms that did not require sifting through previous communications were also seen as more convenient:

> *"If I had previously spoken to someone about a problem that I've had, or a problem that they have had. Then, that post is now I don't know where, and I feel that I have an additional bit of information that they could use, then I might tag them on that. I write it on my timeline and just say, "So-and-so, I just thought I would share this with you."* [3, F, Asian, 50-64]

Private messages were also mentioned as a means of reducing discomfort that could come from confronting someone in a public forum such as a Facebook timeline:

> *"If I see somebody giving [misleading health information] or leading a person, I feel, astray, then I will send them a private message versus putting somebody on the spot right there, calling someone out."* [114, F, Black or African American, 50-64]

Connect With Others

The ability of the Facebook mechanism to allow for a message to reach the intended audience was a motivating factor for use by some participants. When considering the ability to obtain responses quickly, one participant explained:

> *"The groups are between 500 to 3,000 people…there's going to be more people online at any given time to give you instant feedback."* [33, F, White, 50-64]

The participant also stated that having more group members helped him find:

> *"...somebody to sympathize with or to share information with or to get information from."* [33, F, White, 50-64]

Another participant posted on Facebook immediately after obtaining the result of an exam that indicated potential need for retinal surgery, so that his doctor would receive the message:

> *"I did that primarily because my eye doctor reads that, so that as soon as he sees it he's gonna call me and we're gonna talk about it on the phone."* [130, M, White, 65+]

Moreover, some participants noted that communicating health information on friends' timelines would be considered a last resort if they could not reach members of their social network by other means:

> *"Maybe if I was dying then I would post something on my best friend's timeline to get her attention if I could not reach her by telephone, or couldn't reach her family or mom or anybody, then maybe I would do that. It would be a last resort."* [90, F, Declined to answer race, 30-49]

Preserve Personal Security

Participants reflected on the goal of maintaining online security when deciding to use one Facebook mechanism over another. In particular, some participants favored use of mechanisms that restricted communication to one or a few individuals in order to maintain security of their information:

> *"If I don't want—if I'm on one of my group pages, and I don't want all those strangers to know what's going on, I would use the private messaging to go back and forth with someone."* [120, F, White, 50-64]

Participants' social network members also eased their online security concerns about communicating through more public forums, such as Facebook groups:

> *"It would depend. This particular instance, your group was referred to me by someone I knew very well. I felt comfortable doing that. I don't necessarily trust other groups. I don't know what their guidelines for their privacy and all that are."* [120, F, White, 50-64]

Manage Content Received

Participants used Facebook mechanisms that would ensure they obtained the content desired. Groups in particular were sources of desired information:

> *"On the groups I'll post something because it's a diabetes type 2 group and I'm sure that somebody is going through the same thing or been through it, and they're going to tell me something that may help me out."* [90, F, Declined to answer race, 30-49]

Obtaining the health information or feedback desired was also achieved using Facebook mechanisms with the ability to target participants' messages to certain social network members:

> *"Again, it's situational. I don't think everybody needs to know everything, so I target my message to who I think is interested or who I'll get sympathy from. I can't put something out there and have everybody go, "Poor baby, suck it up!" I want people to go, "Aww, it's okay. It'll be okay."* [33, F, White, 50-64]

Similarly, participants sought to avoid unwanted input and information by refraining from using certain Facebook mechanisms. One participant knew that if he posted his poor A1C level to his timeline he would:

> *"...get scolded by one friend who's an eye doctor."* [226, M, White, 50-64]

Provide Support

Participants noted that some Facebook mechanisms facilitated their ability to provide instrumental and emotional support. One participant felt that by using the "share" mechanism to post diabetes-related information:

> *"someone else can read it, and can learn from it, and not feel that they're in the daily management of diabetes by themselves."* [173, F, Black or African American, 30-49]

Participants also stressed commenting on posts as a form of providing emotional support.

Provide Response

Wanting or not wanting to address others' inquiries, interests, and/or requests influenced participants' choice of Facebook mechanism. In many cases, participants were comfortable responding to social network members' questions using the mechanism by which the question was asked:

> *"If someone asks me about it, if someone from one of the groups or something or the pages asks me about it, I would probably communicate through that, in that way."* [120, F, White, 50-64]

However, other participants expressed situations in which other considerations outweighed their desire to respond to inquiries from their social network:

> *"Right now, I have an issue with my kidneys…I'm not posting that. It's probably related to diabetes, but I don't know. If it's something that opens myself up to a bunch of questions, I don't post it."* [173, F, Black or African American, 30-49]

Theme 6: Attributes of the Technology

Participants' choice of Facebook communication mechanism reflected characteristics of specific Facebook mechanisms.

Mechanism Reliability

Participants considered the reliability of social network members receiving health information through a specific Facebook mechanism. For example, due to Facebook's structure, depending on the time of day a user is logged into the site and/or the popularity of a particular post, certain posts may or may not appear in a user's newsfeed. Therefore, participants depended on other means of sending a message:

> *"I know with the newsfeed thing, you have most recent and whatever the other one is, most popular, or something. I know that sometimes I could post something on my page, but it doesn't show up in everyone's newsfeed. Depending on when they look at Facebook, like it could be at the bottom, and they don't see it. I will send private message just as confirmation if I wanna make sure that person got the information."* [173, F, Black or African American, 30-49]

Mechanism Integration With External Sites

Facebook is integrated with many external websites through the "share" mechanism. A few participants acknowledged the convenience of using the "share" Facebook mechanism on external websites to post useful health information to their timeline:

> *"They're convenient to share. Usually it's from a website with a share button."* [172, M, Hispanic or Latino, American Indian/Alaskan Native, 30-49]

Mechanism Reach

The structures of the private message and chat mechanisms on Facebook have the ability to easily facilitate communication with groups of people. Several participants acknowledged that these mechanisms were beneficial when they wanted to send out information to a large number of specific people in a short time:

> *"If something was—I needed to get out to a lot of people in a short period of time and sitting down making 20, 30, 40 phone calls was not something that I could do, then I could use Facebook and those friends and send a private message, because you can send a private message to groups of people, I'm able to do it that way."* [114, F, Black or African American, 50-64]

Table 1. Demographic characteristics of participants (N=25).

Sample characteristics		Valid % (n)
Gender		
	Male	44 (11)
	Female	56 (14)
Age		
	18-29	0 (0)
	30-49	44 (11)
	50-64	40 (10)
	65+	16 (4)
Race		
	White	64 (16)
	Black or African American	16 (4)
	American Indian/Alaskan Native	4 (1)
	Asian	4 (1)
	Native Hawaiian/Pacific Islander	0 (0)
	Other	4 (1)
	Declined to answer	8 (2)
Ethnicity		
	Hispanic or Latino	12 (3)
	Not Hispanic or Latino	88 (22)
Education		
	Less than high school	0 (0)
	High school diploma or equivalent	16 (4)
	Some college, but no degree	16 (4)
	Associate’s degree	8 (2)
	Bachelor’s degree	20 (5)
	Some graduate work	12 (3)
	Master’s degree	16 (4)
	Doctoral degree	4 (1)
	Professional degree	8 (2)
Employment status		
	Working full time (≥35 hours/week)	48 (12)
	Working part time	20 (5)
	Looking for work	4 (1)
	Homemaker	0 (0)
	Retired	20 (5)
	Student	0 (0)
	Other (eg, disabled)	8 (2)
Marital status		
	Married	68 (17)
	Widowed	4 (1)
	Divorced	16 (4)
	Separated	4 (1)

Sample characteristics	Valid % (n)
Never married	8 (2)
Yearly household income, USD	
<$30,000	16 (4)
$30,000-$49,999	16 (4)
$50,000-$74,999	12 (3)
$75,000-$149,999	36 (9)
≥$150,000	8 (2)
Decline to answer	12 (3)
Geographic region	
Urban	24 (6)
Suburban	40 (10)
Rural	32 (8)
Other	4 (1)
Self-rated health status	
Poor	16 (4)
Fair	12 (3)
Good	52 (13)
Very good	20 (5)
Excellent	0 (0)
Years with type 2 diabetes	
<1 year	0 (0)
1-2 years	8 (2)
3-5 years	28 (7)
6-10 years	20 (5)
11-19 years	28 (7)
≥20 years	4 (1)
Declined to answer	12 (3)
Facebook use	
More than once a day	56 (14)
Once a day	32 (8)
A few times a week	8 (2)
Once a week	0 (0)
Less than once a week	0 (0)
Decline to answer	4 (1)

Discussion

Principal Results

The results of this study demonstrate the complexity of selecting a mechanism of health information communication on Facebook. Participants considered many factors before making a decision, including their own and the recipient's characteristics and relationships, the content of the information, the communication purpose, and the attributes of a particular communication mechanism. These findings mirror findings of previous studies documenting patients' general rationales for communicating health information [13-15], but add the dimension of the inherent features of specific communication mechanisms. It is thus imperative that consumer health IT designers create solutions responsive to these diverse rationales, better facilitating communication between patients and their social networks.

Our findings partially support previous literature focused on the use of Facebook groups for health information communication [23,35-40]. Similar to previous studies, participants in our study used Facebook groups for acquiring

informational support (eg, explaining symptoms, treatments, medications) and emotional support (eg, providing encouragement, affirmations). Our qualitative interviews yielded discussion focused primarily on these rationales for Facebook group use. Moreover, as reported in a previous study, participants indicated using Facebook groups to express their opinions and experiences with health care centers, providers, and health insurance companies [39]. Other previously documented rationales for using Facebook groups for health information communication such as marketing health services and products [23,41], and supporting fundraising efforts for a health condition [22,37,41] were not explicitly mentioned in our qualitative interviews. This contrast in findings may have been because our interview probes focused on clinical and experiential forms of health information relevant to the individual.

In addition to supporting previous literature on the use of Facebook groups for health information communication, our study provided a novel, in-depth explication of how other communication mechanisms on Facebook are used for this purpose. As a whole, participants reported using the full range of available communication mechanisms, spanning those that are private versus public, synchronous versus asynchronous, and active versus passive to meet a range of communication needs. This general finding, that patients use a variety of mechanisms to communicate health information depending on the situation and intended recipient(s), parallels those of other studies of patients' online and offline health information communication with social network members [17,50,51]. Despite these communication patterns, few existing consumer health IT solutions mimic and integrate this broad range of communication mechanisms. This is particularly true of consumer health IT solutions focusing on personal social networks such as Microsoft HealthVault [52] and CaringBridge [53], which are constrained by the types of communication mechanisms offered to users. Few support communication mechanisms beyond those that, in Facebook terms, equate with posting/commenting on a timeline and giving another individual full access to one's account. There is a need to expand the communication mechanisms offered within consumer health IT to better encompass the range of mechanisms for which participants demonstrated utility in this study.

The use of multiple communication mechanisms on Facebook to communicate health information may arise from participants' need to balance competing priorities. Understanding these balancing acts can provide insight into additional design opportunities for consumer health IT. Patients' use of both active (eg, chat, posting and commenting on timeline) and passive (eg, tagging, sharing) mechanisms reflects their need to variously balance the desire to communicate health information against the effort required [16]. In this decision, the effort in question occurs at the time of the communication act. Thus, passive mechanisms, although not yet incorporated into existing consumer health IT solutions, may serve a unique purpose. Specifically, they would enable users to communicate emotional support (by liking), the relevance of specific information to an individual (by sharing or tagging), or their physical presence at a health-related event (by checking in) without a significant time investment. In contrast, while the communication-effort balance also manifests itself in the decision to use public versus private mechanisms, the effort in question occurs at the time of the response to the initial communication act. For example, participants in our study spoke about the decision not to use public communication mechanisms such as posting on a timeline because of the multitude of responses that would need to be addressed. To mitigate the effort required to sift through and make sense of responses, natural language processing capabilities could be embedded within consumer health IT, grouping and summarizing similar types of feedback. A user could then address a group of similar messages with one response. Participants also noted that the choice between public and private mechanisms requires balancing the desire to obtain the fruits of communicating (eg, information or social support) with the desire to manage self-presentation [30] or the presentation of others. Design opportunities, including providing proactive feedback to dissuade social network members from providing negative responses [30], offering transparent granularity in sharing controls [54], and enabling anonymous posting within one's social network [55], have been offered as means of alleviating the necessity for such balance. These could be augmented with ways of illustrating who in a recipient's social network would have access to posted information so that the communicator can work to preserve not only their own, but also their social network member's self-image.

Beyond the design recommendations discussed above, there are likely opportunities to integrate or extend existing functionality within Facebook to better meet patient needs for consumer health IT supporting health information communication. For example, participants in our study indicated that their reaction to social network members' feedback is dependent on mood. Facebook currently has the capability of allowing users to indicate their mood and to link this mood to a specific post. This type of functionality could be integrated into consumer health IT to allow people to provide their current emotion as related to a piece of information, better enabling them to receive appropriate and supportive feedback. Similarly, our participants stated that their use of certain Facebook mechanisms is motivated by a need to ensure that the social network member will receive the message. Facebook currently informs users if a message was received through the chat or private messaging function (by displaying the word "read") and through groups (by displaying who has seen the post). This functionality could also be integrated into consumer health IT, along with an additional layer that enables users to visualize the reliability of a social network member receiving a message through a given communication mechanism. As a third example, participants expressed the importance of communicating information that was relevant to themselves and to their social network members. At present, Facebook allows users to control input to their newsfeeds by individual (whose posts are prioritized). To meet the needs of our participants, consumer health IT should enable users to prioritize information received by both user and topic. Moreover, to help users not only receive but also send relevant information, text mining capabilities could be extended to offer suggestions on which social network members should be tagged in a post.

Limitations

This study is subject to several limitations. First, the retrospective interview approach to data collection is subject to respondent recall bias. We attempted to mitigate this bias in the qualitative portion of our study by asking participants to review their Facebook communications prior to engaging in the interview. Second, the study focused only on individuals with one diagnosis (ie, type 2 diabetes), although some reported co-morbidities. It is likely, however, that the results have some generalizability to individuals with other chronic health conditions requiring similar levels of daily engagement. Third, participants in this study all lived in or were residents of the United States. It is possible that individuals living in other countries would approach health information communication on Facebook differently. However, our sample contained cultural diversity in the sense that it purposefully oversampled individuals identifying as racial and ethnic minorities. Finally, it is important to note that understanding communication mechanism use on Facebook is only one of multiple ways of understanding patient needs for consumer health IT focused on supporting health information communication between patients and their social networks. The lessons learned from this study should be combined with other needs assessments within a participatory design process when developing a specific consumer health IT solution.

Conclusions

When choosing a mechanism for health information communication, participants consider multiple factors that intersect in complex ways. Factors included what information they intended to share, what they were trying to accomplish, attributes of technology, and attributes and communication practices of their social network. There is a need for consumer health IT communication mechanisms to allow for a range of choices to suit the intersectionality of participants' rationales. Technology that better meets patients' needs will lead to better self-management of health conditions, and therefore, improve overall health outcomes. This study demonstrates that this intersectionality leads to a range of preferred communication mechanisms beyond those commonly available within existing consumer health IT solutions. As design of these systems evolve, there is a need to meet demand for a bundle of mechanisms facilitating communication between an individual and multiple types of social networks—communication mechanisms that vary in terms of level of effort, degree of synchronous communication, and privacy control. Future research could include designing consumer health IT solutions that have these specific communication mechanisms and then testing the solution with patients of varying health conditions. Additionally, by describing participants' rationales, we can infer the underlying dynamics of participants' choice of communication method. However, this interpretation is at an aggregate level across participants. Data from our interviews do not tell us systematically who is making what type of decisions under what circumstances. Another promising direction for future research will be to build a model of health information communication choices. Analysis of the results from the larger study from which data for this study were drawn will contribute to this research.

Acknowledgments

This study was sponsored by the Agency for Healthcare Research and Quality (R03 HS22930-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. We would also like to thank the administrators of the Facebook groups and pages that supported our study and all of our participants. We thank Deborah Rexrode, Morgan Thompson, Jeremiah Reilly, and Claire Wellbeloved-Stone for their contributions to this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

[PDF File (Adobe PDF File), 49KB - jmir_v18i8e218_app1.pdf]

References

1. Office of the National Coordinator for Health Information Technology. Federal health information technology strategic plan 2011 - 2015. 2011. URL: https://www.healthit.gov/sites/default/files/utility/final-federal-health-it-strategic-plan-0911.pdf [accessed 2016-07-28] [WebCite Cache ID 6jLiKZICP]
2. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
3. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
4. Wasson JH, Forsberg HH, Lindblad S, Mazowita G, McQuillen K, Nelson EC. The medium is the (health) measure: patient engagement using personal technologies. J Ambul Care Manage 2012;35(2):109-117. [doi: 10.1097/JAC.0b013e31824a235e] [Medline: 22415284]

5. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
6. Flaherty D, Hoffman-Goetz L, Arocha JF. What is consumer health informatics? A systematic review of published definitions. Inform Health Soc Care 2015 Mar;40(2):91-112. [doi: 10.3109/17538157.2014.907804] [Medline: 24801616]
7. Gibbons MC, Wilson RF, Samal L, Lehman CU, Dickersin K, Lehmann HP, et al. Impact of consumer health informatics applications. Evid Rep Technol Assess (Full Rep) 2009 Oct(188):1-546. [Medline: 20629477]
8. Marquard JL, Zayas-Cabán T. Commercial off-the-shelf consumer health informatics interventions: recommendations for their design, evaluation and redesign. J Am Med Inform Assoc 2012;19(1):137-142 [FREE Full text] [doi: 10.1136/amiajnl-2011-000338] [Medline: 21727206]
9. Montague E. Improving consumer health IT application development: lessons from other industries: findings from key informant interviews.: Agency for Healthcare Research and Quality; 2012. URL: https://healthit.ahrq.gov/sites/default/files/docs/publication/developmentmethodskeyinformantinterviewsreport.pdf [accessed 2016-07-28] [WebCite Cache ID 6jLosm9r7]
10. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
11. Valdez RS, Holden RJ, Novak LL, Veinot TC. Technical infrastructure implications of the patient work framework. J Am Med Inform Assoc 2015 Apr;22(e1):e213-e215 [FREE Full text] [doi: 10.1093/jamia/ocu031] [Medline: 25665705]
12. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
13. Skeels MM, Unruh KT, Powell C, Pratt W. Catalyzing social support for breast cancer patients. CHI Conf Proc 2010:173-182 [FREE Full text] [doi: 10.1145/1753326.1753353] [Medline: 21654894]
14. Valdez RS, Brennan PF. Exploring patients' health information communication practices with social network members as a foundation for consumer health IT design. Int J Med Inform 2015 May;84(5):363-374. [doi: 10.1016/j.ijmedinf.2015.01.014] [Medline: 25704761]
15. Weiss JB, Lorenzi NM. Online communication and support for cancer patients: a relationship-centric design framework. AMIA Annu Symp Proc 2005:799-803 [FREE Full text] [Medline: 16779150]
16. Pratt W, Unruh K, Civan A, Skeels M. Personal health information management. Commun. ACM 2006 Jan 01;49(1):51. [doi: 10.1145/1107458.1107490]
17. Pang C, Neustaedter C, Riecke B, Odour E, Hillman S. Technology preferences routines for sharing health information during the treatment of a chronic illness. 2013 Presented at: SIGHI Conference on Human Factors in Computing Systems; 2013; Paris, France p. 1759-1768. [doi: 10.1145/2470654.2466232]
18. Zhou X, Sun S, Yang J. Understanding diabetes management via a Chinese online community. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2014; Toronto, Canada p. 3997-4006. [doi: 10.1145/2556288.2557344]
19. Myneni S, Cobb NK, Cohen T. Finding meaning in social media: content-based social network analysis of QuitNet to identify new opportunities for health promotion. Stud Health Technol Inform 2013;192:807-811. [Medline: 23920669]
20. Centola D, van de Rijt A. Choosing your network: social preferences in an online health community. Soc Sci Med 2015 Jan;125:19-31. [doi: 10.1016/j.socscimed.2014.05.019] [Medline: 24951403]
21. Aarts JWM, Faber MJ, den Boogert AG, Cohlen BJ, van der Linden PJQ, Kremer JAM, et al. Barriers and facilitators for the implementation of an online clinical health community in addition to usual fertility care: a cross-sectional study. J Med Internet Res 2013;15(8):e163 [FREE Full text] [doi: 10.2196/jmir.2098] [Medline: 23996964]
22. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
23. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
24. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
25. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: a systematic review and novel taxonomy. Am J Med 2015 Dec;128(12):1335-1350. [doi: 10.1016/j.amjmed.2015.06.015] [Medline: 26159633]
26. Tengilimoglu D, Sarp N, Yar CE, Bektaş M, Hidir MN, Korkmaz E. The consumers' social media use in choosing physicians and hospitals: the case study of the province of Izmir. Int J Health Plann Manage 2015 Jul 22:-. [doi: 10.1002/hpm.2296] [Medline: 26198772]
27. Fox S. Pew Research Center's Internet & American Life Project. 2011. The social life of health information URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2016-07-28] [WebCite Cache ID 6jLgXeVDf]

28. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
29. Naftel RP, Safiano NA, Falola MI, Shannon CN, Wellons JC, Johnston JM. Technology preferences among caregivers of children with hydrocephalus. J Neurosurg Pediatr 2013 Jan;11(1):26-36. [doi: 10.3171/2012.9.PEDS12208] [Medline: 23092227]
30. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. 2011 Presented at: ACM Conference on Computer Supported Cooperative Work, CSCW; 2011; Hangzhou, China p. 341-350. [doi: 10.1145/1958824.1958876]
31. Whitehill JM, Brockman LN, Moreno MA. "Just talk to me": communicating with college students about depression disclosures on Facebook. J Adolesc Health 2013 Jan;52(1):122-127 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.09.015] [Medline: 23260845]
32. Duggan M. Pew Research Center. 2015. Mobile messaging and social media 2015 URL: http://www.pewinternet.org/files/2015/08/Social-Media-Update-2015-FINAL2.pdf [accessed 2016-07-28] [WebCite Cache ID 6jLgjZmn4]
33. Makovsky Integrated Communications. Online health research eclipsing patient-doctor conversations. 2013. URL: http://www.makovsky.com/news/online-health-research-eclipsing-patient-doctor-conversations/ [accessed 2016-07-28] [WebCite Cache ID 6jLgsQ6Ue]
34. University of Twente: Center for eHealth Research and Disease Management. CeHRes Roadmap. 2011. URL: http://www.ehealthresearchcenter.org/wiki/index.php/Main_Page [accessed 2016-07-28] [WebCite Cache ID 6jLh589sS]
35. Gajaria A, Yeung E, Goodale T, Charach A. Beliefs about attention-deficit/hyperactivity disorder and response to stereotypes: youth postings in Facebook groups. J Adolesc Health 2011 Jul;49(1):15-20. [doi: 10.1016/j.jadohealth.2010.09.004] [Medline: 21700151]
36. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]
37. De la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]
38. Gaysynsky A, Romansky-Poulin K, Arpadi S. My YAP Family: analysis of a Facebook group for young adults living with HIV. AIDS Behav 2015 Jun;19(6):947-962. [doi: 10.1007/s10461-014-0887-8] [Medline: 25186783]
39. Ahmed S, Haines-Saah RJ, Afzal AR, Tam-Tham H, Al MM, Hemmelgarn BR, et al. User-driven conversations about dialysis through Facebook: A qualitative thematic analysis. Nephrology (Carlton) 2016 Mar 22:-. [doi: 10.1111/nep.12780] [Medline: 27003709]
40. Zhang Y. College students' uses and perceptions of social networking sites for health and wellness information. Information Research 2013 Sep;17(3):1-20 [FREE Full text]
41. Al MM, Ibrahim HM, Turin TC. Social media in communicating health information: an analysis of Facebook groups related to hypertension. Prev Chronic Dis 2015;12:E11 [FREE Full text] [doi: 10.5888/pcd12.140265] [Medline: 25633486]
42. Valdez RS, Guterbock TM, Thompson MJ, Reilly JD, Menefee HK, Bennici MS, et al. Beyond traditional advertisements: leveraging Facebook's social structures for research recruitment. J Med Internet Res 2014;16(10):e243 [FREE Full text] [doi: 10.2196/jmir.3786] [Medline: 25348050]
43. Qualtrics. Provo, UT; 2015. URL: https://www.qualtrics.com/ [accessed 2016-07-28] [WebCite Cache ID 6jLhH1Kmu]
44. Patton MQ. Qualitative research and evaluation methods. Thousand Oaks, CA: Sage Publications; 2001.
45. Pasmore WA. Designing effective organizations: the sociotechnical systems perspective. New York, NY: John Wiley & Sons; 1988.
46. Valdez RS, Zayas-Caban T. Developing hybrid macroergonomic methodologies: a systems foundation for consumer health IT design. 2015 Presented at: 19th Trienial Congress of the International Ergonomics Association; 2015; Melbourne, Australia.
47. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
48. NVivo (Version 10). 2012. QSR International URL: http://www.qsrinternational.com/ [accessed 2016-07-28] [WebCite Cache ID 6jLhn8kgf]
49. Saldaña J. The coding manual for qualitative researchers. Thousand Oaks, CA: SAGE Publications Ltd; 2009.
50. Valdez RS. Creating a Foundation for the Design of Culturally Informed Consumer Health IT. In: Dissertation. Madison, Wisconsin: Industrial and Systems Engineering: University of Wisconsin-Madison; 2012:1-303.
51. Skeels MM. Sharing by Design: Understanding and supporting personal health information sharing and collaboration within social networks. In: Dissertation. Seattle, WA: University of Washington; 2010:1-251.
52. Microsoft Corporation. HealthVault. 2016. URL: https://www.healthvault.com/us/en [accessed 2016-07-28] [WebCite Cache ID 6jLiAEGUT]
53. CaringBridge. 2016. URL: https://www.caringbridge.org/ [accessed 2016-07-28] [WebCite Cache ID 6jLi2Q8jA]

54. Hartzler A, Skeels M, Mukai M, Powell C, Klasnja P, Pratt W. Sharing is caring, but not error free: transparency of granular controls for sharing personal health information in social networks. AMIA Annu Symp Proc 2011;2011:559-568 [FREE Full text] [Medline: 22195111]
55. Valdez R, Guterbock T, Fitzgibbon K. From loquacious to reticent: understanding patient health information communication to guide consumer health IT design. Journal of the American Medical Informatics Association 2017:- This paper is currently under review. It does not make sense to include page numbers here(forthcoming).

Abbreviations

IT: information technology
SNS: social networking site

Edited by G Eysenbach; submitted 06.05.16; peer-reviewed by M Merolli, F Martin-Sanchez, P Fontelo; comments to author 26.05.16; revised version received 07.07.16; accepted 20.07.16; published 11.08.16.

Please cite as:
Menefee HK, Thompson MJ, Guterbock TM, Williams IC, Valdez RS
Mechanisms of Communicating Health Information Through Facebook: Implications for Consumer Health Information Technology Design
J Med Internet Res 2016;18(8):e218
URL: http://www.jmir.org/2016/8/e218/
doi:10.2196/jmir.5949
PMID:27515151

Original Paper

Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection

Didi Surian[1], PhD; Dat Quoc Nguyen[2], MSc; Georgina Kennedy[1], MEng; Mark Johnson[2], BSc, MA, PhD; Enrico Coiera[1], MBBS, PhD; Adam G Dunn[1], PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, New South Wales, Australia

[2]Department of Computing, Faculty of Science and Engineering, Macquarie University, North Ryde, New South Wales, Australia

Corresponding Author:
Didi Surian, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera Road
North Ryde, New South Wales, 2109
Australia
Phone: 61 +61298502455
Fax: 61 +61298502499
Email: didi.surian@mq.edu.au

Abstract

Background: In public health surveillance, measuring how information enters and spreads through online communities may help us understand geographical variation in decision making associated with poor health outcomes.

Objective: Our aim was to evaluate the use of community structure and topic modeling methods as a process for characterizing the clustering of opinions about human papillomavirus (HPV) vaccines on Twitter.

Methods: The study examined Twitter posts (tweets) collected between October 2013 and October 2015 about HPV vaccines. We tested Latent Dirichlet Allocation and Dirichlet Multinomial Mixture (DMM) models for inferring topics associated with tweets, and community agglomeration (Louvain) and the encoding of random walks (Infomap) methods to detect community structure of the users from their social connections. We examined the alignment between community structure and topics using several common clustering alignment measures and introduced a statistical measure of alignment based on the concentration of specific topics within a small number of communities. Visualizations of the topics and the alignment between topics and communities are presented to support the interpretation of the results in context of public health communication and identification of communities at risk of rejecting the safety and efficacy of HPV vaccines.

Results: We analyzed 285,417 Twitter posts (tweets) about HPV vaccines from 101,519 users connected by 4,387,524 social connections. Examining the alignment between the community structure and the topics of tweets, the results indicated that the Louvain community detection algorithm together with DMM produced consistently higher alignment values and that alignments were generally higher when the number of topics was lower. After applying the Louvain method and DMM with 30 topics and grouping semantically similar topics in a hierarchy, we characterized 163,148 (57.16%) tweets as evidence and advocacy, and 6244 (2.19%) tweets describing personal experiences. Among the 4548 users who posted experiential tweets, 3449 users (75.84%) were found in communities where the majority of tweets were about evidence and advocacy.

Conclusions: The use of community detection in concert with topic modeling appears to be a useful way to characterize Twitter communities for the purpose of opinion surveillance in public health applications. Our approach may help identify online communities at risk of being influenced by negative opinions about public health interventions such as HPV vaccines.

(J Med Internet Res 2016;18(8):e232) doi:10.2196/jmir.6045

KEYWORDS

topic modelling; graph algorithms analysis; social media; public health surveillance

Introduction

The human papillomavirus (HPV) vaccine was first introduced to reduce the incidence of HPV and the majority of cervical cancers [1]. Despite evidence of its safety and efficacy [2-5], the quality of information about the vaccine on the Web is varied [6,7], and coverage of the vaccine is low in some countries, including the United States [8]. For vaccines generally, there is evidence to suggest that negative information about vaccines from celebrities, health practitioners, and news media can increase vaccine hesitancy and refusal [9-11]. Although the HPV vaccine is still a recent addition to the armament of public health, it is important to perform surveillance on social media to understand various opinions about vaccination.

The use of social media information in public health applications has previously centered on forecasting clinical outcomes that have traditionally been measured using surveys and registries. Applications of data mining using Twitter have included influenza surveillance [12-14] and measuring spatial differences in language or mood [15,16]. Sentiment and language analyses on Twitter have been used as indicators for the geographical variation in heart disease mortality [17]. Examples that are relevant to our research include the use of topic modeling to extract tobacco-related tweets in the United States [18] and the surveillance of information about the misunderstanding and misuse of antibiotics from online media [19]. There is a growing area of research considering the spread of information, news, and opinions about vaccines [20-24], and research in this area focuses on measuring associations between misinformation, beliefs, and decision making across a range of community health practices [25].

People interact and form relationships to each other on social media. With these relationships, communities are formed. The structure of online communities influences—and can be influenced by—the information that enters and is diffused through them. Studies examining the spread of information through online communities, social media, and news media have shown that the heterogeneous social structure of the network and external factors can play a role in how far and fast new information spreads [26,27]. That competition between memes and natural attrition may affect how far and fast they can spread and how quickly they decay [28-30], and that topics of interest in a community can also influence the formation and persistence of social structure [31].

Differences in the content of the information posted in online news and social media influence the variation in opinions and beliefs in online communities. While the opinions held by community members and the content they post are not equivalent, we assume that the content provides a reasonable proxy for the opinions that might influence decision making. Topic modeling methods are appropriate for identifying thematic structure (topic) within a set of tweets because they can be applied to an unstructured corpus of documents and there is no requirement that the topic be defined in advance [32]. The primary challenge associated with applying topic modeling to tweets comes from the short length of tweets (140 characters). Despite this challenge, topic modeling has been used to examine topics across a range of subjects on Twitter—by pooling tweets to produce longer documents to analyze [33-35], or applying extensions or alternatives to existing models that work better on shorter documents [36-40].

We expected to find that homophily and contagion would lead to a clustering of opinions in online communities, but to date there has been little research done to measure this important information for vaccines. Our aim was to evaluate the combination of community structure and topic modeling methods in measuring the distribution of topics about HPV vaccines from the tweets posted by users within communities on Twitter, with the broader goal of evaluating a new process for characterizing online communities by the public health information expressed by the community members.

Methods

Study Data

Using repeated searches via the Twitter Search Application Programming Interface, we collected tweets about HPV vaccines between October 1, 2013, and October 29, 2015, using the keywords shown in Table 1. For each tweet labeled as English language by Twitter, we stored the text of the tweet and the related metadata. Each time a new user tweeted about HPV vaccines for the first time in the period, we additionally collected the lists of users they followed and who followed them. This information on relationships was used later to construct the network of users for our analysis. At the conclusion of the data collection period, there were 302,856 tweets (including retweets) and 112,944 users. Following the data collection, we removed users who were suspended, protected, or deleted, which left 101,519 users and 285,417 tweets for the analysis.

Table 1. Search keywords used to collect tweets about HPV vaccines for our analysis.

No.	Keywords
1.	“HPV” and “vaccine”
2.	“HPV” and “vaccination”
3.	“gardasil”
4.	“cervical” and “vaccination”
5.	“cervical” and “vaccine”
6.	“cervarix”

We pre-processed the tweets prior to topic modeling. For words that were hashtags (beginning with "#") or usernames (beginning with "@"), we made no further modifications. The remaining words in the text of the tweet were converted to lowercase, and we removed stop words, the word "RT" (which represents a retweet), and any numerical values. We then applied the Porter stemmer [41]. We excluded URLs (uniform resource locator) generated from generic URL shortening services (eg, "http://bit.ly") and included the domain of any full URLs identified from the list of expanded URLs. Document size plays an important role in topic modeling methods [42], so we chose to assign all of the tweets with fewer than three words to a single extra topic (1114 tweets), leaving 284,303 tweets.

We ran Latent Dirichlet Allocation (LDA) and the Dirichlet Mixture Model (DMM) to infer the topics of the 125,003 unique tweets that were identified within the set of 284,303 tweets after pre-processing the text. After inferring the topics using LDA and DMM, we mapped those topics back onto the full set, so that each tweet was associated with a single topic.

We constructed the network from the 4,387,524 follower connections among the 101,519 users using an undirected graph. A node represents a user, and an edge between two users is established if one was found to be following the other. The network included 100,826 (99.32%) users comprising the single largest connected component, 500 (0.49%) users who formed smaller islands disconnected from the largest connected component, and 193 (0.19%) disconnected users with no connections to the core. Within the largest connected component, the average number of social connections was 86.98 and the largest number of connections was 18,635. We measured the alignment between topics and communities for the users who were part of the largest connected component. More details on network construction are provided in Multimedia Appendix 1.

Community Detection

Community detection algorithms aim to find sets of nodes in a graph that have a greater density of connections within their set compared to across sets. Traditionally, community detection algorithms produced a hard clustering—where each node belongs to only one community [43-47]. Some of the more recent methods have considered overlapping communities [48]. In this work, we chose two algorithms that assign each node to a single community, are known to produce reliable results, and work efficiently in large networks.

The Infomap algorithm was developed to extract community structure in large complex networks [49]. Using random walks as a proxy for the way that information flows through a system, the method first determines the probability of visiting each node in the network and then characterizes the community and node structure of the network as a Huffman code. By progressively modifying the community affiliation, the aim is to compress the code describing the network to its smallest size. We used the implementation of Infomap from igraph [50].

The Louvain algorithm is a relatively fast community detection algorithm to compute and can therefore scale to large networks [51]. The algorithm is agglomerative—nodes are initialized to belong to a community of size one and sequentially aggregated with the neighboring community that produces the greatest gain in modularity (if a positive gain exists). Communities detected in this first phase become nodes in a new network with edge weights determined by the number of connections between the communities from the first phase. The algorithm therefore constructs a hierarchical representation of the network and proceeds until no more modularity gains can be identified. The final clustering that results from this procedure is used to define the community structure. We used the code released by the author of Louvain from MapEquation [52].

Topic Inference

Topic modeling is used to find natural clusters based on the co-occurrence of words. We used the Latent Dirichlet Allocation (LDA) model [53] and the Dirichlet Multinomial Mixture (DMM) model [54]. The LDA model is a standard method for topic modeling, and the DMM model is a variant especially developed for short documents such as tweets. When applying the DMM model, only one topic is assigned to each document, so we labeled each tweet according to the topic inferred by the DMM model. For LDA—where a probability topic distribution is produced for each document—tweets were labeled using the topic with the largest probability [33,40,55]. We used the implementation of LDA from gensim [56] and the jLDADMM implementation of DMM [57]. For both methods, we used standard settings for each model [39,55] and did not attempt to optimize the parameters further. Details of the formal specification and notations for the methods are provided in Multimedia Appendix 1.

Alignment Measures

The aim of measuring the alignment between topics and communities is to determine if the topics appear more frequently within some communities relative to all others. Since each tweet was associated with a single topic, we represented communities by the distribution of topics in the tweets posted by the users in that community. We adapted measures of alignment that are typically used to quantify the quality of an estimated clustering against an observed clustering to compare between the clustering methods that use the observed structure (social connections) and the clustering methods that use the observed content (topics in tweets). There are several appropriate metrics for assessing cluster quality in this scenario, including purity, normalized mutual information (NMI), and the adjusted Rand index (ARI) [39,58] (see Multimedia Appendix 1 for definitions).

While these typical metrics provide a general measure of the alignment between community structure and the topics of the tweets posted by users in those communities, they were not useful for summarizing how topics may be disproportionately represented within a small subset of the communities. We therefore additionally considered a measure of topic concentration (*TC*). We defined a *TC* value by the smallest number of communities required to cover a specified percentage of the tweets about a given topic, so TC_{95} is the number of communities required to cover 95% of the tweets in that topic, and TC_{100} is the number of communities that covers every tweet labeled with that topic. A lower TC_{95} value therefore implies a

higher concentration of topics within a small number of communities.

When comparing measures of topic concentration across multiple networks to determine alignment, the differences in the number of tweets associated with each topic can influence the measures independently of the alignment, so we used permutation tests to produce a fair comparison. The permutation tests create a baseline distribution of *TC* values that may occur in the absence of any real alignment, which can then be used to establish the level of alignment relative to the levels of alignment that could be produced by chance [59]. To do this, we randomly permuted the topics associated with each tweet such that the distributions of tweets per topic and tweets per community remained the same as the observed network. We then compared the observed TC_{95} values against the distribution of TC_{95} values produced in the permutation tests. Typical permutation tests report the percentile of a single observed value within the distribution of values produced after permutation. In the permutation tests we applied, distributions of TC_{95} values (one for each topic) were produced rather than single values, so we used a two-sample Kolmogorov Smirnov test to compare the distributions. The Kolmogorov-Smirnov test statistic varies between 0 and 1, and a higher test statistic means that the topics were more concentrated within individual communities than would be expected if the same number of tweets per topic were randomly distributed across the communities.

Manual Intrusion Tests

We performed intrusion tests on the topics from the tweets. One investigator, blinded to the results of the topic modeling, was presented with sets of five test cases per pairwise combination of topics. Each test case included the text of five tweets chosen at random from one topic and one tweet chosen at random from a different topic. The investigator was tasked with identifying the tweet that did not belong to the topic. The results of these intrusion tests indicated how well the topic modeling was able to capture semantic differences in the tweets. We additionally used the results of the intrusion tests to construct a hierarchy of topics based on their semantic dissimilarity by applying multidimensional scaling [60-62]. The method produces a distance between every pair of topics, which is then used to merge the closest topics to construct the hierarchy.

Results

Community Detection and Topic Modeling

The two community detection algorithms were applied to the largest connected component of 100,826 users. Applying the Louvain algorithm, we identified 38 distinct communities of sizes between 3 and 21,733 users. The Infomap algorithm identified 1334 distinct communities, ranging in size from 2 to 18,974 users.

We constructed a series of LDA and DMM models by varying the number of topics between 5 and 200. From the purity, NMI, and ARI scores, we found that the alignment between the community structure and the topics was higher across all measures for DMM compared to LDA. The highest purity score (0.495) and the highest ARI scores (0.166) were found when applying the DMM model with the Louvain algorithm. The highest NMI score (0.185) was found when applying the DMM model with the Infomap algorithm. The results of these experiments suggest that the DMM topic model may have produced a more realistic clustering of the tweets by topic.

The TC_{95} scores were consistently higher when using the DMM model compared to the LDA model (see Multimedia Appendix 1 for detailed results). In combination with the Infomap algorithm, TC_{95} scores were highest between 10 and 25 topics, and in combination with the Louvain algorithm, TC_{95} scores were highest between 20 to 30 topics. Considering these results, we used the DMM model (with 30 topics) and the Louvain algorithm to demonstrate the characterization of the communities by topic in what follows.

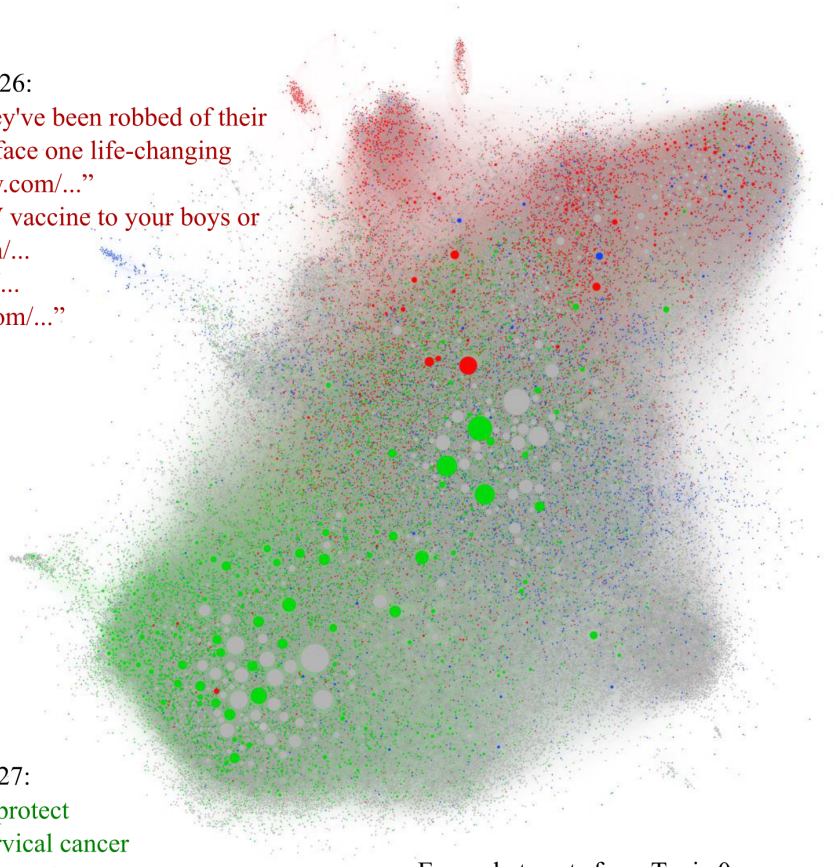
To illustrate how the topics tend to cluster within communities, we selected three representative topics and visualized them in the network constructed from the set of followers among the 100,826 users (Figure 1). The topics include one of the topics that captured clinical and scientific evidence (Topic 27), the topic comprising experiential tweets (Topic 0), and one of the topics describing side effects and harms (Topic 26).

Topic 27 includes words that are common to published studies about the efficacy of the vaccine such as “prevent,” “protect,” “study,” “news,” and “research.” Links to news media alongside other published articles and related media tended to be grouped within this topic, and the topic is broadly represented throughout the majority of the core network, including among the users with the greatest number of connections (typically news organizations in the center, and news organizations, health-related magazines, and scientific journals to one side).

Topic 0 captures a large number of tweets from users describing their own experiences with the vaccine, including temporal words such as “today,” “get,” “got,” and “go.” Tweets including phrases such as “my arm hurts like a...” were commonly assigned to this topic, and these users appeared to share fewer connections with other users posting about HPV vaccines.

In Topic 26, emotive words like “kill,” “victim,” and “death” are common. Tweets that include links to specific antivaccine websites were commonly assigned to this topic, and users posting tweets in Topic 26 appeared to cluster with different densities in three distinct groups that were separated from the groups of users posting tweets labeled as Topic 27.

Figure 1. A network of 100,826 users (nodes) who posted tweets about HPV vaccines in the period. The sizes of the nodes are proportional to the number of social connections they have in this network. Nodes are colored if they posted tweets labeled as Topic 0 (blue), Topic 26 (red), or Topic 27 (green). Node position was determined by a heuristic that attempts to locate connected nodes closer together, partially revealing the community structure.



Topic Grouping

We measured the quality of the topic modeling using the manual intrusion tests. Overall, the correct intruder was identified in 63.7% of the 4650 tests, which is a clear departure from the 16.7% that would be expected by chance. The hierarchy constructed from the manual intrusion tests revealed the semantically similar topics (Figure 2). The topic groups were (1) media debates, (2) politics and policy debates, (3) scandals and conspiracies, (4) side effects and harms, (5) public health advocacy, (6) clinical evidence, and (7) experiences. When measured across the groups of topics, the intrusion test accuracy was 76% and when measured within the groups of topics, the intrusion test accuracy was 49%. These results suggest that the separation among topic groups is clear (high score for intergroup accuracy and low score for intragroup accuracy).

Using the topics groups, we were then able to characterize the communities by the distribution of topics among the set of tweets posted by the users in those communities. Figure 3 details the topic distributions for three selected communities, notable because they illustrate the concentration of vaccine harms/conspiracies, evidence/advocacy, and experiential themes within different communities. Note also that the number of tweets per user is highest for users in the community that posts tweets labeled mostly among the vaccine harms/conspiracies theme and lowest for users in the community posting mostly about their experiences with the HPV vaccine.

Across the set of all communities, we found that users posting about their own experiences with the HPV vaccine belonged to communities for which the majority of tweets were related to evidence and advocacy. Of the 4548 users who posted tweets labeled as experiential, 3449 (75.84%) belonged to communities for which the majority of tweets were related to evidence/advocacy, 674 (14.8%) belonged to communities for which the majority of tweets were related to harms/conspiracies, 196 (4.3%) belonged to communities for which the majority of tweets were experiential, and 229 (5.0%) belonged to the group of users who were not connected to the core of the network. Figures 4 and 5 detail the distribution of themes within the communities.

Figure 2. A dendrogram of 30 topics (Topic 0-29) from the Dirichlet Mixture Model and one separate topic (Topic 30) for the tweets with fewer than 3 words. The groups were identified post-hoc and the colors represent themes—harms/conspiracies (red), evidence/advocacy (green), and experiential (blue).

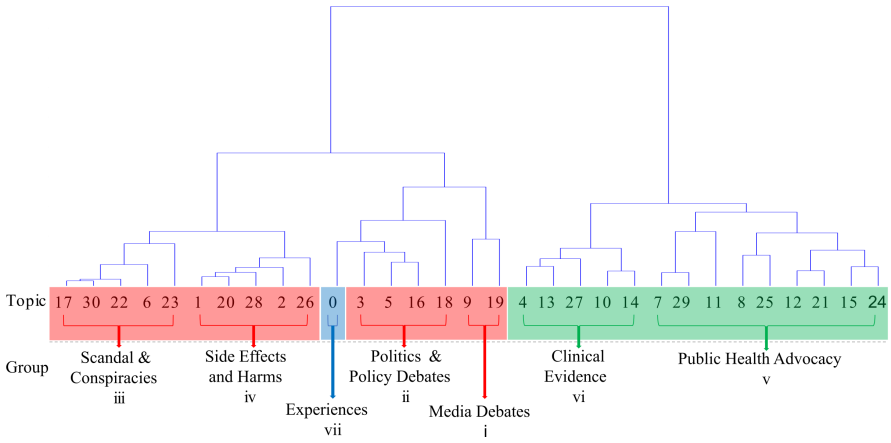


Figure 3. Topic distributions for 3 selected communities ordered by group and theme: (1) community 24 included 5275 users and an average of 2.46 tweets per user; (2) community 20 included 11,047 users and an average of 1.96 tweets per user; and (3) community 34 included 187 users and an average of 1.16 tweets per user.

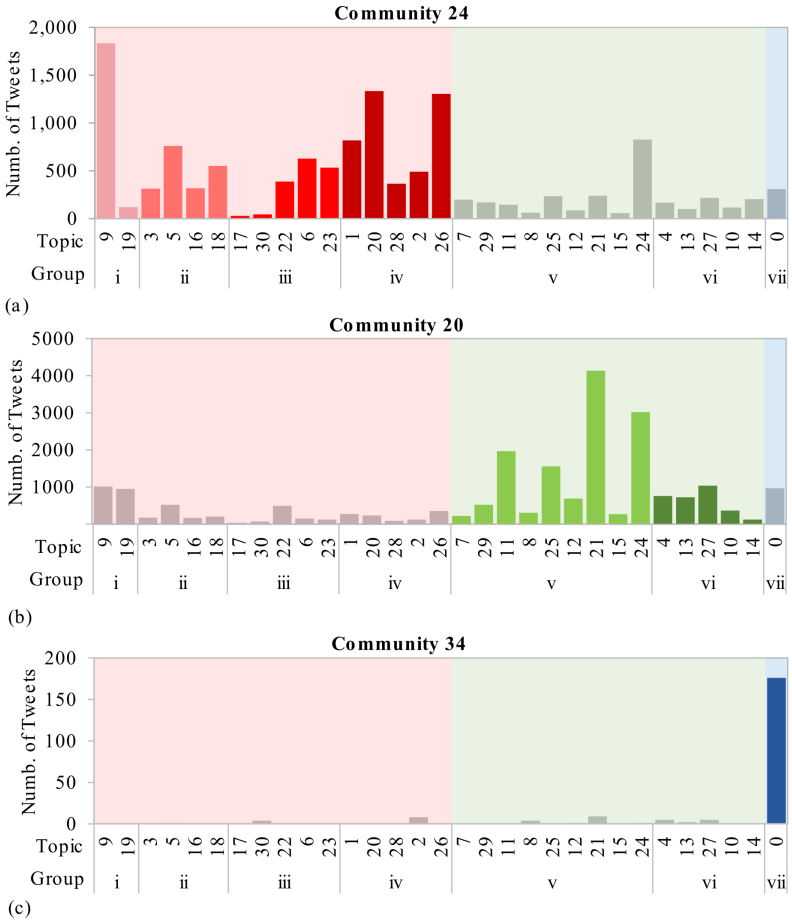


Figure 4. The proportion of evidence/advocacy and harms/conspiracies themes for the identified 39 communities. Each circle represents a community and the size is proportional to the number of tweets in the respective community. Communities further from the diagonal include greater proportions of experiential theme tweets.

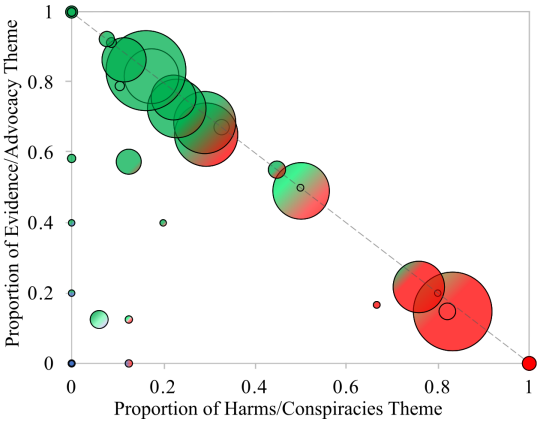
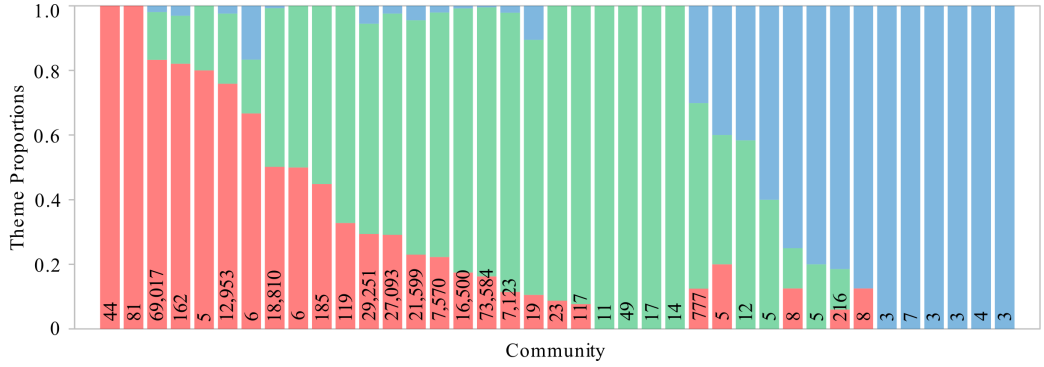


Figure 5. Theme proportions across the 39 communities, representing the proportion of tweets that were assigned to the themes of harms/conspiracies (red), evidence/advocacy (green), and experiential (blue). Values along the horizontal axis are the total number of tweets in the community.



Discussion

Principal Results

In this study, we sought to measure the alignment between the community structure implicit among the follower network of Twitter users posting tweets about HPV vaccines and the topics about which they posted. Given what is already known about the variable quality of information about HPV vaccines on Twitter [22,23], we expected to find that some communities would more often perpetuate negative opinions about HPV vaccines and that these communities would be distinct from the communities describing the favorable evidence or advocating for its uptake. Using a statistical measure quantifying the strength of the topic concentration, we found that some topics were heavily concentrated within a small number of communities, which was consistent with our expectations. Compared to our previous work in this application domain [22,23], the process described here provides a more nuanced view of the specific concerns about HPV vaccines expressed on social media and the ability to identify communities in which these concerns were the predominant topics. Using the combination of topic modeling and community structure to characterize communities, we were able to identify communities in which specific concerns about safety or politics were predominant, as well as identify younger Twitter users who posted experiential tweets and were at risk of greater exposure to safety concerns than to evidence and advocacy, which may occur between the first and subsequent doses of the vaccine. Analysis of tweets for public health where opinions and experiences are mixed have been investigated previously for influenza, where some tweets may help identify influenza incidence and others represent evidence dissemination or opinion [21].

Comparison With Prior Work

A growing set of methods has been developed using either structural information to improve inferences about the content of a corpus, or the information characterizing the nodes in a network to improve the analysis of the structure. Those aiming to understand the content of tweets have used social connections to improve tweet classification [22,63-65]. These studies have considered mentions, retweets, and other forms of interaction that are available on Twitter, but the use of information about followers generally produced the highest levels of performance. Other researchers have proposed methods for incorporating network structure into topic modeling approaches in networks other than Twitter [66,67]. Conversely, some studies have considered the use of content associated with nodes in networks to improve the quality of community detection [68,69]. Among the studies examining documents and the structure between them—such as emails [70], co-authorship [71], and Wikipedia [72]—one study produced topic profiles for communities in a similar fashion to the way we have done in Figure 3 [73]. The approach we presented here differs from these studies because we applied community detection and topic modeling independently, rather than attempting to leverage the information available about social connections to improve the quality of the

topic modeling process, or to use content information to improve community structure or predict new connections.

Limitations

A limitation of this work is that we considered a single application domain. While the uptake of vaccines is of critical importance to public health, further testing on other application domains would be required to determine the generalizability of assessing topic concentrations as a way of characterizing Twitter communities. A further limitation in our work is that we did not consider the temporal dynamics of the topics or the community structure in any detail. Given that the topics related to HPV vaccines are likely to produce similar temporal patterns to those observed by Leskovec et al [30], future work in this area may benefit from further analysis of the relationship between the temporal dynamics of the topics and the economy of attention within communities, which has been explored elsewhere [28,29]. Finally, we considered the follower network as an undirected network and did not incorporate weights or directionality based on mutual followers, or the presence of retweets and mentions, which would have provided a more nuanced representation of the social connections and may have produced a different community structure.

Future Directions

Our work here has potential implications for public health practices. Applying topic modeling and community detection methods in concert to a corpus of tweets about HPV vaccines, we found that it was possible to characterize online communities by the topics that are most heavily concentrated among their tweets. One way of translating these methods into public health practice would be to use these methods in combination with new spatial and demographic estimation methods [74-76], to produce spatiotemporal indicators that determine where and when the growth of specific concerns may lead to increased vaccine hesitancy or refusal. We think that these indicators may have a future role in helping public health organizations design interventions and communication strategies that are better targeted and thus more efficient.

Conclusions

In this work, we demonstrated a novel process for characterizing the concentration of certain opinions in online communities by independently applying existing community detection and topic modeling methods, and quantifying the differences in the topic distributions across communities. Among tweets about HPV vaccines, we found that there were clear differences in the distribution of topics across communities defined by the follower network. In practice, public health organizations may wish to consider identifying the locations and demographics of communities that are at risk of exposure to antivaccine information in order to intervene with positive messages targeting the specific concerns identified through topic modeling. The value of this work in public health includes a more nuanced representation of the variety of concerns expressed about HPV vaccines online and some practical steps towards the development of an automated system for the surveillance of public opinions with the purpose of understanding localized differences in decision making and health behaviors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Network construction, formal specification/notations, and detailed experiment results.

[PDF File (Adobe PDF File), 981KB - jmir_v18i8e232_app1.pdf]

References

1. Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. Vaccine 2012 Nov 20;30 Suppl 5:F12-F23. [doi: 10.1016/j.vaccine.2012.07.055] [Medline: 23199955]
2. Crowe E, Pandeya N, Brotherton JML, Dobson AJ, Kisely S, Lambert SB, et al. Effectiveness of quadrivalent human papillomavirus vaccine for the prevention of cervical abnormalities: case-control study nested within a population based screening programme in Australia. BMJ 2014;348:g1458 [FREE Full text] [Medline: 24594809]
3. Skinner SR, Szarewski A, Romanowski B, Garland SM, Lazcano-Ponce E, Salmerón J, et al. Efficacy, safety, and immunogenicity of the human papillomavirus 16/18 AS04-adjuvanted vaccine in women older than 25 years: 4-year interim follow-up of the phase 3, double-blind, randomised controlled VIVIANE study. Lancet 2014 Dec 20;384(9961):2213-2227. [doi: 10.1016/S0140-6736(14)60920-X] [Medline: 25189358]
4. Tabrizi SN, Brotherton JML, Kaldor JM, Skinner SR, Liu B, Bateson D, et al. Assessment of herd immunity and cross-protection after a human papillomavirus vaccination programme in Australia: a repeat cross-sectional study. Lancet Infect Dis 2014 Oct;14(10):958-966. [doi: 10.1016/S1473-3099(14)70841-2] [Medline: 25107680]
5. Brotherton JML, Fridman M, May CL, Chappell G, Saville AM, Gertig DM. Early effect of the HPV vaccination programme on cervical abnormalities in Victoria, Australia: an ecological study. Lancet 2011 Jun 18;377(9783):2085-2092. [doi: 10.1016/S0140-6736(11)60551-5] [Medline: 21684381]
6. Madden K, Nan X, Briones R, Waks L. Sorting through search results: a content analysis of HPV vaccine information online. Vaccine 2012 May 28;30(25):3741-3746. [doi: 10.1016/j.vaccine.2011.10.025] [Medline: 22019758]

7. Robbins SCC, Pang C, Leask J. Australian newspaper coverage of human papillomavirus vaccination, October 2006-December 2009. J Health Commun 2012;17(2):149-159. [doi: 10.1080/10810730.2011.585700] [Medline: 22136302]
8. Printz C. HPV vaccination rates remain low: cancer prevention community needs to continue promoting the vaccine's safety and efficacy, say experts. Cancer 2015 May 1;121(9):1341-1343 [FREE Full text] [doi: 10.1002/cncr.28993] [Medline: 25902756]
9. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. J Epidemiol Community Health 2000 Jun;54(6):473-474 [FREE Full text] [Medline: 10818125]
10. Hoffman SJ, Tan C. Following celebrities' medical advice: meta-narrative analysis. BMJ 2013 Dec 17;347(dec17 14):f7151. [doi: 10.1136/bmj.f7151]
11. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
12. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
13. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014;6:- [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
14. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. In: 1st Workshop on Social Media Analytics. USA: ACM Press; 2010 Presented at: The 1st Workshop on Social Media Analytics; July 25, 2010; Washington, DC p. 115-122 URL: http://snap.stanford.edu/soma2010/papers/soma2010_16.pdf [WebCite Cache ID 6jyz7IJSE] [doi: 10.1145/1964858.1964874]
15. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. The Twitter of Babel: mapping world languages through microblogging platforms. PLoS One 2013;8(4):e61981 [FREE Full text] [doi: 10.1371/journal.pone.0061981] [Medline: 23637940]
16. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
17. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
18. Prier KW, Smith MS, Giraud-Carrier C, Hanson CL. Identifying health-related topics on Twitter: an exploration of tobacco-related tweets as a test topic. In: Proceedings of the 4th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. Heidelberg: Springer-Verlag; 2011 Presented at: The 4th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction; March 29-31, 2011; College Park, Maryland p. 18-25 URL: http://mail.smithworx.com/publications/SBP11.pdf [WebCite Cache ID 6jzLj57N1]
19. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
20. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
21. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
22. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong M, Dunn AG. Using social connection information to improve opinion mining: identifying negative sentiment about HPV vaccines on Twitter. Stud Health Technol Inform 2015;216:761-765. [Medline: 26262154]
23. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about Human Papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
24. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The digital distribution of public health news surrounding the Human Papillomavirus Vaccination: a longitudinal infodemiology study. JMIR Public Health Surveill 2015;1(1):e2. [doi: 10.2196/publichealth.3310] [Medline: 27227125]
25. Oliver JE, Wood T. Medical conspiracy theories and health behaviors in the United States. JAMA Intern Med 2014 May;174(5):817-818. [doi: 10.1001/jamainternmed.2014.190] [Medline: 24638266]
26. Romero DM, Tan C, Ugander J. On the interplay between social and topical structure. In: 7th International Conference on Weblogs and Social Media. Palo Alto, California: AAAI Press; 2013 Presented at: The 7th International Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, Massachusetts p. 516-525 URL: http://www.cs.cornell.edu/~chenhao/pub/social-topical-structure.pdf [WebCite Cache ID 6jyzAnVnX]
27. Weng L, Menczer F, Ahn Y. Virality prediction and community structure in social networks. Sci Rep 2013;3:2522 [FREE Full text] [doi: 10.1038/srep02522] [Medline: 23982106]

28. Weng L, Flammini A, Vespignani A, Menczer F. Competition among memes in a world with limited attention. Sci Rep 2012;2:335 [FREE Full text] [doi: 10.1038/srep00335] [Medline: 22461971]
29. Lehmann J, Gonc¸alves B, Ramasco JJ, Cattuto C. Dynamical classes of collective attention in Twitter. In: 21st International Conference on World Wide Web. USA: ACM; 2012 Presented at: The 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 251-260. [doi: 10.1145/2187836.2187871]
30. Leskovec J, Backstrom L, Kleinberg J. Meme-tracking and the dynamics of the news cycle. In: 15th ACM International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2009 Presented at: 15th ACM International Conference on Knowledge Discovery and Data Mining; June 28-July 1, 2009; Paris, France p. 497-506 URL: https://cs.stanford.edu/people/jure/pubs/quotes-kdd09.pdf [WebCite Cache ID 6jz03c62k] [doi: 10.1145/1557019.1557077]
31. Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group formation in large social networks: Membership, growth, and evolution. In: 12th ACM International Conference on Knowledge Discovery and Data Mining.: ACM; 2006 Presented at: The 12th ACM International Conference on Knowledge Discovery and Data Mining; August 20-23, 2006; Philadelphia, USA p. 44-54 URL: http://www.cs.cornell.edu/~lars/kdd06-comm.pdf [WebCite Cache ID 6jz066dOT] [doi: 10.1145/1150402.1150412]
32. Blei DM. Probabilistic topic models. Commun. ACM 2012 Apr 01;55(4):77. [doi: 10.1145/2133806.2133826]
33. Mehrotra R, Sanner S, Buntine W, Xie L. Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In: 36th International Conference on Research and development in Information Retrieval. USA: ACM; 2013 Presented at: 36th International Conference on Research and development in Information Retrieval; July 28-August 1, 2013; Dublin, Ireland p. 889-892 URL: http://users.cecs.anu.edu.au/~ssanner/Papers/sigir13.pdf [WebCite Cache ID 6jz08zvPi] [doi: 10.1145/2484028.2484166]
34. Weng J, Lim EP, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers. In: 3rd ACM International Conference on Web Search and Data Mining. USA: ACM; 2010 Presented at: 3rd ACM International Conference on Web Search and Data Mining; 2010; New York, USA p. 261-270 URL: http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1503&context=sis_research [WebCite Cache ID 6jz0CEI6Q] [doi: 10.1145/1718487.1718520]
35. Balasubramanyan R, Kolcz A. “w00t! feeling great today!” chatter in Twitter: identification and prevalence. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. USA: ACM; 2013 Presented at: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 25-29, 2013; Niagara, Canada p. 312-316. [doi: 10.1145/2492517.2492611]
36. Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, et al. Comparing Twitter and traditional media using topic models. In: 33rd European Conference on Advances in Information Retrieval. Heidelberg: Springer-Verlag; 2011 Presented at: 33rd European Conference on Advances in Information Retrieval; April 18-21, 2011; Dublin, Ireland p. 338-349 URL: http://www.mysmu.edu/faculty/jingjiang/papers/ECIR'11.pdf [WebCite Cache ID 6jz0FUxJj] [doi: 10.1007/978-3-642-20161-5_34]
37. Ramage D, Dumais S, Liebling D. Characterizing microblogs with topic models. In: 4th International AAAI Conference on Weblogs and Social Media. USA: AAAI Press; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC p. 130-137 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1528/1846 [WebCite Cache ID 6jz0IYjJQ]
38. Yang S, Kolcz A, Schlaikjer A, Gupta P. Large-scale high-precision topic modeling on Twitter. In: 20th ACM International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2014 Presented at: 20th ACM International Conference on Knowledge Discovery and Data Mining; August 24-27, 2014; New York, USA p. 1907-1916 URL: http://cobweb.cs.uga.edu/~squinn/mmd_s15/papers/p1907-yang.pdf [WebCite Cache ID 6jzMST8PR] [doi: 10.1145/2623330.2623336]
39. Yin J, Wang J. A Dirichlet Multinomial Mixture model-based approach for short text clustering. In: 20th ACM International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2014 Presented at: 20th ACM International Conference on Knowledge Discovery and Data Mining; August 24-27, 2014; New York, USA p. 233-242 URL: http://dbgroup.cs.tsinghua.edu.cn/wangjy/papers/KDD14-GSDMM.pdf [WebCite Cache ID 6jz0NhILQ] [doi: 10.1145/2623330.2623715]
40. Nguyen DQ, Billingsley R, Du L, Johnson M. Improving topic models with latent feature word representations. Transactions of the Association for Computational Linguistics. 2015. p. 299-313 URL: https://transacl.org/ojs/index.php/tacl/article/viewFile/582/132 [WebCite Cache ID 6jz0QYY58]
41. Porter MF. An algorithm for suffix stripping. In: Readings in Information Retrieval (Morgan Kaufmann Series in Multimedia Information and Systems). San Francisco, USA: Morgan Kaufmann; 1997:313-316.
42. Jian T, Zhaoshi M, Xuanlong N, Qiaozhu Mei MZ. Understanding the limiting factors of topic modeling via posterior contraction analysis. In: 31st International Conference on Machine Learning.: International Machine Learning Society (IMLS); 2014 Presented at: 31st International Conference on Machine Learning; June 21-26, 2014; Beijing, China p. 190-198 URL: http://jmlr.org/proceedings/papers/v32/tang14.pdf [WebCite Cache ID 6jz0T9bdl]
43. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci U S A 2002 Jun 11;99(12):7821-7826 [FREE Full text] [doi: 10.1073/pnas.122653799] [Medline: 12060727]
44. Newman MEJ. Detecting community structure in networks. The European Physical Journal B - Condensed Matter 2004 Mar 1;38(2):321-330. [doi: 10.1140/epjb/e2004-00124-y]
45. Newman MEJ. Modularity and community structure in networks. Proc Natl Acad Sci U S A 2006 Jun 6;103(23):8577-8582 [FREE Full text] [doi: 10.1073/pnas.0601602103] [Medline: 16723398]

46. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW. Statistical properties of community structure in large social and information networks. In: 17th International Conference on World Wide Web. USA: ACM; 2008 Presented at: 17th International Conference on World Wide Web; April 21-25, 2008; Beijing, China p. 695-704 URL: https://cs.stanford.edu/people/jure/pubs/ncp-www08.pdf [WebCite Cache ID 6jz18eFhf] [doi: 10.1145/1367497.1367591]
47. Fortunato S. Community detection in graphs. Physics Reports 2010 Feb 17;486(3-5):75-174. [doi: 10.1016/j.physrep.2009.11.002]
48. Ahn Y, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. Nature 2010 Aug 5;466(7307):761-764. [doi: 10.1038/nature09182] [Medline: 20562860]
49. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 2008 Jan 29;105(4):1118-1123 [FREE Full text] [doi: 10.1073/pnas.0706851105] [Medline: 18216267]
50. Python igraph (version 0.7.0). URL: http://igraph.org/python/ [accessed 2016-05-26] [WebCite Cache ID 6hmWgjDT3]
51. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
52. Edler D, Rosvall M. Source code for multilevel community detection with Infomap (version Sep 07). 2015. URL: http://www.mapequation.org/code.html [accessed 2016-05-26] [WebCite Cache ID 6hmXBQpp5]
53. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3:993-1022.
54. Nigam K, McCallum AK, Thrun S, Mitchell T. Text classification from labeled and unlabeled documents using EM. Machine Learning - Special Issue on Information Retrieval 2000;39(2-3):103-134 [FREE Full text] [doi: 10.1023/A:1007692713085]
55. Lu Y, Mei Q, Zhai C. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. Inf Retrieval 2010 Aug 5;14(2):178-203. [doi: 10.1007/s10791-010-9141-9]
56. Řehůřek R. Software framework for topic modelling with large corpora. In: LREC 2010 Workshop on New Challenges for NLP Frameworks. Malta: ELRA; 2010 Presented at: LREC 2010 Workshop on New Challenges for NLP Frameworks; May 22, 2010; Valletta, Malta p. 45-50 URL: https://radimrehurek.com/gensim/lrec2010_final.pdf [WebCite Cache ID 6jz2QxAvb]
57. Nguyen DQ. jLDADMM: A Java package for the LDA and DMM topic models (version 1.0, 2015-07-06). 2015. URL: http://jldadmm.sourceforge.net/ [accessed 2016-05-26] [WebCite Cache ID 6hmXKnvie]
58. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: 22nd International Conference on World Wide Web. USA: ACM; 2013 Presented at: 22nd International Conference on World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil p. 1445-1456. [doi: 10.1145/2488388.2488514]
59. Dunn AG, Westbrook JI. Interpreting social network metrics in healthcare organisations: a review and guide to validating small networks. Soc Sci Med 2011 Apr;72(7):1064-1068. [doi: 10.1016/j.socscimed.2011.01.029] [Medline: 21371798]
60. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 1964 Mar;29(1):1-27. [doi: 10.1007/BF02289565]
61. Cox TF, Cox MAA. Multidimensional scaling. In: Handbook of Data Visualization. Heidelberg: Springer-Verlag; 2008:315-347.
62. Borg I, Groenen PJF. Modern multidimensional scaling: Theory and applications. New York: Springer-Verlag; 2005.
63. Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P. User-level sentiment analysis incorporating social networks. In: 17th ACM International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2011 Presented at: 17th ACM International Conference on Knowledge Discovery and Data Mining; August 21-24, 2011; San Diego, CA p. 1397-1405. [doi: 10.1145/2020408.2020614]
64. Hu X, Tang L, Tang J, Liu H. Exploiting social relations for sentiment analysis in microblogging. In: 6th ACM International Conference on Web Search and Data Mining. USA: ACM; 2013 Presented at: 6th ACM International Conference on Web Search and Data Mining; February 4-8, 2013; Rome, Italy p. 537-546 URL: http://faculty.cs.tamu.edu/xiahu/papers/wsdm13Hu.pdf [WebCite Cache ID 6jz1POZoe] [doi: 10.1145/2433396.2433465]
65. Speriosu M, Sudan N, Upadhyay S, Baldridge J. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-31, 2011; Edinburgh, Scotland p. 53-63 URL: http://anthology.aclweb.org/W/W11/W11-2207.pdf [WebCite Cache ID 6jz1RI5q5]
66. Mei Q, Cai D, Zhang D, Zhai CX. Topic modeling with network regularization. In: 17th International Conference on World Wide Web. USA: ACM; 2008 Presented at: 17th International Conference on World Wide Web; April 21-25, 2008; Beijing, China p. 101-110 URL: http://www-personal.umich.edu/~qmei/pub/www08-netplsa.pdf [WebCite Cache ID 6jz1e33Vf] [doi: 10.1145/1367497.1367512]
67. Sun YZ, Han JW, Gao J, Yu YT. iTopicModel: Information network-integrated topic modeling. In: 9th IEEE International Conference on Data Mining.: IEEE; 2009 Presented at: 9th IEEE International Conference on Data Mining; December 6-9, 2009; Miami, FL p. 493-502. [doi: 10.1109/ICDM.2009.43]
68. McAuley J, Leskovec J. Learning to discover social circles in ego networks. In: Advances in Neural Information Processing Systems 25. 2013 Presented at: 26th Annual Conference on Neural Information Processing Systems 2012; December 3-8,

2012; Nevada, USA p. 548-556 URL: http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2012_0272.pdf [WebCite Cache ID 6jz1n4wak]
69. Yang J, McAuley J, Leskovec J. Community detection in networks with node attributes. In: 13th IEEE International Conference on Data Mining. 2013 Presented at: 13th IEEE International Conference on Data Mining; December 7-10, 2013; Dallas, TX p. 1151-1156 URL: https://cs.stanford.edu/people/jure/pubs/cesna-icdm13.pdf [WebCite Cache ID 6jz1pWpAH] [doi: 10.1109/ICDM.2013.167]
70. Zhou D, Manavoglu E, Li J, Giles CL, Zha H. Probabilistic models for discovering e-communities. In: 15th International World Wide Web Conference. USA: ACM; 2006 Presented at: 15th International World Wide Web Conference; May 22-26, 2006; Edinburgh, Scotland p. 173-182. [doi: 10.1145/1135777.1135807]
71. Chang J, Blei DM. Relational topic models for document networks. In: 12th International Conference on Artificial Intelligence and Statistics. 2009 Presented at: 12th International Conference on Artificial Intelligence and Statistics; April 16-18, 2009; Florida, USA p. 81-88.
72. Chang J, Blei DM. Connections between the lines: Augmenting social networks with text. In: 15th International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2009 Presented at: 15th International Conference on Knowledge Discovery and Data Mining; June 28-July 1, 2009; Paris, France p. 169-178. [doi: 10.1145/1557019.1557044]
73. Pathak N, DeLong C, Erickson K, Banerjee A. Social topic models for community extraction. In: 2nd SNA-KDD Workshop. 2008 Presented at: 2nd SNA-KDD Workshop; August 24, 2008; Las Vegas, NV URL: http://www-users.cs.umn.edu/~banerjee/papers/08/snakdd08.pdf [WebCite Cache ID 6jz1tbTTy]
74. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on Twitter. In: Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-31, 2011; Edinburgh, Scotland p. 1301-1309.
75. Jurgens D, Finnethy T, McCorriston J, Xu YT, Ruths D. Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice. In: 9th International AAAI Conference on Web and Social Media. California: The AAAI Press; 2015 Presented at: 9th International AAAI Conference on Web and Social Media; May 26-29, 2015; Palo Alto, CA p. 188-197 URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10584/10502 [WebCite Cache ID 6jz1xM3Dd]
76. Jurgens D, Allen D. Geotagging one hundred million Twitter accounts with total variation minimization. In: IEEE International Conference on Big Data. USA: IEEE; 2014 Presented at: IEEE International Conference on Big Data; October 27-30, 2014; Washington, DC p. 393-401.

Abbreviations

ARI: adjusted Rand index
DMM: Dirichlet Multinomial Mixture
HPV: human papillomavirus
LDA: Latent Dirichlet Allocation
NMI: normalized mutual information

Edited by G Eysenbach; submitted 30.05.16; peer-reviewed by D Arachi, A MacKinlay; comments to author 20.07.16; accepted 03.08.16; published 29.08.16.

Please cite as:
Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG
Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection
J Med Internet Res 2016;18(8):e232
URL: http://www.jmir.org/2016/8/e232/
doi:10.2196/jmir.6045
PMID:27573910

Original Paper

Seeing the "Big" Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data

Jordan Carpenter[1], PhD; Patrick Crutchley[1,2], BSE; Ran D Zilca[3], MSc; H Andrew Schwartz[2,4], PhD; Laura K Smith[1], BA; Angela M Cobb[5], BA; Acacia C Parks[3,5], PhD

[1]Positive Psychology Center, University of Pennsylvania, Philadelphia, PA, United States

[2]Penn Medicine Social Media and Health Innovation Lab, Penn Medicine Center for Health Care Innovation, University of Pennsylvania, Philadelphia, PA, United States

[3]Happify, New York, NY, United States

[4]Computer Science, Stony Brook University, Stony Brook, NY, United States

[5]Department of Psychology, Hiram College, Hiram, OH, United States

Corresponding Author:
Acacia C Parks, PhD
Department of Psychology
Hiram College
Bates Hall 215
11715 Garfield Rd
Hiram, OH,
United States
Phone: 1 330 569 5229
Fax: 1 330 569 5398
Email: parksac@hiram.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2017/12/e347/

Abstract

Background: Assessing the efficacy of Internet interventions that are already in the market introduces both challenges and opportunities. While vast, often unprecedented amounts of data may be available (hundreds of thousands, and sometimes millions of participants with high dimensions of assessed variables), the data are observational in nature, are partly unstructured (eg, free text, images, sensor data), do not include a natural control group to be used for comparison, and typically exhibit high attrition rates. New approaches are therefore needed to use these existing data and derive new insights that can augment traditional smaller-group randomized controlled trials.

Objective: Our objective was to demonstrate how emerging big data approaches can help explore questions about the effectiveness and process of an Internet well-being intervention.

Methods: We drew data from the user base of a well-being website and app called Happify. To explore effectiveness, multilevel models focusing on within-person variation explored whether greater usage predicted higher well-being in a sample of 152,747 users. In addition, to explore the underlying processes that accompany improvement, we analyzed language for 10,818 users who had a sufficient volume of free-text response and timespan of platform usage. A topic model constructed from this free text provided language-based correlates of individual user improvement in outcome measures, providing insights into the beneficial underlying processes experienced by users.

Results: On a measure of positive emotion, the average user improved 1.38 points per week (SE 0.01, t122,455=113.60, P<.001, 95% CI 1.36–1.41), about a 27% increase over 8 weeks. Within a given individual user, more usage predicted more positive emotion and less usage predicted less positive emotion (estimate 0.09, SE 0.01, t6047=9.15, P=.001, 95% CI .07–.12). This estimate predicted that a given user would report positive emotion 1.26 points higher after a 2-week period when they used Happify daily than during a week when they didn't use it at all. Among highly engaged users, 200 automatically clustered topics

showed a significant (corrected $P<.001$) effect on change in well-being over time, illustrating which topics may be more beneficial than others when engaging with the interventions. In particular, topics that are related to addressing negative thoughts and feelings were correlated with improvement over time.

Conclusions: Using observational analyses on naturalistic big data, we can explore the relationship between usage and well-being among people using an Internet well-being intervention and provide new insights into the underlying mechanisms that accompany it. By leveraging big data to power these new types of analyses, we can explore the workings of an intervention from new angles, and harness the insights that surface to feed back into the intervention and improve it further in the future.

(J Med Internet Res 2016;18(8):e241) doi:10.2196/jmir.5725

KEYWORDS

well-being intervention; big data; qualitative analysis; linguistic analysis; word cloud; multilevel modeling

Introduction

As Internet interventions become increasingly popular—in research settings, but even more so in industry—they yield large datasets that can be used for research purposes. These datasets often fall under the umbrella of "big data," where big data is defined as a dataset so large and complex that traditional data analytic approaches cannot easily handle them [1]. While these datasets are often not designed upfront to answer research questions, and therefore often do not have control groups, they are large in size, are rich in content, and offer a view into the intervention that allows users to interact with an intervention naturalistically. Big intervention data, therefore, offer the opportunity to test interventions in the real world, where interventions are actually being found by individuals and used naturalistically. Moreover, since the data are produced during normal usage, they can be immediately useful as a means for assessing not only whether the intervention is effective, but also which aspects and parts of it are more effective than others and how the intervention can be modified to become more effective. Such iterative processes take advantage of the short implementation cycles of software as opposed to more traditional medical interventions.

Specialized analytic methods are necessary to handle the amount and frequency of data given for a particular user and to accommodate a very large number of users. Despite the extra care needed to analyze big data, there are some substantial benefits to using a big data approach. First, whereas more traditional intervention evaluation studies are usually constrained by budgetary concerns (each participant costs money in order to encourage retention), openly available products generally aim to acquire as many users as possible. The resulting potential sample size is massive and grants the power to do a class of analyses that could not be considered in a typical study with 100–200 (or even 500–600) users per cell. The large number of participants also provides enough power for moderator analyses, which is a capability often available only to meta-analyses. Second, the sample derived from an existing intervention potentially has much greater external validity because of the lack of a highly controlled experimental setting. Third, by analyzing data outside of the restrictions of a traditional randomized, controlled design (including unstructured data such as text produced by users), researchers are opened up to a new variety of potential research questions [2] and can directly examine the relationship between spontaneous usage and outcome.

Of course, the analysis of naturalistic big data has its own problems, especially in an industry context. Researchers often have less flexibility in what they can ask participants, because excessive in-product assessment can reduce retention [1]. researchers are therefore limited to knowing only what participants tell them over the course of naturally using the product. Perhaps most problematically, there is no formal control group. It is easy, therefore, to relegate any data analysis that springs from a real-world Internet intervention to a lower status than that from a randomized controlled trial (RCT) [3]. However, some of these issues are addressable with the right kind of data.

Although there have been write-ups of psychological interventions that explore efficacy or effectiveness without a control group, these are largely treated as uncontrolled pilot studies [4] where no special approach was taken to offer an alternative to the control group. One common study design in this context, the cohort study, is essentially a longitudinal study in which a sample is followed and tracked over time. Watching outcomes unfold naturally in this way allows researchers to establish temporal precedence. The hypothesized cause comes before the hypothesized effect, which is superior to a cross-sectional study when establishing causality [5]. This type of design can often yield a more externally valid sample, with fewer screening criteria or other restrictions to sample membership, which makes generalization more appropriate. However, cohort studies are vulnerable to problems with bias, because the lack of random assignment may mean systematic differences between those who do well and those who do not. As such, there is a substantial gap in the literature regarding rigorous approaches to testing interventions in uncontrolled settings [2]. There are not yet best practices for doing so.

One possible alternative approach stems from the self-controlled case series approach, most commonly used in medicine [6,7]. Like in a cohort study, participants are tracked over time in a self-controlled case series study, but the emphasis is not on averaging individuals together or on making comparisons between individuals. Instead, the emphasis is on within-person changes on (and potentially interactions between) observed variables [6]. In this study, we applied some of the strengths of a cohort study with a statistical approach that might be used in a self-controlled case series study (study 1). Specifically, we

used ongoing data about user engagement available through a website and app. Our objective was to conceptualize usage not as a trait or static variable (eg, user 1 was high usage, user 2 was low usage) but as a dynamic, constantly changing variable that may be tied to higher or lower outcomes as it varies. Whereas many RCTs aim to standardize or maximize engagement (ideally, every participant would show 100% engagement) [8], multilevel models can harness that variation, creating a dose-response relationship between behavior and outcome.

An additional goal was to explore an exciting aspect of big data: its potential to fuel linguistic analysis of the text produced by users and entered into a site (study 2). Large amounts of language can be mined to automatically reveal latent psychological processes [9,10]. Such analyses can permit us to examine not only effectiveness, but also processes in ways that would not be possible without large volumes of text. Commercial platforms can allow natural language data to accrue daily, and the text can be mined for broad-scale patterns in language that reflect a person's process of improvement. In this way, our work is an example of an existing literature that uses passive forms of data collection, such as search behavior and phone sensors [11], to measure and characterize psychological constructs and to develop opportunities for targeted intervention [12].

To summarize, the goal of our study was to broach several interesting research questions that become possible to ask when working with a massive, real-world intervention dataset. We began in study 1 by using multilevel modeling to establish the overall effects of a Web-based self-help platform on well-being, tracking the improvement over time as it varied within an individual [5,13]. By focusing on within-person variation, it was possible to address a common problem in uncontrolled data analysis, that is, systematic biases between different users who exhibit generally high or generally low usage, by emphasizing how each person's improvement varied based on his or her usage. The question shifted from "Did users who get the intervention improve?" to "Did users feel better during weeks that they used the intervention more and worse during weeks that they used it less?"

In line with previous work finding that Internet interventions are able to improve well-being effectively, we expected that users would exhibit higher well-being during periods in which they used the intervention more frequently [14,15]. Consistent with previous work on the relationship between effort and outcomes, within any given user, more usage would be correlated with higher well-being [16]. Furthermore, consistent with moderator analyses from meta-analyses of behavioral well-being interventions, we expected that this effect would be magnified for users who began with lower well-being [17,18].

In study 2, we then used linguistic analysis to gain a descriptive picture of this effect, that is, to visualize what people were saying (and, we might extrapolate, what psychological processes they were experiencing) that might help explain why they experienced improvements. While our overall hypothesis was that certain patterns of word usage would be related to well-being, we did not have specific predictions about which words might be most strongly associated.

Methods

We drew all data from the user base of Happify, a Web-based platform that offers techniques grounded in positive psychology, cognitive behavioral therapy, and mindfulness. Happify can be used on the Internet, via an app (Android and iOS), or both. People find Happify through media coverage, word of mouth, social media, and paid advertising across the Internet.

Consistent with previous research using commercial well-being apps [19], participants gave semipassive consent by accepting a user agreement that explained that their data may be used for research. Specifically, the terms and conditions of Happify stated that "Information that we collect about you also may be combined by us with other information available to us through third parties for research and measurement purposes, including measuring the effectiveness of content, advertising, or programs. This information from other sources may include age, gender, demographic, geographic, personal interests, product purchase activity or other information."

We assessed 3 demographic questions—age, employment status, and sex—in all users. A fourth question about number of children was added later, and therefore only asked of a subset of users. The dataset contained data from users whose accounts were created between December 1, 2014 and May 1, 2016, and the 8-week intervention period began at the time they completed their first assessment.

Content of Happify

We organized activities into the following 5 categories using the acronym STAGE: savor (mindfulness activities) [20,21], thank (gratitude activities) [22,23], aspire (optimism, best possible selves, goal setting, and meaning or purpose activities) [24,25], give (kindness, prosocial spending, and forgiveness activities) [26], and empathize (self-compassion and perspective-taking activities) [27]. Not all users used the same activities, as there were many possible ways to progress through the site. Users selected from many possible tracks, which were collections of activities that targeted a particular goal or problem, such as coping better with stress or improving one's romantic relationship. Any given track drew activities from several areas of STAGE, and the track recommended activities to the user on a daily basis that were customized to his or her particular goal. Users also received automated reminder emails and mobile phone notifications (if they used the app). However, users were not necessarily constrained by a track, either; some chose to also use a free play section, where they could pick and choose individual activities as they liked.

Users were able to select from hundreds of variations of 58 core activities spanning the STAGE categories. However, some activities were used more frequently than others and were therefore more likely to be the activities used by those in our subsamples. One commonly selected activity, "Thx Thx Thx" (a thank task), asked the user to write down 3 good things that happened to them that day. One variation, "What Went Smoothly Today?," instructed users as follows:

> *Think of three things about your day that went better than usual—maybe your commute to work was seamless, or you sent your kids off to school without a fight, or you simply had a little extra time for yourself. It can be anything—big or small. Jot them down and add a sentence or two describing why they made you feel grateful and what, if any[,] role you played in the experience.*

Another activity, called "Savor the Small Stuff" (a savor task), instructed users to set aside a few moments to give their full attention to a sensory or cognitive experience. One variation, called "Smell the Roses," instructed users thusly:

> *Indulge each of your senses by savoring something that's right in front of you. For example, instead of walking past the local park in your neighborhood, sit down on a bench and be mindful as you take in the surroundings. What do you hear? Are there sights you've never noticed before? What does the air smell like? You can also practice mindfulness while savoring an indulgent dessert, or while looking at some of your favorite photos from good times past. How did you feel at the time? What did you talk about? Just be in the moment.*

"I Think I Can" (an aspire activity) combined goal setting with behavioral activation research to get the user working on a goal. One variation of this activity, called "Mission: Possible," instructed the user to

> *Think hard and narrow in on a goal that will be reasonable to achieve this week. No more excuses—aim to complete it this week, but make sure you pick something that terrifies you just a bit. Look at whatever's cramming your ultimate to-do list, from a home renovation project to updating a blog or website or charging after new business. Then, decidedly look that fear in the eye, and without blinking, roll up your sleeves and get to the task at hand.*

While a few activities on Happify could be done sitting at a computer (meditation, for example), many activities asked users to try a new behavior in their everyday lives, then report back on how it went. Users described what they did and how they felt about it in a textbox. Some activities specifically asked users to do the activity on Happify (eg, if a user was supposed to write down 3 good things that happened to them before bed, then Happify asked them to actually enter those three things into the site or app). Others, however, didn't ask for the results of the activity itself (eg, if a user wrote a letter to someone expressing their gratitude, the letter itself would not go into the textbox, just their reflection on the experience of writing and delivering the letter). The text we had from users, therefore, was a mixture of words they used when doing activities and words they used when talking about their experiences with activities.

Well-Being Assessment

Our primary outcome was well-being, which, consistent with current thinking on subjective well-being, we split into two components: positive emotion and satisfaction with life [28]. Due to the proprietary nature of Happify, a new measure (the Happify Scale) was developed and validated by the last author (AP) to measure well-being in users. Users were prompted to take the well-being questionnaire the day after registering, and again every 2 weeks after that.

The positive emotion subscale of the Happify Scale was developed based on the Positive and Negative Affect Schedule [29], which is a self-report survey that measures the extent to which a person has experienced a variety of both activated (high arousal) and deactivated (low arousal) positive and negative emotions. We shortened the survey for practical purposes to 4 sets of emotions: (1) joyous, exuberant, inspired, and awestruck, (2) serene, grateful, and relaxed, (3) sad, guilty, and lonely, and (4) angry, anxious, and afraid. For example, a user would be asked "In the past month, how often have you felt joyous, exuberant, inspired, or awestruck?" We added the 2 positively valenced items to the 2 negatively valenced items, reversed scored, to generate the positive emotion measure. In an internal validation study on a sample of 559 participants recruited via Amazon's Mechanical Turk (Amazon.com, Inc, Seattle, WA, USA), the positive emotion subscale had acceptable internal consistency (alpha = .72) and was strongly positively correlated with the Positive and Negative Affect Schedule (r=.76, P<.001).

We modeled the life satisfaction subscale of the Happify Scale after the Satisfaction With Life Scale [30] but adjusted it to ask users about satisfaction with different domains of their life, including work, leisure, and relationships. For example, a user would be asked "How satisfied do you feel with the relationships in your life?" We computed this score as a simple sum. In the previously mentioned internal validation study, the life satisfaction subscale had acceptable internal consistency (alpha=.88) and was strongly correlated with the Satisfaction With Life Scale (r=.80, P<.001).

Study 1

Previous research has found that continued practice of happiness activities leads to better outcomes among happiness seekers compared with those who do not practice on a regular basis [16]. Therefore, we hypothesized that usage is tied to improvement, such that users experience higher well-being scores during periods when they use the site more and lower well-being scores during periods when they use the site less. We chose to focus our analysis on the positive emotion subscale of the Happify Scale, as previous research has shown that life satisfaction is relatively stable in the short term, while positive emotion is more susceptible to change [30,31]. Consistent with previous research, we also expected to see greater improvement in users with lower well-being to start.

Study 1 Methods

Participants

The sample comprised 152,747 users who completed at least two well-being assessments. Users were asked to complete a well-being assessment the day after registration, but some users did not return after their first visit and therefore were never offered the assessment. Others chose not to complete the assessment but kept using the site. Therefore, the sample

contained users who were moderately interested in the platform and who were interested in tracking their own well-being.

To examine the possibility of sample bias as a result of excluding participants with <2 assessments, we compared those who completed ≥2 assessments (n=152,747) with those who completed only 1 (n=568,205) on the measure of positive emotion. Table 1 shows the results. There was no statistically significant difference between users who completed 1 assessment and users who completed ≥2, and the effect size for the difference between the means was very small per Cohen *d* guidelines [32]. However, when it came to positive emotion, we did not have the ability to compare users who completed any number of assessments with users who completed none. It is still possible (perhaps even likely) that our sample was not representative of the overall user base in terms of well-being levels.

We also compared these 2 groups on demographic variables using chi-square tests and, for the most part, found statistically significant but practically small differences (see Table 2). Specifically, compared with users who did not complete ≥2 assessments, our sample had more women, fewer people aged 18–24 years and more people aged 35–44 years and 45–54 years, fewer students and more people who were employed, more people without children, and fewer people with children ≥19 years old and with children of different ages. However, most of these differences are in the 1% to 3% range, and were likely significant only because of the very large sample size. The only substantial, and possibly quite important, difference was in the age of the ≥2 assessment sample. Users who completed ≥2 assessments were significantly older than the overall user base, with 6% fewer people in the 18- to 24-year age range.

Table 1. Baseline differences on positive emotion measure between the study 1 Happify user sample (completed ≥2 assessments) and users who did not complete ≥2 assessments.

Number of assessments	No.	Mean score[a]	SD	*t*	*df*	*P* value	*d*
1	568,205	38.75	19.80	3.39	720950	.99	.00
≥2	152,747	38.56	19.38				

[a]Scored on a scale of 1–100 in the Happify Scale.

Table 2. Differences in demographic variables between the study sample (≥2 assessments, n=1,925,376) and those not included in the analysis for study 1 (1 assessment only, n=152,747).

Characteristics	1 Assessment, % (n)	≥2 Assessments, % (n)	χ^2	Cramer *V*	*df*	*P* value
Sex			1371.56	.05	2	<.001
Male	13% (19,857.11)	10% (192,537.60)				
Female	87% (132,889.89)	90% (1,732,838.40)				
Age range (years)			4075.98	.07	5	<.001
18–24	20% (30,549.40)	13.9% (267,627.26)				
25–34	30% (45,824.10)	30% (577,612.80)				
35–44	24% (36,659.28)	28% (539,105.28)				
45–54	17% (25,966.99)	19% (365,821.44)				
55–64	8% (12,219.76)	8% (154,030.08)				
≥65	1.5% (2291.21)	1.5% (28,880.64)				
Employment status			1804.80	.12	5	<.001
Retired	3% (4,582.41)	3% (57,761.28)				
Self-employed	12% (18,329.64)	12% (23,1045.12)				
Unemployed	6% (9,164.82)	6% (115,522.56)				
Student	14% (21,384.58)	11% (21,1791.36)				
Employed	57% (87,065.79)	62% (1,193,733.12)				
Homemaker	7% (10,692.29)	7% (134,776.32)				
Parental status			1714.74	.05	5	<.001
Children ≥19 years	7% (10,692.29)	5.4% (103,970.30)				
Children 13–18 years	2% (3,054.94)	2% (38,507.52)				
Children 0–12 years	5% (7,637.35)	5% (96,268.80)				
Children of different ages	5% (7,637.35)	4% (77,015.04)				
No children	15% (22,912.05)	18% (346,567.68)				

In summary, when considering to whom this research is generalizable, it is important to remember that the subsample we drew from was biased in one key way that may limit generalizability: our participants were older than the overall Happify user base. Furthermore, our sample may be biased when it comes to users' well-being levels; the data available suggested not, but we did not have data for users completing no assessments. Based on previous research, it is likely that those users were different from our sample in some way.

Baseline Well-Being as a Moderator

On a scale of 0–100, the average positive emotion score among study 1 users at baseline was 39.03 (SD 19.45) and average life satisfaction was 52.00 (SD 22.78). However, previous research suggests that there are two distinct types of happiness seekers: those who are relatively distressed and those who are relatively nondistressed [19]. Other work replicates this 2-cluster structure, typically derived from a positive emotion measure, a life satisfaction measure, and a depression measure, and suggests that these different groups may respond differently to happiness interventions [17]. Specifically, some evidence suggests that happiness seekers who are more distressed may experience greater benefit [18]. Therefore, following a similar procedure used in previous work clustering happiness seekers, we performed a 2-step cluster analysis in IBM SPSS (version 19, IBM Corporation) using baseline positive emotion and life satisfaction scores to sort participants in our sample. Previous research has found this approach to be robust for use in large datasets [33,34]. Although we did not have a measure of depressive symptoms, we hypothesized that using 2 of the 3 measures used previously would still yield a 2-cluster pattern of division, with one group showing overall higher than average well-being and the other showing overall lower than average well-being.

Even without a measure of depressive symptoms, we found the expected cluster structure. As anticipated, the model yielded two distinct types of users in the sample: low well-being (n=69,474), whose mean scores for positive emotion (23.64) and life satisfaction (32.63) were lower than the sample average; and high well-being (n=83,273), whose mean scores for positive emotion (51.05) and life satisfaction (68.01) were higher than the sample average. Multimedia Appendix 1 displays the silhouette diagram for the model [35], which describes the model fit as good. We used this clustered variable as a moderator in

our effectiveness analysis to see whether Happify affected distressed users differently from nondistressed users.

Analytic Strategy

We analyzed data in IBM SPSS using a multilevel modeling procedure originally designed for use in diary data, where multiple assessments on both the independent and dependent variables are taken for each individual participant [13]. Multilevel modeling is an advanced form of linear regression that is ideal for assessing longitudinal data because it is able to use however few or many assessment points a user has provided (in other words, it does not discard users with missing data and instead plots a line for them using whatever data they gave).

In this particular variation of multilevel modeling [36], there is—like any multilevel model for an intervention—a main effect for time, which shows how well-being changed over the course of the 8-week intervention period. We included time in the model because many intervention studies explore psychological changes in this way: incrementally, in sequence, over time. Furthermore, including time is important because it controls for the possibility that the relationship between usage and outcome was due only to the passage of time [36]. However, since a main effect of time was uncontrolled, and is subject to criticisms such as regression to the mean and confoundedness with usage and dropout rate, we were especially interested in analyses that look at well-being as a function of usage. When analyzing usage, we separated within-person and between-persons variation, and we assessed both between- and within-person terms separately in the model. The model, then, yielded estimates for time, within-person variation (ie, whether a person's well-being varied was a function of their using Happify more or less), and between-persons variation (ie, whether users with an overall pattern of high usage differed from users with an overall pattern of low usage) with well-being as the dependent variable. Therefore, analyses involving usage (within or between) focused on the short-term impact of usage—during each individual 2-week window—rather than taking place over the entire course of the 8-week intervention period.

For between-persons terms, we compared people who tended to use the app more with people who tended to use the app less. Usage is continuous, not categorical, in this term. A significant between-persons term would indicate that an overall pattern of higher usage predicts higher well-being. Between-persons differences in overall usage patterns were not of primary interest to us but, rather, were important potential confounding variables to control for as we explored the impact of within-person variation. For the within-person terms, a dose-response line could be calculated for each individual participant, not over time, but over levels of usage. It examined the relationship between usage and mood for that individual. Analysis of within-group effects provided a way to see how a person did at different "dosages," and the focus was on their change within-person, at these different dosages, rather than on differences between people who got one dose or another. Such an analysis is not as vulnerable as a between-persons comparison based on usage (splitting by heavy users vs light users) would be, as there would likely be systematic differences between the people in the 2 groups. Because a participant was being compared with him- or herself, concerns about differences between users became less salient.

In smaller subsamples, the participants' individual lines could be visualized in a "spaghetti plot," which shows individual differences in the role of usage on well-being. Spaghetti plots are a useful adjunct to a statistical model, as they help to visualize the extent to which the model's overall slope is representative of the slope of each individual in the sample. More simply, they indicate whether the overall slope is representative of what is happening in the sample. The analysis set well-being (positive emotion and life satisfaction were examined separately) as the dependent variable and usage as the predictor.

All participants had at least two well-being assessments, but some had up to 5 assessments, spanning 8 weeks. The 8-week assessment period began from the time users completed their first well-being assessment. For any time that well-being was assessed, we calculated a "usage" variable, which we operationalized as the number of visits to the site in which the user completed an activity that took place between the last assessment and the current one. The average number of activities completed prior to the first assessment was 5 (SD 5.11). Additionally, we used a baseline clustering for well-being as a moderator (see above). We tested the predictive power of usage, as well as the potential interaction of baseline well-being with usage, with between- and within-persons variation separated. The analysis, therefore, yielded traditional estimates of how much the average person improved (a between-persons approach), but also looked at how usage variation for each individual predicted his or her well-being (a within-person approach), generatin a dose-response relationship. In short, it asked "For any given person, how was their well-being during time periods that they used Happify more, and during time periods where they used Happify less?"

Study 1 Results

Table 3 contains the descriptive statistics for the positive emotion measure, as well as for usage, across each of the time points. Positive emotion increased by 10.47 points (10.47%) over the course of the 8-week study period. Usage, which started relatively high (about 5 visits per week), decreased over time and by 8 weeks was between 0 and 1. The sample size for positive emotion changed over time as users dropped out, but because usage was observed rather than self-reported, data on usage were available for everyone in the sample, regardless of compliance with the well-being assessment.

Table 3. Positive emotion[a] and usage among a sample of Happify users over the course of 8 weeks.

Time point	No.	Mean	SD
Positive emotion scores			
Baseline	152,747	38.56	19.38
2 weeks	148,740	42.46	19.68
4 weeks	52,177	45.29	19.80
6 weeks	25,435	47.46	19.73
8 weeks	15,140	49.03	19.63
Usage (visits/week)			
Baseline	152,747	5.19	5.11
2 weeks	152,747	4.39	11.40
4 weeks	152,747	2.06	8.26
6 weeks	152,747	1.25	6.45
8 weeks	152,747	0.85	5.15

[a]Scored on a scale of 1–100 in the Happify Scale.

There was a main effect for time that echoed the observed mean increases in positive emotion. As time passed, positive emotion improved at a rate of about 1.38 points per week for the average user (estimate 1.38, SE 0.01, $t_{122,455}$=113.60, P<.001, 95% CI 1.36–1.41). This suggests an average overall improvement of 11.04 points, or about 27%, over the course of 8 weeks.

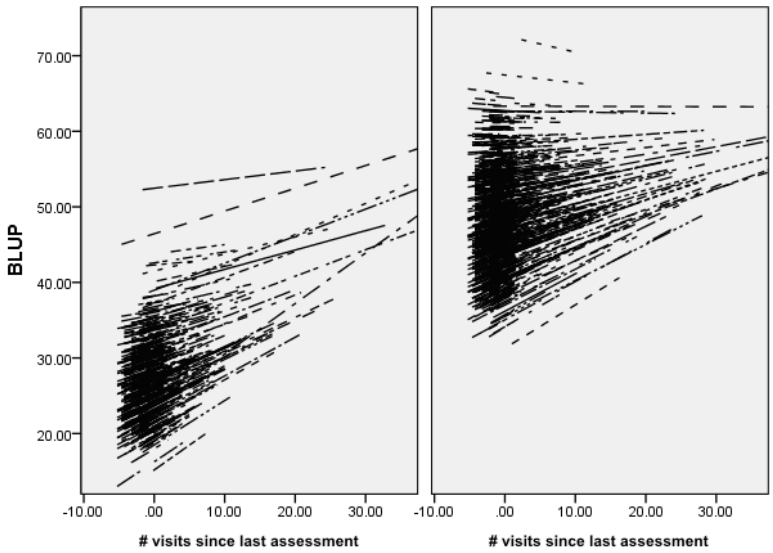
There was also a significant impact of usage on overall well-being compared with users who did not use the platform as often. On average, high-usage users experienced more positive emotion (estimate 0.20, SE 0.02, $t_{85,929}$=11.63, P<.001, 95% CI 0.17–0.23). The significance of this term suggests the importance of including it as a control variable. However, due to the way usage was measured (segmented into 2-week chunks, rather than being cumulative), the estimate yielded by this term is not meaningful or interpretable for practical purposes.

The results for the within-person terms revealed that for any given user, more usage predicted more positive emotion and less usage predicted less positive emotion (estimate 0.09, SE 0.01, t_{6047}=9.15, P=.001, 95% CI 0.07–0.12). This estimate predicted that a given user would report positive emotion that would be 1.26 points higher after a 2-week period when they used Happify daily than after a week when they didn't use it at all. On a week-to-week basis, it seems that users got more out of the site during weeks when they used it often, and less when they used it rarely.

There was also a significant between-persons usage × baseline well-being interaction such that people with lower baseline well-being and who used the site more often experienced the most improvement compared with high well-being users, or low well-being users who did not use the site often (estimate 0.20, SE 0.02, $t_{85,929}$=18.60, P<.001, 95% CI 0.18–0.22). Both baseline well-being level and usage seemed to interact to determine improvement. Although, for reasons described above, the estimate yielded here is not meaningful or interpretable, the results can be seen in Figure 1, which shows spaghetti plots for low well-being users (left) and high well-being users (right). Due to computer memory limitations in rendering the graph, we created the plot on a randomly selected subset of the data (n=1505). There were no statistically significant differences from the overall sample on baseline well-being or demographic variables, and the pattern of results observed in the sample was the same. The x-axis is usage (number of visits since the last assessment), grand mean centered. The y-axis is a best linear unbiased prediction [37], which is intended to capture random effects for positive emotion. Each line on the plot represents the dose-response curve for an individual user; longer lines indicate users with greater variability in usage. The overall positive relationship between usage and well-being is visible in both plots. Furthermore, the graphs indicate that, although more usage seems to lead to higher well-being for users starting with low well-being, the benefits of usage on well-being seem to level off at higher usage levels for some users in the group with higher starting well-being.

There was not a significant within-person usage × baseline well-being interaction. Low well-being or high well-being participants did not show differing levels of sensitivity to usage (P=.28). The relationship between usage and well-being was the same regardless of a person's baseline well-being.

Figure 1. Spaghetti plot illustrating the impact of usage (number of visits since last assessment, grand mean centered) on positive emotion for Happify users with low baseline well-being (left) and high baseline well-being (right). Positive emotion is represented with best linear unbiased prediction (BLUP) scores. Illustrated using a randomly selected subsample of n=1505.



Study 1 Discussion

The results suggest overall improvement across users of Happify regardless of baseline well-being, with higher usage generally associated with higher well-being across the board. The average person using the site improved, with more improvement the more they used it. However, that improvement seems to be greater among those who started with lower well-being, especially on the measure of positive emotion. Thus, users with more room to grow due to their low well-being experienced greater changes in well-being when using the site than did their high well-being counterparts.

It is important to consider some limitations to the study design, and to temper interpretation of the findings accordingly. The first is the bias we observed in our sample. Users who completed ≥2 questionnaires, a requirement for inclusion, were substantially older and may have otherwise been different in ways we could not assess from users who completed only 1 questionnaire. While we did not know the well-being level of the great majority of users who were excluded from the sample (they did not complete even 1 well-being measure), other studies have suggested that dropouts may be biased toward lower well-being, and we also observed bias toward an older demographic. We also observed substantial drop off over the course of the study, with only 10% of users completing all assessments; while these dropout rates are typical in naturalistic Internet intervention, they also press the limits of what regression can reliably do to accommodate missing data. We also acknowledge that our operationalization of usage—the number of visits to the site in which an activity was completed—is only one of many ways to quantify usage. For the purposes of our analysis, it was important to quantify usage every 2 weeks so that each 2-week period could be analyzed separately. However, in an analysis that looked at change over time continuously, usage could be conceptualized cumulatively. For example, a user could have an ongoing tally of "number of activities completed" that grows over time. This type of approach would be beneficial for looking at total usage, something we were unable to do in this study.

Nevertheless, we were able to observe these effects outside of a controlled research setting, which was one of our goals, since control groups are often impractical in large, naturalistic data sets. The use of multilevel models offered us a way to sidestep some (but not all) problems with an uncontrolled design. Within-person terms of the model (focusing on whether usage at different time points led to changes in well-being for any given person) are less vulnerable to traditional criticisms of an uncontrolled study. However, between-persons terms (finding that some users do better than others) are still vulnerable to concerns about sample bias. This approach to evaluating effectiveness, while not at all the same as conducting an RCT, may allow researchers to get an estimate of effectiveness when an RCT is not possible.

Study 2

Study 1 provided evidence that users who interacted with the site more often also reported higher levels of well-being. However, it did not suggest the *ways* of engaging with the site that were associated with improvement. An important means of addressing this question is by analyzing the language expressed by users as they complete activities. By using big data techniques to determine the topics associated with user improvement, we can provide insight to users' spontaneous behavior; that is, what behaviors are beneficial when users are free to engage with an intervention however they wish. Because our methods were data driven, they allowed the data to speak for themselves. We had no specific, pre hoc hypotheses about our results. Our goals were to develop insight into the kinds of spontaneous engagement that are associated with successful outcomes in large, open Internet well-being interventions such as Happify. Looking only at users who were engaged to begin with, we isolated the language associated with improvement, giving insight into the maximally effective ways to be engaged.

Study 2 Methods

Materials

For all linguistic analyses, our dependent variables of interest were the same pair of self-report surveys used in study 1. We derived predictor variables from task-based text that participants wrote over the course of their time (see the general Methods section above for more description of the activities on the site) using natural language processing methods described below.

Participants

To construct stable outcome variables based on the well-being scales, we limited linguistic analysis to users who completed the scales at least twice and who had a span of at least 30 days between first and last self-report. Table 4 lists differences between those who met these criteria and those who did not. Note that demographic information was not available for all participants with ≥500 words, so the users analyzed here are only a subset of those used in our analysis. In addition, because large amounts of an individual's language are required for reliable language analysis [38], we only considered users who wrote at least 500 words across all free-text tasks. This left us with a final sample of 10,818 users. Participants in the sample used Happify over an average of 168 days, and they wrote an average of 51.23 words for each task they completed.

Table 4. Differences in demographic variables between the sample (<500 words, n=2,073,333) and those not included in the analysis for study 2 (≥500 words, n=4790).

Characteristics	<500 words, % (n)	≥500 words % (n)	χ^2	Cramer *V*	*df*	*P* value
Sex			169.19	.02	2	<.001
Male	12% (248799.96)	9% (431.10)				
Female	87% (1803799.71)	90% (4311.00)				
Age range (years)			381.55	.04	6	<.001
18–24	19% (393933.27)	19% (910.10)				
25–34	30% (621,999.90)	37% (1,772.30)				
35–44	24% (497,599.92)	24% (1,149.60)				
45–54	17% (352,466.61)	14% (670.60)				
55–64	8% (165,866.64)	6% (287.40)				
≥65	2% (41,466.66)	1% (47.90)				
Employment status			155.44	.03	5	<.001
Retired	3% (62,199.99)	2% (95.80)				
Self-employed	12% (248,799.96)	12% (574.80)				
Unemployed	6% (124,399.98)	8% (383.20)				
Student	14% (290,266.62)	16% (766.40)				
Employed	58% (1,202,533.14)	52% (2,490.80)				
Homemaker	7% (145,133.31)	7% (335.30)				
Parental status			8951.68	.03	5	<.001
Children ≥19 years	7% (145,133.31)	7% (335.30)				
Children 13–18 years	2% (41,466.66)	3% (143.70)				
Children 0–12 years	5% (103,666.65)	10% (479.00)				
Children of different ages	5% (103,666.65)	5% (239.50)				
No children	15% (310,999.95)	45% (2,155.50)				

Table 5 shows descriptive statistics for the sample, as well as an analysis of baseline differences between users in our sample and the remaining users in the user base. Users in the study 2 sample had significantly higher well-being at their first assessment than did users who did not write ≥500 words. Not surprisingly, this subsample is not a random subset of the overall user base—a very specific set of users engaged with the site frequently enough to yield the amount of text needed by this analysis.

Table 5. Baseline characteristics for the study 2 sample on both dependent variables, and analysis of the difference between users in the study sample (who wrote ≥500 words) and users who wrote <500 words.

Dependent variables	No.	Mean score[a]	SD	*t*	*df*	*P* value	*d*
Positive emotion							
<500 words	710,348	39.43	19.97	–47.41	721,164	<.001	.46
≥500 words	10,818	48.63	20.11				
Life satisfaction							
<500 words	710,348	52.31	23.36	–44.39	721,164	<.001	.42
≥500 words	10,818	62.39	24.26				

[a]Scored on a scale of 1–100 in the Happify Scale.

This sample was clearly not randomly taken from the population of Happify users, and is therefore likely not representative of the whole user base. However, focusing specifically on highly active users let us explore the *kind* of activity associated with improvement. Even among highly engaged users, there was variance in improvement experienced. Therefore, simply using an intervention often may not be enough for it to be effective. There are likely specific behaviors and psychological orientations associated with improvement. Our linguistic analysis used data-driven techniques to reveal these factors.

Procedures

Text Preprocessing

Some Happify tasks contained multiple text fields. We combined all text for a given task instance into a single document for topic modeling. Hard returns were replaced with a "<newline>" placeholder. Tokenization, feature extraction (other than topic modeling), regression, and correlation analysis were performed in Python version 2.7 (Python Software Foundation).

Topic Modeling

We clustered users' free text using a latent Dirichlet allocation (LDA), a topic modeling technique [39]. The LDA technique assumes that documents (in our case, text from a single task instance) comprises a combination of topics, and that each topic is a cluster of words. Using the words found in each document, the makeup of each topic is estimated using Gibbs sampling [40,41]. We used the Mallet implementation of the LDA algorithm [40] to produce 200 topics, adjusting the alpha parameter (alpha=5) to tune for fewer topics per document due to each document's short length compared with typical applications of LDA (eg, encyclopedia or news articles). Previous work by members of our research group [42] used a larger number of topics, 2000, but the task-directed language in the study 2 data set led to reduced variability in language, able to be captured with a smaller number of topics. To use topics as features, we calculated the probability of a user's topic usage, p(topic | user), using LDA outputs and user word probabilities (see [41] for details). This gives a 200-dimensional vector representing the language of each user, where each dimension maps onto a discrete topic word cluster.

Factor Analysis of Topics

LDA produces topics that can be strongly intercorrelated over users [43,44]; that is, if topics are correlated, when users write about certain topics, they tend to also write about other specific topics that are related to them. For example, a user who writes about cooking is more likely to also write about dessert than about schoolwork, even though "cooking" and "dessert" constitute 2 separate topics.

Multimedia Appendix 2 shows the intertopic correlation matrix and scree plot for the 200 topics. While most topic correlations were weak or very weak, 41 topic pairs exhibited moderate correlations, and 6 topic pairs exhibited strong or very strong correlations. To account for topic correlation, we ran exploratory factor analysis on topic scores using varimax rotation. Based on the scree plot, we decided to use a 50-factor solution that exhibited a close fit (root mean square error of approximation of 0.01807) explaining 32.5% of overall variation. The resulting factors may be viewed as patterns of user behavior.

In summary, the LDA topics clustered together aspects of messages that likely had some similarity. The factor analysis clustered together topics that tended to co-occur within the same user. This difference allowed us to answer similar but distinct questions: (1) What do people write about that is related to effective engagement with the intervention? (2) What types of general user behavior are related to effective engagement with the intervention?

Outcome Variables

We constructed ordinary least squares regressions with time (days since first response) as the independent variable and an overall well-being measure constructed by aggregating the positive emotion score and the overall life satisfaction score as a dependent variable. Our outcome for each participant was the ordinary least squares regression slopes of users' total score (positive emotion plus life satisfaction) change over time. Positive emotion and life satisfaction subscales were correlated at r=.61, and analyzing them separately did not produce meaningfully different results (see Multimedia Appendix 3). Therefore, for simplicity, we combined these variables into a single score of subjective well-being [45].

Correlation Analysis

Using the technique of differential language analysis [9], we correlated users' well-being change over time with language rates on two levels: LDA topics and user-level factors.

Because this analysis involves multiple, independent tests, which increases the possibility of type I error, we applied

Benjamini-Hochberg false discovery rate correction to our results, which adjusts *P* values based on the number of tests run [46].

A positive correlation indicated that the language feature was associated with improvement in well-being. If a particular topic positively correlated with the outcome, then it meant that writing more often about that topic was associated with higher levels of improvement in well-being over time. Because of our specific interest in users' improvement, we only considered positive correlations between language features and outcomes.

Study 2 Results

Topics

A total of 14 topics significantly predicted increased well-being. The topics with the strongest relationship tended to be about directly engaging with negative thoughts and emotions, but some also included descriptions of positive experiences. Figure 2 displays examples of each kind. All topics with a significant effect are included in Multimedia Appendix 4.

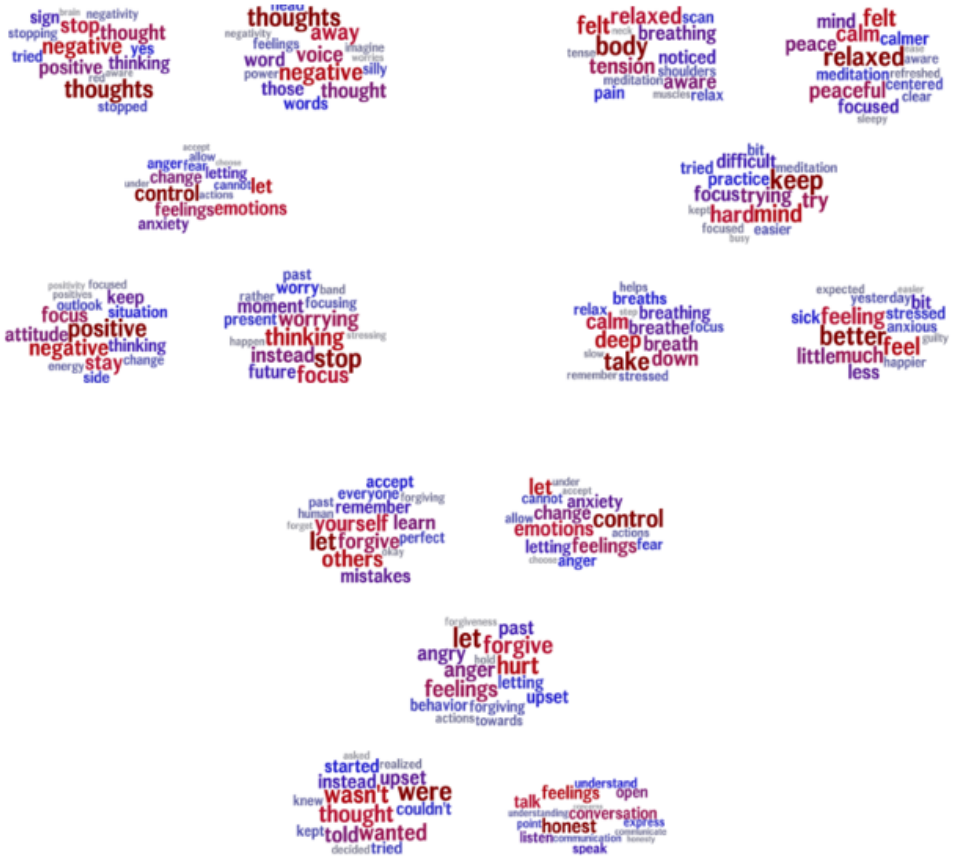
Figure 2. Example topics predicting increased well-being.

We reran the topic analysis only for users with initially high well-being to address the possibility that the results were influenced by a ceiling effect. The results are similar to those of the overall sample (see Multimedia Appendix 5).

Factors

A total of 3 factors significantly predicted increased well-being, and they also primarily centered on engaging actively with negative experiences. For ease of interpretability, we labelled these factors with 3 distinct aspects of this process: (1) restructuring negative thoughts, (2) controlling anxiety, and (3) coming to terms with interpersonal strife. These labels, while open to interpretation and derived post hoc (not from the language analysis), support the general pattern of an actively positive orientation toward life difficulty. Figure 3 illustrates the 5 highest-loading topics for these 3 factors. These factors describe general user behavior, providing a more contextualized view of the language associated with successful engagement.

Figure 3. The 5 highest-loading topics within the 3 factors predicting improved well-being. The top left factor contains topics about negative thoughts (restructuring negative thoughts). The top right factor contains topics about dealing with anxiety (controlling anxiety). The bottom factor contains topics about past experiences, and conflicts and interactions with other people (coming to terms with interpersonal strife).



Study 2 Discussion

Automatic language analysis revealed the topics and factors associated with successful engagement with Internet well-being interventions. Of note is that increased positive affect and life satisfaction were associated with negatively valenced topics containing such words as “negative,” “anxiety,” and “worrying.” Among those who were already highly engaged, the most successful users addressed their unpleasant thoughts, anxieties, and difficulties. This result is consistent with previous, experimentally based theories. Cognitive therapy techniques focus on intervening in maladaptive thought patterns [47]. Also, previous research has shown health and well-being benefits specifically from writing about negative experiences [48,49].

One reason why users may have found themselves discussing negative content was the structure of Happify. As discussed above, users’ experiences on Happify are framed by tracks, or sets of activities that are built around a common goal or theme. The overwhelming majority of the user base chose problem-focused tracks such as “conquering negative thoughts” and “cope better with stress.” Therefore, while the tasks they did were quite varied and were generally focused on positive emotion, they also referred to negative aspects of the user’s life. In that sense, we may also have been detecting adherence and compliance wherein users who stayed on point with the goals of the track by discussing negative content benefitted the most.

Successful users also focused on a variety of positive emotional experiences. Notably, many of these experiences referred specifically to positive emotions, such as pride, enjoyment, appreciation, and celebration, which are reactions to specific objects or events rather than general moods.

In general, users who increased their well-being wrote frankly about their unpleasant or unadaptive thoughts and feelings, addressing these negative experiences with direct strategy. However, they did not write solely about unpleasant things, and therefore they balanced both positive and negative topics. These qualitatively induced conclusions are based on real-world data and capture important trends in people’s spontaneous attempts to increase their own well-being. In other words, even among users who engaged strongly with the interventions, there was variance in how much benefit they received. Users who put in the effort of working through their unpleasant emotions and thoughts had the greatest improvement in well-being. Probably due to the relative sparseness of language for each user, correlations between these topics and outcomes were rather weak (see Multimedia Appendixes for details). This fact reveals another important feature of big data methods: they allow researchers to uncover subtle effects such as these that would have been hidden in smaller samples. Though small, these results help explain well-being improvement across a large number of people in an uncontrolled setting, and therefore potentially have dramatic real-world impact. Furthermore, the open nature of these platforms means that data will continue to accrue, allowing increasingly powerful and nuanced language analyses.

Discussion

The goal of this study was to showcase the advantages of analyzing intervention data in a massive but uncontrolled dataset. Our goal was to try two novel methods for making sense of the chaos in uncontrolled and potentially overwhelming intervention data. In study 1, consistent with hypotheses and with previous research [16,18], we found that the average Happify user improved in well-being by about 11% over the course of 8 weeks. However, without a control group, it is difficult to infer that the intervention, and not the passage of time or other factors, led to the observed improvement. Thus, we also explored within-person dose-response relationships and found that users reported higher well-being during periods of time when they used Happify more frequently. While this type of design is weaker than its more controlled counterparts when inferring overall effectiveness, it does allow us to explore what happens to users while they use an intervention at a finer level of granularity, week by week, as their usage levels vary. It may also help us to determine what level of usage is optimal or whether, at some point, more usage stops being better.

The within-person analysis with usage predicting well-being week to week is interesting in that it looked not only at differences between users, but also differences within users at different dosages of the intervention. While this type of approach is typical in diary and experience sampling research, it is seen less frequently in intervention research. Baseline well-being was a moderator, such that users with lower starting well-being improved more. This is also consistent with previous work [17,18]. Using natural language analysis, a procedure only possible in large datasets, study 2 yielded a snapshot of what processes may be going on when users successfully practice activities on Happify. Rather than focusing solely on the positive, users appeared to be using happiness-focused activities as a way of working through problems. In some sense, this is surprising, because the activities in the STAGE model are all focused on positive experiences and cognitions. However, there is evidence that writing about the negative can help individuals create a meaningful narrative, which may also be happening [48].

When evaluating effectiveness, many researchers may avoid these kinds of data because the lack of a control group makes it difficult to establish the effectiveness of an intervention. To address this problem, we also examined within-person variation, watching how each user's well-being varied as a function of their usage frequency. Control groups typically exist to account for systematic differences between participants in each group, as well as natural change over time. A within-person analysis is less subject to these concerns, since each data point comes from the same person. Using this approach, our results indicated that, for most users, usage and well-being went hand in hand, with greater usage during a given 2-week period predicting higher well-being at the end of that period. We were also able to include the effects of individual differences by visualizing these data using spaghetti plots. The plots indicated that, although a small number of outliers existed, the positive relationship between usage and well-being held true for most of the sample.

Another hesitation that many researchers might have is the inability to explore theoretical mechanisms. However, we were able to create insight into mechanisms with the use of linguistic analysis to identify topics that were more likely to be used by users whose well-being improved. Discussion of negative topics, especially negative thoughts, was associated with better outcomes. Users acknowledging and engaging with their anxieties and unadaptive cognitive patterns reaped benefits. While these results are ultimately just correlational, they provide a descriptive snapshot of what distinguishes effective use of the intervention from less effective use. Furthermore, as research converges on the types of language use that are most strongly related to improvement, this kind of fine-grained data can provide intervention designers with feedback on how to present an activity. For example, if focusing on negative topics continues to emerge as a common factor among people who benefit from Happify, then Happify would do well to revise activity instructions to encourage users to focus on negative topics.

The nature of the dataset also carried specific advantages over more structured intervention research study designs. Most apparent is the large sample size, which was several orders of magnitude greater than all but the most ambitious randomized studies. This large sample size allowed us to generate parameters with a high amount of confidence, including the standard errors generally in the 0.01–0.03 range, with tight 95% confidence intervals. It also allowed us to do a moderator analysis that examined differences between people whose baseline well-being was high or low, without hindering our ability to detect effects. Often, moderator analyses are reserved for meta-analyses, where the data of multiple studies are available. Also, automatic language analysis requires large amounts of data to provide valid results, rendering it an impossible tool in most controlled intervention studies. Furthermore, the very nature of these datasets means that they are continually expanding. As the amount of data increases, so too does the ability of researchers to measure and isolate nuanced effects.

The dearth of experimental control over variables within our data, while in some ways a liability, also provided access to important research questions. In the real world, users are free to engage with interventions whenever and however they wish; these variables are therefore important to consider in studying general effectiveness. Our data describe participants' spontaneous behavior; that is, their levels of engagement in the midst of their everyday lives and their freely chosen strategies when using the intervention. We therefore provide results that are not only externally valid but also engage with variables that would be "controlled away" in many other designs.

Many authors have noted the need for new methodologies to accommodate the rapidly changing landscape of Internet intervention research [2,3]. We modeled the use of some alternative strategies that yield valid information about outcome and process in Internet interventions. Thus far, we have focused on the benefits of using real-world intervention data. It is, however, worthwhile to discuss some limitations as well.

Limitations

Real-world interventions are constantly changing by releasing new content, revamping existing features, and providing

interactions with other users via discussion forums, and that presents a challenge to researchers who are trained to keep as many factors as constant as possible [50]. Furthermore, the effectiveness evaluation approach we used relies on within-person variation in usage, which is not as clean an independent variable as randomly assigned group membership. Users are self-selected on several levels from their decision to try Happify at all, and even more so in their continued use. The nature of our analytic strategy—using usage, text input, and well-being data—restricted us to users who provided that type of data in sufficient quantity. For study 1, we were limited to users who completed well-being assessments. For study 2, we were limited to users who wrote ≥500 words. For both studies, we had demographic data only for users who reported them. In all of these cases, many, many users were eliminated from our sample due to insufficient data. We can't analyze data that we don't have and, as a result, our samples were biased. We took some measures to explore the nature of that bias, but missing data makes it difficult to be sure that we understand every potential way that the data are not representative of the overall user base.

While temporal precedence of usage behavior compared with well-being scores allowed for some measure of ability to infer causality, there were also potential third variables that could have influenced both usage and well-being. One major concern was the role of dropouts in our findings. Dropouts are common, and dropout rates are high, even in controlled research studies where some incentive for participation may be provided [51]. In a consumer environment, where users are customers and it is a buyer's market, the frequency of dropout becomes even greater [51]. This was apparent in our sample, where the rate of attrition mirrors that observed in other published analyses of uncontrolled intervention datasets. While high attrition rates are certainly cause for concern and caution when interpreting findings, they do not render a study without value [51]. Nevertheless, it is important to consider potential sources of bias that may be introduced with so large a percentage of the sample missing and, as much as possible, to account for those missing data using appropriate statistical approaches.

We made some attempts to uncover any possible differences between people who did and those who did not engage in Happify enough to meet our inclusion criteria, and we did find some key differences between users who were highly engaged (completing ≥2 assessments or writing ≥500 words) and those who were not highly engaged. Available demographic data were, however, limited and many other factors that we did not measure may have played a role in dropouts. Therefore, there remain other potential biases in our samples that are difficult to determine. The linguistic analysis has its own limitations in terms of control. Due to the large amounts of language required for analysis, we were unable to observe the specific pattern of changes in language use over time as a result of intervention use.

We also acknowledge that our mode of measuring well-being using two self-report measures, based on the subjective well-being model, is just one of many possible ways to measure well-being. Self-report measures exist for other conceptualizations, including psychological well-being [52] and mental well-being [53]. Beyond that, it is possible to measure emotions on a more day-to-day basis using experience sampling methods, a methodology that is becoming more and more possible as more people begin to carry mobile phones. Furthermore, recent technological advances in wrist-worn technology bring variables such as heart rate variability within the grasp of researchers. There are many possible frontiers for advancing the assessment of well-being in big data research; ours was just a first step.

Another issue with real-world intervention data is a preponderance of statistical power. Study 1 analyzed the data of over 150,000 participants; at that sample size, nearly everything is significant, and statistical significance therefore cannot be relied on as an indication of which findings are "important." The potential for overpowering is just as great in the linguistic analysis, where many effects, while statistically significant, were very small. As linguistic analysis of this kind is relatively new, it is hard to know what constitutes a meaningful effect. Future research is needed to explore the predictive power of word and topic usage on future behavior to clarify how much this particular process matters for well-being.

Finally, data from real-world interventions are problematic in that they have a reputation for being inaccessible to researchers. It is not always of interest to a company to produce scientific research, and it is not always of interest to researchers to have conversations with companies. We would encourage both groups to reach out to one another and forge mutually beneficial relationships. Published research using company data can strengthen the company's legitimacy. Researchers can access massive, free (to them!) datasets containing outcomes that often are not available in more controlled settings, which is, we hope we have convinced the reader, worth pursuing.

Ethics: An Outstanding Issue

Although we aimed in this study to help establish some new and interesting approaches to analyzing big intervention data, there are some issues that our research raises, but does not address. One important topic for further discussion is the ethics of consent in the context of commercial products. Is passive consent by endorsing a user agreement enough, or should there be more active consent procedures? For example, one might imagine implementing a "use my data for research" dialogue box on each user's profile that is opt-in, so data are only analyzed for users who have considered the risks and decided to make their data available. At the same time, doing so would greatly restrict both the sample size and potentially the sample's validity; what systematic differences might exist between users who do not opt into research? Wouldn't a user need to be relatively engaged in the first place in order to opt in to research? And, if so, wouldn't it be impossible to examine differences between users who engage highly and users who do not, since researchers would not have access to the data of those who do not? It seems clear that in a research context, where users would be randomly assigned between conditions and one or more of those conditions may be intended as "inert," informed consent is a necessity. However, more discussion is needed to develop standards for consent in a context that lacks random assignment, where everyone gets the "best" possible product, and where the

purpose of the site is not to collect data or perform an experiment, but to deliver that product to the public.

Conclusions

We have provided some novel insight into the relationship between general usage of a Web- and app-based platform, as well as the use of specific topics, and change in well-being. We have not, of course, provided an exhaustive or comprehensive review of big data methodologies that might be appropriate for Internet interventions. However, we hope we have effectively argued that analytic methods for unstructured intervention data are needed; without them, entire, massive, naturalistic datasets will remain out of reach. We also hope that our initial attempt at exploring such an unstructured dataset will spark others to experiment with these approaches and improve them.

Acknowledgments

This work was partially supported by a grant from the Templeton Religion Trust (ID #TRT0048).

Conflicts of Interest

RZ and AP are equity holders of Happify, which served as the data source for all analyses in this paper. RZ is a full-time employee and AP is a half-time employee of Happify. JC, PC, HAS, LS, and AC have no financial relationship with Happify.

Multimedia Appendix 1

Cluster quality chart for split between distressed and nondistressed users in Happify.

[PDF File (Adobe PDF File), 15KB - jmir_v18i8e241_app1.pdf]

Multimedia Appendix 2

Heat map for intercorrelations among 200 topics and scree plot. These patterns describe largely small intercorrelations, but with a not insubstantial number of moderate, positive correlations, suggesting that factor analysis may reveal meaningful factors.Eigenvalues of the Correlation Matrix of Topic Scores.

[PDF File (Adobe PDF File), 77KB - jmir_v18i8e241_app2.pdf]

Multimedia Appendix 3

The five topics most strongly correlated with positive emotion and life satisfaction.

[PDF File (Adobe PDF File), 956KB - jmir_v18i8e241_app3.pdf]

Multimedia Appendix 4

All topics significantly correlated with well-being improvement and reduction. Topics predicting increased well-being.

[PDF File (Adobe PDF File), 138KB - jmir_v18i8e241_app4.pdf]

Multimedia Appendix 5

Topics predicting increased well-being among users in the top tercile of initial well-being.

[PDF File (Adobe PDF File), 44KB - jmir_v18i8e241_app5.pdf]

References

1. Halevi G. Research Trends: Special Issue on Big Data. Amsterdam: Elsevier; 2012 Sep 30. URL: https://www.researchtrends.com/wp-content/uploads/2012/09/Research_Trends_Issue30.pdf [accessed 2016-06-17] [WebCite Cache ID 6iLACPaoo]
2. Schueller SM, Munoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 03;22(6):478-483. [doi: 10.1177/0963721413495872]
3. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. Am J Prev Med 2007 Aug;33(2):155-161. [doi: 10.1016/j.amepre.2007.04.007] [Medline: 17673104]
4. Munoz RF, Bunge EL, Chen K, Schueller SM, Bravin JI, Shaughnessy EA, et al. Massive open online interventions: a novel model for delivering behavioral-health services worldwide. Clin Psychol Sci 2015 May 13:1-12. [doi: 10.1177/2167702615583840]

5. Thadhani R, Tonelli M. Cohort studies: marching forward. Clin J Am Soc Nephrol 2006 Sep;1(5):1117-1123 [FREE Full text] [doi: 10.2215/CJN.00080106] [Medline: 17699334]
6. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med 2006 May 30;25(10):1768-1797. [doi: 10.1002/sim.2302] [Medline: 16220518]
7. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res 2009 Feb;18(1):7-26. [doi: 10.1177/0962280208092342] [Medline: 18562396]
8. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: Access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
9. Schwartz H, Eichstaedt J, Dziurzynski L, Kern M, Blanco E, Kosinki M, et al. Toward personality insights from language exploration in social media. 2013 Presented at: AAAI Spring Symposium Series: Analyzing Microtext; March 15, 2013; Stanford, CA, USA URL: http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5764/5915 [WebCite Cache ID 6k9bsqzZc]
10. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
11. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
12. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence, review, and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
13. Bolger N, Laurenceau J. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research (Methodology in the Social Sciences). New York, NY: The Guilford Press; 2013.
14. Schueller SM, Parks AC. Disseminating self-help: positive psychology exercises in an online trial. J Med Internet Res 2012 Jun;14(3):e63 [FREE Full text] [doi: 10.2196/jmir.1850] [Medline: 22732765]
15. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005 Aug;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
16. Lyubomirsky S, Dickerhoof R, Boehm JK, Sheldon KM. Becoming happier takes both a will and a proper way: an experimental longitudinal intervention to boost well-being. Emotion 2011 Apr;11(2):391-402 [FREE Full text] [doi: 10.1037/a0022575] [Medline: 21500907]
17. Sergeant S, Mongrain M. Distressed users report a better response to online positive psychology interventions than nondistressed users. Can Psychol 2015;56(3):322-331. [doi: 10.1037/cap0000034]
18. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
19. Parks AC, Della Porta MD, Pierce RS, Zilca R, Lyubomirsky S. Pursuing happiness in everyday life: the characteristics and behaviors of online happiness seekers. Emotion 2012 Dec;12(6):1222-1234. [doi: 10.1037/a0028587] [Medline: 22642345]
20. Aikens KA, Astin J, Pelletier KR, Levanovich K, Baase CM, Park YY, et al. Mindfulness goes to work: impact of an online workplace intervention. J Occup Environ Med 2014 Jul;56(7):721-731. [doi: 10.1097/JOM.0000000000000209] [Medline: 24988100]
21. Niemiec RM. Mindful living: character strengths interventions as pathways for the five mindfulness trainings. Int J Wellbeing 2012 Apr 28;2(1):22-33. [doi: 10.5502/ijw.v2i1.2]
22. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003 Feb;84(2):377-389. [Medline: 12585811]
23. Fagley N. Appreciation uniquely predicts life satisfaction above demographics, the Big 5 personality factors, and gratitude. Pers Individual Differences 2012 Jul;53(1):59-63. [doi: 10.1016/j.paid.2012.02.019]
24. Layous K, Katherine Nelson S, Lyubomirsky S. What is the optimal way to deliver a positive activity intervention? The case of writing about one's best possible selves. J Happiness Stud 2012 May 6;14(2):635-654. [doi: 10.1007/s10902-012-9346-2]
25. Ho MY, Cheung FM, Cheung SF. The role of meaning in life and optimism in promoting well-being. Pers Individual Differences 2010 Apr;48(5):658-663. [doi: 10.1016/j.paid.2010.01.008]
26. Thoits PA, Hewitt LN. Volunteer work and well-being. J Health Soc Behav 2001 Jun;42(2):115-131. [Medline: 11467248]
27. Allen AB, Leary MR. Self-compassion, stress, and coping. Soc Personal Psychol Compass 2010 Feb 1;4(2):107-118 [FREE Full text] [doi: 10.1111/j.1751-9004.2009.00246.x] [Medline: 20686629]
28. Diener E, Diener C. Most people are happy. Psychol Sci 1996 May;7(3):181-185. [doi: 10.1111/j.1467-9280.1996.tb00354.x]
29. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [Medline: 3397865]

30. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
31. Crawford JR, Henry JD. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004 Sep;43(3):245-265. [doi: 10.1348/0144665031752934] [Medline: 15333231]
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Routledge; 1988.
33. Chiu T, Fang D, Chen J. A robust scalable clustering algorithm for mixed type attributes in large database environment. 2001 Presented at: Seventh ACM SIGKDD international conference on Knowledge discovery and data mining; August 26-29, 2001; San Francisco, CA, USA. [doi: 10.1145/502512.502549]
34. Okazaki S. What do we know about mobile Internet adopters? A cluster analysis. Inform Manage 2006 Mar;43(2):127-141. [doi: 10.1016/j.im.2005.05.001]
35. Norusis M. IBM SPSS Statistics 19 Statistical Procedures Companion. 1st edition. Upper Saddle River, NJ: Prentice Hall; 2012.
36. Bolger N, Laurenceau J. Intensive Longitudinal Methods: An introduction to Diary and Experience Sampling Research. Chapter 5: Modeling the Within-Subject Causal Process. URL: http://www.intensivelongitudinal.com/ch5/ch5index.html [accessed 2016-06-10] [WebCite Cache ID 6iAhW63F8]
37. Robinson GK. That BLUP is a good thing: the estimation of random effects. Stat Sci 1991 Feb;6(1):48-51. [doi: 10.1214/ss/1177011933]
38. Sap M, Par G, Eichstaedt J. Developing age and gender predictive lexica over social media. 2014 Presented at: Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar p. 1146-1151 URL: http://wwbp.org/papers/emnlp2014_developingLexica.pdf [WebCite Cache ID 6k9cOGU2b]
39. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Machine Learning Res 2003;3(4/5):993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]
40. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990 Jun;85(410):398-409. [doi: 10.1080/01621459.1990.10476213]
41. Steyvers M, Griffiths T. Probabilistic topic models. In: Landauer TK, McNamara DS, Dennis S, Kintsch W, editors. Handbook of Latent Semantic Analysis. 1st edition. Mahwah, NJ: Lawrence Erlbaum Associates; 2007:427-448.
42. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 2013;8(9):e73791 [FREE Full text] [doi: 10.1371/journal.pone.0073791] [Medline: 24086296]
43. Blei D, Lafferty J. Correlated topic models. In: Weiss YB, Schölkopf B, Platt J, editors. Advances in Neural Information Processing Systems. 18th edition. Cambridge, MA: MIT Press; 2006.
44. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
45. Diener E. Subjective well-being. Psychol Bull 1984 May;95(3):542-575. [Medline: 6399758]
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57(1):289-300 [FREE Full text] [doi: 10.2307/2346101]
47. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, NY: Guilford Press; 1979.
48. Pennebaker JW, Beall SK. Confronting a traumatic event: toward an understanding of inhibition and disease. J Abnorm Psychol 1986 Aug;95(3):274-281. [Medline: 3745650]
49. Spera SP, Buhrfeind ED, Pennebaker JW. Expressive writing and coping with job loss. Acad Manage J 1994 Jun 01;37(3):722-733. [doi: 10.2307/256708]
50. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
51. 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]
52. Ryff CD, Keyes CL. The structure of psychological well-being revisited. J Pers Soc Psychol 1995 Oct;69(4):719-727. [Medline: 7473027]
53. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5:63-76 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]

Abbreviations

LDA: latent Dirichlet allocation
RCT: randomized controlled trial
STAGE: savor, thank, aspire, give, and empathize

Edited by G Eysenbach; submitted 05.03.16; peer-reviewed by A AL-Asadi, D Meyer, M Schotanus-Dijkstra, E Yom-Tov; comments to author 28.03.16; revised version received 17.05.16; accepted 28.07.16; published 31.08.16.

Please cite as:
Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, Parks AC
Seeing the "Big" Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data
J Med Internet Res 2016;18(8):e241
URL: http://www.jmir.org/2016/8/e241/
doi:10.2196/jmir.5725
PMID:27580524

Original Paper

Spanish-Language Consumer Health Information Technology Interventions: A Systematic Review

Alexis V Chaet[1]; Bijan Morshedi[1]; Kristen J Wells[2], MPH, PhD; Laura E Barnes[3], PhD; Rupa Valdez[1], PhD

[1]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

[2]Department of Psychology, San Diego State University, San Diego, CA, United States

[3]Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Rupa Valdez, PhD
Department of Public Health Sciences
University of Virginia
PO Box 800717
Charlottesville, VA, 22908
United States
Phone: 1 4349822510
Fax: 1 4342433418
Email: rsv9d@virginia.edu

Abstract

Background: As consumer health information technology (IT) becomes more thoroughly integrated into patient care, it is critical that these tools are appropriate for the diverse patient populations whom they are intended to serve. Cultural differences associated with ethnicity are one aspect of diversity that may play a role in user-technology interactions.

Objective: Our aim was to evaluate the current scope of consumer health IT interventions targeted to the US Spanish-speaking Latino population and to characterize these interventions in terms of technological attributes, health domains, cultural tailoring, and evaluation metrics.

Methods: A narrative synthesis was conducted of existing Spanish-language consumer health IT interventions indexed within health and computer science databases. Database searches were limited to English-language articles published between January 1990 and September 2015. Studies were included if they detailed an assessment of a patient-centered electronic technology intervention targeting health within the US Spanish-speaking Latino population. Included studies were required to have a majority Latino population sample. The following were extracted from articles: first author's last name, publication year, population characteristics, journal domain, health domain, technology platform and functionality, available languages of intervention, US region, cultural tailoring, intervention delivery location, study design, and evaluation metrics.

Results: We included 42 studies in the review. Most of the studies were published between 2009 and 2015 and had a majority percentage of female study participants. The mean age of participants ranged from 15 to 68. Interventions most commonly focused on urban population centers and within the western region of the United States. Of articles specifying a technology domain, computer was found to be most common; however, a fairly even distribution across all technologies was noted. Cancer, diabetes, and child, infant, or maternal health were the most common health domains targeted by consumer health IT interventions. More than half of the interventions were culturally tailored. The most frequently used evaluation metric was behavior/attitude change, followed by usability and knowledge retention.

Conclusions: This study characterizes the existing body of research exploring consumer health IT interventions for the US Spanish-speaking Latino population. In doing so, it reveals three primary needs within the field. First, while the increase in studies targeting the Latino population in the last decade is a promising advancement, future research is needed that focuses on Latino subpopulations previously overlooked. Second, preliminary steps have been taken to culturally tailor consumer health IT interventions for the US Spanish-speaking Latino population; however, focus must expand beyond intervention content. Finally, the field should work to promote long-term evaluation of technology efficacy, moving beyond intermediary measures toward measures of health outcomes.

(J Med Internet Res 2016;18(8):e214) doi:10.2196/jmir.5794

KEYWORDS

health information technology; consumer health information; consumer health informatics; health education; health promotion; health care quality, access, and evaluation; patient compliance; patient participation; patient satisfaction; patient preference; patient education; preventive health services; Hispanic; Latinos; cultural characteristics; cultural competency; ethnicity

Introduction

Patients are at the heart of the health care system. As primary stakeholders, they are not only affected by national and local policy, medical services, and the health care workforce, but also have the ability to affect health care cost, quality, and access through individual and community engagement. Patient engagement is a broadly defined term used to describe patient acquisition of knowledge, skills, ability, and motivation to participate in positive health behaviors and the interventions increasing these attributes [1,2]. A wide range of factors including literacy level, personal interest, information quality and access, and knowledge impact a patient's ability to engage in their health and health care [3-7]. Engagement not only increases a patient's overall satisfaction with the health care experience but also directly impacts health outcomes [8,9]. Engaged patients are more proactive in using preventative health resources such as screenings and immunizations, more willing to ask questions of their provider, more effective in managing chronic conditions, and less likely to participate in unhealthy lifestyle behaviors such as smoking and drug use [10-12]. As the complexity of medical technology and the burden of chronic disease grow, empowering patients to engage on all levels of care will be critical to reducing health care costs, improving population health, and maintaining patient satisfaction [13].

Consumer health information technology (IT) is increasingly being used to engage patients in shared decision making, self-management, and disease prevention through facilitation of health information access, social and clinical support, and electronic communication [14,15]. Consumer health IT may be conceptualized within the field of wellness informatics, which Grinter et al define as "a human-centered computing science focused on the design, deployment, and evaluation of human-facing technological solutions to promote and manage wellness acts, such as the prevention of disease and the management of health" (pg. 78) [16]. Although consumer health IT has varied definitions [15,17-19], there is a consensus that it is a form of electronic technology used by lay people to support health and health care management. Consumer health IT platforms include a variety of electronic systems such as desktop or laptop computers, touchscreen kiosks, personal wireless devices, and mass media [16,18,20,21]. A growing body of research has explored the feasibility, acceptability, and efficacy of these tools in promoting health [15,22-26]. While challenges remain in the standardized evaluation and cost-benefit analysis of consumer health IT [27,28], initial studies have shown potential for these engagement-oriented tools to positively impact health outcomes, quality of life, hospital readmission rates, and mortality [29-34].

Critical to the design of technologies that facilitate health and health care management is the consideration of population needs and characteristics [35-38]. A patient-centered design approach aims to systematically partner with patients in the creation and tailoring of technologies to best suit their unique needs, skills, environments, and preferences [36,39-41]. Design of technologies without consideration of the end user can lead to user frustration, error and misuse, and technology abandonment [41-44].

As consumer health IT becomes more thoroughly integrated into patient care, it is critical that these tools are appropriate for the diverse patient populations whom they are intended to serve [45-48]. Cultural differences associated with ethnicity are one aspect of diversity that may play a role in user-technology interactions [30,46,49]. Many studies that focus on consumer health IT within the literature do not have a diverse population sample [18]. It is difficult to determine whether lessons learned from heterogeneous populations may be universally applicable to minority ethnic groups. Instead there may be unique lessons to be learned. While there has been growing recognition of the importance of providing culturally competent health care [50], most of these initiatives have focused on providers, health care organizations, and public health interventions [51]. Given the potential for consumer health IT to enhance patient well-being both within and outside the institutional health care setting, it is critical that we view the design of these tools through the same conscientious lens of cultural competency [52].

Research within this field is both timely and important because of existing health disparities faced by ethnic minority populations and the national priority to decrease these disparities [53]. Designing appropriate consumer health IT interventions for ethnic minorities offers an opportunity to bridge existing health disparities faced by these populations [53]. In contrast, failure to design consumer health IT appropriately for ethnic minority groups risks exacerbating this divide [54]. Consequently, there is a need to assess our current understanding of the intersection between culture, ethnicity, and consumer health IT and to elucidate areas for future research within the field.

This paper focuses on a single ethnic group: the US Spanish-speaking Latino population, a heterogeneous group consisting of numerous subpopulations. This ethnic population was chosen for its current and increasing prominence within the United States. The Latino population represents the nation's largest ethnic minority, numbering over 54 million [55]. In accordance with other ethnic minorities, Latinos face stark disparities in health and health literacy, influenced by poverty, institutional racism, and linguistic barriers, among other factors [56]. English-speaking Latinos show higher levels of technology ownership than their Spanish-speaking counterparts [57], and the latter group experiences worse self-reported health status and access to care [58]. This study aims to evaluate the current scope of consumer health IT targeted to the US Spanish-speaking Latino population, characterizing interventions in terms of technological attributes, health domains, cultural tailoring, and evaluation methods. In doing so, it identifies gaps

in and limitations of interventions for this specific population and offers considerations for research within the broader field of culturally informed consumer health IT [46].

Methods

Design

A narrative synthesis [59] was conducted of existing Spanish language consumer health IT interventions indexed within health and computer science databases. A narrative review is a qualitative systematic review, which looks for themes or constructs present within a body of studies. In contrast to meta-analysis, the aim of a narrative review is to create a broad understanding of a particular phenomenon [59]. The aim of this study was to synthesize existing literature and to systematically assess gaps in consumer health IT interventions tailored to US Spanish-speaking Latinos.

Search Strategy

Searches were first conducted in August 2014 within four health sciences (ie, PubMed, Web of Science, CINAHL, Cochrane Central Register of Controlled Trials [Cochrane]) and three computer sciences and engineering databases (Compendex, IEEE Xplore, and the Computers and Applied Sciences Complete [CASC]). A second search was run in September 2015 within these databases to capture additional articles published during the screening process. All databases were accessed via the University of Virginia libraries. A third search was run in June 2016 to expand the search to a 25-year time span from 1990-2015. Search terms were divided into three clusters referencing technology, ethnicity, and patient-centeredness (see Table 1). Terms were adapted for each unique database in consultation with a University of Virginia librarian. The search was limited to English language articles with human subjects. Database-specific Boolean search strings can be seen in the Multimedia Appendix 1.

Table 1. Search terms for PubMed (terms were adapted for each database).

Technology	Ethnicity	Patient-centeredness
cellular phone[a]	Hispanic Americans	consumer health information
mobile phone	Hispanic	health education[a]
mobile computing	Spanish Americans	health promotion[a]
mobile health	Latino(a)	health care quality, access, and evaluation[a]
text messaging	Spanish-speaking	patient compliance
internet[a]		patient participation
ehealth		patient satisfaction
blogging		patient preference
social media		patient education
facebook		preventive health services[a]
twitter		
telemedicine[a]		
audio player		
audiovisual aids[a]		
multimedia		
health records, personal		
computer systems[a]		
tablet computer		
computer/utilization[a]		
user-computer interface[a]		
computer user		
television[a]		
radio[a]		
soap opera		
reminder system		
educational technology[a]		
medical informatics		
health information technology		

[a]Medical Subject Headings (MeSH) term.

The combined electronic searches identified 2742 records. Records were divided as follows: PubMed (1798 citations), Web of Science (42 citations), CINAHL (717 citations), Cochrane (87 citations), Compendex (6 citations), CASC (136 citations), and IEEE Xplore (0 citations). After removal of duplicates, a combined total of 2626 unique records was compiled for preliminary abstract review.

Inclusion Criteria

The search was limited to full-text, English language articles published between January 1990 and September 2015, with additional inclusion and exclusion criteria described in Textbox 1. If inclusion could not be determined by information provided within the abstract, the article was included for full-text review. Full-text inclusion criteria were identical to those used in abstract screening.

Textbox 1. Inclusion criteria for abstract screening and full-text review.

Article characteristics:

- Article must be published between January 1990 and September 2015.
- Article must be in the English language.

Population characteristics:

- Participants must live within the United States, defined by the 48 contiguous states, Alaska, and Hawaii.
- If participants lived in both the United States and abroad, article must analyze US participants as cohesive population subset.

Intervention characteristics:

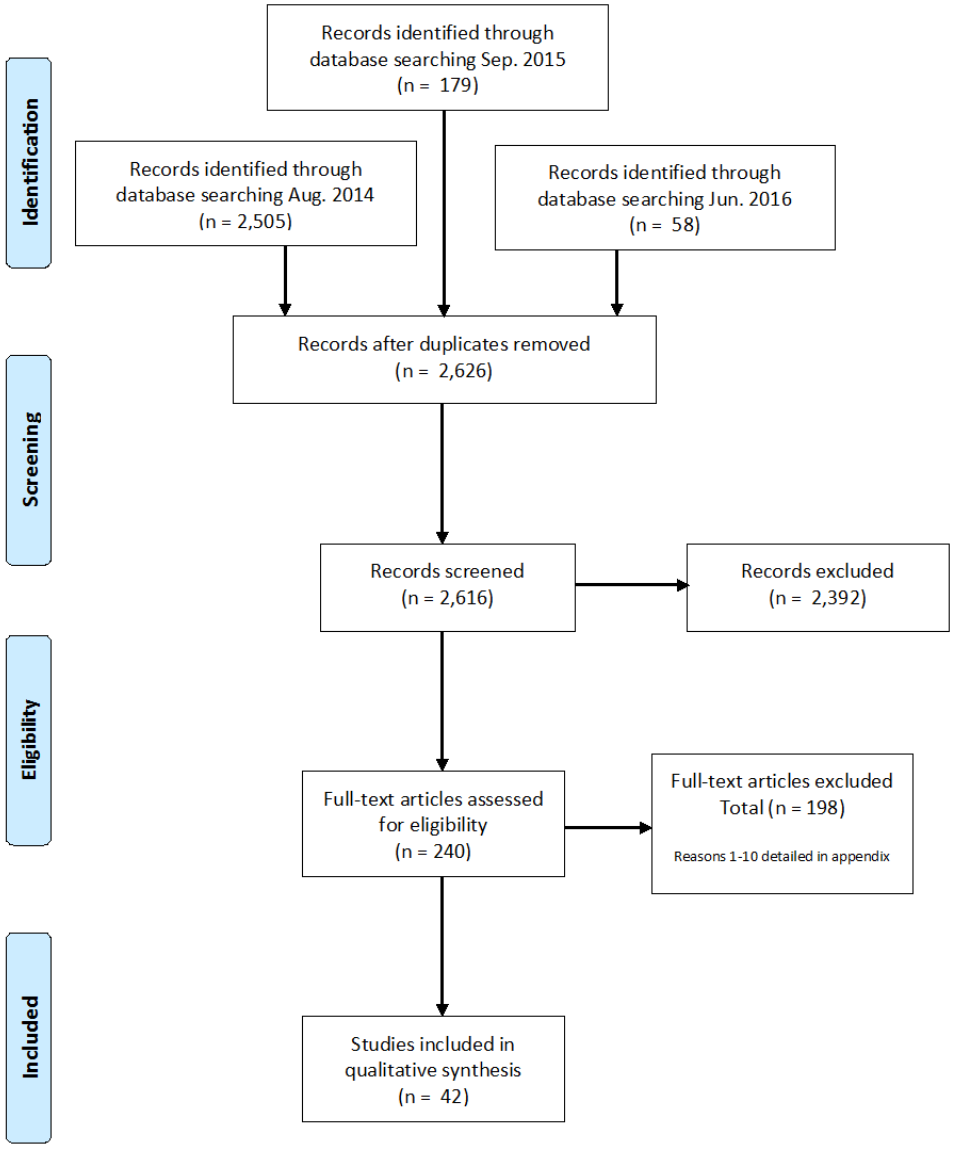
- Intervention must involve electronic technology (technology using electricity). This includes radio, television, mobile phone, computer, tablet, MP3 player, etc.
- Information delivered through the intervention must be available in Spanish.
- Intervention must target health and include topics pertaining to one or more of the following:
  - disease treatment
  - disease prevention
  - health education
  - personal safety
  - access to care
  - personal wellness
  - mental health
  - well-being
  - care of dependents
- Patient and/or the patient's legal guardian must be the end user and direct benefactor of the device.
- Interventions targeting providers are excluded.
- Interventions consisting only of phone calls were excluded.

Study Selection

Study selection consisted of two steps: abstract screening and full-text review. Abstracts were independently screened by authors AC and BM and compared at intervals of 50-100 articles until a Cohen's kappa score of .95, indicating near-perfect agreement [60], was reached after three rounds. Discrepancies were discussed and eligibility criteria were refined between intervals when necessary. After final inclusion criteria were agreed upon, the remaining abstracts were divided between AC and BM and screened independently.

In total, 240 articles were returned for full-text review. AC reviewed all full-text articles using inclusion criteria. Reasons for article exclusion are detailed in the Multimedia Appendix 2, and excluded articles are listed in Multimedia Appendix 3. If an article failed multiple inclusion criteria, the first criterion identified upon a linear read of the article from introduction to conclusion was documented. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [61] flowchart representing the study selection process is shown in Figure 1.

Figure 1. PRISMA schematic of study selection.



Data Extraction

We identified 42 articles for full-text extraction. AC and BM independently reviewed and extracted data from all 42 articles (see Multimedia Appendix 4). A standardized template within Microsoft Excel was used to ensure systematic extraction of data. Extraction data were compared, and all discrepancies led to a joint article review and discussion. The following data were collected: (1) first author's last name, (2) publication year, (3) population characteristics (eg, sample size, gender, age, percent Latino, ethnicity), (4) journal domain, (5) health domains, (6) technology platform, (7) technology functionality, (8) available languages of consumer health IT intervention, (9) US region, (10) cultural tailoring category, (11) population density (eg, urban versus rural), (12) intervention delivery location, (13) evaluation metrics, (14) evaluation results, and (15) study design.

Deductively Derived Categories

Categories for journal domain [62], US region [63], study design [64], cultural tailoring [46], and technology functionality [65] were deductively derived using external frameworks.The external frameworks for technology functionality and cultural tailoring categories are summarized in Table 2. Evaluation metrics were categorized according to a deductive framework developed by the collective expertise of the study team (see Table 4).

Table 2. Technology functionality framework [61].

Functionality subcategory	Definition
Inform	Provide information in a variety of formats (text, photo, video)
Instruct	Provide instructions to the user
Record	Capture user-entered data
Display	Graphically display user-entered data/ output user-entered data
Guide	Provide guidance based on user-entered information (eg, recommend a physician consultation or course of treatment)
Remind/Alert	Provide reminders to the user
Communicate	Provide communication with health care provider/ patients and/or provide links to social networks

Table 3. Culturally-informed design framework [46].

Cultural tailoring category	Description	Examples of cultural considerations
Content	Message being delivered through the technology	Origins and consequences of health conditions
		Norms related to diet, religion, and division of labor
		Alternative medicine
Functionality	Array of actions performed by the technology	Culturally specific health management behaviors
		Perception of privacy and health care decision making
		Preferences for information delivery and communication (eg, voice communication)
Technology platform	Technology hardware used to deliver the health intervention	Access and exposure to technology
		Use of hardware within target population
		Role of Internet
User interface	Presentation and organization of the content and functionality	Cultural symbols
		Language and dialect
		Spatial orientation
		Colors

Table 4. Evaluation metrics categories.

Category	Description
Behavior/ attitude change	Changes in the lifestyle, disease management, or attitude toward a health topic or behavior. These include measurements such as changes in disease screening rates, treatment compliance, medical care utilization, performance of self-care tasks, attitude toward organ donation, attitude toward breast cancer screening, and attitude toward alcohol use.
Knowledge retention	Any measurement of information taught through technology intervention. These include measurements such as knowledge of diabetes care, knowledge of disease prevention techniques, or knowledge of vaccination schedules.
Self-reported health marker	Self-reported health measures including depression scale rankings, pain rankings, self-efficacy, psychosocial functioning, or quality of life.
Biometric health marker	Quantitative measures of body function including HbA1c levels, blood pressure, glycemic control, and body mass index. Both clinic-generated and self-reported biometric health markers were included within this category.
Usability	Specific feedback regarding physical characteristics of technology, user interface, acceptability of technology, and perceived utility. These include measures such as ease of use, readability, ability of patient to relate to video characters, acceptability of video length, emotional appeal, and satisfaction with device.

Inductively Derived Categories

Using the data abstracted from the individual studies, the research team inductively derived categories for the following variables: technology platform, population density, health domains, evaluation results, and intervention delivery location (see Table 5). Additionally, we followed the procedure below:

Technology platform: When it was not possible to infer the specific technology platform, the article was coded as "unspecified." This was most common for video interventions that did not specify whether the video was delivered through a

computer, digital video disc (DVD) player, videocassette recorder (VCR), or tablet.

Population density: If it was possible to infer an urban or rural location from article language or study location, the article was coded accordingly. For example, an intervention conducted within a "city center" was coded as "urban."

Health domain: Given the numerous health domains investigated by studies, we engaged in a second round of coding to reduce the number of health domain categories. For example, "infant immunization" was coded as "infant, child health, or maternal health."

Intervention delivery location: Similarly, some intervention delivery locations were coded at a higher level of abstraction. For example, "mass media" and "internet" interventions were ultimately coded as "ubiquitous environment."

Evaluation results: Results were drawn directly from the text of each article. If multiple outcomes were reported, only primary outcomes were included.

Table 5. Inductively derived categories.

Variable	Description
Population density	Article was categorized as "rural" only if authors specified a rural community. If authors did not use the term "urban" but specified a city or county that was predominantly urban, the article was classified as urban. If technology use occurred within an urban hospital center, the article was classified as urban. Articles that were unclear or did not specify any location were classified as "Did Not Specify (DNS)."
Intervention delivery locations	Clinic: Intervention delivery within a clinic, hospital, or medical center. Includes clinic waiting room or medical encounter.
	Ubiquitous environment: Intervention delivery could occur in multiple physical environments. Technologies accessed by patients through personal devices such as mobile phones, desktop computers, or radio, or through public mass media. This includes all interventions accessed through the Internet.
	Community center: Intervention delivery in any public gathering space that does not formally provide medical care (ie, not a clinic). This includes churches, schools, pharmacies, cafes, libraries, and other community centers.

Data Analysis

Data were coded manually into numerical categories, and basic statistics were computed using Microsoft Excel Version 14.1. All percentages were calculated out of the total number of included studies (N=42). If a study did not report on a given category, the study was coded as "did not specify." Studies could be categorized in multiple subcategories for the following: cultural tailoring category, evaluation metrics, technology platform and functionality, population density, intervention delivery location, and journal domain.

Results

General Study Characteristics

All studies were published between 1990 and 2015, with the majority of studies (30/42, 71%) published between 2009 and 2015. All 42 articles detailed distinct consumer health IT interventions [66-100, 122-124, 136-139]. Studies were published across 26 unique journal domains. Medical Sciences—general (15/42, 36%) and Medical Sciences—specialty (17/42, 40%) were the most prevalent subcategories, followed by Public Health and Safety (11/42, 24%). The mean sample size was 416 participants (SD 603; 42/42 articles). This value was inflated by the inclusion of four mass media studies that contained more than 1000 participants [92-95]. Randomized controlled trial (RCT; 18/42, 43%) and non-experimental (19/42, 45%) were the most commonly used study designs. The percentage of female study participants in the abstracted studies ranged from 49.5-100% (40/42), and the mean age of participants reported in the included studies ranged from 15.0-72.1 years (21/42). Across study samples, the percentage of Latino participants ranged from 51-100% (40/42). However, to be included in the study, it was required that studies have a majority of Latino participants. Country of origin was reported within half of the articles (21/42, 50%). Given inconsistent reporting strategies, however, it is difficult to make representative generalizations. Most studies focused within the western region of the United States (17/42, 41%) and on an urban population center (35/42, 83%).

Intervention Characteristics

A wide variety of consumer health IT interventions have been used to target health within the Latino population. These interventions have focused most commonly on chronic diseases and included some degree of cultural tailoring. Computer, radio, and television were the most commonly used technology platforms; however, a fairly even distribution across all technologies was noted. Nearly all interventions had the functionality of informing the end user (38/42, 91%), and nearly one half of studies employed more than one functionality (19/42, 45%). The large majority of technology interventions (32/42, 76%) specified availability in English in addition to Spanish. Cancer (10/42, 24%), diabetes (9/42, 21%), and child, infant, or maternal health (9/42, 21%) were the most commonly addressed health domains. As shown in Table 6, a number of health domains were targeted by only one study; these included topics such as sexual health, appointment reminders, anesthesia, and health care utilization. The most common intervention delivery location was within a ubiquitous environment (19/42, 45%) or clinic setting (17/42, 41%). More than half of the interventions (25/42, 60%) were culturally tailored.

Table 6. Frequencies of selected intervention characteristics of included studies (for all domains, articles may be included within multiple subcategories).

		Frequency, n	Percentage, %
Technology platform			
	Computer	8	19
	Radio	8	19
	Television	8	19
	Kiosk	7	17
	Unspecified	5	12
	Mobile phone text message	5	12
	VCR	3	7
	DVD	2	4
	Tablet	2	4
	Not reported	0	0
Intervention delivery location			
	Ubiquitous environment	19	45
	Clinic setting	17	40
	Community center	5	12
	Not reported	2	5
Population density			
	Urban	35	83
	Rural	5	12
	Not reported	4	10
Cultural tailoring characteristics			
	Content	21	50
	User interphase	6	14
	Functionality	2	5
	Technology platform	1	2
	Not reported	17	40
Technology functionality			
	Inform	38	91
	Communicate	6	14
	Guide	5	12
	Instruct	4	10
	Record	4	10
	Remind/Alert	4	10
	Display	3	7
	Not reported	0	0
Health domain			
	Cancer	10	24
	Child, infant, or maternal health	9	21
	Diabetes	9	21
	Cardiovascular disease	3	7
	Organ donation	3	7
	Physical activity	2	5

	Frequency, n	Percentage, %
General adult health	2	5
Sexual health	1	2
Anesthesia	1	2
Appointment reminder	1	2
Driving under influence recidivism	1	2
Pain	1	2
Health care utilization	1	2
Patient safety	1	2
Not reported	0	0

Evaluation and Results

Evaluation metrics included both intermediate health-related measures and measures of technology usability. Most articles (28/42, 67%) used more than one evaluation metric. The most commonly used evaluation metric subcategory was behavior/attitude change (31/42, 74%), followed by usability (20/42, 48%), and knowledge retention (22/42, 52%). Behavior/attitude change included metrics such as medication adherence [96,97], intention to become an organ donor [93], and colorectal cancer screening rate [98]. Knowledge retention included metrics such as knowledge of cardiovascular disease risk factors and prevention [99], knowledge of basic child immunization schedule [100], and knowledge of general recommendations for breast cancer treatment [95]. Specific details for each study, including evaluation results, can be found in Multimedia Appendix 7.

Discussion

Principal Findings

In summary, the Spanish language consumer health IT interventions targeting US Spanish-speaking Latinos published within the past 25 years (1990-2015) are characterized by a great amount of diversity with regards to technology platform and study design (eg, sample size and evaluation metrics); however, similarities can be seen in the technology functionality, specific populations, and health domains addressed by these interventions. Interventions most commonly focused on urban population centers, the western United States, and chronic health domains including cancer and diabetes. While sample size varied tremendously across studies, study samples were largely female. Behavior/attitude change, knowledge retention, and usability were the most commonly used evaluation metrics. Just over half of studies detailed some type of cultural tailoring, with content tailoring being the most common.

Intervention Characteristics

Technology Platform and Functionality

A wide distribution of technology platforms to engage Spanish-speaking patients in their health and health care is seen within the literature. Computer, radio, and television were the most common platforms used across interventions. As we continue to explore newer technologies, it is important to understand what aspects of these intervention designs are applicable to other technology modalities. Moreover, as the technology platforms used to engage consumers evolve, researchers should be cognizant of disparities in technology exposure and access across subpopulations. For example, within the Latino population computer ownership as well as Internet usage, mobile phone ownership, smartphone ownership, and social media use vary significantly across age, socioeconomic status, and language dominancy [52]. For some subpopulations, traditional technologies may be more appropriate given varying levels of access and exposure to newer technology platforms.

Nearly all interventions served to inform or educate the end users while fewer interventions incorporated other functionalities such as delivering a direct service or treatment to patients. This is not surprising given that health education aims to equip patients with the knowledge and skills they need to manage their disease while promoting behavior change [101,102]. Nevertheless, future interventions could strive to incorporate additional functionalities to address specific user needs. For example, consumer health IT (whether publicly or privately disseminated) could be used to mitigate communication barriers between digitally underserved Latinos and health care providers or social resources [103].

Cultural Tailoring

This study reveals an initial movement toward integrating culturally tailored features into consumer health IT for the US Spanish-speaking Latino population. More than half of the articles mentioned some form of cultural tailoring, suggesting awareness of the interaction between culture and user interactions with technology. Intervention content was the predominant mechanism of tailoring, with fewer articles tailoring functionality, technology platform, or user interface. While there is abundant literature on culturally competent health care for Latinos in the past two decades [104-110], the cultural competency movement within the health sciences has focused predominantly on language, cultural traditions, and cultural differences in health beliefs within this population [111]; these domains are predominantly pertinent to technology content. In contrast, there have been some initial efforts within the fields of engineering, marketing, and health to shed light on the influence of cultural preferences not only on content, but also user interface, functionality, and technology platform [47,112-116]. Some of these studies have focused specifically

on the Latino population [114-116]. Although these studies face some limitations in sample size, reliance on a conceptualization of cultural context as one broadly defined cultural identity, and lack a systematic understanding of the design space, they nonetheless provide initial design constructs to aid future consumer health IT designers in generating a more holistic understanding of culturally competent technology form and function preferences.

Health Domains and Population Target

Health Domains

This study reveals a focus within the literature on chronic diseases and a need for future consumer health IT interventions to target two areas: underrepresented health domains within Latino subpopulations and challenges faced by Latinos in health care access and utilization. Given the disproportionate prevalence of diabetes, obesity, and cancer within the Latino population [112,117], focus on chronic disease prevention and management offers significant opportunity for social and economic incentives. Nonetheless, population health statistics can mask wider variation in Latino subpopulations. As a relevant example, Latino migrant farmworkers face higher concerns related to pesticide exposure given their living or work environment [118,119]. No articles included within this review targeted pesticide exposure, yet technology access within this population suggests health IT interventions may be feasible to address this topic. For example, mobile phone penetration among farmworkers has been found to be comparable to the general Latino population [120]. Furthermore, in cases where there is a lack of technology access, there are examples in the literature where technology was incorporated at the community level [28]. To avoid widening intragroup health disparities, future research should ensure that the needs of these subpopulations are met. In addition, beyond disease-specific health domains, future consumer health IT interventions for the US Spanish-speaking Latino population should target health care utilization. Lack of trust, economic barriers, lack of knowledge of services, immigration status, and linguistic and cultural differences contribute to a sense of disconnect between Latinos and the US health care system [121]. To mitigate these systemic barriers, future research should explore how consumer health IT interventions might be used to facilitate peer-support systems, connect patients to financial and linguistic services, or assist patients in health care system navigation.

Gender

It is important to note that the majority of participants across studies were female. This may be influenced by the fact that a large number of interventions focused within the category of "Child, Infant, or Maternal Health," targeting health concerns such as breast cancer [122,123] or cervical cancer [124]. However, it is not uncommon for women to be more represented in scientific research [125,126].

Location

This study offers two principle findings regarding the location of intervention delivery. First, the majority of studies were conducted in urban settings, likely reflecting the location of academic institutions. Given that nearly 12% of the Latino population lives within a rural area [127] and that rural Latinos are less likely to receive preventative care, to be screened for certain cancers, and to meet vigorous physical activity recommendations than their urban counterparts [128], future intervention should engage rural communities to avoid exacerbating existing intragroup health disparities. Second, within urban populations, the vast majority of interventions were delivered through technology platforms that could be used in ubiquitous environments as opposed to commercial locations such as clinics or pharmacies. This likely reflects the increasing popularity of Internet-capable personal devices. Accelerating consumer health IT interventions built upon these platforms might allow for greater reach given the growing adoption of mobile phones and smartphone technologies within the Latino population [57]. Researchers, however, should be cognizant of wireless access limitations or unique structures of service provision within rural communities.

Evaluation Metrics and Study Design

The majority of included studies focused on intermediate measures, or those that are conceptualized as precursors to predicting health outcomes, namely, knowledge retention and behavior/attitude change. Although studies have shown that positive behaviors can affect significant changes in chronic disease outcomes, these behaviors must be sustained in the long term for significant changes in health status [129]. Given documentation of high rates of attrition in the use of consumer health IT [130], future studies should address the longevity of these behavior/attitude changes. In addition, future studies should use both validated measures particular to a given health condition as well as validated measures that have relevance across health conditions such as the Patient Activation Measure (PAM) [12]. PAM assesses a patient's knowledge, skill, and confidence for self-management of disease, which predicts health behaviors, self-management behaviors, and consumeristic type behaviors [12]. This will facilitate comparison of outcomes across consumer health IT interventions targeted to diverse health conditions within the US Spanish-speaking Latino population.

Given the interdisciplinary nature of studies focused on consumer health IT interventions for the US Spanish-speaking Latino population, diversity was seen both in types of evaluation metrics and combinations of metrics used by studies. Although this approach allows for multiple perspectives on effective intervention development, a key limitation is the ability to conduct meta-analyses and to compare findings across studies. Future research should synthesize various perspectives from relevant disciplines to create a framework for evaluation of consumer health IT for the US Spanish-speaking Latino population. A crosswalk approach might then be used to identify connections between various evaluation metrics [131]. The large number of studies that used an RCT design is a promising finding given that RCTs are considered to be the strongest form of clinical evidence within intervention-based studies [132]. The use of RCT study design is not uncommon within in the evaluation of consumer health IT interventions [31].

Considerations for Reporting Future Studies

Characterizing studies was challenging because of lack of detail and vague or incomplete descriptions in study reporting. Lack of detail was evident in the large number of articles that did not specify the technology platform used. This required an "unspecified" category for technology platform to be made. Vague or incomplete intervention descriptions were evident in our classification of cultural tailoring. Future studies should explicitly detail cultural tailoring processes and should cite feasibility studies or other evidence-based rationales to substantiate these tailoring choices. Ultimately, our ability to report frequency statistics was limited by a lack of standardized reporting methods across studies.

A consensus statement for reporting consumer health IT studies would improve the prospects for valuable meta-analyses to be conducted in the future. Consensus statements have been developed for many other types of studies, such as RCTs [133], systematic reviews [134], and Internet surveys [135]. Some of these guidelines are applicable to the consumer health IT studies reported in this paper. For example, several studies use an RCT design [96,98,136-139] and therefore may be reported using Consolidated Standards of Reporting Trials (CONSORT) guidelines [140]. However, the guidelines for CONSORT focus on study design elements and do not provide guidance for reporting aspects related to technology design. The development of a consensus statement for reporting culturally informed health IT studies, which includes both experimental and technology design elements, is important for advancing our understanding of how culture might be integrated into health information technologies. Such a consensus statement might include technology design elements such as technology functionality, user interface, content, and technology platform, as well as demographic information including country of birth, self-reported ethnicity, and user language preference.

Limitations

Several study limitations warrant mention. Given restrictions in database access and time, a limited number of databases was chosen and additional mechanisms for hand searching were not undertaken. Furthermore, only English language articles were included. This likely contributed to selection bias and limitations in the scope of studies compiled for screening. Nonetheless, the authors used a wide range of databases from both health and computer sciences to capture a rich body of articles. Limitations were additionally faced in the classification of cultural tailoring categories. Some studies justified selecting a particular technology platform; in other cases, it was unclear whether selection of a technology platform was founded upon population-specific needs assessments or usage statistics. In these latter cases, the intervention was not considered to be culturally tailored in terms of technology platform. This approach likely led to a conservative estimate of cultural tailoring and reveals a need for more explicit descriptions of decision-making approaches used to design and develop consumer health IT interventions. Finally, the inductively derived categories were representative only of characteristics present within the studies. The scope of these categories does not give a sense of characteristics that should ideally be included but were not represented.

Conclusions

In this study, we have characterized the growing body of consumer health IT interventions targeted toward the US Spanish-speaking Latino population. In doing so, three primary needs have been identified within this field. First, while the increase in studies targeting the Latino population in the last decade is a promising advancement, future research is needed that focuses on subpopulations previously overlooked in designing interventions within this space. For example, the Latino migrant farmworker community faces acute health conditions such as pesticide exposure, which may pose a more immediate health threat than the chronic diseases plaguing the statistical majority of the Latino demographic. Second, preliminary steps have been taken to culturally tailor consumer health IT interventions for the US Spanish-speaking Latino population; however, focus has remained predominantly on intervention content. Interdisciplinary fieldwork between the health sciences and engineering is needed to understand how to create technology culturally tailored in terms of platform, user interface, and functionality preferences. Finally, the majority of studies used intermediary measures such as knowledge retention and behavior/attitude change to evaluate technology efficacy. Given the immense financial investment and potential social benefits of consumer health IT, it is critical that research within the field engages patients long enough to begin measuring health outcomes.

Acknowledgments

Research reported in this publication was supported by National Cancer Institute of the National Institutes of Health under Award Number R21CA167418. Author AC was supported by a Diversity Training Branch Supplement under this parent grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would also like to thank Voto Latino for the funding support provided for this publication through their Innovators Challenge. We would like to thank the University of Virginia librarians for their immense help in database selection and refinement of our search strategy.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Boolean search strings across databases.

[PDF File (Adobe PDF File), 248KB - jmir_v18i8e214_app1.pdf]

Multimedia Appendix 2

Full-text exclusion justification breakdown.

[PDF File (Adobe PDF File), 159KB - jmir_v18i8e214_app2.pdf]

Multimedia Appendix 3

List of excluded studies.

[PDF File (Adobe PDF File), 141KB - jmir_v18i8e214_app3.pdf]

Multimedia Appendix 4

General study characteristics of included studies.

[PDF File (Adobe PDF File), 70KB - jmir_v18i8e214_app4.pdf]

Multimedia Appendix 5

Participant demographics of included studies.

[PDF File (Adobe PDF File), 67KB - jmir_v18i8e214_app5.pdf]

Multimedia Appendix 6

Intervention characteristics of included studies.

[PDF File (Adobe PDF File), 63KB - jmir_v18i8e214_app6.pdf]

Multimedia Appendix 7

Outcome metrics, study design, and evaluation results of included studies.

[PDF File (Adobe PDF File), 82KB - jmir_v18i8e214_app7.pdf]

References

1. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
2. James J. Health Affairs. 2013 Feb 14. Health Policy Brief: Patient Engagement URL: http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=86 [accessed 2016-07-20] [WebCite Cache ID 6j8rLWyeF]
3. Coulter A. Patient engagement--what works? J Ambul Care Manage 2012;35(2):80-89. [doi: 10.1097/JAC.0b013e318249e0fd] [Medline: 22415281]
4. Ishikawa H, Yano E. Patient health literacy and participation in the health-care process. Health Expect 2008 Jun;11(2):113-122. [doi: 10.1111/j.1369-7625.2008.00497.x] [Medline: 18494956]
5. Maurer M, Dardess P, Carman K, Frazier K, Smeeding L. Guide to Patient and Family Engagement: Environmental Scan Report. AHRQ Publication No. 12-0042-EF 2012 May;315(12):1-89 [FREE Full text]
6. Dy SM, Purnell TS. Key concepts relevant to quality of complex and shared decision-making in health care: a literature review. Soc Sci Med 2012 Feb;74(4):582-587. [doi: 10.1016/j.socscimed.2011.11.015] [Medline: 22236643]
7. Parsons S, Winterbottom A, Cross P, Redding D. The quality of patient engagement and involvement in primary care. An Inquiry into the Quality of General Practice in England 2010:1-99 [FREE Full text]
8. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
9. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214 [FREE Full text] [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]

10. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
11. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is patient activation associated with outcomes of care for adults with chronic conditions? J Ambul Care Manage 2007;30(1):21-29. [Medline: 17170635]
12. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
13. Dentzer S. Rx for the 'blockbuster drug' of patient engagement. Health Aff (Millwood) 2013 Feb;32(2):202 [FREE Full text] [doi: 10.1377/hlthaff.2013.0037] [Medline: 23381509]
14. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
15. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
16. Grinter R, Siek K, Grimes A. Is wellness informatics a field of human-centered health informatics? interactions 2010 Jan 01;17(1):76. [doi: 10.1145/1649475.1649494]
17. Flaherty D, Hoffman-Goetz L, Arocha JF. What is consumer health informatics? A systematic review of published definitions. Inform Health Soc Care 2015 Mar;40(2):91-112. [doi: 10.3109/17538157.2014.907804] [Medline: 24801616]
18. Gibbons M, Wilson RS, Samal L. AHRQ Evidence Reports. In: Impact of Consumer Health Informatics Applications Internet. Rockville, MD: Agency for Healthcare Research and Quality; Oct 2009:A.
19. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. J Am Med Inform Assoc 2008;15(4):473-483 [FREE Full text] [doi: 10.1197/jamia.M2744] [Medline: 18436895]
20. Eichner J, Dullabh P. Accessible Health Information Technology (IT) for Populations with Limited Literacy: A Guide for Developers and Purchasers of Health IT Internet. Rockville, MD: Agency for Healthcare Research and Quality; 2007:3-11.
21. Montague E, Perchonok J. Health and wellness technology use by historically underserved health consumers: systematic review. J Med Internet Res 2012;14(3):e78 [FREE Full text] [doi: 10.2196/jmir.2095] [Medline: 22652979]
22. Handel MJ. mHealth (mobile health)-using Apps for health and wellness. Explore (NY) 2011;7(4):256-261. [doi: 10.1016/j.explore.2011.04.011] [Medline: 21724160]
23. López L, Grant RW. Closing the gap: eliminating health care disparities among Latinos with diabetes using health information technology tools and patient navigators. J Diabetes Sci Technol 2012 Jan;6(1):169-176 [FREE Full text] [Medline: 22401336]
24. van den Berg N, Schumann M, Kraft K, Hoffmann W. Telemedicine and telecare for older patients--a systematic review. Maturitas 2012 Oct;73(2):94-114. [doi: 10.1016/j.maturitas.2012.06.010] [Medline: 22809497]
25. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
26. Gibbons MC, Wilson RF, Samal L, Lehmann CU, Dickersin K, Lehmann HP, et al. Consumer health informatics: results of a systematic evidence review and evidence based recommendations. Transl Behav Med 2011 Mar;1(1):72-82 [FREE Full text] [doi: 10.1007/s13142-011-0016-4] [Medline: 24073033]
27. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul;7(4):1039-1056 [FREE Full text] [Medline: 23911188]
28. Wells KJ, Vasquez-Otero C, Bredice M, Meade CD, Chaet A, Rivera MI, et al. Hispanic Healthcare International. Acceptability of a Virtual Patient Educator for Hispanic Women 2015;13(4):85. [doi: 10.1891/1540-4153.13.4.179]
29. Sands DZ, Wald JS. Transforming health care delivery through consumer engagement, health data transparency, and patient-generated health information. Yearb Med Inform 2014;9:170-176 [FREE Full text] [doi: 10.15265/IY-2014-0017] [Medline: 25123739]
30. Christopher GM. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011;8:1f [FREE Full text] [Medline: 21307989]
31. Or CKL, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]
32. 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]
33. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
34. Mignerat M, Lapointe L, Vedel I. Using telecare for diabetic patients: A mixed systematic review. Health Policy and Technology 2014 Jun;3(2):90-112. [doi: 10.1016/j.hlpt.2014.01.004]

35. Valdez RS, Brennan PF. Exploring patients' health information communication practices with social network members as a foundation for consumer health IT design. Int J Med Inform 2015 May;84(5):363-374. [doi: 10.1016/j.ijmedinf.2015.01.014] [Medline: 25704761]
36. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
37. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10 [FREE Full text] [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
38. Moen A, Brennan PF. Health@Home: the work of health information management in the household (HIMH): implications for consumer health informatics (CHI) innovations. J Am Med Inform Assoc 2005;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.M1758] [Medline: 16049230]
39. Gasson S. Human-Centered Vs. User-Centered Approaches to Information System Design. The Journal of Information Technology Theory and Application 2003;5(2):29-46 [FREE Full text]
40. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
41. Marquard JL, Zayas-Cabán T. Commercial off-the-shelf consumer health informatics interventions: recommendations for their design, evaluation and redesign. J Am Med Inform Assoc 2012;19(1):137-142 [FREE Full text] [doi: 10.1136/amiajnl-2011-000338] [Medline: 21727206]
42. Stead WW, Lin H, editors. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Washington, DC: National Academies Press; 2009.
43. Zayas-Cabán T, Dixon B. Considerations for the design of safe and effective consumer health IT applications in the home. Qual Saf Health Care Internet 2010;19:61-67 [FREE Full text] [doi: 10.1136/qshc.2010.041897]
44. Karsh B, Weinger MB, Abbott PA, Wears RL. Health information technology: fallacies and sober realities. J Am Med Inform Assoc 2010;17(6):617-623 [FREE Full text] [doi: 10.1136/jamia.2010.005637] [Medline: 20962121]
45. Forsythe DE. New bottles, old wine: hidden cultural assumptions in a computerized explanation system for migraine sufferers. Med Anthropol Q 1996 Dec;10(4):551-574. [Medline: 8979239]
46. Valdez R, Gibbons M, Siegel E, Kukafka R, Brennan P. Designing consumer health IT to enhance usability among different racial and ethnic groups within the United States. Health Technol 2012 Jul 13;2(4):225-233. [doi: 10.1007/s12553-012-0031-6]
47. Choong Y, Salvendy G. Design of icons for use by Chinese in mainland China. Interacting with Computers 1998 Feb;9(4):417-430 [FREE Full text] [doi: 10.1016/S0953-5438(97)00026-X]
48. Eagle PF, Marcos LR. Factors in medical students' choice of psychiatry. Am J Psychiatry 1980 Apr;137(4):423-427. [doi: 10.1176/ajp.137.4.423] [Medline: 7361926]
49. Gibbons MC. eHealth solutions for healthcare disparities. New York: Springer; 2008.
50. Koh H, Gracia N, Alvarez M. Culturally and Linguistically Appropriate Services: Advancing Health with CLAS. N Engl J Med Internet 2014 Jul 17;371:198-201 [FREE Full text] [doi: 10.1056/NEJMp1404321]
51. Valdez R. Creating a Foundation for the Design of Culturally-Informed Health Information Technology. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
52. López L, Green AR, Tan-McGrory A, King R, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011 Oct;37(10):437-445. [Medline: 22013816]
53. US Department of Health and Human Services. National Partnership for Action to End Health Disparities. 2011. HHS Action Plan to Reduce Racial and Ethnic Health Disparities URL: http://www.minorityhealth.hhs.gov/npa/files/Plans/HHS/HHS_Plan_complete.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFbvXla1]
54. Cooper LA, Hill MN, Powe NR. Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. J Gen Intern Med 2002 Jun;17(6):477-486 [FREE Full text] [Medline: 12133164]
55. United States Census Bureau. American Fact Finder. 2016 Jun. Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States, States, and Counties: April 1, 2010 to July 1, 2015 URL: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk [accessed 2016-07-24] [WebCite Cache ID 6jFcQpvvg]
56. Agency for Healthcare Research and Quality. 1975 Dec. 2015 National Healthcare Quality and Disparities Report and 5th Anniversary Update on the National Quality Strategy URL: http://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/nhqdr15/2015nhqdr.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFcg1sAx]
57. Lopez M, Gonzalez-Barrera A, Patten E. Pew Research Center. 2013 Mar 07. Closing the Digital Divide: Latinos and Technology Adoption URL: http://www.pewhispanic.org/2013/03/07/closing-the-digital-divide-latinos-and-technology-adoption/ [accessed 2016-07-24] [WebCite Cache ID 6jFcrCjfP]
58. DuBard A, Gizlice Z. Language Spoken and Differences in Health Status, Access to Care, and Receipt of Preventive Services Among US Hispanics. Am J Public Health 2008 Nov;98(11):2021-2028. [doi: 10.2105/AJPH.2007.119008]

59. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
60. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
61. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
62. Ulrichsweb: Global serials directory.: ProQuest, LLC URL: https://ulrichsweb.serialssolutions.com/ [accessed 2016-07-25] [WebCite Cache ID 6jGYakFaM]
63. US Census Bureau. Census Regions and Divisions of the United States. URL: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFdhL7bf]
64. Trochim W, Donnelly J. Research Methods Knowledge Base. 2006 Oct 20. The Research Methods Knowledge Base: Types of designs URL: http://www.socialresearchmethods.net/kb/ [accessed 2016-07-24] [WebCite Cache ID 6jFdo4XjM]
65. Aitken M. IMS Health Incorporated. 2013 Oct 30. Patient Apps for Improved Healthcare: From Novelty to Mainstream URL: http://www.imshealth.com/en/thought-leadership/ims-institute/reports/patient-apps-for-improved-healthcare [accessed 2016-07-24] [WebCite Cache ID 6jFfc35Az]
66. Arora S, Burner E, Terp S, Nok LC, Nercisian A, Bhatt V, et al. Improving attendance at post-emergency department follow-up via automated text message appointment reminders: a randomized controlled trial. Acad Emerg Med 2015 Jan;22(1):31-37 [FREE Full text] [doi: 10.1111/acem.12503] [Medline: 25388481]
67. Brown SA, Duchin SP, Villagomez ET. Diabetes education in a Mexican-American population: pilot testing of a research-based videotape. Diabetes Educ 1992;18(1):47-51. [Medline: 1370266]
68. Calderón JL, Bazargan M, Sangasubana N, Hays RD, Hardigan P, Baker RS. A comparison of two educational methods on immigrant Latinas breast cancer knowledge and screening behaviors. J Health Care Poor Underserved 2010 Aug;21(3 Suppl):76-90 [FREE Full text] [doi: 10.1353/hpu.0.0364] [Medline: 20675947]
69. Collins TC, Dong F, Ablah E, Parra-Medina D, Cupertino P, Rogers N, et al. Text messaging to motivate exercise among Latino adults at risk for vascular disease: a pilot study, 2013. Prev Chronic Dis 2014;11:E192 [FREE Full text] [doi: 10.5888/pcd11.140219] [Medline: 25357260]
70. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
71. Frates J, Bohrer GG, Thomas D. Promoting organ donation to Hispanics: the role of the media and medicine. J Health Commun 2006;11(7):683-698. [doi: 10.1080/10810730600934542] [Medline: 17074735]
72. Freda MC, Damus K, Andersen HF, Brustman LE, Merkatz IR. A "PROPP" for the Bronx: preterm birth prevention education in the inner city. Obstet Gynecol 1990 Jul;76(1 Suppl):93S-96S. [Medline: 2359588]
73. Gilliam M, Tapia V, Goldstein C. Increasing Contraceptive Use Among Sexually Active Latinas: Evaluation of a Self-Efficacy Enhancing Videotape. Hispanic Health Care International 2003;2(2):63-72.
74. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464. [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
75. Gerber B, Lawless K, Smolin L, Brodsky I, Girotti M, Pelaez L, et al. Diabetes and Your Eyes: A Pilot Study on Multimedia Education for Underserved Populations. In: American Medical Informatics Association. 2002 Presented at: American Medical Informatics Association; November 9, 2002; San Antonio, TX.
76. Makoul G, Cameron KA, Baker DW, Francis L, Scholtens D, Wolf MS. A multimedia patient education program on colorectal cancer screening increases knowledge and willingness to consider screening among Hispanic/Latino patients. Patient Educ Couns 2009 Aug;76(2):220-226. [doi: 10.1016/j.pec.2009.01.006] [Medline: 19250791]
77. Matthews PH, Darbisi C, Sandmann L, Galen R, Rubin D. Disseminating health information and diabetes care for Latinos via electronic information kiosks. J Immigr Minor Health 2009 Dec;11(6):520-526. [doi: 10.1007/s10903-008-9134-6] [Medline: 18392935]
78. McDonald DD, Gifford T, Walsh S. Effect of a virtual pain coach on older adults' pain communication: a pilot study. Pain Manag Nurs 2011 Mar;12(1):50-56. [doi: 10.1016/j.pmn.2009.10.001] [Medline: 21349449]
79. Osilla KC, D'Amico EJ, Díaz-Fuentes CM, Lara M, Watkins KE. Multicultural web-based motivational interviewing for clients with a first-time DUI offense. Cultur Divers Ethnic Minor Psychol 2012 Apr;18(2):192-202 [FREE Full text] [doi: 10.1037/a0027751] [Medline: 22506822]
80. Porter S, Chapman-Novakofski K, Scherer J. Your Guide to Diet and Diabetes: web-based diabetes education tailored to Hispanics. Journal of Nutrition Education and Behavior 2009;41(5):374-376. [doi: 10.1016/j.jneb.2009.04.003]
81. Quinn GP, Thomas KB, Hauser K, Rodríguez NY, Rodriguez-Snapp N. Evaluation of educational materials from a social marketing campaign to promote folic acid use among Hispanic women: insight from Cuban and Puerto Rican ethnic subgroups. J Immigr Minor Health 2009 Oct;11(5):406-414. [doi: 10.1007/s10903-008-9162-2] [Medline: 18561023]

82. Reuland DS, Ko LK, Fernandez A, Braswell LC, Pignone M. Testing a Spanish-language colorectal cancer screening decision aid in Latinos with limited English proficiency: results from a pre-post trial and four month follow-up survey. BMC Med Inform Decis Mak 2012;12:53 [FREE Full text] [doi: 10.1186/1472-6947-12-53] [Medline: 22691191]
83. Rosas LG, Trujillo C, Camacho J, Madrigal D, Bradman A, Eskenazi B. Acceptability of health information technology aimed at environmental health education in a prenatal clinic. Patient Educ Couns 2014 Nov;97(2):244-247 [FREE Full text] [doi: 10.1016/j.pec.2014.07.018] [Medline: 25085548]
84. Scheinmann R, Chiasson MA, Hartel D, Rosenberg TJ. Evaluating a bilingual video to improve infant feeding knowledge and behavior among immigrant Latina mothers. J Community Health 2010 Oct;35(5):464-470. [doi: 10.1007/s10900-009-9202-4] [Medline: 20039195]
85. Stockwell MS, Hofstetter AM, DuRivage N, Barrett A, Fernandez N, Vargas CY, et al. Text message reminders for second dose of influenza vaccine: a randomized controlled trial. Pediatrics 2015 Jan;135(1):e83-e91 [FREE Full text] [doi: 10.1542/peds.2014-2475] [Medline: 25548329]
86. Suarez-Balcazar Y, Friesema J, Lukyanova V. Culturally competent interventions to address obesity among African American and Latino children and youth. Occup Ther Health Care 2013 Apr;27(2):113-128. [doi: 10.3109/07380577.2013.785644] [Medline: 23855570]
87. Thompson DA, Joshi A, Hernandez RG, Bair-Merritt MH, Arora M, Luna R, et al. Nutrition education via a touchscreen: a randomized controlled trial in Latino immigrant parents of infants and toddlers. Acad Pediatr 2012;12(5):412-419. [doi: 10.1016/j.acap.2012.03.020] [Medline: 22682718]
88. Valdez A, Banerjee K, Ackerson L, Fernandez M. A multimedia breast cancer education intervention for low-income Latinas. J Community Health 2002 Feb;27(1):33-51. [Medline: 11845940]
89. Vaughn S. Stroke and heart disease prevention education via telenovela: a focus group's evaluation. Rehabil Nurs 2012;37(5):215-219. [doi: 10.1002/rnj.53] [Medline: 22949274]
90. West AM, Bittner EA, Ortiz VE. The effects of preoperative, video-assisted anesthesia education in Spanish on Spanish-speaking patients' anxiety, knowledge, and satisfaction: a pilot study. J Clin Anesth 2014 Jun;26(4):325-329. [doi: 10.1016/j.jclinane.2013.12.008] [Medline: 24882604]
91. Zyskind A, Jones K, Pomerantz K, LaFaye BA. Exploring the use of computer-based patient education resources to enable diabetic patients from underserved populations to self-manage their disease. Information Services and Use 2009;29:29-43. [doi: 10.3233/ISU-2009-0592]
92. Alvaro EM, Jones SP, Robles ASM, Siegel J. Hispanic organ donation: impact of a Spanish-language organ donation campaign. J Natl Med Assoc 2006 Jan;98(1):28-35. [Medline: 16532975]
93. Alvaro EM, Siegel JT, Crano WD, Dominick A. A mass mediated intervention on Hispanic live kidney donation. J Health Commun 2010 Jun;15(4):374-387. [doi: 10.1080/10810731003753133] [Medline: 20574876]
94. Lalonde B, Rabinowitz P, Shefsky ML, Washienko K. La Esperanza del Valle: alcohol prevention novelas for Hispanic youth and their families. Health Educ Behav 1997 Oct;24(5):587-602. [Medline: 9307895]
95. Wilkin HA, Valente TW, Murphy S, Cody MJ, Huang G, Beck V. Does entertainment-education work with Latinos in the United States? Identification and the effects of a telenovela breast cancer storyline. J Health Commun 2007;12(5):455-469. [doi: 10.1080/10810730701438690] [Medline: 17710596]
96. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-754.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
97. Heisler M, Choi H, Palmisano G, Mase R, Richardson C, Fagerlin A, et al. Comparison of community health worker-led diabetes medication decision-making support for low-income Latino and African American adults with diabetes using e-health tools versus print materials: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S13-S22 [FREE Full text] [doi: 10.7326/M13-3012] [Medline: 25402398]
98. Aragones A, Schwartz M, Shah N, Gany F. A randomized controlled trial of a multilevel intervention to increase colorectal cancer screening among Latino immigrants in a primary care facility. J Gen Intern Med 2010 Jun;25(6):564-567 [FREE Full text] [doi: 10.1007/s11606-010-1266-4] [Medline: 20213208]
99. Alcalay R, Alvarado M, Balcazar H, Newman E, Huerta E. Salud para su Corazón: a community-based Latino cardiovascular disease prevention and outreach model. J Community Health 1999 Oct;24(5):359-379. [Medline: 10555925]
100. de Nuncio ML, Price SA, Tjoa T, Lashuay N, Jones MC, Elder JP. Pretesting Spanish-language educational radio messages to promote timely and complete infant immunization in California. J Community Health 1999 Aug;24(4):269-284. [Medline: 10463471]
101. Steckler A, Allegrante JP, Altman D, Brown R, Burdine JN, Goodman RM, et al. Health education intervention strategies: recommendations for future research. Health Educ Q 1995 Aug;22(3):307-328. [Medline: 7591787]
102. Chen W. The Relationship between Health Education and Health Promotion: A Personal Perspective. American Journal of Health Education 2001 Dec;32(6):369-370. [doi: 10.1080/19325037.2001.10603499]
103. Peña-Purcell N. Hispanics' use of Internet health information: an exploratory study. J Med Libr Assoc 2008 Apr;96(2):101-107 [FREE Full text] [doi: 10.3163/1536-5050.96.2.101] [Medline: 18379664]

104. Jackson K. Building cultural competence: A systematic evaluation of the effectiveness of culturally sensitive interventions with ethnic minority youth. Children and Youth Services Review 2009 Nov;31(11):1192-1198. [doi: 10.1016/j.childyouth.2009.08.001]
105. Teal CR, Street RL. Critical elements of culturally competent communication in the medical encounter: a review and model. Soc Sci Med 2009 Feb;68(3):533-543. [doi: 10.1016/j.socscimed.2008.10.015] [Medline: 19019520]
106. Kohli H, Huber R, Faul A. Historical and Theoretical Development of Culturally Competent Social Work Practice. J Teach Soc Work 2010 Aug 17;30(3):252-271. [doi: 10.1080/08841233.2010.499091]
107. Whittemore R. Culturally competent interventions for Hispanic adults with type 2 diabetes: a systematic review. J Transcult Nurs 2007 Apr;18(2):157-166. [doi: 10.1177/1043659606298615] [Medline: 17416718]
108. Chin JL. Culturally competent health care. Public Health Rep 2000;115(1):25-33 [FREE Full text] [Medline: 10968582]
109. Guarnaccia RO, Rodriguez O. Concepts of Culture and Their Role in the Development of Culturally Competent Mental Health Services. Hispanic Journal of Behavioral Sciences 1996 Nov 01;18(4):419-443. [doi: 10.1177/07399863960184001]
110. Saha S, Beach MC, Cooper LA. Patient centeredness, cultural competence and healthcare quality. J Natl Med Assoc 2008 Nov;100(11):1275-1285 [FREE Full text] [Medline: 19024223]
111. Marcus A, Gould E. Crosscurrents: cultural dimensions and global Web user-interface design. Interactions 2000;7(4):32-46. [doi: 10.1145/345190.345238]
112. Livingston G, Minushkin S, Cohn D. Pew Research Center. 2008. Hispanics and Health Care in the United States: Access, Information and Knowledge URL: http://www.pewhispanic.org/2008/08/13/hispanics-and-health-care-in-the-united-states-access-information-and-knowledge/ [accessed 2016-07-24] [WebCite Cache ID 6jFgVOVFu]
113. Choi B, Lee I, Kim J. Culturability in Mobile Data Services: A Qualitative Study of the Relationship Between Cultural Characteristics and User-Experience Attributes. International Journal of Human-Computer Interaction 2006 Jul;20(3):171-203. [doi: 10.1207/s15327590ijhc2003_2]
114. Singh N, Baack D, Kundu S, Hurtado C. U.S. Hispanic consumer E-commerce preferences: Expectations and attitudes toward Web content. Journal of Electronic Commerce Research 2008;9(2):162-175 [FREE Full text]
115. Victorson D, Banas J, Smith J, Languido L, Shen E, Gutierrez S, et al. eSalud: designing and implementing culturally competent ehealth research with latino patient populations. Am J Public Health 2014 Dec;104(12):2259-2265. [doi: 10.2105/AJPH.2014.302187] [Medline: 25320901]
116. Sachau L, Hutchinson S. Trends in culturally relevant interface design features for Latino Web site users. Education Tech Research Dev 2012 Aug 14;60(6):1033-1050. [doi: 10.1007/s11423-012-9270-5]
117. Siegel R, Miller K, Jemal A. American Cancer Society. 2015. Cancer Facts & Figures for Hispanics/Latinos 2015-2017 URL: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046405.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFhP8Cbp]
118. Arcury TA, Nguyen HT, Summers P, Talton JW, Holbrook LC, Walker FO, et al. Lifetime and current pesticide exposure among Latino farmworkers in comparison to other Latino immigrants. Am J Ind Med 2014 Jul;57(7):776-787 [FREE Full text] [doi: 10.1002/ajim.22324] [Medline: 24737498]
119. Grzywacz JG, Quandt SA, Vallejos QM, Whalley LE, Chen H, Isom S, et al. Job demands and pesticide exposure among immigrant Latino farmworkers. J Occup Health Psychol 2010 Jul;15(3):252-266 [FREE Full text] [doi: 10.1037/a0019303] [Medline: 20604632]
120. Sandberg J, Spears Johnson CR, Nguyen HT, Talton JW, Quandt SA, Chen H, et al. Mobile and Traditional Modes of Communication Among Male Latino Farmworkers: Implications for Health Communication and Dissemination. J Immigr Minor Health 2016 Jun;18(3):522-531 [FREE Full text] [doi: 10.1007/s10903-015-0299-5] [Medline: 26463228]
121. National Council of La Raza. An Inside Look at Chronic Disease and Health Care among Hispanics in the United States. 2014 Jul 14. URL: http://www.lchc.org/wp-content/uploads/Chronic_Disease_Report_2014.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFi1aroe]
122. Goel MS, Gracia G, Baker DW. Development and pilot testing of a culturally sensitive multimedia program to improve breast cancer screening in Latina women. Patient Educ Couns 2011 Jul;84(1):128-131. [doi: 10.1016/j.pec.2010.05.014] [Medline: 20638219]
123. Calderón JL, Shaheen M, Hays RD, Fleming ES, Norris KC, Baker RS. Improving Diabetes Health Literacy by Animation. Diabetes Educ 2014 Mar 27;40(3):361-372. [doi: 10.1177/0145721714527518] [Medline: 24676274]
124. Byrd TL, Wilson KM, Smith JL, Coronado G, Vernon SW, Fernandez-Esquer ME, et al. AMIGAS: a multicity, multicomponent cervical cancer prevention trial among Mexican American women. Cancer 2013 Apr 1;119(7):1365-1372 [FREE Full text] [doi: 10.1002/cncr.27926] [Medline: 23280399]
125. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
126. Meinert CL, Gilpin AK, Unalp A, Dawson C. Gender representation in trials. Control Clin Trials 2000 Oct;21(5):462-475. [Medline: 11018563]
127. US Census Bureau. QuickFacts United States URL: http://www.census.gov/quickfacts/ [accessed 2016-07-24] [WebCite Cache ID 6jFiDmJ13]

128. Bennett K, Olatosi B, Probst J. South Carolina Rural Health Research Center. Health Disparities: A Rural – Urban Chartbook URL: http://rhr.sph.sc.edu/report/(7-3)%20Health%20Disparities%20A%20Rural%20Urban%20Chartbook%20-%20Distribution%20Copy.pdf [accessed 2016-07-24] [WebCite Cache ID 6jFiQ3exw]
129. Ory MG, Lee SM, Mier N, Wernicke MM. The science of sustaining health behavior change: the health maintenance consortium. Am J Health Behav 2010;34(6):647-659 [FREE Full text] [Medline: 20604691]
130. 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]
131. O'Sullivan R. Improving evaluation design and use through the "Evaluation Crosswalk" method. National Forum of Applied Educational Research Journal 1991;4:43-49. [doi: 10.1016/j.evalprogplan.2011.12.007]
132. Misra S. Randomized double blind placebo control studies, the "Gold Standard" in intervention based studies. Indian J Sex Transm Dis 2012 Jul;33(2):131-134 [FREE Full text] [doi: 10.4103/0253-7184.102130] [Medline: 23188942]
133. Schulz KF, Altman DG, Moher D. [CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials (Chinese version)]. J Chinese Integr Med 2010 Jul;8(7):604-612 [FREE Full text] [Medline: 20619135]
134. 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]
135. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
136. Bolin JN, Ory MG, Wilson AD, Salge L. Diabetes education kiosks in a latino community. Diabetes Educ 2013;39(2):204-212. [doi: 10.1177/0145721713476346] [Medline: 23435404]
137. Calles-Escandón J, Hunter JC, Langdon SE, Gómez EM, Duren-Winfield VT, Woods KF. La Clínica del Pueblo: a model of collaboration between a private media broadcasting corporation and an academic medical center for health education for North Carolina Latinos. J Immigr Minor Health 2009 Dec;11(6):513-519. [doi: 10.1007/s10903-008-9223-6] [Medline: 19156524]
138. Jerant A, Kravitz RL, Sohler N, Fiscella K, Romero RL, Parnes B, et al. Sociopsychological tailoring to address colorectal cancer screening disparities: a randomized controlled trial. Ann Fam Med 2014;12(3):204-214 [FREE Full text] [doi: 10.1370/afm.1623] [Medline: 24821891]
139. Leeman-Castillo B, Beaty B, Raghunath S, Steiner J, Bull S. LUCHAR: using computer technology to battle heart disease among Latinos. Am J Public Health 2010 Feb;100(2):272-275 [FREE Full text] [doi: 10.2105/AJPH.2009.162115] [Medline: 20019305]
140. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c332-c332. [doi: 10.1136/bmj.c332]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
IEEE: Institute of Electrical and Electronics Engineers
IT: information technology
MeSH: Medical Subject Headings
PAM: Patient Activation Measure
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.04.16; peer-reviewed by D Tao, C Or, T Veinot, N Vandenberg; comments to author 11.05.16; revised version received 22.06.16; accepted 08.07.16; published 10.08.16.

Please cite as:
Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R
Spanish-Language Consumer Health Information Technology Interventions: A Systematic Review
J Med Internet Res 2016;18(8):e214
URL: http://www.jmir.org/2016/8/e214/
doi:10.2196/jmir.5794
PMID:27511437

Original Paper

An Evaluation and Ranking of Children's Hospital Websites in the United States

Timothy R Huerta[1], MS, PhD; Daniel M Walker[1], MPH, PhD; Eric W Ford[2], MPH, PhD

[1]The Ohio State University, College of Medicine, Department of Family Medicine, Columbus, OH, United States

[2]Johns Hopkins Bloomberg School of Public Health, Department of Health Policy and Management, Baltimore, MD, United States

Corresponding Author:
Timothy R Huerta, MS, PhD
The Ohio State University, College of Medicine
Department of Family Medicine
2231 North High St.
Rm 265
Columbus, OH, 43201
United States
Phone: 1 6143666636
Fax: 1 6142932715
Email: timothy.huerta@osumc.edu

Abstract

Background: Children's hospitals are faced with the rising need for technological innovation. Their prospective health care consumers, who increasingly depend on the Web and social media for communication and consumer engagement, drive this need. As patients and family members navigate the Web presence of hospitals, it is important for these specialized organizations to present themselves and their services efficiently.

Objective: The purpose of this study was to evaluate the website content of children's hospitals in order to identify opportunities to improve website design and create benchmarks to judge improvement.

Methods: All websites associated with a children's hospital were identified using a census list of all children's hospitals in the United States. In March of 2014, each website and its social media were evaluated using a Web crawler that provided a 5-dimensional assessment that included website accessibility, marketing, content, technology, and usability. The 5-dimensional assessment was scored on a scale ranging from 0 to 10 with positive findings rated higher on the scale. Websites were ranked by individual dimensions as well as according to their average ranking across all dimensions.

Results: Mean scores of 153 websites ranged from 5.05 to 8.23 across all 5 dimensions. Results revealed that no website scored a perfect 10 on any dimension and that room exists for meaningful improvement.

Conclusions: Study findings allow for the establishment of baseline benchmarks for tracking future website and social media improvements and display the need for enhanced Web-based consumer engagement for children's hospitals.

(J Med Internet Res 2016;18(8):e228) doi:10.2196/jmir.5799

KEYWORDS

pediatric hospital; Internet; social media; patient participation; patient education

Introduction

More and more, patients seek health-related information from Internet sources [1]. The trend is particularly strong among teens and young adults, as these demographics show the highest rate of Internet usage both in general and specifically for health information [2-4]. The increased use of Web-based technology to seek health information provides opportunities for health care organizations to create innovative platforms to educate and engage their consumers. Despite the promise of greater information availability, consumer-focused Web interfaces have not advanced as quickly in the health care industry as in other sectors [5].

By virtue of the fact that children's hospitals serve younger demographics, they may be particularly well positioned to capitalize from an effective Web presence to attract new patients and provide high-fidelity information [6-8]. For example, a study of diabetic children's parents found that they have

significant desire for a Web-based platform to augment clinical services. However, participants expressed concerns, including whether the websites' information was current and their own abilities to discern whether information was credible [9,10]. Therefore, information sources may not be reliable and children may lack the health literacy needed to assess content credibility [11].

Children's hospital websites are well positioned to explore concerns about consumer-targeted information's reliability and relevance [12]. In addition, they can act as a portal for patient-specific health-related information and serve as a comprehensive source to engage patients with the health care system early in their lives. Thus, it is important to understand the current state of the art in children's hospital websites for consumers, providers, and health information content creators. The purpose of this short paper is to identify the degree to which children's hospital websites adhere to Internet industry standards for usability [13]. Using Internet industry standards provides a reference point against which to assess the children's hospital websites and enables the establishment of benchmarks that can help children's hospitals identify areas for improvement as they aim to optimize the use of Web-based technology to engage their patients.

Methods

A list of 157 US children's hospitals with unique website domains linked to the named institutions was compiled from the Children's Hospital Association's membership in 2014. For websites that were part of a general hospital domain, only the sections of the website pertaining specifically to the children's hospital were included in the analysis. Website analyses were simultaneously performed in March of 2014.

Procedures for the website analytic tool paralleled previous health care website assessments [14,15]. The automated tool, termed the "Web crawler," was created to navigate and assess websites' functionalities. The Web crawler assessed 24 metrics about each website (eg, broken links, content readability). A complete list of these indices and their description is provided in Multimedia Appendix 1. Different subsets of the 24 metrics were weighted and combined to create 5 indices: accessibility, content, marketing, technology, and usability (see Multimedia Appendix 1 for weighting) [15,16].

The analysis algorithm begins with the Web crawler developing a map of the website from its front or "top" page down through all pages associated with the children's hospital Web domain to a maximum of 1000 pages' breadth. Once the sampled pages for each hospital are mapped, the systematic tool surveys the pages for each component that comprises each of the 5 dimensions. The dimension of accessibility refers to the ability of those with low levels of computer literacy to access and navigate the hospital's Web presence. The marketing dimension is composed of items such as a website's ability to be found through search engines, examining the relevance of descriptions to the links provided, and ranking of the website in performed searches. Content dimensionality refers to items such as the grammar of website text, the frequency of information updates, material relevancy, and readability metrics. The technology dimension is scored by how quickly the website downloads, the quality of the programming code, the website's infrastructure, and how it performs. The fifth dimension of website assessment is usability. The usability dimension provides an overall assessment of website quality and is formed as a composite of the metrics used in the other indices, rather than a composite of the other indices themselves. Please see Huerta et al [15] for a more detailed breakdown and itemization of the process.

Websites were assigned a score for each of the 5 dimensions, ranging from 0 to 10, with a comparatively better rating reflected with a higher score. After dimension scores were determined, hospitals were assigned a rank for their score in each dimension. An overall rank was determined by averaging the ranks of all 5 dimensions. For the overall rank, we used the average rank from the 5 dimensions rather than the average score to keep the relative nature of the scores intact in our overall ranking.

Results

During the analytic process, technical issues prevented running the Web crawler on 4 of the 157 sites, leaving them unanalyzed and prompting their removal from the test sample. These technical issues were likely due to tools on the website servers that prevented "pings" against their website. However, the exact cause of these technical issues remains unknown, as they arose from issues with the server rather than the Web crawler. Of the 5 dimensions evaluating Web domains for the included 153 children's hospitals in the United States, the dimension of content had the highest average hospital website score at 8.23 on a scale from 1 to 10, as well as the largest range. The accessibility dimension had the lowest average assessment value at 5.05. Mean scores for the other dimensions ranged from 5.36 to 6.7. Summary statistics for the 5 dimensions are presented in Table 1.

Table 1. Children's hospital website summary statistics for each dimension assessed by the Web crawler.

Dimension	Mean	Standard deviation	Minimum	Maximum
Accessibility	5.05	1.29	0.7	7.4
Content	8.23	1.15	1.2	9.7
Marketing	6.73	1.17	1.7	8.8
Technology	5.36	1.87	0	7.7
Usability	6.13	1.39	1.7	8.1

The overall rankings, and the rankings and scores for each dimension, for the top 100 overall rated children's hospital websites are reported in Multimedia Appendix 2. On the basis of rankings, the top overall performing children's hospital website was the Children's Hospital of Philadelphia. This hospital had the top performing website for both marketing and technology, the 2nd highest score for usability, the 6th highest for content, and the 10th best for accessibility.

Discussion

Principal Findings

This study provides a comprehensive and thorough assessment of children's hospital websites. The growing demand for technological innovation combined with an increasing reliance on Web-based information among the age demographic most heavily represented in children's hospitals is essential for these hospitals to bear in mind. A well-designed website that adheres to national standards and demonstrated best practices is of greater importance for issues related to care access, as infirm children and their families may be ill-equipped to overcome barriers to receiving care [17,18].

The results of this study can help children's hospitals to assess their own Web presence relative to their peers, providing a basis to improve Web services. Hospitals ranked lower on the list within each domain, or overall, can look to the top performers for guidance on how best to design their websites. The difference between low and high rankings may not require complete redesign, but rather may entail straightforward solutions, such as fixing broken links, correcting misspelled items, adding Twitter links, or offering Spanish-language services—all common yet easily fixable pitfalls. Moreover, the high-ranking hospitals appear to use their websites to orient, engage, and inform their visitors, rather than solely to provide visitors with basic information. These high-ranking hospitals view their websites as critical to an ongoing relationship with patients. The lesson for the lower-ranked hospitals is that a website can enlist features that promote health and wellness and also engage patients in their care. For example, websites can host Web-based interventions [19] or offer useful patient education materials and disease management materials [20].

Comparison With Prior Work

The authors are unaware of any similar previous studies done on children's hospitals. As a result, no prior work exists that provides a useful comparison for the findings in this study. However, earlier in 2014, a study with similar methodology was performed on 2407 acute care general hospital website domains [15]. For a point of relative comparison, a side-by-side comparison of the two studies is presented in Table 2. The study found a mean usability score of 5.16 for US hospitals, which is comparably lower to the score found in the usability dimension in our study. This difference may stem from a higher likelihood of children's hospitals to be connected with large, academic facilities. Both large and academic medical centers may have more resources available for Web design, thereby producing higher scores for children's hospitals on average. Children's hospitals could also recognize the opportunity to use their websites to educate and engage patients in their target demographic relatively more than general hospitals.

Table 2. Comparison of Web crawler analytic scores for websites of US acute care general hospitals [15] and children's hospitals.

Dimension	US hospitals[a] Mean (SE[b])	Children's hospitals Mean (SE)
Accessibility	5.08 (0.05)	5.05 (0.10)
Content	6.49 (0.02)	8.23 (0.09)
Marketing	5.03 (0.03)	6.73 (0.09)
Technology	4.43 (0.04)	5.36 (0.15)
Usability	5.16 (0.03)	6.13 (0.11)

[a]Acute care general hospitals.

[b]SE: standard error.

Limitations

Several important limitations affect this study. Previous studies of this nature have excluded facilities associated with an education top-level domain (.edu) [14,15]; however, this domain is prevalent enough in children's hospitals to warrant inclusion. As a result, there is a risk of inclusion of academic departments or other such pages unrelated to patient care or the hospital that could skew assessment. It should be noted that the hospitals are generally confined to a smaller subsection of a .edu domain, decreasing the risk of an unrelated academic side being evaluated.

The second limitation influencing this study relates to website size. A website can have several pages associated with it or can be more limited in scope. The Web crawler does not directly adjust for this size component; this results in websites with a few Web pages and narrowly focused high-quality content potentially being rated the same as websites with numerous pages and equivalently high-quality content that is broader in scope. Thus, interpreting a single overall score, although convenient, may negate the diversity of hospitals' Web design.

Third, our study assessed websites in the spring of 2014. The Internet is a rapidly evolving landscape, websites are continuously improving, and the best demonstrated practices are evolving. As a result, our findings may not represent the

current status of each hospital's websites. Nonetheless, we view our assessment of children's hospital website design as establishing a baseline with which future comparisons can be made and against which progress can be measured.

Finally, our use of a Web crawler as the key measurement tool in this study biases our results toward items discernible by the technology. More nuanced website features may be indiscernible by the Web crawler: for example, the Web crawler may miss website layout issues that influence usability. Alternatively, the use of an established Web crawler does ensure high levels of validity and reliability to our assessment. Ultimately, our study achieves its aim to evaluate the websites to better understand the development of this infrastructure but does not evaluate end-user satisfaction and engagement that may lead to website use and impact. Understanding how specifically consumers are interacting with technology is a ripe area for future research.

Conclusions

This analysis presents a systematic assessment of children's hospital Web and social media presence. Giving consideration to the increasing societal expectations for technology, in particular those of the younger demographic, the number of poorly performing facilities across all the calculated scores is a cause for concern in the near term. The social media and Web presence of children's hospitals may represent the first contact patients make with an organization. This initial contact must engage the patient, or else they may explore alternative treatment options or not recognize care services available to them. This consequence may not only have a negative impact on the hospital, particularly in more competitive markets, but also lead to suboptimal care processes and outcomes for the patient. Improving website design, complying with Internet industry standards, and optimizing search engine performance can maximize the potential power of the Internet to engage and inform patients.

Conflicts of Interest

The authors are faculty employed by their respective universities, some of which maintain a Web presence that was assessed in this study.

Multimedia Appendix 1

Scale components and weightings.

[PDF File (Adobe PDF File), 27KB - jmir_v18i8e228_app1.pdf]

Multimedia Appendix 2

Ranking of the top 100 children's hospital websites for each dimension and an average ranking across dimensions.

[PDF File (Adobe PDF File), 102KB - jmir_v18i8e228_app2.pdf]

References

1. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
2. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Research Center. 2010. Social Media and Young Adults: Acknowledgements and Methodology URL: http://www.pewinternet.org/2010/02/03/acknowledgements-and-methodology/ [accessed 2016-08-16] [WebCite Cache ID 6joGSPDJ4]
3. Fox S. Pew Research Center. 2011. Health Topics URL: http://www.pewinternet.org/2011/02/01/health-topics-2/ [accessed 2016-08-16] [WebCite Cache ID 6joGUYHyW]
4. Atkinson N, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
5. LaPenna A. New sources of revenue in the age of connectivity. J Healthc Manag 2009;54(1):11-14. [Medline: 19227850]
6. Pinsker J, Nguyen C, Young S, Fredericks GJ, Chan D. A pilot project for improving paediatric diabetes outcomes using a website: the Pediatric Diabetes Education Portal. J Telemed Telecare 2011;17(5):226-230. [doi: 10.1258/jtt.2010.100812] [Medline: 21565846]
7. Wagner S. Healthcare Informatics. 2010. More than just social. Hospital leaders discover the revenue cycle gains from social media URL: http://www.healthcare-informatics.com/article/more-just-social [accessed 2016-08-16] [WebCite Cache ID 6joGaiuEL]
8. Wong C, Merchant RM, Moreno MA. Using social media to engage adolescents and young adults with their health. Healthc (Amst) 2014 Dec;2(4):220-224 [FREE Full text] [doi: 10.1016/j.hjdsi.2014.10.005] [Medline: 25984444]
9. Nicholas D, Gutwin C, Paterson B. Examining preferences for website support to parents of adolescents with diabetes. Soc Work Health Care 2013;52(9):862-879. [doi: 10.1080/00981389.2013.827144] [Medline: 24117033]
10. Hanberger L, Ludvigsson J, Nordfeldt S. Use of a web 2.0 portal to improve education and communication in young patients with families: randomized controlled trial. J Med Internet Res 2013;15(8):e175 [FREE Full text] [doi: 10.2196/jmir.2425] [Medline: 23973555]

11. McPherson A, Gofine ML, Stinson J. Seeing is believing? A mixed-methods study exploring the quality and perceived trustworthiness of online information about chronic conditions aimed at children and young people. Health Commun 2014;29(5):473-482. [doi: 10.1080/10410236.2013.768325] [Medline: 24099647]
12. Lonzer J, Lonzer D, Medina M, Piedimonte G. Social media in pediatrics: a call for guidelines. J Pediatr 2015 Mar;166(3):511-512. [doi: 10.1016/j.jpeds.2014.11.032] [Medline: 25722261]
13. W3C. 2016. Standards URL: https://www.w3.org/standards/ [accessed 2016-08-16] [WebCite Cache ID 6joGcowl9]
14. Ford E, Huerta TR, Schilhavy RA, Menachemi N. Effective US health system websites: establishing benchmarks and standards for effective consumer engagement. J Healthc Manag 2012;57(1):47-64; discussion 64. [Medline: 22397104]
15. Huerta T, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
16. Huerta T, Walker DM, Ford EW. Cancer Center Website Rankings in the USA: Expanding Benchmarks and Standards for Effective Public Outreach and Education. J Cancer Educ 2015 Oct 16:- Epub ahead of print. [doi: 10.1007/s13187-015-0931-z] [Medline: 26472325]
17. Beckjord E, Finney RL, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
18. Zeng X, Parmanto B. Web content accessibility of consumer health information web sites for people with disabilities: a cross sectional evaluation. J Med Internet Res 2004 Jun 21;6(2):e19 [FREE Full text] [doi: 10.2196/jmir.6.2.e19] [Medline: 15249268]
19. Kelders S, 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. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]

Edited by G Eysenbach; submitted 25.03.16; peer-reviewed by N Diviani, C Kruse, P Schulz; comments to author 28.06.16; revised version received 13.07.16; accepted 03.08.16; published 22.08.16.

Please cite as:
Huerta TR, Walker DM, Ford EW
An Evaluation and Ranking of Children's Hospital Websites in the United States
J Med Internet Res 2016;18(8):e228
URL: http://www.jmir.org/2016/8/e228/
doi:10.2196/jmir.5799
PMID:27549074

Original Paper

Datathons and Software to Promote Reproducible Research

Leo Anthony Celi[1,2*], MSc, MPH, MD; Sharukh Lokhandwala[1,2,3*], MD; Robert Montgomery[1,2*], MD; Christopher Moses[1*], SB; Tristan Naumann[1*], MS; Tom Pollard[1*], PhD; Daniel Spitz[1*], BS; Robert Stretch[1,2*], MD

[1]Critical Data, Massachusetts Institute of Technology, Cambridge, MA, United States

[2]Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States

[3]Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Sharukh Lokhandwala, MD
Critical Data
Massachusetts Institute of Technology
77 Massachusetts Avenue
E25-505
Cambridge, MA, 02139
United States
Phone: 1 617 253 7937
Fax: 1 617 258 7859
Email: sharukh.lokhandwala@gmail.com

Abstract

Background: Datathons facilitate collaboration between clinicians, statisticians, and data scientists in order to answer important clinical questions. Previous datathons have resulted in numerous publications of interest to the critical care community and serve as a viable model for interdisciplinary collaboration.

Objective: We report on an open-source software called Chatto that was created by members of our group, in the context of the second international Critical Care Datathon, held in September 2015.

Methods: Datathon participants formed teams to discuss potential research questions and the methods required to address them. They were provided with the Chatto suite of tools to facilitate their teamwork. Each multidisciplinary team spent the next 2 days with clinicians working alongside data scientists to write code, extract and analyze data, and reformulate their queries in real time as needed. All projects were then presented on the last day of the datathon to a panel of judges that consisted of clinicians and scientists.

Results: Use of Chatto was particularly effective in the datathon setting, enabling teams to reduce the time spent configuring their research environments to just a few minutes—a process that would normally take hours to days. Chatto continued to serve as a useful research tool after the conclusion of the datathon.

Conclusions: This suite of tools fulfills two purposes: (1) facilitation of interdisciplinary teamwork through archiving and version control of datasets, analytical code, and team discussions, and (2) advancement of research reproducibility by functioning postpublication as an online environment in which independent investigators can rerun or modify analyses with relative ease. With the introduction of Chatto, we hope to solve a variety of challenges presented by collaborative data mining projects while improving research reproducibility.

(J Med Internet Res 2016;18(8):e230) doi:10.2196/jmir.6365

KEYWORDS

reproducibility of findings; big data; database; Internet; medical informatics

Introduction

A growing body of evidence suggests that high-quality data are lacking to guide clinician decision making. A systematic review of the joint American College of Cardiology and American Heart Association clinical practice guidelines revealed that only 314 of the 2711 recommendations were based on high-quality evidence [1]. Most of what clinicians do in practice has not been evaluated and is not covered by existing guidelines.

Furthermore, it is unlikely that there will be prospective randomized controlled trials, the gold standard of evidence-based medicine, to address all the information gaps in clinical practice.

Furthermore, the reliability of published research has been called into question. One study found that among 49 highly cited articles, 16% were subsequently contradicted, and another 16% were found to have significantly smaller effect sizes in subsequent studies [2]. In order to address some of these issues, a greater emphasis has been placed on sharing data, as well as fostering a more open peer review process.

The near ubiquitous implementation of electronic health records has allowed investigators to address treatment and diagnostic dilemmas where evidence-based guidelines are lacking; however, clinicians are often lacking the increasing expertise in data science required to acquire and synthesize such dat. In order to tackle this new frontier of medical research, it is paramount that clinicians and data scientists work together to harness the power of electronic health records.

In previous publications, we described the "datathon" model, which brings together the requisite experts from different scientific disciplines in a venue that supports collaboration, group learning, error checking, and methodological review during the initial design and subsequent phases of research [3,4].

In September 2015, the second international Critical Care Datathon was held simultaneously at the Massachusetts Institute of Technology (MIT) in Boston, USA, and in London, UK and Paris, France. The event coincided with the launch of Medical Information Mart for Intensive Care (MIMIC-III), the successor to Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-II), and an open-access database of patients admitted to an intensive care unit (ICU) at Beth Israel Deaconess Medical Center in Boston, MA, USA [5]. MIMIC-III spans the period from 2002 through 2012 and contains data on over 60,000 ICU admissions. Previous datathons have resulted in numerous publications of interest to the critical care communities. Using a diverse range of methodologies, groups have investigated the association between elevated central venous pressure and acute kidney injury [6], proton pump inhibition and cardiac arrhythmias [7], hyperdynamic left ventricular ejection fraction and mortality [8], diuretic use and obesity [9], and hypermagnesemia and blood pressure [10]. Pereira et al used fuzzy modeling to predict severely depressed left ventricular ejection fraction following ICU admission [11].

All participants at the 2015 datathon were encouraged to use open-source software called Chatto [12] during the event. This suite of tools was created by members of our group to fulfill two purposes: (1) facilitation of interdisciplinary teamwork through improved communication, as well as archiving and version control of datasets, analytical code, and team discussions, and (2) advancement of research reproducibility by functioning postpublication as an online environment in which independent investigators can rerun or modify analyses with relative ease. Chatto allows individual data scientists and teams to rapidly begin to explore, visualize, extract, and analyze data. Perhaps its most important role is in its simplification of the data analysis pipeline and subsequent capacity to improve reproducibility without requiring additional effort outside of the usual research workflow.

Chatto is composed of five key components: a project website, integration with a group chat service called Slack [13], integration with GitHub [14] for source code control, a Jupyter [15,16] notebook for interactive code development, and an open-source library [12] for connecting to data sources and transforming data (Figure 1, Figure 2, Figure 3). Figure 1 shows a sample workbook. Figure 2 and Figure 3 demonstrate a scatterplot and a histogram, respectively, which were made using the provided data transformations.

Figure 1. Sample workbook from Jupyter notebooks, which provide an interface for documentation of the research.

Let's plot number of days on ventilator vs days until death

```
In [11]: from chatto_transform.lib.chunks import left_join

In [12]: vent_death = left_join(vent_days, death_days, 'icustay_id')

In [13]: vent_death
```

Out[13]:

	subject_id_x	hadm_id_x	icustay_id	ventdays	subject_id_y	hadm_id_y	hosdead	time_of_death	time_until_death
0	3	145834	211552	4 days	3	145834	False	2102-06-14	225 days 10:02:00
1	9	150750	220597	5 days	9	150750	True	2149-11-14	-1 days +13:45:00
2	12	112213	232669	4 days	12	112213	True	2104-08-20	-1 days +21:03:00
3	13	143045	263738	2 days	13	143045	False	NaT	NaT
4	31	128652	254478	8 days	31	128652	True	2108-08-30	-1 days +09:00:00
...	...	...	...	...	...	...	...	...	...
20124	99939	159023	252442	1 days	99939	159023	False	NaT	NaT
20125	99955	108494	243255	5 days	99955	108494	True	2139-07-19	-1 days +18:40:00
20126	99982	151454	221194	1 days	99982	151454	False	2157-03-04	85 days 10:15:00
20127	99985	176670	279638	10 days	99985	176670	False	NaT	NaT
20128	99991	151118	226241	1 days	99991	151118	False	NaT	NaT

20129 rows × 9 columns

Figure 2. Generation of figures within Jupyter notebooks: sample scatterplot.

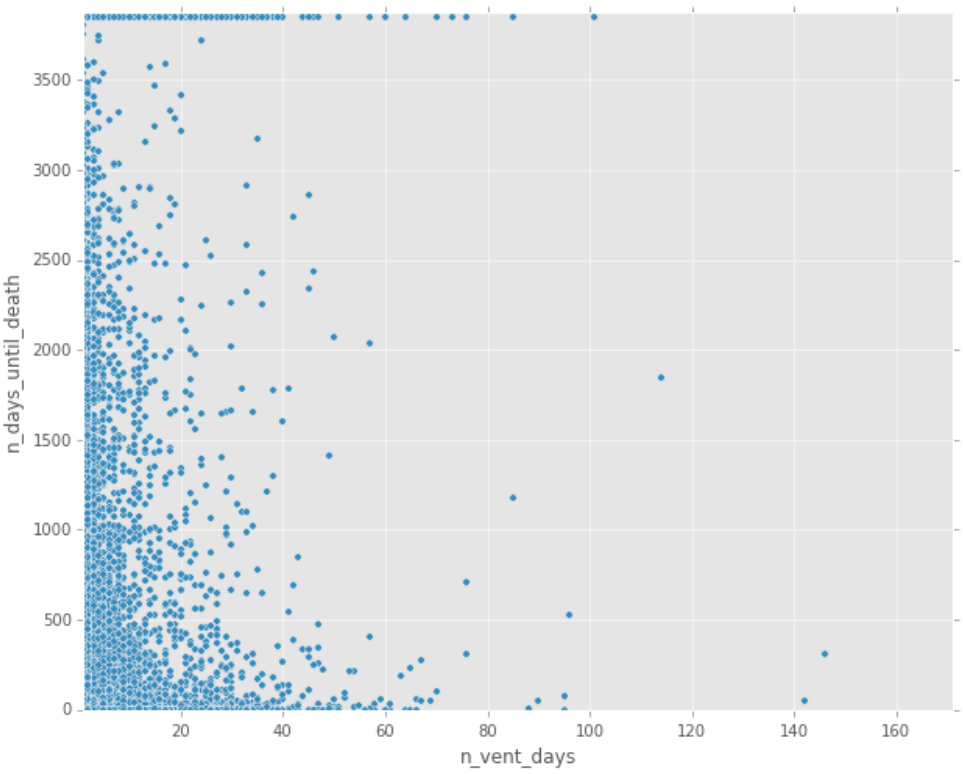
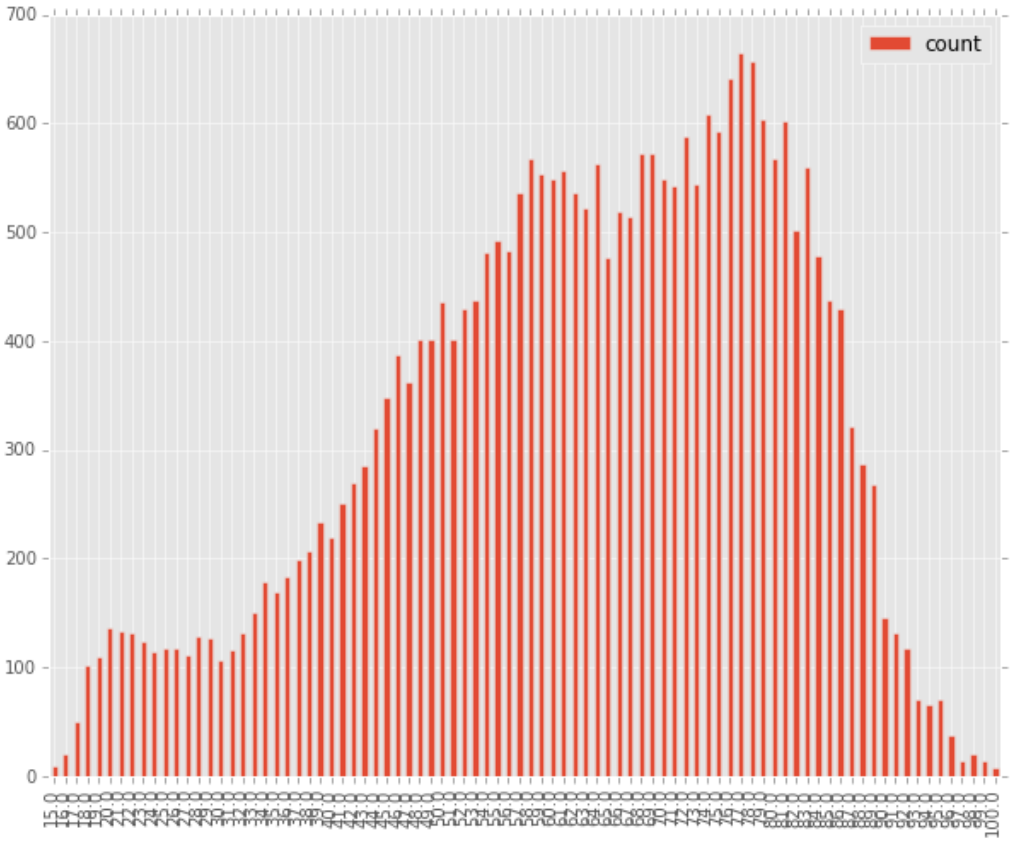


Figure 3. Generation of graphs within Jupyter notebooks: sample histogram.

Plot the histogram

(Axes need work...)

```
In [8]: age_count_df.plot(kind='bar')
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x10da75f28>
```



Importing the Blood Urea Nitrogen transform

Capturing the 'bun' recipe from the MIMIC cookbook:

```
select bucket, count(*) from (
  select width_bucket(valuenum, 0, 280, 280) as bucket
    from mimic2v26.labevents le,
```

Methods

Similar to prior datathons, the event began with introductory presentations outlining the program and sharing lessons learned from prior datathons. Afterward, participants pitched various clinical questions that could be addressed using the MIMIC-III database. Teams formed organically as clinicians and data scientists discussed potential research questions and the methods required to address them. Each multidisciplinary team spent the next 2 days with clinicians working alongside data scientists to write code, extract and analyze data, and reformulate their queries in real time as needed. All projects were then presented on the last day of the datathon to a panel of judges that consisted of clinicians and scientists.

Results

Event participants were exceptionally diverse. Of the 48 attendees at MIT, 23 (48%) were clinicians, of whom 18 (78%) were fellows, residents, or medical students. Of the 25 data scientists, 5 were postdoctoral associates (20%) and another 5 were graduate students (20%). Before the event, 32 participants (67%) had never used MIMIC and only 10 (21%) had used SQL queries frequently in their research work. Of the 8 teams, 6 subsequently submitted, and had accepted, abstracts to the American Thoracic Society international conference.

At the start of the event, all datathon participants were invited to visit a project website called Chatto Hub to register their team's project in a listing alongside other datathon projects. Chatto Hub users were authenticated using GitHub user

credentials. Integration with the GitHub application program interface (API) enabled a public code repository to be automatically created for each team upon project registration, based on a clone of a common parent repository that included code examples and documentation. Using the Slack API, a group chat "channel" was also automatically created for each team upon project registration, providing a means for each team to communicate and share files. Lastly, private Jupyter notebooks—browser-based interactive coding environments for data analysis and documentation—were automatically created for each user on each registered project.

The Jupyter notebooks were preconfigured to connect to a cloud-based instance of the MIMIC-III database hosted by Amazon Web Services (Amazon.com, Seattle, WA, USA). Each notebook also included MIMIC-specific data transformations—modular, reusable pieces of code—as well as the tools to create new transformations that could later be shared with the broader research community through GitHub. The result was a project listing on the Chatto Hub website that included the project title, short description, and links to both the project's Slack channel and GitHub repository.

Use of Chatto was particularly effective in the datathon setting, enabling teams to reduce the time spent configuring their research environments to just a few minutes—a process that would normally take hours to days. Chatto continued to serve as a useful research tool after the conclusion of the datathon. The Slack implementation allowed teams to continue to collaborate in a manner that was automatically documented and accessible to all team members. Furthermore, simplification of the data analysis pipeline through the use of Jupyter notebooks means that code published alongside each study is more easily interpretable. Independent researchers will also be able to rerun or modify the original analytical code with minimal effort.

Discussion

The issue of reproducibility among scientific publications has generated substantial concern in both the research and lay communities. Increased use of large, publicly accessible datasets in medical research has, to some extent, facilitated recognition of this problem by enabling independent investigators to reanalyze data at minimal cost. Nevertheless, the majority of published studies are not externally validated in this manner due to the persistence of significant technical and cultural barriers. These impediments include time-intensive configuration, inadequate documentation, code rot, "dependency hell," and a research environment that insufficiently rewards efforts to reproduce the work of other investigators.

Breaking down the technical barriers that stand in the way of research reproducibility is a ripe goal. For research involving large datasets, we have developed an online open-source tool that serves as a development environment leading up to the time of publication, while functioning postpublication as a playground for independent investigators to rerun the analysis, en bloc or piecemeal, with or without modification. As a proof-of-concept, this tool was used successfully in tandem with the MIMIC-III database during the most recent MIT Critical Care Datathon (September 2015), although connectors to any large research database can be easily generated.

As tools such as the one we have described continue to evolve, the responsibility of ensuring that the analytical code related to a given study is accessible, interpretable, and "runnable" with minimal effort should increasingly fall on the original investigators. The results of studies where authors have made inadequate efforts to enable and encourage others to examine, rerun, and modify their analysis should be viewed with prejudice, and their suitability for publication should be questioned.

Limitations

Despite the many benefits of integrating the services outlined above into a single software package, not all issues related to collaborative data mining were solved. Teams developed different methods for sharing the results of their analyses, including exporting intermediate results to comma separated value (CSV) files for distribution through the team's Slack channel. This highlighted that teams often divided their labor based on expertise, requiring each team member's results to be shared and then combined by one individual for more advanced analysis. In the future, enabling multiple team members to simultaneously edit a shared Jupyter notebook might help to resolve this problem. Jupyter notebooks were also not automatically source controlled. In the future, GitHub integration for hosted notebooks would dramatically simplify the research workflow and sharing of analytical code between team members. Lastly, although documentation existed for both MIMIC-III and each individual service integrated with Chatto, the lack of a single centralized location for this documentation presented a problem for some participants. This issue was mitigated during the event by datathon staff actively helping participants to troubleshoot technical problems, but this should be addressed through centralized, simplified documentation in future datathons.

Conclusion

The product of scientific research is not a number with an accompanying *P* value, but rather a thorough demonstration of the method through which a conclusion was reached from a given set of data with the ultimate goal of improving patient outcomes and quality of care. As analytical methods becoming increasingly sophisticated and datasets grow in size and complexity, we must not lose sight of the importance of enabling independent researchers to validate the findings of their peers without requiring them to reinvent the wheel.

Acknowledgments

All authors participated in the drafting of the manuscript. All authors read and approved the final manuscript. Data for this study were collected from the Medical Information Mart for Intensive Care (MIMIC-III) database. The creation, maintenance, and use

of the MIMIC-III database was approved by the institutional review boards of the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center. LAC is funded by the National Institutes of Health (NIH) through National Institute of Biomedical Imaging and Bioengineering grant R01 EB017205-01A1 and by Philips. The MIMIC database is funded by the National Institutes of Health Grant R01 EB001659, while the MIT Critical Care Datathons were funded by the MIT International Science and Technology Initiatives.

Conflicts of Interest

None declared.

References

1. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb 25;301(8):831-841. [doi: 10.1001/jama.2009.205] [Medline: 19244190]
2. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005 Jul 13;294(2):218-228. [doi: 10.1001/jama.294.2.218] [Medline: 16014596]
3. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, MIT Critical Data Conference 2014 Organizing Committee. Making big data useful for health care: a summary of the inaugural mit critical data conference. JMIR Med Inform 2014;2(2):e22 [FREE Full text] [doi: 10.2196/medinform.3447] [Medline: 25600172]
4. Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al. A "datathon" model to support cross-disciplinary collaboration. Sci Transl Med 2016 Apr 6;8(333):333ps8. [doi: 10.1126/scitranslmed.aad9072] [Medline: 27053770]
5. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. Crit Care Med 2011 May;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]
6. Chen KP, Cavender S, Lee J, Feng M, Mark RG, Celi LA, et al. Peripheral edema, central venous pressure, and risk of AKI in critical illness. Clin J Am Soc Nephrol 2016 Apr 7;11(4):602-608. [doi: 10.2215/CJN.08080715] [Medline: 26787777]
7. Chen KP, Lee J, Mark RG, Feng M, Celi LA, Malley BE, et al. Proton pump inhibitor use is not associated with cardiac arrhythmia in critically ill patients. J Clin Pharmacol 2015 Jul;55(7):774-779 [FREE Full text] [doi: 10.1002/jcph.479] [Medline: 25655574]
8. Paonessa JR, Brennan T, Pimentel M, Steinhaus D, Feng M, Celi LA. Hyperdynamic left ventricular ejection fraction in the intensive care unit. Crit Care 2015;19:288 [FREE Full text] [doi: 10.1186/s13054-015-1012-8] [Medline: 26250903]
9. de Louw EJ, Sun PO, Lee J, Feng M, Mark RG, Celi LA, et al. Increased incidence of diuretic use in critically ill obese patients. J Crit Care 2015 Jun;30(3):619-623 [FREE Full text] [doi: 10.1016/j.jcrc.2015.01.023] [Medline: 25721030]
10. Celi LA, Scott DJ, Lee J, Nelson R, Alper SL, Mukamal KJ, et al. Association of hypermagnesemia and blood pressure in the critically ill. J Hypertens 2013 Nov;31(11):2136-41; discussion 2141. [doi: 10.1097/HJH.0b013e3283642f18] [Medline: 24029865]
11. Pereira RD, Salgado CM, Dejam A, Reti SR, Vieira SM, Sousa JM, et al. Fuzzy modeling to predict severely depressed left ventricular ejection fraction following admission to the intensive care unit using clinical physiology. ScientificWorldJournal 2015;2015:212703 [FREE Full text] [doi: 10.1155/2015/212703] [Medline: 26345130]
12. Chatto Transform. San Francisco, CA: GitHub, Inc; 2016. URL: https://github.com/smartscheduling/chatto-transform-open-source/ [accessed 2016-08-16] [WebCite Cache ID 6jo9MCM4B]
13. Slack. San Franciso, CA: Slack Technologies, Inc URL: https://slack.com/ [accessed 2016-08-16] [WebCite Cache ID 6jo9OCvzW]
14. GitHub. San Francisco, CA: GitHub, Inc; 2016. URL: https://github.com/ [accessed 2016-08-16] [WebCite Cache ID 6jo9QA0km]
15. Jupyter.: Project Jupyter; 2016 Jul 01. URL: http://jupyter.org/ [accessed 2016-08-16] [WebCite Cache ID 6jo9Si7SH]
16. JupyterHub.: Project Jupyter; 2016. URL: https://jupyterhub.readthedocs.io/en/latest/ [accessed 2016-08-16] [WebCite Cache ID 6jo9UZNmY]

Abbreviations

API: application program interface
CSV: comma separated value
ICU: intensive care unit
MIMIC-II: Multiparameter Intelligent Monitoring in Intensive Care
MIMIC-III: Medical Information Mart for Intensive Care
MIT: Massachusetts Institute of Technology

Edited by G Eysenbach; submitted 17.07.16; peer-reviewed by G Ajuwon, A Benis; comments to author 10.08.16; revised version received 11.08.16; accepted 12.08.16; published 24.08.16.

Please cite as:
Celi LA, Lokhandwala S, Montgomery R, Moses C, Naumann T, Pollard T, Spitz D, Stretch R
Datathons and Software to Promote Reproducible Research
J Med Internet Res 2016;18(8):e230
URL: http://www.jmir.org/2016/8/e230/
doi:10.2196/jmir.6365
PMID:27558834

Original Paper

Reliability and Validity of Assessing User Satisfaction With Web-Based Health Interventions

Leif Boß[1], MSc; Dirk Lehr[2], PhD; Dorota Reis[3], PhD; Christiaan Vis[4,5], MA; Heleen Riper[4,5], PhD; Matthias Berking[6], PhD; David Daniel Ebert[6], PhD

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

[2]Department of Health Psychology and Applied Biological Psychology, Institute of Psychology, Leuphana University of Lueneburg, Lueneburg, Germany

[3]Section of Personality, Psychological Assessment, and Psychological Methods, Department of Psychology, University of Koblenz-Landau, Koblenz, Germany

[4]Section of Clinical Psychology, Department of Clinical, Developmental and Neuro Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[5]EMGO Institute for Health and Care Research, VU Medical Centre, Amsterdam, Netherlands

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

Corresponding Author:
Leif Boß, MSc
Division of Online Health Training
Innovation Incubator
Leuphana University of Lueneburg
Scharnhorststr 1
C113
Lueneburg, 21335
Germany
Phone: 49 41316771710
Fax: 49 41316771717
Email: boss@leuphana.de

Abstract

Background: The perspective of users should be taken into account in the evaluation of Web-based health interventions. Assessing the users' satisfaction with the intervention they receive could enhance the evidence for the intervention effects. Thus, there is a need for valid and reliable measures to assess satisfaction with Web-based health interventions.

Objective: The objective of this study was to analyze the reliability, factorial structure, and construct validity of the Client Satisfaction Questionnaire adapted to Internet-based interventions (CSQ-I).

Methods: The psychometric quality of the CSQ-I was analyzed in user samples from 2 separate randomized controlled trials evaluating Web-based health interventions, one from a depression prevention intervention (sample 1, N=174) and the other from a stress management intervention (sample 2, N=111). At first, the underlying measurement model of the CSQ-I was analyzed to determine the internal consistency. The factorial structure of the scale and the measurement invariance across groups were tested by multigroup confirmatory factor analyses. Additionally, the construct validity of the scale was examined by comparing satisfaction scores with the primary clinical outcome.

Results: Multigroup confirmatory analyses on the scale yielded a one-factorial structure with a good fit (root-mean-square error of approximation =.09, comparative fit index =.96, standardized root-mean-square residual =.05) that showed partial strong invariance across the 2 samples. The scale showed very good reliability, indicated by McDonald omegas of .95 in sample 1 and .93 in sample 2. Significant correlations with change in depressive symptoms (r=−.35, P<.001) and perceived stress (r=−.48, P<.001) demonstrated the construct validity of the scale.

Conclusions: The proven internal consistency, factorial structure, and construct validity of the CSQ-I indicate a good overall psychometric quality of the measure to assess the user's general satisfaction with Web-based interventions for depression and stress management. Multigroup analyses indicate its robustness across different samples. Thus, the CSQ-I seems to be a suitable measure to consider the user's perspective in the overall evaluation of Web-based health interventions.

(J Med Internet Res 2016;18(8):e234) doi:10.2196/jmir.5952

KEYWORDS

Internet; mental health; evaluation; clinical effectiveness; personal satisfaction

Introduction

State of Research on Web-Based Health Interventions

In recent years, development and usage of Web-based health interventions have been on the rise, and these interventions show potential in expanding upon established services for preventing and treating impaired health [1]. Many studies so far have shown that Web-based interventions are effective for various conditions, including depression [2,3], anxiety [4,5], sleep disorders [6], alcohol consumption [7], and stress [8-11]. However, there is a lack of published knowledge regarding aspects of effectiveness that are related to external validity, such as the applicability of proven interventions in routine care practice [12]. There are different reasons why evaluators should also examine how affected users directly evaluate the intervention. First, adding to the discussion on the importance of external validity [12,13], the evaluation should go beyond clinical effects that are assessed by health care professionals using observer- or self-rated health measures (eg, assessing depressive symptoms) [14]. The users' satisfaction with their intervention can be an important source for this metric. We adapted the definition given by Ware and colleagues [15] in our study, which states that satisfaction is a user's evaluation of the received Web-based intervention. Thus, it provides information beyond what is assessed by health care professionals. Second, it also provides information beyond the design qualities of a Web-based intervention that can be assessed by usability experts [16]. Thus, it delivers important information to service providers so that they can improve their interventions. Third, studying satisfaction can help to successfully implement and disseminate clinically effective Web-based interventions as a part of routine health care [17,18]. Fourth, there are ongoing debates on the relationship between user satisfaction and clinical intervention outcomes [14]. Previous studies found significant correlations between satisfaction with face-to-face interventions and psychological health [19-22] insofar as people with better health were more satisfied. One problem in Web-based interventions is the number of users who do not fully adhere to the intervention protocol [23]. Satisfaction with the delivered intervention may play an important role in understanding adherence to Web-based interventions and vice versa. Some studies, most of which focused on inpatient settings, found that patients who adhere to the intervention are more satisfied than patients who stop participating in the intervention [24]. Investigating the user satisfaction in Web-based health interventions could therefore add to the understanding of such relations. However, there is a strong need for thoroughly studied Web-based measures [25]. To the best of our knowledge, there is yet no validated measure for the assessment of user satisfaction with Web-based interventions.

Review of Established Satisfaction Measures

Various satisfaction measures have already been developed, such as the Patient Satisfaction Questionnaire Short Form (PSQ-18) [26], the Service Satisfaction Scale-30 (SSS-30) [27], the Satisfaction with Stroke Care Questionnaire (SASC) [28], and the Client Satisfaction Questionnaire (CSQ-8) [29]. Most of these instruments were designed to evaluate health care in hospitals or the general practitioner's office. In these settings, satisfaction ratings evaluate some dimensions that are not or less relevant to Web-based interventions, including satisfaction with the clinical staff, for example, "I have been treated with kindness and respect by the staff at the hospital," SASC; the waiting time, for example, "waiting time between asking to be seen and the appointment (date and time) given," SSS-30; the time spent with a doctor, for example, "Doctors usually spend plenty of time with me," PSQ-18; or the technical quality, for example, "I think my doctor's office has everything needed to provide complete care," PSQ-18. Therefore, satisfaction measures for Web-based interventions must be modified to address their unique characteristics that are not represented in traditional health services. For example, Web-based interventions can be delivered with or without direct contact with health care professionals or can be accessed immediately after registration without any waiting time. In order to address all of these variations, any developed generic measure should be applicable to a wide range of Web-based interventions. Moreover, especially for ease of usage in routine care settings, it is important to have an economically efficient instrument that requires little time to administer by staff and to complete by users. The CSQ seems to be a feasible candidate for adaptation and application to Web-based interventions. The original CSQ has shown good psychometric properties in a study (N=45) to investigate the effects of pretherapy orientation on psychotherapy outcome [30]. The German adaptation has been validated in a sample (N=300) of patients undergoing inpatient treatment within a psychosomatic clinic [31] and has already been integrated as a measure of routine monitoring in inpatient rehabilitation (N=53,177) [32].

The CSQ has also become a widely used instrument for assessing user satisfaction with Web-based health interventions [33-36]. It has been used as secondary outcome in a study comparing Web-based interpersonal psychotherapy and Web-based cognitive behavioral therapy for adults with depressive symptoms, indicating that participants of the first intervention group were more satisfied than the second group [33]. In another study, the CSQ was used as a secondary outcome to compare a Web-based intervention for depression with and without weekly therapist support, indicating that participants of the supported intervention were more satisfied than participants without support [34]. A modified version of the scale was also used in a pilot study of a Web-based screening and brief intervention for student marijuana use, where 95 out of 123 participants (77.2%) were at least moderately satisfied with the intervention [35]. In a previous study by our research group, the CSQ was used as secondary outcome of a Web-based recovery training for employees and indicated that 44 out of 49 participants (89.8%) would recommend the training to a friend

in need (item 4 of the CSQ) [36]. However, to our knowledge, there is yet to be a study evaluating the psychometric quality of assessed user satisfaction that tests its factorial structure and its association with indicators of effectiveness such as training adherence and health outcomes in Web-based interventions.

Aim of the Study

This study aimed to validate an adapted version of the German CSQ to assess the user satisfaction with Web-based health interventions (Client Satisfaction Questionnaire adapted to Internet-based interventions, CSQ-I). First, we examined the internal consistency of the scale, particularly whether the measurement model underlying its 8 items is at least essentially tau equivalent [37], which means that each item measures the same latent variable but with possibly different degrees of precision. Second, considering previous findings [29,31], we expected a single-factorial structure of user satisfaction that would be invariant across both samples in this study. The measurement model and the factorial structure of the scale were cross-validated in 2 independent samples to increase the generalizability of the findings. We further evaluated the validity of the scale by analyzing its correlation with other indicators of effectiveness. The overall evaluation of the psychometric quality of the scale was conducted according to recommendations derived from the COSMIN (Consensus-Based Standards for the Selection of Health Status Measurement Instruments) checklist [38,39].

Methods

Study Design

The CSQ-I was evaluated across 2 randomized controlled trials. The first trial was conducted to evaluate the efficacy of a Web-based intervention in preventing the onset of major depressive disorder (Trial Registration: German Clinical Trial Registry DRKS00004709) [40,41]. Participants were recruited from 2013 to 2014 from the general population via newspaper articles, on-air media, and through a campaign of a large insurance company. After completing a Web-based screening questionnaire and telephone interview, individuals aged at least 18 years with elevated depressive symptoms (Center for Epidemiologic Studies Depression Scale, CES-D, ≥16), not having a major depressive episode, were randomly assigned to either the intervention or a control group. The full inclusion and exclusion criteria are described in the efficacy paper of this trial [41]. The intervention consisted of 6 modules that were based on cognitive behavioral therapy. The participants were expected to complete each module within 1 week. All study outcomes were assessed using self-report measures at baseline (T1) and in postintervention assessment after 7 weeks (T2). Study outcomes relevant for the CSQ validation were the reduction of depressive symptoms between T1 and T2, adherence to the intervention, negative side effects, and user satisfaction at T2. The second trial was conducted to evaluate a Web-based stress management intervention in employees (Trial Registration: German Clinical Trial Registry DRKS00004749) [42,43]. Participants of this trial were recruited in 2013 from the general population via newspaper articles, on-air media, and through a campaign of a large insurance company. After completing a Web-based screening questionnaire, individuals aged at least 18 years with elevated symptoms of stress (Perceived Stress Scale, PSS-10, ≥22) were randomly assigned to either the Web-based intervention or a control group. The full inclusion and exclusion criteria are described in the efficacy paper of this trial [43]. The stress management intervention consisted of 7 modules, each to be completed within 1 week. Outcome assessments took place at baseline (T1) and after 7 weeks (T2). The primary outcome was symptom reduction of perceived stress between T1 and T2. Secondary outcomes included adherence to the intervention, negative side effects, and user satisfaction at T2.

Measures

Client Satisfaction Questionnaire Adapted to Internet-Based Interventions

The CSQ-I consists of 8 items measuring global satisfaction with the Web-based intervention (Table 1). On the original German scale [31], respondents rate each of the items on a 4-point Likert-type scale, but wording of response categories differed between the items. For example, the responses for item 1, “How would you rate the quality of service you received?” are rated between 4=“Excellent” and 1=“Poor,” and the responses for item 2, “Did you get the kind of service you wanted?” are rated between 1=“No, definitely not” and 4=“Yes, definitely.” Furthermore, 4 items are scored in inverse to minimize stereotypic response sets. However, after pilot testing (N=15) of the scale, discussion of the results in a focus group consisting of members of our research group, and obtaining advice from experts in the field of Web-based research, we decided to adapt the questionnaire in the following way. First, all questions were rephrased as statements to have constant response scales across the items, ranging from 1= “Does not apply to me” to 4=“Does totally apply to me.” In addition, we replaced the word “service” with “training” in all items because we expected that this wording would be more precise and common for users of Web-based interventions. As in the original version, the scores from all 8 items can be summed to a total score that ranges from 8 to 32. On 5 items (items 1, 2, 3, 5, and 7) participants rate the degree to which the intervention fulfilled their general satisfaction with the quality, kind of training, and amount of help they received. On item 6, respondents rate the degree to which the intervention helped them to deal with their problems. Item 4 assesses the degree to which respondents would recommend the intervention to others. Finally, on item 8, respondents rate their likelihood of using the intervention for themselves again.

Clinical Outcome

Depressive symptoms were assessed using the German version of the CES-D [44]. The CES-D is a self-report scale and consists of 20 items (eg, “During the past week I felt sad”), wherein each scored from 0 to 3. The total score ranges from 0 to 60, with a higher score indicating more severe depressive symptoms. A cutoff of 16 is usually regarded as indicating clinically relevant depressive symptom severity.

Symptoms of stress were assessed using the German version of the PSS-10 [45,46]. The PSS-10 assesses the degree to which

people perceive their lives as stressful. Participants are asked to answer questions regarding the previous week (eg, "In the past week, how often have you felt that you were unable to control the important things in your life?") on a 5-point Likert-type scale, with responses ranging from 0 to 4. The total score ranges from 0 to 40, with a higher score indicating higher perceived stress.

Adherence

The number of training modules that participants completed was used as the definition of adherence to the intervention in both trials. Full adherence was achieved when participants completed all 6 modules in the depression prevention intervention or all 7 modules in the stress management intervention.

Side Effects

The side effects of participating in the training modules were measured with the Inventory for the Assessment of Negative Effects of Psychotherapy (INEP) [47] that was adapted to the training settings. The adapted version consists of 15 items assessing negative changes participants experienced after completing the Web-based training in their social or work environments that they directly attribute to their participation in the training (eg, "During or after finishing the training, I got worse in making important decisions by myself" or "Compared to the time before the training, my relationship to my family is worse"). For the analysis, the negative side effects are counted and summed to a total amount of negative effects. The total score ranges from 0 to 15, with a higher score indicating more negative side effects.

Table 1. Item labels and descriptive analysis of the Client Satisfaction Questionnaire adapted to Internet-based interventions.

Item[a]	Sample 1				Sample 2			
	Mean (SD)	n_{low}[b] (%)	n_{high}[c] (%)	L[d]	Mean (SD)	n_{low} (%)	n_{high} (%)	L
1. The training I attended was of high quality or Das Training, an dem ich teilgenommen habe, hatte eine hohe Qualität.	3.48 (0.66)	3 (1.7)	94 (54.0)	0.73	3.45 (0.57)	0	54 (48.6)	0.67
2. I received the kind of training I wanted or Ich habe die Art von Training erhalten, die ich wollte.	3.11 (0.72)	5 (2.9)	54 (31.0)	0.83	3.09 (0.72)	3 (2.7)	31 (27.9)	0.82
3. The training has met my needs or Das Training hat meinen Bedürfnissen entsprochen.	3.13 (0.73)	5 (2.9)	58 (33.3)	0.83	3.05 (0.72)	3 (2.7)	3 (26.1)	0.80
4. I would recommend this training to a friend, if he or she were in need of similar help or Ich würde einem Freund or einer Freundin dieses Training empfehlen, wenn er or sie eine ähnliche Hilfe benötigen würde.	3.36 (0.73)	6 (3.4)	87 (50.0)	0.77	3.50 (0.67)	2 (1.8)	65 (58.6)	0.80
5. I am satisfied with the amount of help I received through the training or Ich bin zufrieden mit dem Ausmaß der Hilfe, die ich durch das Training erhalten habe.	3.18 (0.81)	9 (5.2)	71 (40.8)	0.82	3.12 (0.88)	6 (5.4)	44 (39.6)	0.78
6. The training helped me deal with my problems more effectively or Das Training hat mir dabei geholfen, angemessener mit meinen Problemen umzugehen.	3.25 (0.78)	9 (5.2)	76 (43.7)	0.82	3.09 (0.84)	7 (6.3)	37 (33.3)	0.82
7. In an overall, general sense, I am satisfied with the training or Im Großen und Ganzen bin ich mit dem Training zufrieden.	3.43 (0.80)	8 (4.6)	99 (56.9)	0.90	3.40 (0.70)	1 (0.9)	57 (51.4)	0.89
8. I would come back to such a training if I were to seek help again or Ich würde ein solches Training wieder nutzen, wenn ich Hilfe bräuchte.	3.36 (0.84)	8 (4.6)	101 (58.0)	0.89	3.35 (0.88)	6 (5.4)	63 (56.8)	0.81

[a] Item scoring: 1=does not apply to me or trifft nicht zu; 2=does rather not apply to me or trifft eher nicht zu; 3=does partly apply to me or trifft teilweise zu; 4=does totally apply to me or trifft voll und ganz zu.

[b] n_{low}=number of participants achieving the lowest possible score.

[c] n_{high}=number of participants achieving the highest possible score.

[d] L=standardized factor loadings.

Data Analyses

The analyses were conducted through structural equation modeling using the R package *lavaan* [48]. The covariance matrix was analyzed using the maximum likelihood method with robust (Huber-White) standard errors (MLR), which is asymptotically equivalent to the Yuan-Bentler T2* test statistic [49] recommended for nonnormally distributed data. In the first step, to estimate the internal consistency of the scale, we examined the underlying measurement model of the scale. Essential tau equivalency [37] of the scale indicates that all items can be assumed to assess the same latent variable with the same units of measurement (ie, equal factor loadings). Essential tau equivalency is a necessary assumption for the use of the Cronbach alpha index; if the underlying model violates this assumption, Cronbach alpha will underestimate the

reliability of the scale, and McDonald omega should be used instead as a more precise estimate [50]. In the second step, we evaluated whether the one-factor structure proposed by the authors of the original CSQ [30,31] holds across our 2 training samples by conducting multigroup confirmatory factor analyses (CFAs). The idea underlying the estimation of multigroup CFAs is that mean scores of different samples can only be compared in a meaningful way when the requirements of measurement invariance across groups are satisfied. Additionally, multigroup CFAs allow to test if participants in different training groups interpret and respond to the items in the same way. The procedure to establish measurement invariance involves several steps [51,52], which can be described as follows. Configural invariance means that the same common factor structure is shared across groups. Weak invariance indicates that all participants, regardless of their group membership (ie, the received training), respond to the scale items in the same way. Thus, to achieve weak invariance in addition to configural invariance, equivalent factor loadings across the groups are required. Next, we tested for strong measurement invariance by imposing additional constraints on the intercepts of the items (ie, the intercepts of the items were set to be equal across the groups). Strong invariance implies that individuals who have the same score on the latent variable (true score) will obtain the same score on the observed variable regardless of their training group membership.

To assess the fit of the models to the data, we used the following measures: the chi-square statistic, the relative chi-square (χ^2/ *df*), the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). In general, a χ^2/ *df* value of ≤3.00, a CFI value ≥.95, and RMSEA and SRMR values ≤.08 indicate an acceptable fit to the data. Because the chi-square difference test commonly used to compare nested models is sensitive to sample size, we used additional criteria for model comparisons. For the Akaike information criterion (AIC), models with lower AIC fit the data better than those with higher AIC values. In addition, we used ΔCFI, ΔRMSEA, and ΔSRMR to evaluate test invariance. Considering the number of items and the size of our samples, the following criteria proposed by Chen [53] were applied: ΔCFI<−.005, ΔRMSEA<.010, and ΔSRMR<.005 for testing strong measurement invariance.

To test the convergent validity of the scale, we correlated the CSQ-I scores with the primary clinical outcomes in terms of symptom status at T2 and change of symptoms between T1 and T2. In addition, we compared participants with and without reliable symptom reductions. To assess symptom reductions on an individual level, we examined the number of participants who were classified as having reliably changed according to the reliable change index described by Jacobson and Truax [54]. Participants were defined as having reliably changed if their symptoms declined from T1 to T2 with a reliable change index of greater than 1.96 (8.65 points on the CES-D and 5.16 points on the PSS-10). Furthermore, we conducted explorative subgroup analyses on gender, adherence, and negative side effects from the intervention.

Discriminant validity was evaluated by means of the average variance extracted (AVE). Fornell and Larcker [55] introduced the AVE as an extension of chi-square–based statistics for measuring the goodness of fit between theoretical models with unobservable variables and the empirical data. The index also provides a procedure for establishing discriminant validity. The use of the AVE for this purpose has shown to be robust in various studies, primarily in the field of marketing research. For example, McKinney and colleagues [56] used the AVE as an additional measure of the reliability for the evaluation of a Web-customer satisfaction questionnaire and compared the AVE values with other established measures, such as Cronbach alpha and the composite factor reliability. Liao and colleagues [57] used the AVE as an additional measure to estimate the validity of a planned behavior model, including consumer satisfaction, to predict the customer's intention toward continued use of Web-based services. In this study, we compared the AVE values for the CSQ-I with the squared correlation estimate between satisfaction and clinical outcome at T2. The AVE was calculated as the total of all squared standardized factor loadings divided by the number of items [58]. Assuming discriminant validity, the AVE should be greater than the squared correlation estimate. This would indicate that user satisfaction is a separate construct, distinguishable from clinical symptoms.

All correlations and subgroup analyses were conducted using IBM SPSS version 22 (SPSS Inc, Chicago, IL, USA). All analyses were done only on complete data samples. Cases with missing values in the outcome variables were excluded from the analyses.

Results

Samples' Characteristics

The depression intervention sample consisted of 201 adults from the German population with clinically relevant levels of depression (CES-D≥16). Complete data on the study outcome variables were available for 174 participants (174/201, 86.6%). The participants were on average 45 years of age (SD 11.84), and 130 were female (130/174, 74.7%; Table 2).

Satisfaction with the Web-based intervention to prevent major depression ranged from mean 3.11 (SD 0.72) on item 2 to mean 3.48 (SD 0.66) on item 1 (Table 1). Each of the items showed a ceiling effect, as at least 54 participants (54/174, 31.0%) achieved the highest possible score (4="does totally apply to me") on the items (eg, "I got the kind of training I wanted"). In terms of quality criteria for measurement properties [38], ceiling effects are considered to be present if more than 15% of respondents achieved the highest possible score. The average total CSQ-I score was 26.26 (SD 5.34), with 23 participants (23/174, 13.2%) achieving the highest possible total score and only 1 participant achieving the lowest possible total score. The sample showed a negatively skewed distribution of satisfaction scores (skewness=−1.294, SE=0.184).

The stress management intervention sample consisted of 132 employees from the German population with elevated symptoms of stress (PSS-10 ≥22). Complete data of primary and secondary outcome variables were available for 111 participants (111/132,

84.0%). The participants were on average 42 years of age (SD 9.79). The majority of participants (94/111, 84.7%) were female (Table 2).

The satisfaction ranged from mean 3.05 (SD 0.72) on item 3 to mean 3.50 (SD 0.67) on item 4. At least 29 participants (29/111, 26.1%) achieved the highest possible score on the items (Table 1). The total CSQ-I score was mean 26.05 (SD 4.96) with 14 participants (14/111, 12.6%) achieving the highest possible total score, whereas none of the participants achieved the lowest possible score. The sample showed a distribution that skewed to the left (skewness=−0.909, SE=0.211).

Table 2. Samples' description.

Characteristics	Sample 1 (N=174), n (%)	Sample 2 (N=111), n (%)
Gender		
Female	130 (74.7)	94 (84.7)
Male	16 (14.4)	
44 (25.3)		
Other[a]	-	1 (0.9)
Marital status		
Single	55 (31.6)	33 (29.7)
Married or partnership	87 (50.0)	52 (46.8)
Divorced or separated	32 (18.4)	10 (14.4)
Education		
High school	77 (42.0)	48 (43.3)
College or university	91 (54.6)	59 (53.1)
PhD	6 (3.4)	4 (3.6)
Occupation		
Employed full-time	92 (52.9)	85 (76.6)
Employed part-time	55 (31.6)	25 (22.5)
Not employed	21 (12.1)	-
Unemployed	3 (1.7)	-
Unable to work owing to illness	3 (1.7)	1 (0.9)
History of psychotherapy		
Have not been in psychotherapy	101 (58.0)	72 (64.9)
Have been in psychotherapy	73 (42.0)	39 (35.1)

[a]Participants who wanted to specify their gender as neither female nor male.

Test of the CSQ-I Measurement Model

The tau congeneric model indicated that the one-factor model proposed for the original instrument [30,31] was supported in our adapted version. We found that, in both samples, the one-factor model showed an acceptable fit to the data with CFI=.96 and SRMR=.029 in sample 1 and CFI=.95 and SRMR=.035 in sample 2, but the RMSEA values were questionable. In sample 1, RMSEA was .10 (*P*=.002), and in sample 2, RMSEA was .10 (*P*=.02), indicating that the tested model did not perfectly fit the data in sample 2. In the next step, our results revealed that the essentially tau-equivalent model was rejected by the data in both samples. In sample 1, ΔCFI was −.032, ΔRMSEA=.021, and ΔSRMR=.105. In sample 2, ΔCFI was −.052, ΔRMSEA=.031, and ΔSRMR=.140. Also, in both samples, the AICs were lower for the tau congeneric model. Hence, the assumptions for the computation of Cronbach alpha indices were not met [50], and the more precise McDonald omegas were computed instead. Using this metric, the CSQ-I showed very good reliability in both samples, where sample 1 had omega=.95 (bias-corrected and accelerated, BCa, CI .93-.96) and sample 2 had omega=.93 (BCa CI .91-.95).

CSQ-I Structure Across the 2 Study Samples

In the next step, we examined whether the test scores of the CSQ-I were comparable across the 2 study samples. To do so, we performed multigroup CFAs to test for configural, weak, and strong measurement invariance. The unconstrained model (M1) we used to test for configural invariance fit the data well across sample 1 and sample 2 (Table 3; an extended version of results of this analysis can be found in appendix 1). Furthermore, the model that was used to test for weak invariance (M2) also fit the data well, and the differences in the relevant indices (ΔCFI, ΔRMSEA, ΔSRMR) showed that the additional constraints imposed on the data (ie, equal factor loadings) did not significantly alter the fit of the model. Subsequently, we

tested for strong measurement invariance by imposing equality on the intercepts (M3). Our results indicated that our data did not support strong measurement invariance, although the corresponding cutoffs were only slightly missed (ΔCFI=−0.006, ΔRMSEA=0.001, ΔSRMR=0.005) and the AIC was almost the same: AIC=3581.5 for the model with weak invariance versus AIC=3583.1 for the model with strong invariance. Given that full strong invariance was not supported, we tested for partial strong measurement invariance. The inspection of the means residual matrix revealed a substantial standardized residuum for item 4 ("I would recommend that training to a friend, if he or she were in need of similar help.") with a value of 4.84. This indicated a lack of invariance for this item across the 2 training samples. When the intercept of item 4 was freely estimated, partial strong invariance was supported (M4).

Table 3. Tests of invariance for the proposed one-factor structure of the Client Satisfaction Questionnaire adapted to Internet-based interventions between sample 1 (N = 174) and sample 2 (N = 111): results of multigroup confirmatory factor analyses with MLR estimator.

Model	χ^2	*df* [a]	χ^2/ *df*	CFI[b]	RMSEA[c]	SRMR[d]
M1 configural invariance	99.2	40	2.5	.964	.102	.031
M2 weak invariance	104.5	47	2.2	.965	.093	.043
M3 strong invariance	120.2	54	2.2	.959	.093	.048
M4 partial strong invariance	111.3	53	2.1	.964	.088	.045

[a]*df*: degrees of freedom.

[b]CFI: comparative fit index.

[c]RMSEA: root-mean-square error of approximation.

[d]SRMR: standardized root-mean-square residual.

Convergent Validity

In sample 1, the CSQ-I score was significantly correlated with depressive symptoms at T2 (r=−.35, P<.001; Table 4), indicating that higher satisfaction corresponded to a lower score of depressive symptoms after the intervention. The CSQ-I score was also significantly correlated with a change of depressive symptoms between T1 and T2 (r=.27, P<.001), meaning that, on average, participants with larger reductions in depressive symptoms appeared to be more satisfied with the intervention compared with those with smaller symptom reductions. The group of participants who met the criteria for reliable reduction of depressive symptoms (102/174, 58.6%) showed significantly greater satisfaction (mean 27.45, SD 4.45) than the group without reliable symptom reduction (mean 24.58, SD 6.03; t_{172}=3.609, P<.001; Cohen's d=0.52).

Most of the participants (134/174, 77.0%) fully adhered to the training protocol by completing all 6 training modules. The participants who fully adhered (n=134, mean 26.48, SD 5.38) did not appear to be more or less satisfied with the training than participants who did not fully adhere (n=40, mean 25.55, SD 5.21; t_{172}=0.964, P=.34). There was no meaningful difference in satisfaction between women (mean 26.17, SD 5.51) and men (mean 26.30, SD 4.87; t_{172}=0.086, P=.93). In this sample, 36 out of 174 participants (21%) reported negative side effects due to the training. In terms of satisfaction scores, participants who reported side effects (mean 26.64, SD 6.26) did not significantly differ from participants without such negative effects (mean 26.17, SD 5.09; t_{172}=0.472, P=.64).

Table 4. Means, standard deviations, and intercorrelations of relevant outcomes in sample 1.

Outcome	Mean (SD)	Intercorrelations (P)			
		1	2	3	4
1. Satisfaction at T2	26.26 (5.34)	-			
2. Depressive symptoms at T1	26.29 (7.66)	−.04 (.59)	-		
3. Depressive symptoms at T2	17.10 (8.89)	−.35 (<.001)	.35 (<.001)	-	
4. Depressive symptoms T1-T2	−9.19 (9.50)	.27 (<.001)	.49 (<.001)	−.58 (<.001)	-

In sample 2, the CSQ-I score significantly correlated with perceived stress at T2 (r=−.48, P<.001; Table 5), indicating that higher satisfaction corresponded to a lower score of stress symptoms after the intervention. The CSQ-I score was also significantly correlated with change in perceived stress between T1 and T2 (r=.52, P<.001), meaning that, on average, participants with larger reductions of perceived stress appeared to be more satisfied with the intervention compared with those with smaller symptom reductions. The group of participants who met the criteria for reliable reduction in perceived stress (60/111, 54.1%) showed significantly greater satisfaction (mean 28.12, SD 3.8) than the group without reliable symptom reduction (mean 23.63, SD 5.10; t_{109}=5.302, P<.001; d=1.01). Most of the participants (90/111, 81%) fully adhered to the training protocol by completing all 7 training modules. The participants who fully adhered (n=90, mean 26.81, SD 4.55) were more satisfied with the training compared with the participants who did not fully adhere (n=21, mean 22.81, SD

5.44; t_{109}=3.492, P<.001; d=0.85). There was no meaningful difference in satisfaction between women (mean 26.37, SD 4.76) and men (mean 24.13, SD 5.95; t_{109}=1.681, P=.10). In this study, 12 out of 111 participants (11%) reported they experienced at least one negative side effect due to the training. In terms of satisfaction scores, participants who reported side effects (mean 27.75, SD 3.25) did not significantly differ from participants without such negative effects (mean 25.85, SD 5.11; t_{109}=1.257, P=.21).

Table 5. Means, standard deviations, and intercorrelations of relevant outcomes in sample 2.

Outcome	Mean (SD)	Intercorrelations (P)			
		1	2	3	4
1. Satisfaction at T2	26.05 (4.96)	-			
2. Stress at T1	25.28 (4.60)	.08 (.42)	-		
3. Stress at T2	18.68 (6.27)	−.48 (<.001)	.22 (.02)	-	
4. Stress T1-T2	−6.60 (7.00)	.52 (<.001)	.45 (<.001)	−.74 (<.001)	-

Discriminant Validity

In sample 1, the AVE values for both measures (CSQ-I AVE=0.681, CES-D AVE=0.446) were greater than the squared correlation between these outcomes (R^2=.123), indicating that the CSQ-I construct explained more of the variance in its items than it shared with the CES-D. In sample 2, the AVE values for both measures (CSQ-I AVE=0.512, PSS-10 AVE=0.411) were also greater than the squared correlation between the CSQ-I and the PSS-10 (R^2=.230), indicating that the adapted CSQ-I explained more of the variance in its items than it shared with the PSS-10.

Discussion

Principal Findings

In the evaluation of Web-based health interventions, the user's perspective should be taken into account [15,24]. For this purpose, Web-based measures with proven psychometric quality are needed [25]. In this study, we investigated the factorial structure, the measurement model, and construct validity of an adapted version of the CSQ in 2 samples of adults who had participated in Web-based health interventions for either preventing major depression or improving stress management.

Multigroup factor analyses on the CSQ-I confirmed the proposed one-factorial structure [29,31] of the original scale across 2 independent samples. Our results showed that, although the assumptions needed for Cronbach alpha were not met, the scale demonstrated excellent reliability through McDonald omega [50] with omega=.95 in sample 1 and omega=.93 in sample 2. These findings correspond to previous studies that showed a very good reliability of the original scale, indicated by Cronbach alpha indices of alpha=.93 [30], alpha=.87 [31], and alpha=.90 [32]. The results on measurement invariance across the groups imply that the factor structure was replicated between the 2 samples; however, this should be interpreted with caution because the differences found in the latent means were due to partial rather than full strong measurement invariance. Although some researchers argue that in order to test for latent means between two samples at least two items must have invariant loadings and intercepts [59], Thompson and Green [60] reason that "in models with equivalent factor loadings but differing intercepts, differences in the means on that measure are a function of both the latent factors and the varying intercepts which can be interpreted in terms of a biased measure" (p149). However, we stress that the differences in the indices comparing weak and strong invariance were very small, indicating that the lack of invariance was marginal.

In line with previous findings [30-32,61], the satisfaction scores were on average very high, indicating that the participants tended to be very satisfied with the delivered intervention. This result may be restricted owing to a ceiling effect [38] because many participants achieved the highest possible satisfaction score in both samples. However, the results showed that participants with reliable symptom reductions due to the received intervention were more satisfied than those without reliable reductions. Thus, these findings indicate the ability of the scale to discriminate between more and less satisfied intervention users, despite potential ceiling effects. Nevertheless, some studies suggest that modifying the response choice pattern from a 4-point format to a 5-point format with three positive choices and two negative choices can increase the variability of satisfaction scores [62-64]. Thus, testing a further adaptation of the CSQ-I response format may be useful in the future. The content validity of the scale was primarily investigated in relation to clinical outcomes in terms of psychopathological symptoms after the intervention and change of symptoms between baseline and postintervention assessment. The correlations between satisfaction and symptoms at the postintervention assessment were r=−.35 in sample 1 and r=−.48 in sample 2. These results are in line with findings from the original CSQ version with correlation coefficients of r=−.34 for satisfaction × psychosomatic symptoms at postintervention assessment [30], r=.40 for satisfaction × health condition at discharge [31], and r=.40 for satisfaction × health condition at discharge [32]. The correlation between satisfaction and change of symptom severity from baseline to postintervention assessment was r=.27 in sample 1 and r=.52 in sample 2 in our study. These results also correspond to findings from the original CSQ regarding correlations between satisfaction and psychosomatic symptom reduction of r=−.40 [30] and health condition improvement of r=.52 [31] and r=.60 [32]. In summary, the results of the content validity analysis can lead to the assumption that participants might have rated their satisfaction to be high merely because they felt better after the

training. In this case, the satisfaction score would display a proxy measure for the clinical outcome. However, the discriminant validity analyzed in terms of the AVE values of the CSQ-I indicates that the satisfaction measure and the clinical outcome measure assessed different constructs. The relation between satisfaction and adherence remains unclear. In the second sample only, we found a marginal but statistically significant difference between the satisfaction scores of participants who did and did not fully adhere to the intervention. In general, low satisfaction with the intervention is assumed to be associated with low adherence [25]; notwithstanding that individuals experiencing a high burden (eg, due to depressive symptoms) may be under considerable pressure to find relief. Those individuals may also adhere to an existing intervention, although they do not evaluate all aspects of the intervention as favorable. This might rather apply to the participants in the depressive intervention sample than to those in the stress intervention sample.

Nevertheless, some limitations of this study should be taken into account. First, the study dropout rates were very low in both samples (27/201, 13.4% in sample 1, 22/132, 16.0% in sample 2), corresponding to dropout rates in other validation studies [31,61,65]. However, it is possible that participants who did not complete posttreatment assessments may have rated their satisfaction lower than participants who attended the posttreatment assessment [65]. Second, we were not able to control for the interference between satisfaction and postintervention health state in terms of psychopathological symptoms because both variables were assessed at the same time point. Thus, we could not exclude the possibility that the participants' health state, after participating in the training modules, biased their satisfaction rating. Further experimental studies are needed to investigate the clinical effects of Web-based interventions on satisfaction, using different time points for the assessment of user satisfaction and health outcomes, and also consider follow-up assessments. Third, because we used the same clinical outcome measure for state and for change of psychological health, it was not possible to estimate the predictive impact of both health outcomes on satisfaction independently. Thus, it may be beneficial to use different measures for (1) health condition at the postintervention assessment and (2) change of health over time. Fourth, it was not possible to analyze the impact of adherence on satisfaction. Adherence was operationalized by the number of completed training modules. In both samples most of the participants completed all modules (134/174, 77.0% in sample 1 and 90/111, 81% in sample 2), so that it would not have been of value to determine the correlation between adherence and satisfaction. Future research should use additional measures of adherence (eg, login counts or time spent on the training website) to investigate the construct validity of the scale. Unfortunately, such data were not available for our study. Fifth, the subgroup analyses of gender and negative side effects from the intervention were underpowered. Hence, future studies are needed to explore potential relevant subgroup effects such as gender and negative intervention effects that may influence satisfaction ratings. Finally, one theoretical limitation must be taken into account when using the CSQ-I for the evaluation of Web-based health interventions. There have been previous discussions regarding the usefulness of user satisfaction in assessing quality of health care interventions, mainly because of its construct validity and unclear evidence for its association with other health outcomes [14]. It is important to note that the CSQ-I covers the user's satisfaction with Web-based health interventions in a broader sense rather than focusing on specific intervention aspects. Thus, it is not clear on which aspects of the intervention the participants actually based their satisfaction rating. Most of the CSQ-I items cover the fulfilled expectancy in terms of the general quality of the intervention, their intention to use it again, or their likeliness of recommending it to other affected people. The items do not cover specific aspects and surrounding conditions of the intervention (eg, usability of the Web-based program, registration and login procedures, psychological and technical guidance) that may also be relevant for clinical success [1] and adherence [16] in Web-based health interventions. Thus, it may be valuable to evaluate additional, more specific quality dimensions of Web-based interventions (eg, technical support, usability, simplicity of the intervention content).

Conclusions

In this study the CSQ-I has shown to be a robust measure with a clear factorial structure across different samples. Thus, the CSQ-I seems to be a suitable measure to consider the user's satisfaction in the overall evaluation of Web-based health interventions. It can provide an important source of information for service providers who wish to improve or implement their interventions into routine health care. Furthermore, satisfaction scores derived from the CSQ-I may serve as a useful reference for other people who are seeking help via the Web.

Acknowledgments

We would like to acknowledge the European Union for funding this project within the Lueneburg Innovation Incubator, TM 1.1 (project number: CCI 2007DE161PR001).

Authors' Contributions

DE, DL, and MB obtained funding for this study. LB and DE drafted the study design and the adaptation of the questionnaire. DL contributed to the final study design. LB drafted the manuscript. DL and DE supervised the writing process. LB and DR conducted the analyses. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

[PDF File (Adobe PDF File), 283KB - jmir_v18i8e234_app1.pdf]

References

1. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
2. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869. [doi: 10.1586/ern.12.63] [Medline: 22853793]
3. 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]
4. Mayo-Wilson E, Montgomery P. Media-delivered cognitive behavioural therapy and behavioural therapy (self-help) for anxiety disorders in adults. Cochrane Database Syst Rev 2013;9:CD005330. [doi: 10.1002/14651858.CD005330.pub4] [Medline: 24018460]
5. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
6. Cheng SK, Dizon J. Computerised cognitive behavioural therapy for insomnia: a systematic review and meta-analysis. Psychother Psychosom 2012;81(4):206-216. [doi: 10.1159/000335379] [Medline: 22585048]
7. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
8. 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]
9. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
10. 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]
11. Morledge TJ, Allexandre D, Fox E, Fu AZ, Higashi MK, Kruzikas DT, et al. Feasibility of an online mindfulness program for stress management--a randomized, controlled trial. Ann Behav Med 2013 Oct;46(2):137-148 [FREE Full text] [doi: 10.1007/s12160-013-9490-x] [Medline: 23632913]
12. Glasgow RE, Green LW, Klesges LM, Abrams DB, Fisher EB, Goldstein MG, et al. External validity: we need to do more. Ann Behav Med 2006 Apr;31(2):105-108. [doi: 10.1207/s15324796abm3102_1] [Medline: 16542124]
13. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006 Mar;29(1):126-153. [doi: 10.1177/0163278705284445] [Medline: 16510882]
14. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]
15. Ware JE, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann 1983;6(3-4):247-263. [Medline: 10267253]
16. 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]
17. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
18. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
19. Ankuta GY, Abeles N. Client satisfaction, clinical significance, and meaningful change in psychotherapy. Prof Psychol Res Pr 1993;24(1):70-74. [doi: 10.1037/0735-7028.24.1.70]
20. Puschner B, Bauer S, Kraft S, Kordy H. Zufriedenheit von Patienten und Therapeuten mit ambulanter Psychotherapie [Patient and therapist satisfaction in outpatient psychotherapy]. Psychother Psychosom Med Psychol 2005 Dec;55(12):517-526. [doi: 10.1055/s-2005-867050] [Medline: 16342025]

21. Holcomb WR, Parker JC, Leong GB, Thiele J, Higdon J. Customer satisfaction and self-reported treatment outcomes among psychiatric inpatients. Psychiatr Serv 1998 Jul;49(7):929-934. [doi: 10.1176/ps.49.7.929] [Medline: 9661228]
22. Gebhardt S, Wolak AM, Huber MT. Patient satisfaction and clinical parameters in psychiatric inpatients--the prevailing role of symptom severity and pharmacologic disturbances. Compr Psychiatry 2013 Jan;54(1):53-60. [doi: 10.1016/j.comppsych.2012.03.016] [Medline: 22534034]
23. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
24. Barbosa CD, Balp M, Kulich K, Germain N, Rofail D. A literature review to explore the link between treatment satisfaction and adherence, compliance, and persistence. Patient Prefer Adherence 2012;6:39-48 [FREE Full text] [doi: 10.2147/PPA.S24752] [Medline: 22272068]
25. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 2016;16:45 [FREE Full text] [doi: 10.1186/s12888-016-0735-7] [Medline: 26915661]
26. Marshall G, Hays R. RAND. The Patient Satisfaction Questionnaire Short Form (PSQ-18) URL: http://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf [WebCite Cache ID 6h9lwtac0]
27. Greenfield T, Attkisson C. The UCSF client Satisfcation Scales: II. The Service Satisfaction Scale-30. In: Maruish ME, editor. The use of psychological testing for treatment planning and outcomes assessment. 3rd edition. Mashwah, NJ: Lawrance Erlbaum Associates; 2004:813-837.
28. Pound P, Tilling K, Rudd AG, Wolfe CD. Does patient satisfaction reflect differences in care received after stroke? Stroke 1999 Jan;30(1):49-55 [FREE Full text] [Medline: 9880387]
29. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [Medline: 10245370]
30. 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]
31. Schmidt J, Lamprecht F, Wittmann WW. [Satisfaction with inpatient management. Development of a questionnaire and initial validity studies]. Psychother Psychosom Med Psychol 1989 Jul;39(7):248-255. [Medline: 2762479]
32. Kriz D, Nübling R, Steffanowski A, Wittmann W, Schmidt J. Patientenzufriedenheit in der stationären Rehabilitation: Psychometrische Reanalyse des ZUF-8 auf der Basis multizentrischer Stichproben verschiedener Indikation [Patients' satisfaction in inpatient rehabilitation. Psychometrical evaluation of the ZUF-8 based on a multicenter sample of different indications]. Z Med psychol 2008;17:67-79.
33. Donker T, Bennett K, Bennett A, Mackinnon A, van SA, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
34. 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 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
35. Palfai TP, Saitz R, Winter M, Brown TA, Kypri K, Goodness TM, et al. Web-based screening and brief intervention for student marijuana use in a university health center: pilot study to examine the implementation of eCHECKUP TO GO in different contexts. Addict Behav 2014 Sep;39(9):1346-1352 [FREE Full text] [doi: 10.1016/j.addbeh.2014.04.025] [Medline: 24845164]
36. Ebert DD, Berking M, Thiart H, Riper H, Laferton JA, Cuijpers P, et al. Restoring depleted resources: Efficacy and mechanisms of change of an internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015 Dec;34 Suppl:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
37. Jöreskog KG. Simultaneous factor analysis in several populations. Psychometrika 1971 Dec;36(4):409-426. [doi: 10.1007/BF02291366]
38. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
39. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549 [FREE Full text] [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
40. Buntrock C, Ebert DD, Lehr D, Cuijpers P, Riper H, Smit F, et al. Evaluating the efficacy and cost-effectiveness of web-based indicated prevention of major depression: design of a randomised controlled trial. BMC Psychiatry 2014;14:25 [FREE Full text] [doi: 10.1186/1471-244X-14-25] [Medline: 24485283]
41. Buntrock C, Ebert D, Lehr D, Riper H, Smit F, Cuijpers P, et al. Effectiveness of a web-based cognitive behavioural intervention for subthreshold depression: pragmatic randomised controlled trial. Psychother Psychosom 2015;84(6):348-358. [doi: 10.1159/000438673] [Medline: 26398885]

42. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. J Med Internet Res 2016;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
43. 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]
44. Hautzinger M, Bailer M, Hofmeister D, Keller F. Allgemeine Depressionsskala (ADS) [Center for Epidemiological Studies Depression Scale (CES-D; Radloff, L.S., 1977) - German adaptation.]. Psychiat Prax 2012 Aug 31;39(06):302-304. [doi: 10.1055/s-0032-1326702]
45. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
46. Michalsen A, Jeitler M, Brunnhuber S, Lüdtke R, Büssing A, Musial F, et al. Iyengar yoga for distressed women: a 3-armed randomized controlled trial. Evid Based Complement Alternat Med 2012;2012:408727 [FREE Full text] [doi: 10.1155/2012/408727] [Medline: 23049608]
47. Ladwig I, Rief W, Nestoriuc Y. Welche Risiken und Nebenwirkungen hat Psychotherapie? - Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP) [What are the Risks and Side Effects to Psychotherapy? – Development of an Inventory for the Assessment of Negative Effects of Psychotherapy (INEP)]. Verhaltenstherapie 2014;24:252-263. [doi: 10.1159/000367928]
48. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. J Stat Soft 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
49. Yuan KH, Bentler PM. Normal theory based test statistics in structural equation modelling. Br J Math Stat Psychol 1998 Nov;51 (Pt 2):289-309. [Medline: 9854947]
50. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
51. Widaman K, Reise S. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In: Bryant KJ, Windle M, West SG, editors. The science of prevention: Methodological advances from alcohol and substance abuse research. Washington, DC: American Psychological Association; 1997:281-324.
52. Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika 1993 Dec;58(4):525-543. [doi: 10.1007/BF02294825]
53. Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Struct Equ Modeling 2007 Jul 31;14(3):464-504. [doi: 10.1080/10705510701301834]
54. 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]
55. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
56. McKinney V, Yoon K, Zahedi F. The measurement of web-customer satisfaction: An expectation and disconfirmation approach. Inf Syst Res 2002;13(3):296-315.
57. Liao C, Chen J, Yen DC. Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model. Comput Human Behav 2007 Nov;23(6):2804-2822. [doi: 10.1016/j.chb.2006.05.006]
58. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 7th edition. New Jersey: Pearson Prentice Hall; 2010.
59. Steenkamp J, Baumgartner H. Assessing Measurement Invariance in Cross‐National Consumer Research. J Consum Res 1998 Jun;25(1):78-107. [doi: 10.1086/209528]
60. Thompson M, Green S. Evaluating between-group differences in latent variable means. In: Hancock GR, Mueller RO, editors. Structural equation modeling: A second course. Charlotte, NC: Information Age; 2006:119-169.
61. Hannöver W, Dogs C, Kordy H. Patientenzufriedenheit - ein Maß für Behandlungserfolg? [Patient satisfaction - a measure for success of treatment?]. Psychotherapeut 2000;45(5):292-300.
62. Ware JE, Hays RD. Methods for measuring patient satisfaction with specific medical encounters. Med Care 1988 Apr;26(4):393-402. [Medline: 3352332]
63. Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. Health Technol Assess 2002;6(32):1-244 [FREE Full text] [Medline: 12925269]
64. Moret L, Nguyen J, Pillet N, Falissard B, Lombrail P, Gasquet I. Improvement of psychometric properties of a scale measuring inpatient satisfaction with care: a better response rate and a reduction of the ceiling effect. BMC Health Serv Res 2007;7:197 [FREE Full text] [doi: 10.1186/1472-6963-7-197] [Medline: 18053170]
65. Lebow JL. Research assessing consumer satisfaction with mental health treatment: a review of findings. Eval Program Plann 1983;6(3-4):211-236. [Medline: 10299619]

Abbreviations

AIC: Akaike information criterion
AVE: average variance extracted
BCa: bias-corrected and accelerated
CES-D: Center for Epidemiologic Studies Depression Scale
CFA: confirmatory factor analysis
CFI: comparative fit index
CSQ: Client Satisfaction Questionnaire
CSQ-I: Client Satisfaction Questionnaire adapted to Internet-based interventions
PSQ-18: Patient Satisfaction Questionnaire Short Form
PSS-10: Perceived Stress Scale, 10-item version
RMSEA: root-mean-square error of approximation
SASC: Satisfaction with Stroke Care Questionnaire
SRMR: standardized root-mean-square residual
SSS-30: Service Satisfaction Scale-30

Edited by G Eysenbach; submitted 08.05.16; peer-reviewed by T Kuhlmann, F Drozd; comments to author 02.06.16; revised version received 13.07.16; accepted 27.07.16; published 31.08.16.

Please cite as:
Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, Ebert DD
Reliability and Validity of Assessing User Satisfaction With Web-Based Health Interventions
J Med Internet Res 2016;18(8):e234
URL: http://www.jmir.org/2016/8/e234/
doi:10.2196/jmir.5952
PMID:27582341

Original Paper

What Are We Looking for in Computer-Based Learning Interventions in Medical Education? A Systematic Review

Tiago Taveira-Gomes[1], MD; Patrícia Ferreira[1], MD; Isabel Taveira-Gomes[2], MD; Milton Severo[1,3], PhD; Maria Amélia Ferreira[1], MD, PhD

[1]Department of Medical Education and Simulation, Faculty of Medicine, University of Porto, Porto, Portugal

[2]Department of Clinical Neurosciences and Mental Health, Medical Psychology Unit, Faculty of Medicine, University of Porto, Porto, Portugal

[3]Department of Clinical Epidemiology, Predictive Medicine and Public Health, Faculty of Medicine, University of Porto, Porto, Portugal

Corresponding Author:
Tiago Taveira-Gomes, MD
Department of Medical Education and Simulation
Faculty of Medicine
University of Porto
Alameda Professor Hernani Monteiro
Porto, 4200 319
Portugal
Phone: 351 225513605
Fax: 351 225513604
Email: tiago.taveira@me.com

Abstract

Background: Computer-based learning (CBL) has been widely used in medical education, and reports regarding its usage and effectiveness have ranged broadly. Most work has been done on the effectiveness of CBL approaches versus traditional methods, and little has been done on the comparative effects of CBL versus CBL methodologies. These findings urged other authors to recommend such studies in hopes of improving knowledge about which CBL methods work best in which settings.

Objective: In this systematic review, we aimed to characterize recent studies of the development of software platforms and interventions in medical education, search for common points among studies, and assess whether recommendations for CBL research are being taken into consideration.

Methods: We conducted a systematic review of the literature published from 2003 through 2013. We included studies written in English, specifically in medical education, regarding either the development of instructional software or interventions using instructional software, during training or practice, that reported learner attitudes, satisfaction, knowledge, skills, or software usage. We conducted 2 latent class analyses to group articles according to platform features and intervention characteristics. In addition, we analyzed references and citations for abstracted articles.

Results: We analyzed 251 articles. The number of publications rose over time, and they encompassed most medical disciplines, learning settings, and training levels, totaling 25 different platforms specifically for medical education. We uncovered 4 latent classes for educational software, characteristically making use of multimedia (115/251, 45.8%), text (64/251, 25.5%), Web conferencing (54/251, 21.5%), and instructional design principles (18/251, 7.2%). We found 3 classes for intervention outcomes: knowledge and attitudes (175/212, 82.6%), knowledge, attitudes, and skills (11.8%), and online activity (12/212, 5.7%). About a quarter of the articles (58/227, 25.6%) did not hold references or citations in common with other articles. The number of common references and citations increased in articles reporting instructional design principles (P=.03), articles measuring online activities (P=.01), and articles citing a review by Cook and colleagues on CBL (P=.04). There was an association between number of citations and studies comparing CBL versus CBL, independent of publication date (P=.02).

Conclusions: Studies in this field vary highly, and a high number of software systems are being developed. It seems that past recommendations regarding CBL interventions are being taken into consideration. A move into a more student-centered model, a focus on implementing reusable software platforms for specific learning contexts, and the analysis of online activity to track and predict outcomes are relevant areas for future research in this field.

(J Med Internet Res 2016;18(8):e204) doi:10.2196/jmir.5461

KEYWORDS

medical education; internet-based learning; computer-based learning; e-learning; b-learning; systematic review

Introduction

Medical education is a field that reflects the constant revision of medical knowledge, educational technology, and teaching strategies. For over a century, education in general [1] and medical education in particular [2-4] have been shifting from the traditional instructor-centered model to a learner-centered model, a shift in which the learner has greater control over the learning methodology and the teacher becomes a facilitator of the learning process [5]. This transition was required, since advances in medical knowledge and changes in health care delivery have weighed on the teaching responsibilities of medical schools [6]. The need to review and incorporate emerging fields into the curricula required medical schools to look for means to deliver education with less reliance on instructor availability [6]. The broadening of the setting in which health care is delivered—from the hospital to the community setting—prompted adaptation of these venues to ensure education could be delivered remotely [7]. Digital technology enabled the development of computer-based learning (CBL) and, later, Web-based learning methodologies, which allowed medical schools to cope with the pressing changes in the medical education landscape [4].

The increasing interest in and pervasiveness of CBL and Web-based learning was accompanied by research on how such methods compared with traditional instruction on a wide spectrum of educational end points, leading Friedman in 1994 to reflect on the research we should be doing on CBL [8]. In 2000, Adler and Johnson quantified the medical literature on CBL, concluding that researchers should focus on determining in which settings CBL methods are most adequate, rather than comparing them with the classroom setting [9]. According to these authors, provided that CBL offers tools that cannot be replicated by other means, the typical classroom setting cannot be considered a sound comparison group, as it undermines study internal validity [9,10].

The apparent lack of accommodation of this recommendation in subsequent studies, which kept growing in variety of setting and design, led Cook in 2005 to establish an agenda for research in medical education, suggesting once again that CBL research should look at relative benefits between different CBL methods [11]. In 2008, Cook et al conducted a broad meta-analysis of the effects of CBL in health sciences education, showing that CBL interventions are generally better than no intervention and marginally superior to traditional instruction [12]. Studies using multimedia learning content and student feedback reported the best results [12].

While the issue around CBL arose nearly 22 years ago, and over 8 years have passed since the Cook et al meta-analysis, comparative research between CBL methods is still a contemporary problem [13]. It is relevant to study what features of educational software researchers are reporting, how interventions are being conducted, what end points are being measured, and whether prior recommendations are informing current research. To our knowledge, since 2008 this issue has not been looked at again in a broad and systematic way, and is yet to be carried out specifically in medical education, as opposed to health sciences education in general.

Thus, this work aimed to identify reports of CBL software and CBL interventions, specifically in medical education, and systematically describe features of educational software, instructional design considerations, and the design, setting, and end points of CBL interventions. Finally, we intended to summarize these findings through determining subgroups of similar articles about educational software features and intervention end points, and to understand the extent to which prior work is being taken into consideration by analyzing the reference and citation network of these publications.

Methods

Study Eligibility

We included medical education studies written in English regarding the development of educational software, interventions using educational software, or both. We considered interventions during training or clinical practice that reported effects on learner attitudes, knowledge, and skills, as well as records of online activity. We included pretest-posttest studies, randomized and nonrandomized studies, parallel group and crossover studies, and studies in which a software-based intervention was added to other instructional methods [12].

We did not include studies that exclusively surveyed perceptions and attitudes of students or professionals toward CBL in general, nor studies that solely described course structure or reported how CBL strategies were implemented in medical schools.

Study Identification

We designed a strategy to search MEDLINE, Scopus, Web of Science, and EBSCO databases. Search terms were “medical education,” “medical students,” “e-learning,” “blended learning,” “information technology,” “instructional design,” “software,” and “Web-based platform,” among other terms. The exact queries are available in Multimedia Appendix 1. We established an 11-year period from January 1, 2003 to December 31, 2013. We performed the final database search on January 5, 2015.

Study Selection

Working independently and in duplicate, reviewers (PF, ITG) screened all article titles and abstracts, and in full text all potentially eligible abstracts, abstracts with disagreement, or abstracts with insufficient information. Independently and in duplicate the reviewers considered the eligibility of studies in full text with adequate chance-adjusted interrater agreement of .92 by intraclass correlation (ICC) using the psych package, version 1.5.1 [14] for the R programming language.

Study Analysis

Data Extraction

We conducted data extraction and reporting in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [15,16]. Reviewers abstracted data from each eligible study using a standardized data abstraction spreadsheet. We developed, tested, and revised the spreadsheet based on the review results of the first 30 assessed articles. Conflicts were resolved by consensus with a third reviewer (TTG). We abstracted information on publication year, country, study design, software used, instruction delivery method, CBL interactive features, CBL sharing features, instructional design principles, participant number and training level, study duration, type of comparison between groups, instruments used for assessment of knowledge, attitudes, and skills, correlations between study end points, and records of student online activity. For all categories, information was based on an explicit report of the variables of interest, except for instructional design principles, which we inferred from descriptions and figures using standardized criteria, whenever there were no explicit references [17]. In addition, articles that reported interventions were graded using the Medical Education Research Study Quality Index (MERSQI) for article reporting quality in medical education [18,19].

Data Analysis

We manipulated and prepared data for statistical analysis using NumPy [20] and pandas [21] libraries for the Python language. Latent class analysis uncovered distinct homogeneous groups of articles from the study population, considering that the performance of each article in a set of articles is explained by a categorical latent variable with *k* classes, commonly called latent classes [22]. Interpretation of the model was based on article profiles for each category, obtained from the probability of observing each variable in each class. We defined the number of latent classes according to the Bayesian information criterion (BIC), which is a measurement of model fit that penalizes models with many parameters, preventing model overfit [22]. Starting from a model with 1 class and increasing 1 class at a time, we chose the best model as the one with best interpretability and lowest BIC [22]. We created 2 latent class models, one taking into consideration educational software variables, and the other taking into consideration intervention end point variables. We did not use variables reported in <2% of the studies to compute the classes. Statistical analysis was conducted using the R programming language (The R Foundation). Class models were fitted using the poLCA package [23]. Summary panels were created using the ggplot2 package [24].

Reference and Citation Analysis

Data Extraction

We obtained references of the included papers from Scopus using digital object identifiers (DOIs). We obtained citations of the included papers from Google Scholar by searching for each of the articles by title and abstracting the papers on the "cited by" link. This procedure was carried out using a script built with the WebDriver library [25] for the JavaScript programming language. In order to uniquely identify every reference and citation, we performed a duplicate match and removal procedure by looking for similar matches of the title and authors' names using the fuzzywuzzy library [26] for the Python programming language. We considered 2 references or citations to be the same when the matching probability was >85%. Matching probability was computed using the Levenshtein string distance [27].

Data Analysis

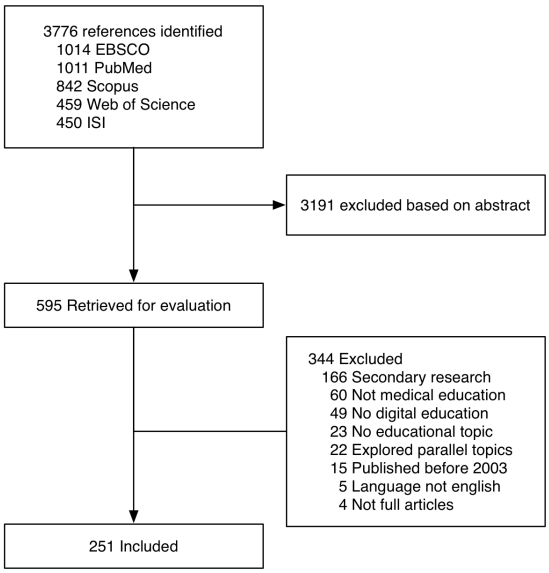
We analyzed the distribution of the total number of references and citations for each paper, and grouped papers based on whether they had ≥1 references or citations in common. We looked for the relationship between the number of citations and interventions comparing traditional instruction versus CBL methods, or CBL versus CBL. In addition, we assessed whether the number of related papers was associated with educational software latent classes or intervention end point latent classes, and with specific references to reviews by Cook and colleagues on CBL [11,12,28]. We used linear models adjusted for article publication year for this purpose. Statistical analysis was performed using the R language. We analyzed the article network using the graph-tool library for the Python programming language [29]. Error plots were created using the ggplot2 package [24] the R programming language.

Results

Study Eligibility, Identification, and Selection

The search strategy yielded 3776 citations, of which we identified 595 potentially eligible articles based on their abstract. Of these, we excluded 344 articles based on a full-text review. In total, we included and analyzed 251 articles. Overall mean ICC was .98. Specific ICCs are reported for variables that were not always explicitly present and relied on reviewer judgment, or when <.95. Figure 1 shows details regarding the study flow.

Figure 1. Flow of a systematic review of the literature published January 1, 2003 to December 31, 2013 regarding either the development of instructional software or interventions using instructional software.



Study Analysis

The number of publications rose over the years, from 13 of the 251 publications in 2003–2004 (5.2%), to 82 in 2012–2013 (32.7%). Medical schools in Germany, the United Kingdom, and the United States contributed more than 30 papers each between 2003 and 2013. Medical schools in Australia, Canada, and Spain contributed more than 10 papers each. Figure 2 presents contributions per medical school nationality. A total of 38 different software platforms were reported, which are listed in Multimedia Appendix 2. Of these, 13 were general educational platforms (34%), the most frequently used being Moodle [30-37] and Blackboard [38-46], mentioned in 8 papers, and WebCT [47-52], mentioned in 6 papers. The online virtual world Second Life [53,54] was mentioned in 2 papers, and 9 additional platforms were mentioned once. Of the 38 platforms, 25 were developed specifically for medical education (66%). Of these platforms, 4 were virtual patient simulators that were mentioned in 3 papers each: CASUS [55-58], HINTS [59-61], INMEDEA [62-64], and Web-SP [34,65,66]. One learning management system named MEFANET [67,68] was mentioned in 2 papers. Finally, 20 other platforms were mentioned once. These platforms were either learning management systems or virtual patient simulators. Of these, 4 systems were specialized in medical fields: a serious 3D game named EMSAVE [69], a system for learning electrocardiography named EKGtolkning [70], a platform entitled Radiology Teacher [71], and a virtual microscope named MyMiCROscope [72].

A total of 146 studies were conducted on clinical specialties (58.2%), 70 studies on basic sciences (27.9%), and 36 studies on surgical specialties (14.3%). Radiology was the clinical specialty with most studies, in 23 articles (9.2%), followed by pediatrics with 13 (5.2%). The basic science subjects with most publications were anatomy with 18 articles (7.2%) and physiology with 9 articles (3.6%). The most studied surgical specialties were urology with 12 studies (4.8%) and general surgery with 10 (4.0%). There was at least one article in most basic sciences and medical specialties, as Figure 3 shows.

Figure 2. Articles published per country of medical school. The article count axis is presented in logarithmic scale for better data representation.

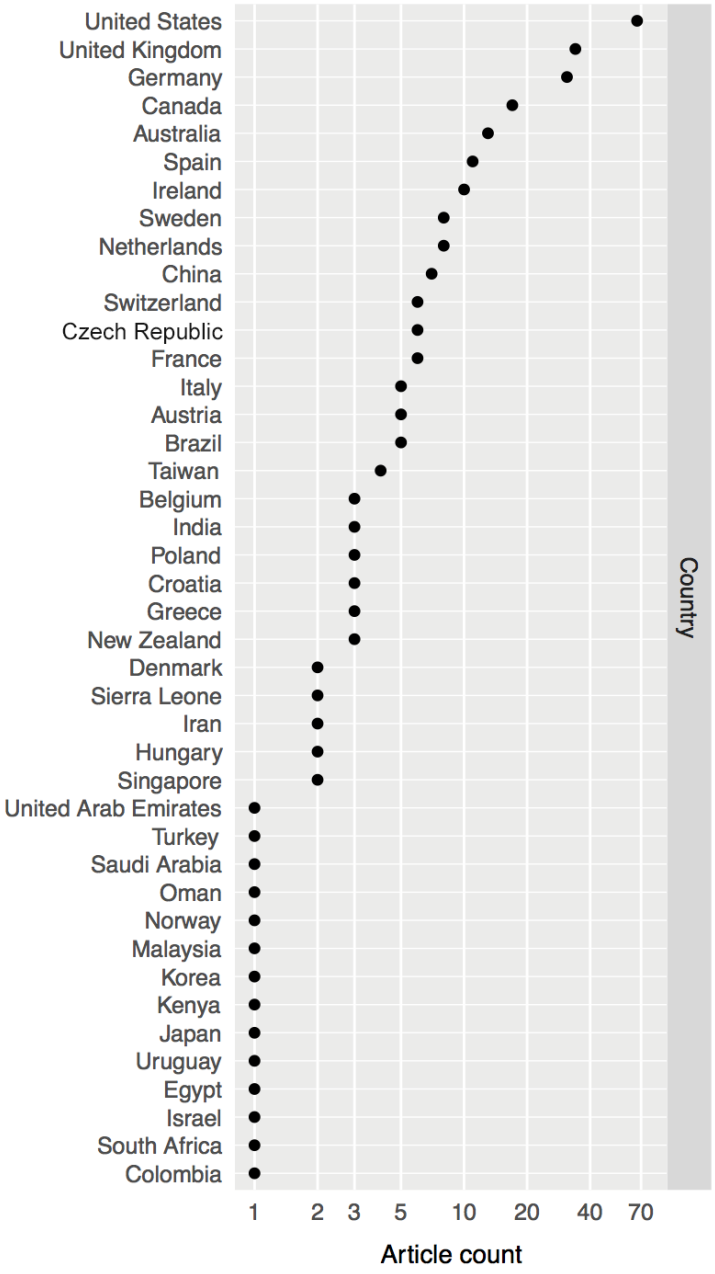


Figure 3. Articles per basic science and clinical subject. The article count axis is presented in logarithmic scale for better data representation.

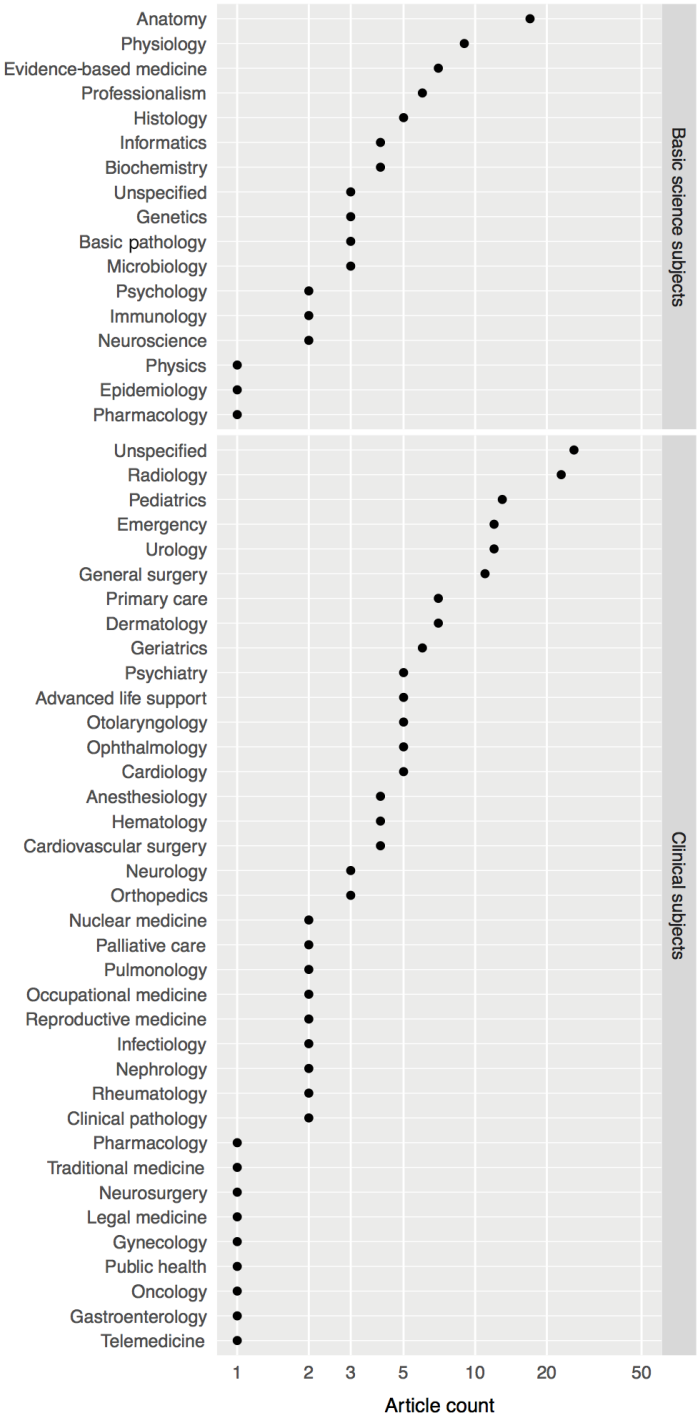
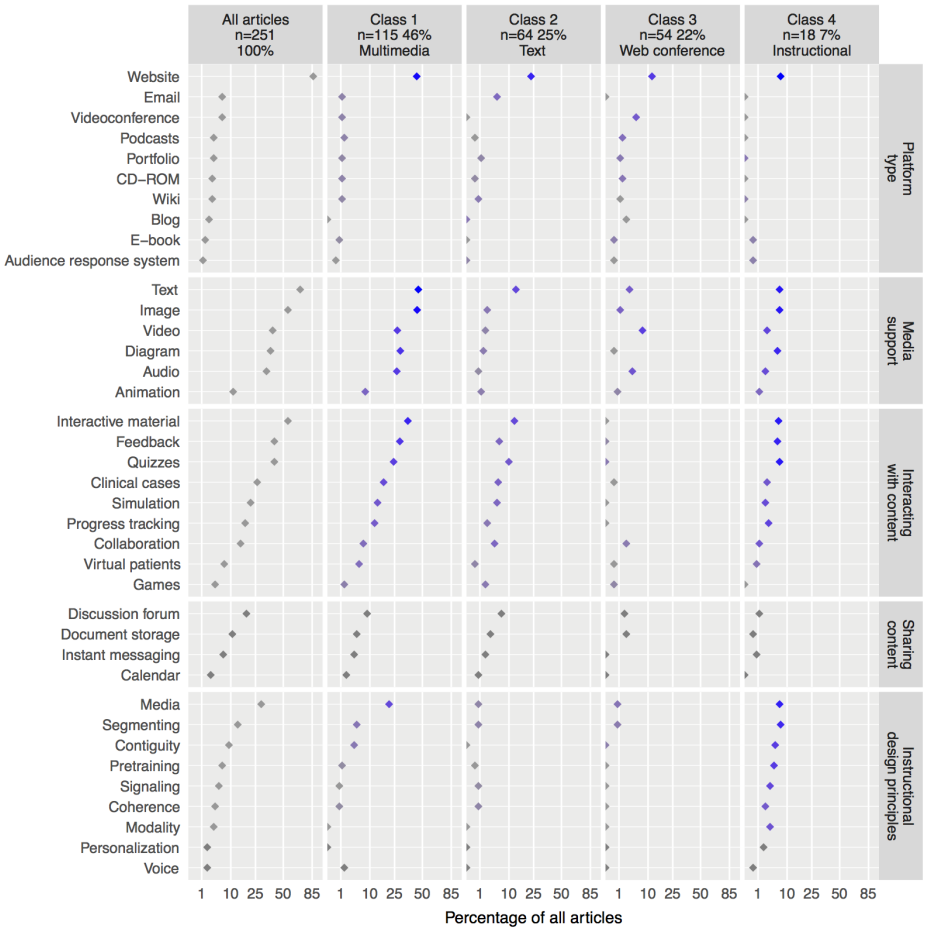


Figure 4. Prevalence of articles per educational software feature and educational software latent class. Horizontal axis ranges between 0 and 100 on a squared root scale. Point color specifies the probability of assigning a paper to each class based on the presence of each variable (gray indicates a probability of 0, ranging to dark blue indicating the highest probability). From the listed variables, those present in more than 2% of all articles were used to determine the educational software latent classes.



Web-Based Learning Software

Of the 251 studies assessed, 113 reported blended learning environments (45.0%, ICC=.98) and 138 reported e-learning environments (55.0%, ICC=.99). Figure 4 summarizes the results for this section and depicts the percentage of studies and relative contribution of each of the learning software variables to the software latent classes described below.

Platform Type

A total of 217 studies used websites (86.5%), 16 used videoconference (6.4%), and 16 other studies used email (6.4%); 9 used podcasts (3.6%) and 9 used portfolios (3.6%). Wikis (3.2%, ICC=.90) and CDs (3.2%, ICC=.83) were both reported in 8 studies. Blogs were reported in 6 studies (2.4%). E-books were reported in 4 studies (1.6%) and audience response systems in 3 articles (1.2%).

Media Support

Of the 251 studies, 174 provided content in text format (69.3%), and 138 used images (55.0%). Video was reported in 99 studies (39.4%) and diagrams in 94 studies (37.5%). Audio was used in 85 articles (33.9%), and animations were reported in 28 articles (11.2%).

Interacting With Content

A total of 138 studies reported unspecified features (55.0%). The software provided feedback to the learner in 103 studies (41.0%); 103 articles reported quizzes (41.0%), 66 reported clinical cases (26.3%), 54 described simulations (21.5%), and 45 tracked learner performance (17.9%). Features allowing collaboration between learners and instructors were reported in 38 studies (15.1%). Virtual patients were reported in 18 studies (7.2%) and games were described in 10 studies (4.0%).

Sharing Content

Of the 251 studies, 47 reported communication and content sharing through discussion forums (18.7%), 27 reported the ability to store documents (10.8%), and 7 used instant messaging communication systems (2.8%). Calendar features were also reported in 7 studies (2.8%).

Instructional Design Principles

The media principle was apparent in 74 studies (29.5%, ICC=.94), followed by the segmenting principle in 34 studies (13.6%, ICC=.98) and the contiguity principle in 23 studies (9.2%, ICC=1.00). The pretraining principle was identified in 16 studies (6.4%, ICC=.98) and the signaling principle in 13 studies (5.2%, ICC=.97). The coherence principle was identified in 10 studies (4.0%, ICC=.97) and the modality principle in 9 studies (3.6%, ICC=1.00). Finally, the personalization principle

and the voice principle were identified in 5 studies each (2.0%, ICC=1.00).

Latent Classes

We considered 4 distinct classes for educational software, according to the model statistics in Table 1. Class 1 was composed of 115 studies (45.8%), mostly of website-based interactive systems presenting content using text, images, audio, and video. Student feedback features were frequently described, namely quizzes and clinical cases. Aside from the multimedia principle, instructional design considerations were rarely present. We thus labeled class 1 *multimedia*.

Table 1. Latent class analysis model fit per number of classes for educational software.

No. of classes	Log likelihood	Parameter number	BIC[a]
1	–2340	21	4797
2	–2017	43	4273
3	–1923	65	4207
4[b]	–1866	87	4214[b]
5	–1854	109	4230

[a]BIC: Bayesian information criterion.

[b]The number of classes selected for the educational software model. This decision was based on picking the model with the best interpretability and lowest BIC.

Class 2 was composed of 64 studies (25.5%) using websites, and to a smaller extent email, to deliver instructional content mostly in the form of text. Interactive features were less frequent than in class 1, and instructional design considerations were scarce. We thus labeled class 2 *text*.

Class 3 was composed of 54 studies (21.5%) making use of websites and videoconference platforms to provide video and audio content. Interactivity and instructional design principles were nearly nonexistent. We thus labeled class 3 *Web conference*.

Class 4 contained 18 studies (7.2%) mostly using Web-based interactive multimedia apps in which the use of multiple instructional principles was frequent. We thus labeled class 4 *instructional*.

The four right-hand columns in Figure 4 depict the composition of each class and the relative weight of each variable on class assignment.

Interventions

Of the 251 articles included in this study, we identified 212 conducting interventions on the end points of interest (84.5%). Figure 5 summarizes the results for this section and depicts the percentage of studies for each intervention characteristic, and the relative contribution of intervention end point variables to the *intervention end point* latent class described below.

Study Design and Study Sample

A total of 81 of 212 studies were conducted with medical students from preclinical years (38.2%) and 56 studies involved students during clinical rotations (26.4%). In addition, 32 studies were conducted with specialist medical doctors (15.1%), and 31 studies were conducted with medical residents (14.6%).

In total, 55 interventions were carried out with <50 participants (25.9%), 97 studies had a sample size ranging between 50 and 200 participants (45.8%), and 59 studies were conducted with >200 students (27.8%).

Of the 212 studies, 54 were conducted over <1 week (25.5%), 90 articles reported interventions lasting between 1 week and 3 months (42.5%), and 50 studies were conducted for >3 months (23.6%).

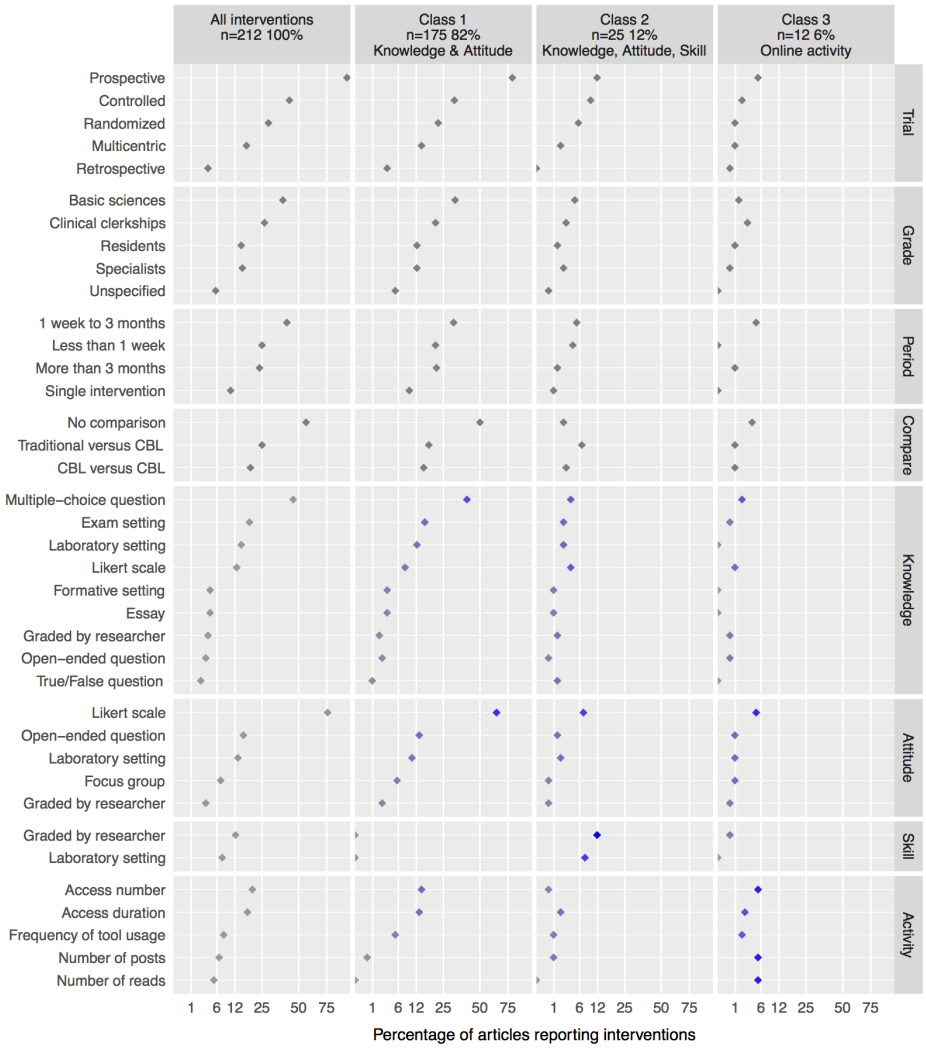
In addition, 84 studies repeatedly tested participants in a pre-post approach (39.6%), and 93 made use of control groups (43.9%). A total of 61 studies were randomized (28.8%) and 37 studies had participants from more than one institution (17.5%); 40 studies compared different CBL approaches (18.9%), while 53 studies compared CBL with traditional methods (25.0%).

The mean MERSQI score for the assessed studies was 9.54 (SD 1.84).

Conducted Comparisons Between Groups

Of the 212 studies, 28 studied controlled interventions between blended learning approaches and traditional lectures (13.2%), while 11 studies compared e-learning approaches with traditional lectures (5.2%). A total of 8 studies compared spaced repetition versus bolus learning (3.8%), and 7 studies compared e-learning *v* ersus no intervention (3.3%). In addition, 5 studies compared the use of 3D models versus 2D images (2.4%). A multitude of other comparisons were performed, such as exploratory versus blocked learning approaches [73-75], complex versus simple user interfaces [73,76,77], immediate versus delayed completion of lectures in CBL systems [78], and multimedia versus text on CBL media [73,79-81]. Multimedia Appendix 3 lists the different comparison groups we identified for each of the 212 articles reporting interventions.

Figure 5. Prevalence of articles per intervention feature and intervention endpoint latent class. Horizontal axis ranges between 0 and 100 on a squared root scale. Point color specifies the probability of assigning a paper to each class based on the presence of each variable (gray indicates a probability of 0, ranging to dark blue indicating the highest probability). Only variables regarding assessment of knowledge, attitudes, skills, and online activity (the 4 last panels) were used to determine intervention end point latent classes. CBL: computer-based learning.



Knowledge End Point

Knowledge outcomes were assessed in 120 of 212 articles (56.6%). Objective knowledge was assessed using multiple choice quizzes in 98 of 120 studies (81.7%), 9 articles used free-text fields (7.5%), and 8 articles used open-ended questions (6.7%, ICC=.89). In addition, 5 studies used true/false questions (4.2%). Judgments of knowledge were collected using Likert scales in 27 studies (22.5%). Researchers directly assessed knowledge in 9 studies (7.5%). A total of 31 studies were conducted in a laboratory setting (25.8%). Knowledge assessment was part of a final examination in 39 articles (32.5%), and in 9 studies assessment was part of a formative assessment (7.5%). Of the 120 studies, 90 reported that interventions improved knowledge acquisition (75.0%), while 27 studies did not find significant effects (22.5%) and 3 multicenter randomized controlled trials reported that interventions did not positively affect knowledge acquisition (2.5%) [66,82,83].

Attitude End Point

Of 212 studies, 172 assessed student attitudes (81.1%); of these, 163 used Likert scales (94.8%) and 34 used free-text fields (19.8%). In 8 articles researchers assessed participants' attitudes directly (4.7%). A total of 29 studies were conducted in a laboratory setting (16.9%) and 16 studies made use of focus groups (9.3%). In addition, 161 studies found positive attitudes toward interventions (93.6%), 8 found neutral attitudes (4.7%), and 3 reported negative attitudes (1.7%) [84-86].

Skill End Point

Of 212 studies, 31 assessed subject skills (14.6%). In 26 of these studies, skills were assessed directly by researchers (84%) and in 16 studies assessment was conducted in a laboratory setting (52%). In addition, 24 studies found positive effects on skills acquisition (77%), 5 reported that the interventions had no effect on assessed skills (16%), and 2 reported that the intervention had negative effects (6%) [82,86].

Online Activity End Point

Online activity was measured in 76 of 212 studies (35.9%). Of these studies, 46 measured total logins to the system (60%), 39

measured time spent in the system (51%), and 18 measured the number of times students used specific learning tools (24%). Further, 16 studies measured the number of student posts (21%), and 12 measured the number of times students viewed the learning materials (16%). A total of 41 studies found no relationship between activity patterns and learning outcomes (54%), 34 articles reported increased activity to have positive effects on learning outcomes (45%), and 1 study found a negative effect (1%) [66].

Intervention End-Point Latent Classes

We considered 3 distinct classes to group the 212 studies taking into consideration intervention end point variables. Class 1 contained 175 articles assessing knowledge and attitudes (82.5%). We labeled class 1 *knowledge and attitude*. Class 2 contained 25 intervention studies (11.8%). In addition to assessing knowledge and attitudes, articles in this class also assessed skills. We labeled class 2 *knowledge, attitude, and skill*. Class 3 contained 12 studies that assessed online activity, specifically through the number of posts and number of reads (5.7%). Attitudes were always assessed, but knowledge and skill assessment were nearly absent. We labeled class 3 *online activity*. Table 2 reports model statistics for the intervention end point latent classes, and Figure 5 depicts the prevalence of articles per intervention feature and intervention end point latent class.

Table 2. Latent class analysis model fit per number of latent classes for intervention end points.

No. of classes	Log likelihood	Parameter number	BIC[a]
1	–1631	22	3382
2	–1510	45	3265
3[b]	–1451	68	3270[b]
4	–1424	91	3268

[a]BIC: Bayesian information criterion.

[b]The number of classes for the intervention end point model. This decision was based on picking the model with the best interpretability and the lowest BIC.

Reported Correlations Between Assessment Outcomes

Of 212 studies, 25 correlated different variables with knowledge outcomes (11.8%). Of these, 1 study correlated system interactivity with knowledge scores and concluded that lower levels of interactivity benefitted knowledge acquisition [73]. Correlations between knowledge gains and time spent using online platforms were also sought. These were found to be positive in 4 studies [49,87-89] and neutral in 1 study [76]. In addition, 1 study described a modest positive correlation between increased knowledge scores on the learning system and an increase in examination scores [90]. Increased learning platform usage was correlated positively with knowledge acquisition in 5 studies [90-94], while 4 found no association [46,95-97]. Other studies found positive relationships between knowledge and the number of posts in online forums [98,99] and comprehensiveness of student study materials [100]. Regarding attitudes, 2 articles found a mild positive correlation between judgments of knowledge and knowledge score [101,102]. Other correlations were assessed, namely confidence and skill [103], study duration and skill [104], and study duration and learning style [105], but did not reach statistical significance.

Reference and Citation Network Analysis

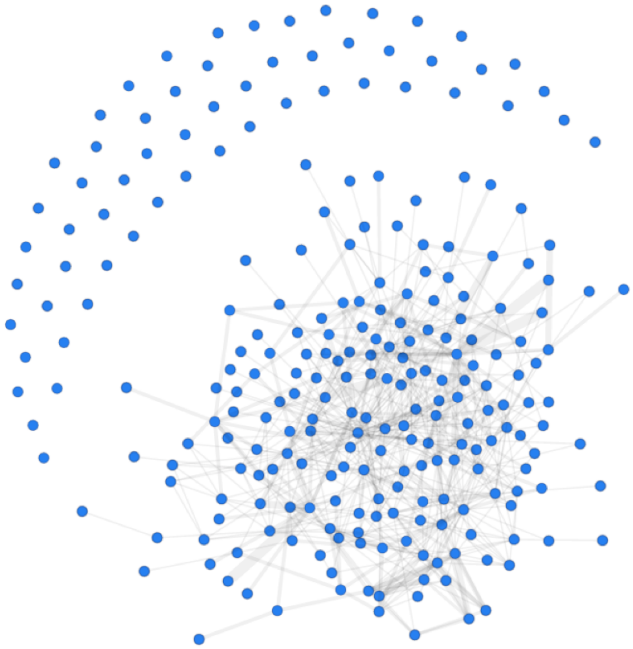
Reference and Citation Analysis

We obtained references and citations for 227 of the 251 articles included in this review (90.4%). The mean number of references was 26.12 (SD 17.41). In total, the abstracted articles had 4010 references to other articles. The most referenced articles were from Ruiz et al [4], Cook et al [12], Chumley-Jones et al [106], Greenhalgh [107], Ward et al [108], Muller [109], and Ellaway and Masters [110]. The mean number of article citations of the 227 abstracted articles was 14.43 (SD 12.12). More than half of the references were common to various abstracted articles, while a smaller percentage of studies had independent sets of references.

Related Article Analysis

Of the 227 articles, 169 had at least one reference or citation in common with other abstracted articles (74.4%), and were thus said to be related, as depicted in Figure 6. A total of 58 articles were not related to any other article, since they did not share references or citations (25.6%). The mean number of related studies for each article included in this review was 4.74 (SD 5.42).

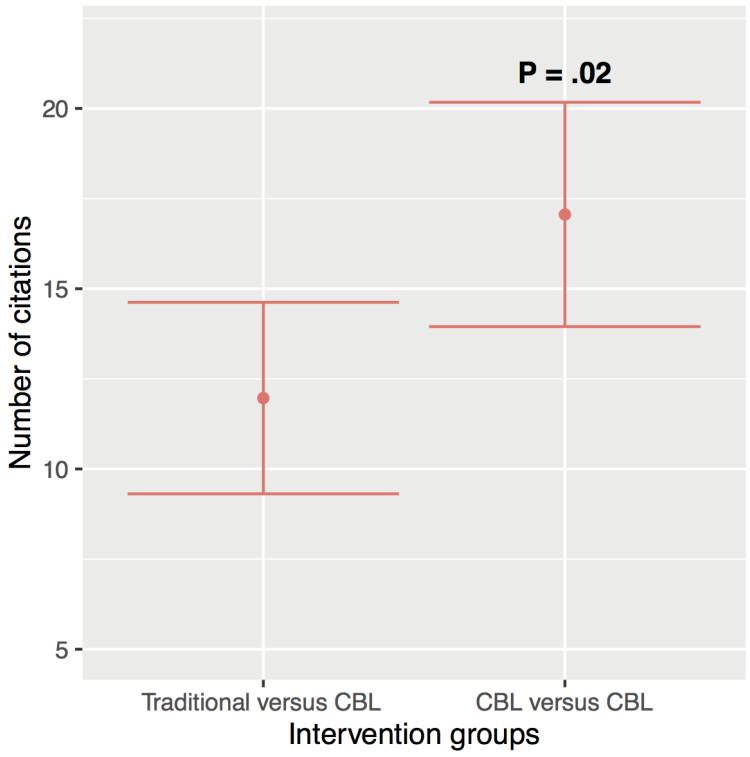
Figure 6. Relationships between articles included in this review (indicated by nodes). Links between nodes indicate that articles have references and citations in common. The width of the link indicates the number of studies in common, ranging from 1 to 5. About a quarter of the studies have no common references or citations. Only 227 of the 251 studies were included in this analysis due to missing information (90.4%).



Citation Differences Between Intervention Group Types

Studies comparing traditional versus CBL methods were cited a mean of 11.92 times (95% CI 9.31–14.6). Studies comparing different CBL methods were cited a mean of 16.71 times, which was statistically significant (95% CI 13.95–20.17, *P*=.02). Figure 7 shows this result.

Figure 7. Mean citation number differences between traditional versus computer-based learning (CBL), and CBL versus CBL, adjusted for publication date. For CBL versus CBL, only 227 of the 251 studies were included in this analysis due to missing information (90.4%). Error bars represent the 95% CI.

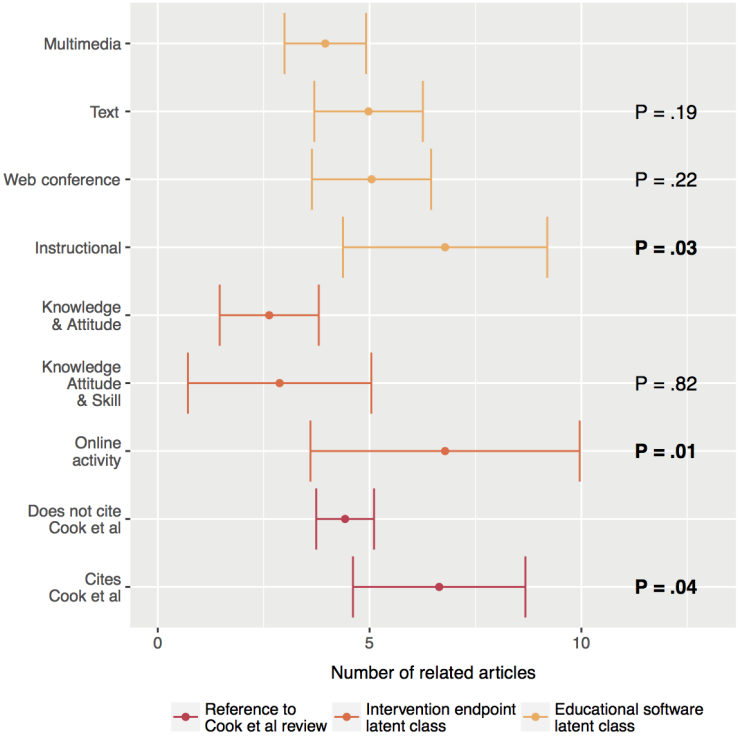


Associations With Latent Classes and the Cook et al Review

Regarding educational software latent classes, articles in the multimedia class had a mean of 3.95 related studies (95% CI 2.99–4.91), while the text class had a mean of 4.98 (95% CI 3.69–6.26, *P*=.19). Articles from the Web conference class had a mean of 5.02 relationships to other studies (95% CI 3.64–6.45, *P*=.22) and articles in the instructional class had a statistically significant mean of 6.78 studies (95% CI 4.37–9.20, *P*=.03) in common. Regarding the intervention end point latent classes, articles in the knowledge and attitude class had a mean of 2.63 related studies (95% CI 1.46–3.80) and the knowledge, attitude, and skill class had a mean of 2.88 studies in common, reaching statistically significance versus the knowledge and attitude class (95% CI 0.71–5.04, *P*=.04). Articles from the online activity class had a mean of 6.78 related studies (95% CI 3.60–9.96, *P*=.01), also reaching a significant value when compared with the knowledge and attitude class.

Finally, articles not citing the Cook et al work had a mean related article count of 4.42 (95% CI 3.74–5.11), while articles citing Cook et al had a mean count of 6.64 (95% CI 4.61–8.68, *P*=.04), which was significantly different. Figure 8 plots the complete results for this section.

Figure 8. Mean number of related articles per latent class and reference to the Cook et al review. Number of related articles is adjusted for publication date. *P* values indicate intraclass pairwise differences from the topmost element of each color-coded class. Significant relationships are marked in bold typeface. Only 227 of the 251 studies were included in this analysis due to missing information (90.4%). Error bars represent the 95% CI.



Discussion

The number of articles on CBL in medical education has been rising, with reports of over 38 different software systems, 25 of which were specifically developed for medical education (66%). Of the 251 studies we analyzed, most used interactive websites making use of text and images (46%) and, to a smaller extent, websites delivering text-based materials (25%). A similar number of reports delivered instruction using Web conferencing systems (22%), and a smaller group of studies reported highly interactive websites with multimedia learning content built according to instructional design principles (7%). Of the 212 interventions, most did not use comparison groups and lasted between 1 week and 3 months. CBL versus CBL studies were less numerous than traditional versus CBL studies. Nearly all studies assessed student attitudes, of which a large fraction also assessed knowledge (82%), and a smaller fraction assessed knowledge and skills (12%). A smaller set of studies looked specifically for patterns of online activity, namely the number of reads and posts (6%). Finally, nearly 75% of articles had references and citations in common, while 25% of the analyzed articles did not have any references in common. Articles comparing different CBL methods were cited more often than were studies comparing traditional versus CBL methods, independent of publication date. Articles reporting instructional design principles, articles measuring online activity, and articles citing the Cook et al CBL reviews had significantly more references and citations in common than did other articles.

Comparison With Previous Reviews

The last systematic review and meta-analysis of this topic encompassed data from 1990 to 2006 and highlighted the problems of intervention variability and lack of evidence for comparative effects of CBL methods [12,13,28]. Recent reviews have also demonstrated that practice exercises, interactivity, feedback, and repetition can favorably influence learning outcomes [13,49]. Other reviews summarized technologies and

methods used [111,112], and addressed specific topics such as the role of blogs [113], wikis [114], portfolios [115], simulations in general [116] and for surgery in particular [117], gastroenterology [118], catheterization [119], and airway management [120]. Other authors focused on specific aspects of the effects of Web-based learning on problem-based learning [121], and the implications of recent Web capabilities, namely Web 2.0 [122,123] and Web 3.0 [124], for medical education. Our study complements previous reviews by encompassing recent work concerning these fields over a large base of abstracted articles. Despite the considerable time overlap with similar reviews, assessments such as latent class analysis and citation network analysis had not yet been conducted during the considered time period [13].

Limitations and Strengths

This study has limitations. We scrutinized databases that frequently index medical education articles. Although we did not query EMBASE, Scopus covers most of the literature indexed in EMBASE and thus Scopus provided a reasonable proxy. However, we did not abstract the gray literature or references from other articles, and thus our article search cannot be considered exhaustive. We narrowed the study participants to medical education only. This can be considered a limitation insofar as these findings cannot be generalized to other health professions. Other reviews have performed similar searches including articles in health professions in general [12]. We performed the article abstraction step manually. While the independent reviewing method and ICC results indicate a low probability of coding error, we cannot completely exclude it. Variables regarding instructional design and assessment outcomes were often not explicitly declared and relied on reviewer judgment. We could not retrieve references and citations for 27 of the 251 articles (10.8%), and unique reference and citation matching relied on probabilistic algorithms that considered a small but nonnegligible error margin.

This study also has strengths. We performed a broad analysis of the literature and accounted for aspects that, to our knowledge, were not previously assessed, such as specific platforms and their features, and correlations assessed between learning end points and types of comparisons. We systematically summarized data using latent class analysis, which, to our knowledge, was for the first time performed in this setting. We described the article citation network and explored relationships between these and the article latent classes and CBL considerations, which, to our knowledge, were also for the first time performed in the field. Finally, we have made these results available through an interactive visualization that allows researchers to deeply explore articles [125].

Implications

CBL Research Should Include Evidence From More Medical Schools

Our findings show that, while CBL in medical education varies significantly, most published articles are from medical schools in a small set of countries. Medical education has geographical specificities, which makes contributions from different geographical areas particularly enriching and should incite more schools to conduct research in this field.

Platform Development Should Avoid Reinventing the Wheel

Over 25 platforms and software projects were built specifically for medical education, despite having significant overlap in goals and features. While a few provided means to interact with learning materials, such as microscopy images [72], in ways not before possible, it would be worthwhile for researchers to try to develop open and generalizable systems addressing specific learning contexts that can be reused by researchers from other medical schools. Initiatives to design pluggable modules for mainstream learning management systems and reusable learning materials, such as learning objects [126], aimed at specific medical contexts should be preferred over building closed systems from scratch.

Instructional Design Considerations Should be Reported

The diversity of methods encompassed by CBL in terms of delivery medium, context, learner, and purpose, without reports of instructional design considerations, obfuscates the effect of different intervention aspects, for which instructional design—or the lack of it—is partly accountable [8,9,13,121]. The value of reporting interactive tools, such as quizzes with feedback, would also increase. Determining which principles best apply to different medical settings and medical knowledge is an issue of interest for future research [8].

Interventions Should Focus on Assessing Unexplored Outcomes

Studies generally report positive outcomes on knowledge, attitudes, and skills. Interestingly, studies that found no positive effect in any of the learning outcomes were often randomized controlled trials [66,83-86], some of them running in multiple institutions [127,128]. Studies with little or no description of the learning and teaching methodology had neutral findings [82,129]. Once again, the lack of comparable arms, such as CBL versus traditional instruction, makes it difficult to assess intervention outcomes. Furthermore, data showing that objective knowledge assessment and skills increase with interventions can be used in deeper ways. Real-time collection of student activity, together with objective performance assessment through multiple choice quizzes, may have predictive value. Judgments of knowledge together with other student activity metrics may provide data for a next generation of intelligent tutoring systems able to track, manage, and predict student performance [130]. An increase in studies reporting online activity measurements and correlations with other learning outcomes using reproducible tools, as described before, would generate useful evidence on the effectiveness of CBL methods in enhancing learning [131]. Metrics could include, for example, student communication style and sentiment [132,133] and time spent on different types of materials [134].

CBL Research Seems to be Progressing on the Right Track

Even though 25% of the articles seemed not to be based on common CBL literature, our findings suggest that research is

moving toward favoring studies comparing CBL methods rather than comparison with traditional methods. Indeed, we found that, while traditional versus CBL articles were more numerous, articles comparing different CBL methods were cited more often than articles comparing CBL versus traditional settings. We take this as a sign that recommendations put forward by previous authors are being taken into consideration [8,9,11]. Articles in the instructional and online activity latent classes, as well as those citing the Cook et al meta-analysis [12], had more references and citations in common with other articles, demonstrating greater awareness of research in this field and possibly indicating future research directions.

A Further Push Into a Student-Centered Models is Key

The shift to student-centered models needs to continue. However, only a few reports put students at the center of the education process, focusing usually on aspects related to teaching [135]. Part of the success of CBL features comes from empowering students to conduct study sessions at their own pace, providing them with richer interactions with learning materials, and facilitating communication, which were not otherwise feasible. Promoting student self-directedness through social media and reward-based systems may lead to increased engagement and improved learning outcomes [136]. Active learning through engagement in collaborative user-generated content, facilitated communication, and feedback in which instructors act as moderators may further promote this change [137]. Engaging students in the creation of content can be a good way to help faculty cope with the increasing demand for learning material [138]. Social media tools such as wikis have been used in the medical context for various purposes [139], but in medical education they still are limited in their format, management, and collaborative features [140]. Other approaches using 3D virtual worlds may offer great potential to learners through immersive exploratory worlds and a rich feedback environment that may be used to engage learners and simulate real-world scenarios of medical doctors [140].

Conclusions

We have come a long way in CBL in medical education. While the field is highly variable and some studies seemed to be unaware of advances in the field, recommendations on comparing different CBL methods seem to have been taken into consideration. Incorporating instructional design principles in the design of learning materials and developing further educational software in ways that can be shared between researchers are paths for further improvement. A focus on measuring online activity and correlating it with other outcomes may provide insights into ways to keep promoting student-centered approaches tailored to specific learning settings.

Authors' Contributions

TTG designed the study, abstracted the articles, conducted statistical analysis, wrote the manuscript, and developed the Web app. PF created the initial data extraction sheet, designed the study, abstracted the articles, and wrote the manuscript. ITG designed the study, abstracted the articles, and wrote the manuscript. MS designed the study and conducted the statistical analysis. MAF designed the study, wrote the manuscript, and gave overall approval and direction.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search queries.

[PDF File (Adobe PDF File), 119KB - jmir_v18i8e204_app1.pdf]

Multimedia Appendix 2

Educational software of abstracted papers.

[PDF File (Adobe PDF File), 64KB - jmir_v18i8e204_app2.pdf]

Multimedia Appendix 3

Complete reference of abstracted papers.

[PDF File (Adobe PDF File), 304KB - jmir_v18i8e204_app3.pdf]

Multimedia Appendix 4

Interactive article explorer.

[ZIP File (Zip Archive), 751KB - jmir_v18i8e204_app4.zip]

References

1. Dewey J. The School and Society and the Child and the Curriculum. Chicago, IL: University of Chicago Press; 2013.

2. Bahner D, Adkins E, Patel N, Donley C, Nagel R, Kman N. How we use social media to supplement a novel curriculum in medical education. Med Teach 2012;34(6):439-444. [doi: 10.3109/0142159X.2012.668245] [Medline: 22449268]
3. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
4. Ruiz J, Mintzer M, Leipzig R. The impact of e-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [Medline: 16501260]
5. Koops W, Van der Vleuten C, De Leng B, Oei SG, Snoeckx L. Computer-supported collaborative learning in the medical workplace: students' experiences on formative peer feedback of a critical appraisal of a topic paper. Med Teach 2011;33(6):e318-e323. [doi: 10.3109/0142159X.2011.575901] [Medline: 21609168]
6. Ozuah PO. Undergraduate medical education: thoughts on future challenges. BMC Med Educ 2002 Jul 30;2:8 [FREE Full text] [Medline: 12149136]
7. Nair BR, Finucane PM. Reforming medical education to enhance the management of chronic disease. Med J Aust 2003 Sep 1;179(5):257-259. [Medline: 12924974]
8. Friedman CP. The research we should be doing. Acad Med 1994 Jun;69(6):455-457. [Medline: 8003158]
9. Adler M, Johnson K. Quantifying the literature of computer-aided instruction in medical education. Acad Med 2000;75(10):1025-1028.
10. Hagler P, Knowlton J. Invalid implicit assumption in CBI comparison research. J Comput-Based Instr 1987:84-88.
11. Cook D. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med 2005 Jun;80(6):541-548. [Medline: 15917356]
12. Cook D, Levinson A, Garside S, Dupras D, Erwin P, Montori V. 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]
13. Cook D, Levinson A, Garside S, Dupras D, Erwin P, Montori V. 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]
14. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, IL: Northwestern University; 2015. URL: http://cran.r-project.org/web/packages/psych/index.html [accessed 2016-07-11] [WebCite Cache ID 6ivuWeLLZ]
15. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
16. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventionsxplanation and elaboration. Ann Intern Med 2009;151(4):e.
17. Mayer R. Applying the science of learning to medical education. Med Educ 2010 Jun;44(6):543-549. [doi: 10.1111/j.1365-2923.2010.03624.x] [Medline: 20604850]
18. Reed D, Cook D, Beckman T, Levine R, Kern D, Wright S. Association between funding and quality of published medical education research. JAMA 2007 Sep 5;298(9):1002-1009. [doi: 10.1001/jama.298.9.1002] [Medline: 17785645]
19. Reed D, Beckman T, Wright S, Levine R, Kern D, Cook D. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. J Gen Intern Med 2008 Jul;23(7):903-907 [FREE Full text] [doi: 10.1007/s11606-008-0664-3] [Medline: 18612715]
20. van der Walt S, Colbert S, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 2011 Mar;13(2):22-30. [doi: 10.1109/MCSE.2011.37]
21. McKinney W. Data structures for statistical computing in Python. 2010 Presented at: Proceedings of the 9th Python in Science Conference; July 3, 2010; Austin, TX p. 51 URL: http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf
22. Nylund K, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equation Modeling 2007;14(4):69.
23. Linzer D, Lewis J. poLCA: an R package for polytomous variable latent class analysis. J Stat Softw 2011;42(10):29.
24. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer-Verlag; 2009.
25. Selenium WebDriver.: Selenium HQ; 2006. URL: http://www.seleniumhq.org/projects/webdriver [accessed 2016-04-10] [WebCite Cache ID 6gfAEsLqI]
26. Cohen A. Fuzzy Wuzzy. 2014. URL: https://pypi.python.org/pypi/fuzzywuzzy [accessed 2016-04-10] [WebCite Cache ID 6gfACTuzC]
27. Bilenko M, Mooney R, Cohen W, Ravikumar P, Fienberg S. Adaptive name matching in information integration. IEEE Intell Syst 2005:23.
28. Cook D. Where are we with Web-based learning in medical education? Med Teach 2006 Nov;28(7):594-598. [doi: 10.1080/01421590601028854] [Medline: 17594549]
29. Peixoto T. The Graph-Tool Python Library.: figshare; 2014. URL: https://figshare.com/articles/graph_tool/1164194 [accessed 2016-04-10] [WebCite Cache ID 6gfAGLrrk]

30. Shah I, Walters M, McKillop J. Acute medicine teaching in an undergraduate medical curriculum: a blended learning approach. Emerg Med J 2008 Jun;25(6):354-357. [doi: 10.1136/emj.2007.053082] [Medline: 18499818]
31. Kukolja-Taradi S, Dogas Z, Dabić M, Drenjancevic Perić I. Scaling-up undergraduate medical education: enabling virtual mobility by online elective courses. Croat Med J 2008 Jun;49(3):344-351 [FREE Full text] [Medline: 18581612]
32. Puljak L, Sapunar D. Web-based elective courses for medical students: an example in pain. Pain Med 2011 Jun;12(6):854-863 [FREE Full text] [doi: 10.1111/j.1526-4637.2011.01104.x] [Medline: 21481172]
33. Sparacia G, Cannizzaro F, D'Alessandro D, D'Alessandro M, Caruso G, Lagalla R. Initial experiences in radiology e-learning 1. Radiographics 2007;27(2):581. [doi: 10.1148/rg.272065077]
34. Botezatu M, Hult HA, Tessma M, Fors UG. Virtual patient simulation for learning and assessment: superior results in comparison with regular course exams. Med Teach 2010;32(10):845-850. [doi: 10.3109/01421591003695287] [Medline: 20854161]
35. Seluakumaran K, Jusof F, Ismail R, Husain R. Integrating an open-source course management system (Moodle) into the teaching of a first-year medical physiology course: a case study. Adv Physiol Educ 2011 Dec;35(4):369-377 [FREE Full text] [doi: 10.1152/advan.00008.2011] [Medline: 22139773]
36. Worm B, Jensen K. Does peer learning or higher levels of e-learning improve learning abilities? A randomized controlled trial. Med Educ Online 2013;18:21877 [FREE Full text] [Medline: 24229729]
37. Iyeyasu JN, Castro S, Sabbatini R, Carvalho K. The development and evaluation of a distance learning system in ophthalmology. Rev Bras Educ Med 2013 Mar;37(1):96-102. [doi: 10.1590/S0100-55022013000100014]
38. Wiecha J, Gramling R, Joachim P, Vanderschmidt H. Collaborative e-learning using streaming video and asynchronous discussion boards to teach the cognitive foundation of medical interviewing: a case study. J Med Internet Res 2003;5(2):e13 [FREE Full text] [doi: 10.2196/jmir.5.2.e13] [Medline: 12857669]
39. de Leng BA, Dolmans MDH, Muijtjens AM, van der Vleuten CP. Student perceptions of a virtual learning environment for a problem-based learning undergraduate medical curriculum. Med Educ 2006 Jun;40(6):568-575. [doi: 10.1111/j.1365-2929.2006.02484.x] [Medline: 16700773]
40. Wiecha JM, Chetty VK, Pollard T, Shaw PF. Web-based versus face-to-face learning of diabetes management: the results of a comparative trial of educational methods. Fam Med 2006 Oct;38(9):647-652 [FREE Full text] [Medline: 17009189]
41. Thakore H, McMahon T. Designing an interactive multimedia rich tutorial for medical students: beyond a 'book on a screen'. J Vis Commun Med 2008 Mar;31(1):4-10. [doi: 10.1080/17453050802040494] [Medline: 18642176]
42. Cobus L. Using blogs and wikis in a graduate public health course. Med Ref Serv Q 2009;28(1):22-32. [doi: 10.1080/02763860802615922] [Medline: 19197741]
43. Gray K, Tobin J. Introducing an online community into a clinical education setting: a pilot study of student and staff engagement and outcomes using blended learning. BMC Med Educ 2010;10:6 [FREE Full text] [doi: 10.1186/1472-6920-10-6] [Medline: 20100354]
44. Desai T, Stankeyeva D, Chapman A, Bailey J. Nephrology fellows show consistent use of, and improved knowledge from, a nephrologist-programmed teaching instrument. J Nephrol 2011;24(3):345-350. [Medline: 20954135]
45. Maggio M, Hariton-Gross K, Gluch J. The use of independent, interactive media for education in dental morphology. J Dent Educ 2012 Nov;76(11):1497-1511 [FREE Full text] [Medline: 23144486]
46. Stebbings S, Bagheri N, Perrie K, Blyth P, McDonald J. Blended learning and curriculum renewal across three medical schools: The rheumatology module at the University of Otago. AJET 2012 Aug 28;28(7):1189. [doi: 10.14742/ajet.795]
47. Naidr J, Adla T, Janda A, Feberová J, Kasal P, Hladíková M. Long-term retention of knowledge after a distance course in medical informatics at Charles University Prague. Teach Learn Med 2004;16(3):255-259. [doi: 10.1207/s15328015tlm1603_6] [Medline: 15388381]
48. 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]
49. Bernhardt J, Hye F, Thallinger S, Bauer P, Ginter G, Smolle J. Simulation of a mycological KOH preparation--e-learning as a practical dermatologic exercise in an undergraduate medical curriculum. J Dtsch Dermatol Ges 2009 Jul;7(7):597-602. [doi: 10.1111/j.1610-0387.2009.07030.x] [Medline: 19298549]
50. Kandasamy T, Fung K. Interactive Internet-based cases for undergraduate otolaryngology education. Otolaryngol Head Neck Surg 2009 Mar;140(3):398-402. [doi: 10.1016/j.otohns.2008.11.033] [Medline: 19248951]
51. Hu A, Wilson T, Ladak H, Haase P, Doyle P, Fung K. Evaluation of a three-dimensional educational computer model of the larynx: voicing a new direction. J Otolaryngol Head Neck Surg 2010 Jun;39(3):315-322. [Medline: 20470679]
52. Tan S, Hu A, Wilson T, Ladak H, Haase P, Fung K. Role of a computer-generated three-dimensional laryngeal model in anatomy teaching for advanced learners. J Laryngol Otol 2012 Apr;126(4):395-401. [doi: 10.1017/S0022215111002830] [Medline: 22075619]
53. Boulos MNK, Hetherington L, Wheeler S. Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Info Libr J 2007 Dec;24(4):233-245 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00733.x] [Medline: 18005298]

54. Andrade A, Cifuentes P, Mintzer M, Roos B, Anam R, Ruiz J. Simulating geriatric home safety assessments in a three-dimensional virtual world. Gerontol Geriatr Educ 2012;33(3):233-252. [doi: 10.1080/02701960.2011.611553] [Medline: 22816973]
55. Fischer MR, Schauer S, Gräsel C, Baehring T, Mandl H, Gärtner R, et al. [CASUS model trial. A computer-assisted author system for problem-oriented learning in medicine]. Z Arztl Fortbild (Jena) 1996 Aug;90(5):385-389. [Medline: 9157728]
56. Abendroth M, Harendza S, Riemer M. Clinical decision making: a pilot e-learning study. Clin Teach 2013 Feb;10(1):51-55. [doi: 10.1111/j.1743-498X.2012.00629.x] [Medline: 23294745]
57. Maier EM, Hege I, Muntau AC, Huber J, Fischer MR. What are effects of a spaced activation of virtual patients in a pediatric course? BMC Med Educ 2013;13:45 [FREE Full text] [doi: 10.1186/1472-6920-13-45] [Medline: 23537162]
58. Kolb S, Wengenroth L, Hege I, Praml G, Nowak D, Cantineau J, NetWoRM Group. Case based e-learning in occupational medicine--a European approach. J Occup Environ Med 2009 Jun;51(6):647-653. [doi: 10.1097/JOM.0b013e3181a90af6] [Medline: 19430311]
59. Chen LS, Cheng YM, Weng SF, Chen YG, Lin CH. Applications of a time sequence mechanism in the simulation cases of a web-based medical problem-based learning system. Educ Technol Soc 2009;12(1):149.
60. Cheng YM, Chen LS, Huang HC, Weng SF, Chen YG, Lin CH. Building a general purpose pedagogical agent in a Web-based multimedia clinical simulation system for medical education. IEEE Trans Learning Technol 2009 Jul;2(3):216-225. [doi: 10.1109/TLT.2009.18]
61. Cheng Y, Kuo S, Lou S, Shih R. The effect of applying online PBL case system to multiple disciplines of medical education. Turk Online J Educ Technol-TOJET 2012;11(4):294.
62. Horstmann M, Horstmann C, Renninger M. Case creation and e-learning in a Web-based virtual department of urology using the INMEDEA simulator. NuMonthly 2011 Nov 18;4(1):356-360. [doi: 10.5812/kowsar.22517006.1503]
63. Funke K, Bonrath E, Mardin W, Becker J, Haier J, Senninger N, et al. Blended learning in surgery using the Inmedea Simulator. Langenbecks Arch Surg 2013 Feb;398(2):335-340. [doi: 10.1007/s00423-012-0987-8] [Medline: 22878596]
64. Wünschel M, Leichtle U, Wülker N, Kluba T. Using a web-based orthopaedic clinic in the curricular teaching of a German university hospital: analysis of learning effect, student usage and reception. Int J Med Inform 2010 Oct;79(10):716-721. [doi: 10.1016/j.ijmedinf.2010.07.007] [Medline: 20732829]
65. Botezatu M, Hult HA, Tessma M, Fors U. Virtual patient simulation: knowledge gain or knowledge loss? Med Teach 2010;32(7):562-568. [doi: 10.3109/01421590903514630] [Medline: 20653378]
66. Gunning W, Fors UG. Virtual patients for assessment of medical student ability to integrate clinical and laboratory data to develop differential diagnoses: comparison of results of exams with/without time constraints. Med Teach 2012;34(4):e222-e228. [doi: 10.3109/0142159X.2012.642830] [Medline: 22455713]
67. Schwarz D, Štourač P, Komenda M, Harazim H, Kosinová M, Gregor J, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res 2013;15(7):e135 [FREE Full text] [doi: 10.2196/jmir.2590] [Medline: 23835586]
68. Komenda M, Schwarz D, Feberová J, Stípek S, Mihál V, Dusek L. Medical faculties educational network: Multidimensional quality assessment. Comput Methods Programs Biomed 2012;108(3):909. [doi: 10.1016/j.cmpb.2012.05.002]
69. Buttussi F, Pellis T, Cabas VA, Pausler D, Carchietti E, Chittaro L. Evaluation of a 3D serious game for advanced life support retraining. Int J Med Inform 2013 Sep;82(9):798-809. [doi: 10.1016/j.ijmedinf.2013.05.007] [Medline: 23763908]
70. Nilsson M, Bolinder G, Held C, Johansson B, Fors U, Ostergren J. Evaluation of a web-based ECG-interpretation programme for undergraduate medical students. BMC Med Educ 2008;8:25 [FREE Full text] [doi: 10.1186/1472-6920-8-25] [Medline: 18430256]
71. Talanow R. Radiology Teacher: a free, Internet-based radiology teaching file server. J Am Coll Radiol 2009 Dec;6(12):871-875. [doi: 10.1016/j.jacr.2009.08.001] [Medline: 19945043]
72. Schmidt C, Reinehr M, Leucht O, Behrendt N, Geiler S, Britsch S. MyMiCROscope: intelligent virtual microscopy in a blended learning model at Ulm University. Ann Anat 2011 Oct 20;193(5):395-402. [doi: 10.1016/j.aanat.2011.04.009] [Medline: 21641191]
73. Levinson A, Weaver B, Garside S, McGinn H, Norman G. Virtual reality and brain anatomy: a randomised trial of e-learning instructional designs. Med Educ 2007 May;41(5):495-501. [doi: 10.1111/j.1365-2929.2006.02694.x] [Medline: 17470079]
74. Sward K, Richardson S, Kendrick J, Maloney C. Use of a Web-based game to teach pediatric content to medical students. Ambul Pediatr 2008;8(6):354-359. [doi: 10.1016/j.ambp.2008.07.007] [Medline: 19084784]
75. Aparicio F, De Buenaga M, Rubio M, Hernando A. An intelligent information access system assisting a case based learning methodology evaluated in higher education with medical students. Computers & Education 2012 May;58(4):1282-1295. [doi: 10.1016/j.compedu.2011.12.021]
76. Woelber J, Hilbert T, Ratka-Krüger P. Can easy-to-use software deliver effective e-learning in dental education? A randomised controlled study. Eur J Dent Educ 2012 Aug;16(3):187-192. [doi: 10.1111/j.1600-0579.2012.00741.x] [Medline: 22783845]
77. Brunetaud J, Leroy N, Pelayo S, Wascat C, Renard J, Prin L. Comparative evaluation of two applications for delivering a multimedia medical course in the French-speaking Virtual Medical University (UMVF). Int J Med Inf 2005;74(2):212. [doi: 10.1016/j.ijmedinf.2004.06.006]

78. Shaw T, Long A, Chopra S, Kerfoot B. Impact on clinical behavior of face-to-face continuing medical education blended with online spaced education: a randomized controlled trial. J Contin Educ Health Prof 2011;31(2):103-108. [doi: 10.1002/chp.20113] [Medline: 21671276]
79. Ridgway P, Sheikh A, Sweeney K, Evoy D, McDermott E, Felle P, et al. Surgical e-learning: validation of multimedia web-based lectures. Med Educ 2007 Feb;41(2):168-172. [doi: 10.1111/j.1365-2929.2006.02669.x] [Medline: 17269950]
80. Steedman M, Abouammoh M, Sharma S. Multimedia learning tools for teaching undergraduate ophthalmology: results of a randomized clinical study. Can J Ophthalmol 2012 Feb;47(1):66-71. [doi: 10.1016/j.jcjo.2011.12.006] [Medline: 22333855]
81. Holzinger A, Kickmeier-Rust M, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education 2009 Feb;52(2):292-301. [doi: 10.1016/j.compedu.2008.08.008]
82. Perkins G, Kimani P, Bullock I, Clutton-Brock T, Davies R, Gale M, Electronic Advanced Life Support Collaborators. Improving the efficiency of advanced life support training: a randomized, controlled trial. Ann Intern Med 2012 Jul 3;157(1):19-28. [doi: 10.7326/0003-4819-157-1-201207030-00005] [Medline: 22751757]
83. Khot Z, Quinlan K, Norman G, Wainman B. The relative effectiveness of computer-based and traditional resources for education in anatomy. Anat Sci Educ 2013;6(4):211-215. [doi: 10.1002/ase.1355] [Medline: 23509000]
84. Aper L, Reniers J, Koole S, Valcke M, Derese A. Impact of three alternative consultation training formats on self-efficacy and consultation skills of medical students. Med Teach 2012;34(7):e500-e507. [doi: 10.3109/0142159X.2012.668627] [Medline: 22746968]
85. Peacock O, Watts E, Foreman D, Lund J, Tierney G. Evaluation of teaching methods for students on hernias: an observational study. ANZ J Surg 2013 Jan;83(1-2):11-14. [doi: 10.1111/j.1445-2197.2012.06280.x] [Medline: 23013104]
86. Stadie A, Degenhardt I, Conesa G, Reisch R, Kockro R, Fischer G. Comparing expert and novice spatial representation on the basis of VR simulation, MRI images, and physical objects. J CyberTherapy Rehabil 2010;3(3):299.
87. Morgulis Y, Kumar R, Lindeman R, Velan G. Impact on learning of an e-learning module on leukaemia: a randomised controlled trial. BMC Med Educ 2012;12:36 [FREE Full text] [doi: 10.1186/1472-6920-12-36] [Medline: 22640463]
88. Kononowicz A, Krawczyk P, Cebula G, Dembkowska M, Drab E, Frączek B, et al. Effects of introducing a voluntary virtual patient module to a basic life support with an automated external defibrillator course: a randomised trial. BMC Med Educ 2012;12:41 [FREE Full text] [doi: 10.1186/1472-6920-12-41] [Medline: 22709278]
89. Boye S, Moen T, Vik T. An e-learning course in medical immunology: does it improve learning outcome? Med Teach 2012;34(9):e649-e653. [doi: 10.3109/0142159X.2012.675456] [Medline: 22497322]
90. Critchley L, Kumta S, Ware J, Wong J. Web-based formative assessment case studies: role in a final year medicine two-week anaesthesia course. Anaesth Intensive Care 2009 Jul;37(4):637-645. [Medline: 19681426]
91. Kraemer D, Reimer S, Hörnlein A, Betz C, Puppe F, Kneitz C. Evaluation of a novel case-based training program (d3web.Train) in hematology. Ann Hematol 2005 Nov;84(12):823-829. [doi: 10.1007/s00277-005-1062-0] [Medline: 15944852]
92. Vollmar H, Mayer H, Ostermann T, Butzlaff M, Sandars J, Wilm S, et al. Knowledge transfer for the management of dementia: a cluster randomised trial of blended learning in general practice. Implement Sci 2010;5:1 [FREE Full text] [doi: 10.1186/1748-5908-5-1] [Medline: 20047652]
93. Smego Jr RA, Herning T, Davis L, Hossain W, bin Mohammed Al-Khusaiby S. A personal computer-based undergraduate medical school curriculum using SOLE. Teach Learn Med 2009;21(1):38-44. [doi: 10.1080/10401330802574009] [Medline: 19130385]
94. Nieder G, Borges N. An eight-year study of online lecture use in a medical gross anatomy and embryology course. Anat Sci Educ 2012;5(6):311-320. [doi: 10.1002/ase.1289] [Medline: 22674561]
95. Varghese J, Faith M, Jacob M. Impact of e-resources on learning in biochemistry: first-year medical students' perceptions. BMC Med Educ 2012;12:21 [FREE Full text] [doi: 10.1186/1472-6920-12-21] [Medline: 22510159]
96. Bacro TR, Gebregziabher M, Fitzharris T. Evaluation of a lecture recording system in a medical curriculum. Anat Sci Educ 2010;3(6):300-308. [doi: 10.1002/ase.183] [Medline: 20954266]
97. Saxena V, Natarajan P, O'Sullivan PS, Jain S. Effect of the use of instructional anatomy videos on student performance. Anat Sci Educ 2008;1(4):159-165. [doi: 10.1002/ase.38] [Medline: 19177403]
98. Silva CS, Souza MB, Silva Filho RS, de Medeiros LM, Criado PR. E-learning program for medical students in dermatology. Clinics (Sao Paulo) 2011;66(4):619-622 [FREE Full text] [Medline: 21655756]
99. Kim K, Han J, Park I, Kee C. Evaluating effectiveness and usability of an e-learning portal: the e-MedEdu site. Med Teach 2010;32(8):702-703. [Medline: 20695029]
100. Woods C, Kemper K. Curriculum resource use and relationships with educational outcomes in an online curriculum. Acad Med 2009 Sep;84(9):1250-1258. [doi: 10.1097/ACM.0b013e3181b188ae] [Medline: 19707065]
101. Kerfoot B. Learning benefits of on-line spaced education persist for 2 years. J Urol 2009 Jun;181(6):2671-2673. [doi: 10.1016/j.juro.2009.02.024] [Medline: 19375095]
102. Ruiz JG, Smith M, Rodriguez O, van Zuilen MH, Mintzer MJ. An interactive e-learning tutorial for medical students on how to conduct the performance-oriented mobility assessment. Gerontol Geriatr Educ 2007;28(1):51-60. [doi: 10.1300/J021v28n01_04] [Medline: 18032182]

103. Yang RL, Hashimoto DA, Predina JD, Bowens NM, Sonnenberg EM, Cleveland EC, et al. The virtual-patient pilot: testing a new tool for undergraduate surgical education and assessment. J Surg Educ 2013;70(3):394-401. [doi: 10.1016/j.jsurg.2012.12.001] [Medline: 23618451]
104. Smolle J, Prause G, Smolle-Jüttner FM. Emergency treatment of chest trauma--an e-learning simulation model for undergraduate medical students. Eur J Cardiothorac Surg 2007 Oct;32(4):644-647. [doi: 10.1016/j.ejcts.2007.06.042] [Medline: 17702590]
105. Groenwold RH, Knol M. Learning styles and preferences for live and distance education: an example of a specialisation course in epidemiology. BMC Med Educ 2013;13:93 [FREE Full text] [doi: 10.1186/1472-6920-13-93] [Medline: 23819522]
106. Chumley-Jones HS, Dobbie A, Alford CL. Web-based learning: sound educational method or hype? A review of the evaluation literature. Acad Med 2002 Oct;77(10 Suppl):S86-S93. [Medline: 12377715]
107. Greenhalgh T. Computer assisted learning in undergraduate medical education. BMJ 2001 Jan 6;322(7277):40-44 [FREE Full text] [Medline: 11141156]
108. Ward JP, Gordon J, Field MJ, Lehmann HP. Communication and information technology in medical education. Lancet 2001 Mar 10;357(9258):792-796. [doi: 10.1016/S0140-6736(00)04173-8] [Medline: 11253986]
109. Muller K. Statistical power analysis for the behavioral sciences. Technometrics 1989;31(4):499-500.
110. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]
111. Chaffin A, Maddux CD. Internet teaching methods for use in baccalaureate nursing education. Comput Inform Nurs 2004;22(3):132-42; quiz 143. [Medline: 15520582]
112. Lau F, Bates J. A review of e-learning practices for undergraduate medical education. J Med Syst 2004 Feb;28(1):71-87. [Medline: 15171070]
113. Sandars J. Twelve tips for using blogs and wikis in medical education. Med Teach 2006 Dec;28(8):680-682. [doi: 10.1080/01421590601106353] [Medline: 17594577]
114. Rasmussen A, Lewis M, White J. The application of wiki technology in medical education. Med Teach 2013;35(2):109-114. [doi: 10.3109/0142159X.2012.733838] [Medline: 23102056]
115. Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, et al. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. Med Teach 2009 Apr;31(4):282-298. [doi: 10.1080/01421590902889897] [Medline: 19404891]
116. Willis RE, Van Sickle KR. Current status of simulation-based training in graduate medical education. Surg Clin North Am 2015 Aug;95(4):767-779. [doi: 10.1016/j.suc.2015.04.009] [Medline: 26210969]
117. Fonseca A, Evans L, Gusberg R. Open surgical simulation in residency training: a review of its status and a case for its incorporation. J Surg Educ 2013;70(1):129-137. [doi: 10.1016/j.jsurg.2012.08.007] [Medline: 23337682]
118. Walsh C, Sherlock M, Ling S, Carnahan H. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2012(6):CD008237. [doi: 10.1002/14651858.CD008237.pub2] [Medline: 22696375]
119. Ma IW, Brindle M, Ronksley P, Lorenzetti D, Sauve R, Ghali W. Use of simulation-based education to improve outcomes of central venous catheterization: a systematic review and meta-analysis. Acad Med 2011 Sep;86(9):1137-1147. [doi: 10.1097/ACM.0b013e318226a204] [Medline: 21785310]
120. Kennedy CC, Cannon EK, Warner DO, Cook DA. Advanced airway management simulation training in medical education: a systematic review and meta-analysis. Crit Care Med 2014 Jan;42(1):169-178. [doi: 10.1097/CCM.0b013e31829a721f] [Medline: 24220691]
121. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. J Med Internet Res 2014;16(12):e251 [FREE Full text] [doi: 10.2196/jmir.3240] [Medline: 25498126]
122. McGee J, Begg M. What medical educators need to know about “Web 2.0”. Med Teach 2008;30(2):164-169. [doi: 10.1080/01421590701881673] [Medline: 18464141]
123. Hollinderbäumer A, Hartz T, Uckert F. Education 2.0: how has social media and Web 2.0 been integrated into medical education? A systematical literature review. GMS Z Med Ausbild 2013;30(1):Doc14 [FREE Full text] [doi: 10.3205/zma000857] [Medline: 23467509]
124. Blaum W, Jarczweski A, Balzer F, Stötzner P, Ahlers O. Towards Web 3.0: taxonomies and ontologies for medical education: a systematic review. GMS Z Med Ausbild 2013;30(1):Doc13 [FREE Full text] [doi: 10.3205/zma000856] [Medline: 23467484]
125. Taveira-Gomes T. A Systematic Review on Computer Based Learning Interventions in Medical Education - What Are We Looking For? An Interactive Visualization. San Francisco, CA: GitHub, Inc; 2016 Jul 16. URL: https://tiagotaveiragomes.github.io/medicaleducationcbl/ [accessed 2016-07-16] [WebCite Cache ID 6j2ysdnFm]
126. Ruiz JG, Mintzer MJ, Issenberg SB. Learning objects in medical education. Med Teach 2006 Nov;28(7):599-605. [doi: 10.1080/01421590601039893] [Medline: 17594550]
127. Perkins G, Fullerton J, Davis-Gomez N, Davies R, 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]

128. Hadley J, Kulier R, Zamora J, Coppus SF, Weinbrenner S, Meyerrose B, et al. Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training. J R Soc Med 2010 Jul;103(7):288-294 [FREE Full text] [doi: 10.1258/jrsm.2010.100036] [Medline: 20522698]
129. Bridge P, Jackson M, Robinson L. The effectiveness of streaming video on medical student learning: a case study. Med Educ Online 2009;14:11 [FREE Full text] [doi: 10.3885/meo.2009.Res00311] [Medline: 20165525]
130. Taveira-Gomes T, Saffarzadeh A, Severo M, Guimarães MJ, Ferreira M. A novel collaborative e-learning platform for medical students - ALERT STUDENT. BMC Med Educ 2014;14:143 [FREE Full text] [doi: 10.1186/1472-6920-14-143] [Medline: 25017028]
131. Taveira-Gomes T, Prado-Costa R, Severo M, Ferreira M. Characterization of medical students recall of factual knowledge using learning objects and repeated testing in a novel e-learning system. BMC Med Educ 2015;15:4 [FREE Full text] [doi: 10.1186/s12909-014-0275-0] [Medline: 25616353]
132. Kimmerle J, Bientzle M, Cress U. Personal experiences and emotionality in health-related knowledge exchange in Internet forums: a randomized controlled field experiment comparing responses to facts vs personal experiences. J Med Internet Res 2014 Dec;16(12):e277 [FREE Full text] [doi: 10.2196/jmir.3766] [Medline: 25486677]
133. Bientzle M, Griewatz J, Kimmerle J, Küppers J, Cress U, Lammerding-Koeppel M. Impact of scientific versus emotional wording of patient questions on doctor-patient communication in an internet forum: a randomized controlled experiment with medical students. J Med Internet Res 2015;17(11):e268 [FREE Full text] [doi: 10.2196/jmir.4597] [Medline: 26607233]
134. Bientzle M, Cress U, Kimmerle J. How students deal with inconsistencies in health knowledge. Med Educ 2013 Jul;47(7):683-690. [doi: 10.1111/medu.12198] [Medline: 23746157]
135. Rezende-Filho FM, da Fonseca LJ, Nunes-Souza V, da Guedes G, Rabelo LA. A student-centered approach for developing active learning: the construction of physical models as a teaching tool in medical physiology. BMC Med Educ 2014;14:189 [FREE Full text] [doi: 10.1186/1472-6920-14-189] [Medline: 25223392]
136. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. Acad Med 2013 Jun;88(6):893-901. [doi: 10.1097/ACM.0b013e31828ffc23] [Medline: 23619071]
137. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Med Educ Online 2010;15 [FREE Full text] [doi: 10.3402/meo.v15i0.5324] [Medline: 20859533]
138. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]
139. Archambault PM, van de Belt TH, Grajales FJ, Faber MJ, Kuziemsky CE, Gagnon S, et al. Wikis and collaborative writing applications in health care: a scoping review. J Med Internet Res 2013 Oct;15(10):e210 [FREE Full text] [doi: 10.2196/jmir.2787] [Medline: 24103318]
140. Brulet A, Llorca G, Letrilliart L. Medical wikis dedicated to clinical practice: a systematic review. J Med Internet Res 2015;17(2):e48 [FREE Full text] [Medline: 25700482]

Abbreviations

BIC: Bayesian information criterion
CBL: computer-based learning
DOI: digital object identifier
ICC: intraclass correlation
MERSQI: Medical Education Research Study Quality Index
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 19.12.15; peer-reviewed by J Kimmerle, D Cook, M Broom; comments to author 03.04.16; revised version received 01.06.16; accepted 21.06.16; published 01.08.16.

Please cite as:
Taveira-Gomes T, Ferreira P, Taveira-Gomes I, Severo M, Ferreira MA
What Are We Looking for in Computer-Based Learning Interventions in Medical Education? A Systematic Review
J Med Internet Res 2016;18(8):e204
URL: http://www.jmir.org/2016/8/e204/
doi:10.2196/jmir.5461
PMID:27480053

Original Paper

Web-Based Virtual Microscopy of Digitized Blood Slides for Malaria Diagnosis: An Effective Tool for Skills Assessment in Different Countries and Environments

Laura Ahmed[1], PhD; Leonard H Seal[1], PhD; Carol Ainley[1], MSc; Barbara De la Salle[2], MSc; Michelle Brereton[3], MSc; Keith Hyde[1,2,3], PhD; John Burthem[4,5], MB ChB; William Samuel Gilmore[1], PhD

[1]Manchester Metropolitan University, School of Healthcare Science, Faculty of Science and Engineering, Manchester, United Kingdom

[2]UK NEQAS Haematology, Watford, United Kingdom

[3]Directorate of Laboratory Medicine, Manchester Royal Infirmary, Manchester, United Kingdom

[4]Department of Haematology, Manchester Royal Infirmary, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom

[5]Institute of Cancer Sciences, Faculty of Medical and Human Sciences, Manchester University, Manchester, United Kingdom

Corresponding Author:
Laura Ahmed, PhD
Manchester Metropolitan University
School of Healthcare Science
Faculty of Science and Engineering
John Dalton Building
Chester Street
Manchester, M1 5GD
United Kingdom
Phone: 44 1612471485
Fax: 44 1612476831
Email: l.ahmed@mmu.ac.uk

Abstract

Background: Morphological examination of blood films remains the reference standard for malaria diagnosis. Supporting the skills required to make an accurate morphological diagnosis is therefore essential. However, providing support across different countries and environments is a substantial challenge.

Objective: This paper reports a scheme supplying digital slides of malaria-infected blood within an Internet-based virtual microscope environment to users with different access to training and computing facilities. The feasibility of the approach was established, allowing users to test, record, and compare their own performance with that of other users.

Methods: From Giemsa stained thick and thin blood films, 56 large high-resolution digital slides were prepared, using high-quality image capture and 63x oil-immersion objective lens. The individual images were combined using the photomerge function of Adobe Photoshop and then adjusted to ensure resolution and reproduction of essential diagnostic features. Web delivery employed the Digital Slidebox platform allowing digital microscope viewing facilities and image annotation with data gathering from participants.

Results: Engagement was high with images viewed by 38 participants in five countries in a range of environments and a mean completion rate of 42/56 cases. The rate of parasite detection was 78% and accuracy of species identification was 53%, which was comparable with results of similar studies using glass slides. Data collection allowed users to compare performance with other users over time or for each individual case.

Conclusions: Overall, these results demonstrate that users worldwide can effectively engage with the system in a range of environments, with the potential to enhance personal performance through education, external quality assessment, and personal professional development, especially in regions where educational resources are difficult to access.

(J Med Internet Res 2016;18(8):e213) doi:10.2196/jmir.6027

KEYWORDS

Malaria; Virtual microscopy; External quality assessment; Internet

Introduction

The ability to make a successful morphological diagnosis of malaria remains an important and cost-effective health intervention worldwide [1]. However, malaria detection and recognition on blood films require individual skills, knowledge, training, and experience. The effects of misdiagnosis can be very serious [2]. Supporting the quality of diagnosis in different countries and settings can be challenging, particularly for those individuals working in isolation [3,4]. Direct practical instruction and education at the microscope are central to training [5], and glass-slide specimens of malaria-infected blood distributed by reference laboratories provide excellent education in diagnosis [6]. External quality assessment (EQA) has been shown to improve accuracy [7,8]; however, it has also been difficult to implement since the collection and provision of clinical materials are labor intensive, and slide numbers may be limited [9-11].

With facilities for Internet access improving globally, it is now possible to deliver digital slides of blood films to a wide range of geographical locations. Web-based virtual microscope systems [12,13] greatly support this process, and the skillsets have been shown to mirror those used in glass slide diagnosis [14]. The virtual microscope has been used in university education [15-18], EQA schemes [19], and for continuing professional development [20,21].

This study aims to demonstrate the feasibility and effectiveness of the virtual microscope to deliver digital slides of malaria-infected blood with high-quality resolution and to assess how the technique can be used to support skills in malaria diagnosis in developing and developed nations.

Methods

Participants for this study were recruited with help of the World Health Organization (WHO) and Liverpool School of Tropical Medicine. We used digital slides prepared from 56 cases that had previously been fully validated and distributed as glass slides through the UK National External Quality Assessment Scheme (UK NEQAS) parasitology scheme, confirmed by molecular techniques and consensus opinion of more than 400 laboratories. The cases were selected to represent different slide preparation techniques, stain quality, malaria species, and parasite density. Images were captured using a Zeiss Axio Imager M1 microscope with HRc camera and 63x Plan Apo Chromat 1.4 Oil immersion lens. At least 40 adjacent fields were acquired to create a single large stitched-image. Subsequent image processing used Zeiss software (Axiovision 4.7), then Photoshop CS3 to ensure sharpness and color balance. Diagnostic features were clearly resolved (contrast mask and detail enhancement using Digital Outback Photo add-in), and image size for upload was around 200 megabytes. Images were uploaded to the viewing software (Digital SlideBox, Leica Biosystems) and assessed to ensure the inclusion and accurate rendition of features required for diagnosis. Registered participants were from five countries (ie, Hong Kong, India, Kenya, Lebanon, and Nigeria), and some received financial support for Internet access. In all cases, the participants were asked to suggest their preferred diagnosis (multiple-choice options). Responses were analyzed using Microsoft Excel and GraphPad Prism software (v6.04).

Results

The virtual microscope environment allowed low-power image scanning, navigation, and high-power view assessment (see Figure 1).

The images showed that the initial high-quality image capture clearly resolved those features required for malaria diagnosis, species assignment, or detection of artifact (see Figures 2 and 3). Images were then reassessed following assembly into a large single slide, and finally as a screen capture of a compressed Web-displayed image. Comparison between initial image and Web-browser images revealed an expected subjective loss of definition, but diagnostic detail was retained (see Figure 4).

The digital slides were then viewed by 38 participants from a range of countries: Hong Kong (n=1), India (n=1), Kenya (n=7), Lebanon (n=6), and Nigeria (n=23). The training and viewing environments differed between groups, and facilities used in different countries were not identical. Although images were viewed predominantly at hospital sites, a significant number used Internet cafés (with financial support). Aspects of the viewing environment were uncontrolled (ie, screen resolution and Internet speed). Although positive experience was reported, the impact of these aspects on results was unclear. Those using Internet cafés had lower rates for sensitivity and specificity for their diagnosis. The mean completion rate was 75% (42/56 cases). Each individual case was completed by a mean of 29 participants (range 22-38). The outcome of slide analysis (indicated by the submitted diagnosis) was compared with parasite density and type. Results indicate that successful recognition and species identification, in keeping with known results of glass slide analysis, were closely linked with the number of malarial parasites present on the slides (see Figure 5). Thick and thin films were assessed with equivalent success, and films containing no parasites were effectively identified. Similarly, in keeping with recognized findings, the recognition that a parasite was *Plasmodium falciparum* or non- *P. falciparum* species was good. However, the precise species-identification for non- *P. falciparum* species was less accurate (see Figure 6).

Finally, results were assessed to see whether they could be used to allow participants to assess their individual performance, comparing their results with other participants with similar training or experience (see Figure 7). The performance data across a consecutive case series were also plotted and used to make a cumulative assessment that allowed individuals to compare their performance trend with other users (see Figure 8).

Figure 1. Blood film of parasites shown within Web viewer (Left panel: low power view of whole digital film, showing information buttons and orientation image at top, and navigation and magnification tools at bottom; Right panel: parasite image shown as high power magnification occupying full screen height for the monitor used).

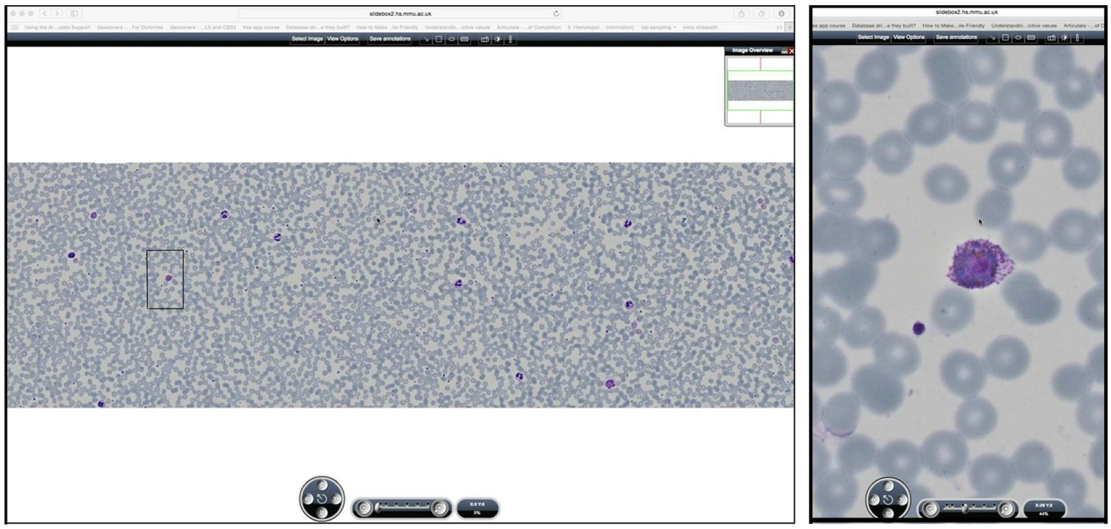


Figure 2. Resolution of features required for parasite detection and species recognition (left panel: digital image of *Plasmodium falciparum* early trophozoites, demonstrating the presence of fine ring together with accole forms with Maurer's dots and clefts; Right panel: *Plasmodium ovale* parasites with coarse ring forms, James' (Schűffner's dots), and cytoplasmic fimbriation of erythrocytes).

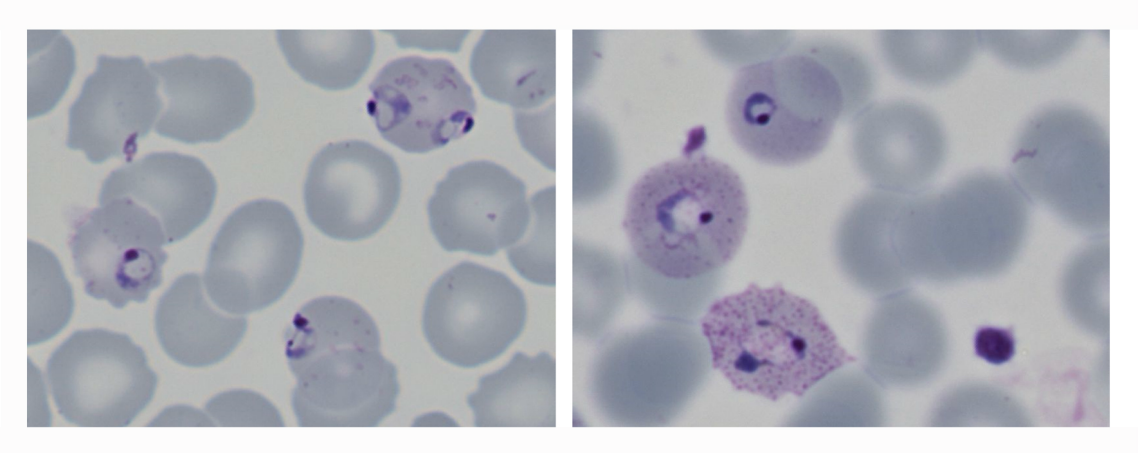


Figure 3. Demonstration of artifact on films presented within Web viewer (Left panel: stain debris visible on a film containing a gametocyte of *Plasmodium malariae*; Right panel: malaria pigment spilling from within a distorted/disrupted parasite [thick film]).

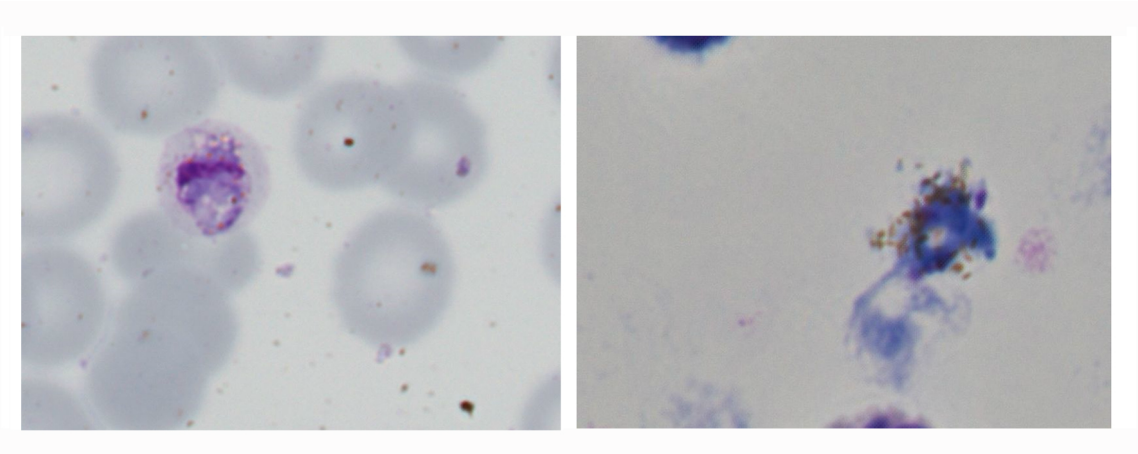


Figure 4. Demonstration of quality considerations applicable to the large digital image: images shown at the time of capture, then subsequently within Web viewer (Left panel: original image quality - resolving parasite, erythrocyte membrane changes, Schűffner's dots, and malarial pigment; Right panel: screenshot of the same cell following compression and Web delivery showing a subjective difference in reproduction, but retention of all significant diagnostic elements).

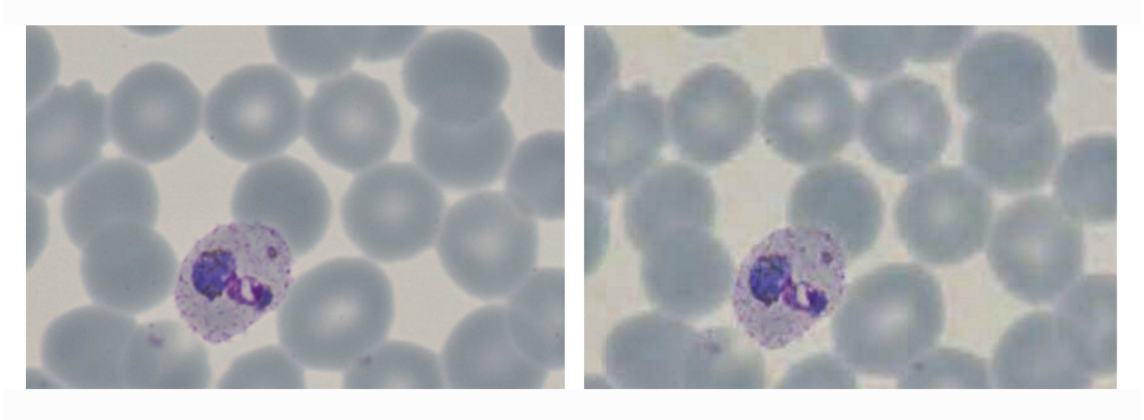


Figure 5. Accuracy of detection and species identification according to number of parasites present on digital blood film. Data groups representing different densities of red cell infection (expressed as parasites per high power field: very low <0.1, low=0.1-0.3, medium=0.3-2, high>2), black=*P. falciparum* cases, gray=non- *P. falciparum*). N= 56 cases, each answered by 22-38 (mean 29) participants. Left panel: parasite detection (mean ±SEM), showing a positive relationship between parasite number and parasite detection (sensitivity=78%) (linear regression analysis of data, r=.48, *P*=.002 for parasite detection); Right panel: species identification (specificity=53%) (r=.54, *P*=.0004).

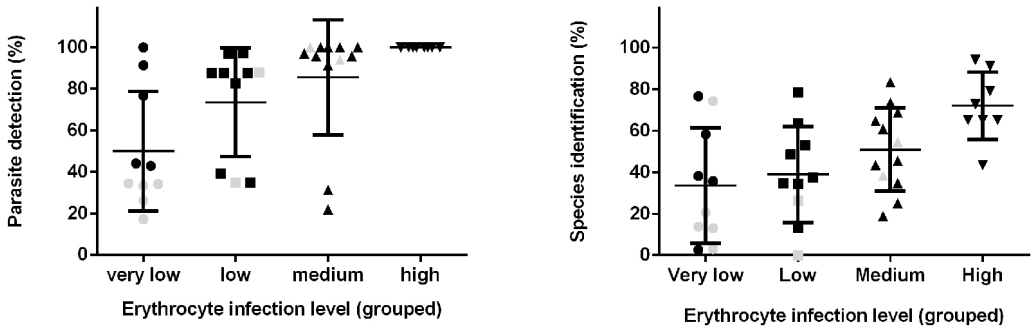


Figure 6. Features determining accuracy of parasite identification by participants using images presented within Web viewer. Left panel: comparison of accuracy of assessment according to type of film or the presence of parasites. Bars represent outcome of detection on thick films or films where no parasites were present compared with the overall responses (bars represent mean ±SEM); n=56 films answered by 22-38 (mean 29) participants. No statistically significant differences were demonstrated. Right panel: comparison of overall parasite detection and detection of non-*P. falciparum* malaria species. Bars represent the mean (±SEM) for each group (*P. falciparum*=31, non-*P. falciparum*=9) analyzed by 38 participants.

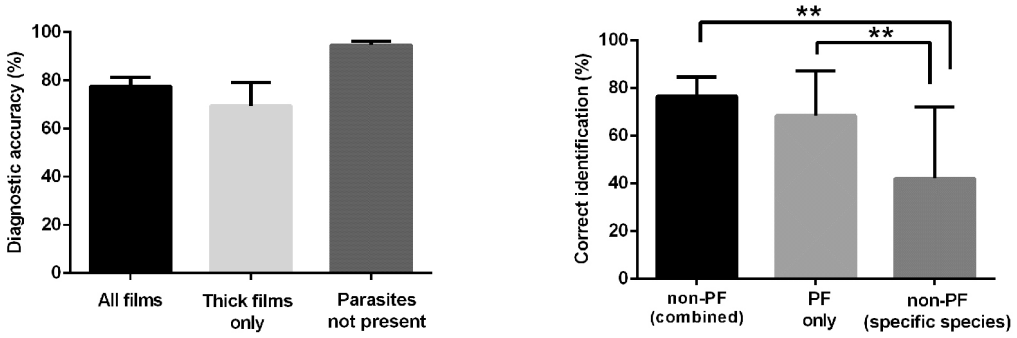


Figure 7. Influence of diagnostic experience performance in parasite identification using consecutive assessments delivered using the image browser. 35 candidates grouped according to length of experience reporting malaria (A: <1yr; B: 1-4yrs; C: 3-9yrs; D: >9yrs). Each bar represents the mean performance for each candidate. The horizontal line represents the performance of UK NEQAS participants using the same case sets on glass slide (solid line=mean, broken line=95% CI, n=435 (mean)). Gray bars represent participants trained at diploma level; all others shown were trained at degree level. Left panel: comparison of individual participants' detection of malaria parasites; Right panel: comparison of individual participants' identification of malaria parasites.

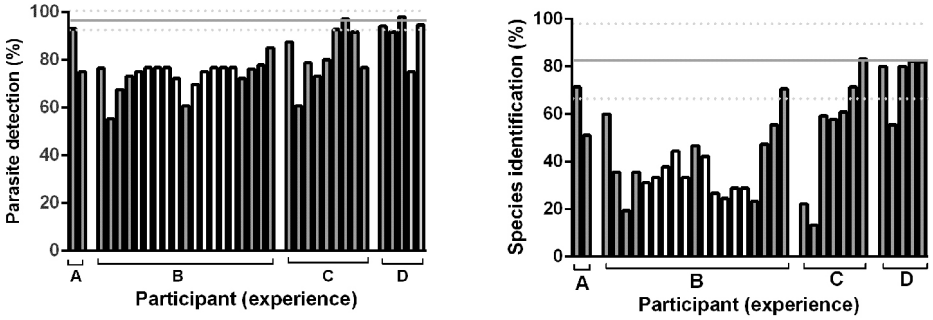
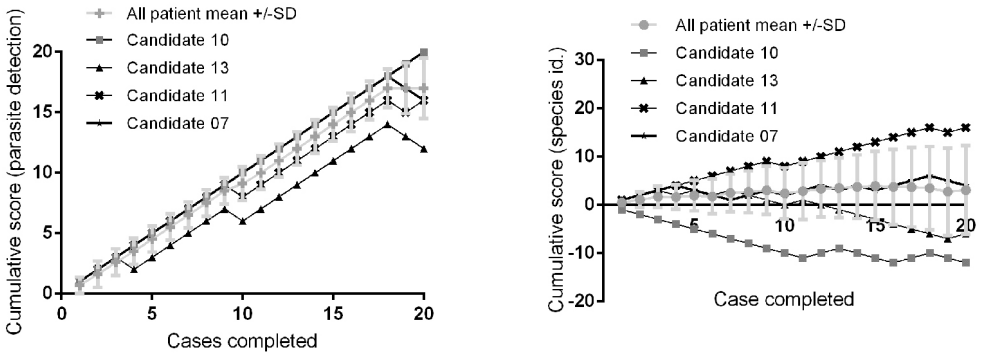


Figure 8. Individual responses compared with overall performance of a cohort over a case series. Left panel (parasite detection): mean cumulative scores ± SD for parasite detection and parasite identification shown for a cohort of 19 participants completing 20 identical cases (correct answer awarded +1 and incorrect is awarded -1). Right panel (parasite identification): representative traces of 4 individual candidates demonstrate tracking of individual performance against the overall mean.



Discussion

Principal Findings

Results show that high-quality digital images of malaria-infected erythrocytes prepared as digital microscopy images effectively resolve those morphological features required for diagnosis of malaria and for species recognition. When converted to a Web-compatible format that supports appropriate download speeds, image quality is retained, and cases can be delivered, viewed, and answered by participants in a range of environments in different countries. Overall, the sensitivity for parasite detection and the specificity for species identification was consistent with reports from others using glass slides for comparable cohorts [4,6,22], suggesting that skills applied with the virtual microscope system reflect those made by conventional microscopy. Similarly, those cases that were poorly recognized using the virtual microscope system generally had low parasite number and/or were non- *falciparum* cases—features associated with less accurate diagnosis using glass slides. Finally, those participants achieving highest diagnostic accuracy in this system showed sensitivity of >90% and specificity of around 80%, comparable to results recorded when the same cases were delivered as glass slides to international laboratories as part of established external quality assessment schemes (96% and 85% respectively), confirming that skilled morphologists could achieve high diagnostic accuracy using the system.

On a practical level, the virtual microscope system received positive reviews from participants. They found the system easy to use and access. In this study, the lessons learned were mainly based around communication with remote participants, where contact was only via email. Increased communication correlated with increased participation.

Conclusion

Web-based delivery allows findings and diagnoses to be rapidly and accurately collected. Analysis of participant responses could be developed to support either individual assessment with longitudinal assessment according to the overall mean, or evaluation of malaria diagnosis skills against a selected peer group. The analysis of performance can be linked to expected performance standards, training level, or personal progression. Linking of this analysis to educational resources presents a real opportunity to support diagnostic skills and to identify sources of error [21], particularly in countries where access to training may be limited. Assessing low parasite density is likely to become more important as many countries move to malaria eradication [23]. It is proposed that the scheme could be adapted and tailored to ensure individual or local relevance (ie, supporting the WHO malaria microscopy quality assurance manual [24]) and can provide opportunities for effective support of malaria diagnosis.

Acknowledgments

We would like to thank:

•participants based in laboratories throughout the world

•UK NEQAS (H) staff for help recruiting participants and providing blood films

•Liverpool School of Tropical Medicine staff for helping recruit participants

•Hospital for Tropical Diseases, London, for supplying blood films

Conflicts of Interest

None declared.

References

1. Endeshaw T, Gebre T, Ngondi J, Graves PM, Shargie EB, Ejigsemahu Y, et al. Evaluation of light microscopy and rapid diagnostic test for the detection of malaria under operational field conditions: a household survey in Ethiopia. Malar J 2008;7:118 [FREE Full text] [doi: 10.1186/1475-2875-7-118] [Medline: 18598344]
2. Amexo M, Tolhurst R, Barnish G, Bates I. Malaria misdiagnosis: effects on the poor and vulnerable. Lancet 2004;364(9448):1896-1898. [doi: 10.1016/S0140-6736(04)17446-1] [Medline: 15555670]
3. Petti CA, Polage CR, Quinn TC, Ronald AR, Sande MA. Laboratory medicine in Africa: a barrier to effective health care. Clin Infect Dis 2006 Feb 1;42(3):377-382 [FREE Full text] [doi: 10.1086/499363] [Medline: 16392084]
4. Kahama-Maro J, D'Acremont V, Mtasiwa D, Genton B, Lengeler C. Low quality of routine microscopy for malaria at different levels of the health system in Dar es Salaam. Malar J 2011;10:332 [FREE Full text] [doi: 10.1186/1475-2875-10-332] [Medline: 22047131]
5. Moura S, Fançony C, Mirante C, Neves M, Bernardino L, Fortes F, et al. Impact of a training course on the quality of malaria diagnosis by microscopy in Angola. Malar J 2014;13:437 [FREE Full text] [doi: 10.1186/1475-2875-13-437] [Medline: 25406586]
6. Ngasala B, Mubi M, Warsame M, Petzold MG, Massele AY, Gustafsson LL, et al. Impact of training in clinical and microscopy diagnosis of childhood malaria on antimalarial drug prescription and health outcome at primary health care level in Tanzania: a randomized controlled trial. Malar J 2008;7:199 [FREE Full text] [doi: 10.1186/1475-2875-7-199] [Medline: 18831737]
7. Kettelhut MM, Chiodini PL, Edwards H, Moody A. External quality assessment schemes raise standards: evidence from the UKNEQAS parasitology subschemes. J Clin Pathol 2003 Dec;56(12):927-932 [FREE Full text] [Medline: 14645352]
8. Thomson S, Lohmann RC, Crawford L, Dubash R, Richardson H. External quality assessment in the examination of blood films for malarial parasites within Ontario, Canada. Arch Pathol Lab Med 2000 Jan;124(1):57-60. [doi: 10.1043/0003-9985(2000)124<0057:EQAITE>2.0.CO;2] [Medline: 10629133]
9. Carter JY, Lema OE, Adhiambo CG, Materu SF. Developing external quality assessment programmes for primary health: care level in resource limited countries. Accred Qual Assur 2014 Mar 18;7(8-9):345-350. [doi: 10.1007/s00769-002-0510-y]
10. Khan MA, Walley JD, Munir MA, Khan MA, Khokar NG, Tahir Z, et al. District level external quality assurance (EQA) of malaria microscopy in Pakistan: pilot implementation and feasibility. Malar J 2011;10:45 [FREE Full text] [doi: 10.1186/1475-2875-10-45] [Medline: 21324210]
11. Mukadi P, Gillet P, Lukuka A, Atua B, Kahodi S, Lokombe J, et al. External quality assessment of malaria microscopy in the Democratic Republic of the Congo. Malar J 2011;10:308 [FREE Full text] [doi: 10.1186/1475-2875-10-308] [Medline: 22008378]
12. Costello SP, Johnston DJ, Dervan PA, O'Shea DG. Development and evaluation of the virtual pathology slide: a new tool in telepathology. J Med Internet Res 2003;5(2):e11 [FREE Full text] [doi: 10.2196/jmir.5.2.e11] [Medline: 12857667]
13. Linder E, Lundin M, Thors C, Lebbad M, Winiecka-Krusnell J, Helin H, et al. Web-based virtual microscopy for parasitology: a novel tool for education and quality assurance. PLoS Negl Trop Dis 2008;2(10):e315 [FREE Full text] [doi: 10.1371/journal.pntd.0000315] [Medline: 18941514]
14. Burthem J, Brereton M, Ardern J, Hickman L, Seal L, Serrant A, et al. The use of digital 'virtual slides' in the quality assessment of haematological morphology: results of a pilot exercise involving UK NEQAS(H) participants. Br J Haematol 2005 Jul;130(2):293-296. [doi: 10.1111/j.1365-2141.2005.05597.x] [Medline: 16029459]
15. Kumar RK, Velan GM, Korell SO, Kandara M, Dee FR, Wakefield D. Virtual microscopy for learning and assessment in pathology. J Pathol 2004 Dec;204(5):613-618. [doi: 10.1002/path.1658] [Medline: 15470688]
16. Dee FR. Virtual microscopy in pathology education. Hum Pathol 2009 Aug;40(8):1112-1121. [doi: 10.1016/j.humpath.2009.04.010] [Medline: 19540551]

17. Koch LH, Lampros JN, Delong LK, Chen SC, Woosley JT, Hood AF. Randomized comparison of virtual microscopy and traditional glass microscopy in diagnostic accuracy among dermatology and pathology residents. Hum Pathol 2009 May;40(5):662-667. [doi: 10.1016/j.humpath.2008.10.009] [Medline: 19144382]
18. Abdul-Ghani R. Towards e-parasitology: making use of virtual microscopy. Trop Med Int Health 2015 Feb;20(2):227-229 [FREE Full text] [doi: 10.1111/tmi.12426] [Medline: 25409835]
19. Intan F, Earl G, Marsden K. Virtual microscopy online pilot survey. RCPA Quality Assurance Programs 2009:1-18.
20. Brereton ML, de la Salle B, Burthem J, Ardern J, Hickman L, Seal L, et al. Review of the UK NEQAS (H) digital morphology pilot scheme for continuing professional development accessed via the internet. Int J Lab Hematol 2008 Oct;30(5):365-371. [doi: 10.1111/j.1751-553X.2008.01086.x] [Medline: 18691344]
21. Brereton M, de la Salle B, Ardern J, Hyde K, Burthem J. Do We Know Why We Make Errors in Morphological Diagnosis? An Analysis of Approach and Decision-Making in Haematological Morphology. EBioMedicine 2015 Sep;2(9):1224-1234 [FREE Full text] [doi: 10.1016/j.ebiom.2015.07.020] [Medline: 26501122]
22. Mbakilwa H, Manga C, Kibona S, Mtei F, Meta J, Shoo A, et al. Quality of malaria microscopy in 12 district hospital laboratories in Tanzania. Pathog Glob Health 2012 Oct;106(6):330-334 [FREE Full text] [doi: 10.1179/2047773212Y.0000000052] [Medline: 23182136]
23. Harris I, Sharrock WW, Bain LM, Gray K, Bobogare A, Boaz L, et al. A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malar J 2010;9:254 [FREE Full text] [doi: 10.1186/1475-2875-9-254] [Medline: 20822506]
24. World Health Organization. Malaria microscopy quality assurance manual, Version 2. 2016. URL: http://apps.who.int/iris/bitstream/10665/204266/1/9789241549394_eng.pdf [accessed 2016-07-22] [WebCite Cache ID 6jfE5E1rX]

Abbreviations

EQA: external quality assessment
UK NEQAS: UK National External Quality Assessment Scheme
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.05.16; peer-reviewed by J Carter, SC Murphy, R Abdul-Ghani; comments to author 22.06.16; revised version received 01.07.16; accepted 04.07.16; published 11.08.16.

Please cite as:
Ahmed L, Seal LH, Ainley C, De la Salle B, Brereton M, Hyde K, Burthem J, Gilmore WS
Web-Based Virtual Microscopy of Digitized Blood Slides for Malaria Diagnosis: An Effective Tool for Skills Assessment in Different Countries and Environments
J Med Internet Res 2016;18(8):e213
URL: http://www.jmir.org/2016/8/e213/
doi:10.2196/jmir.6027
PMID:27515009

Original Paper

Evaluation of a Web-Based E-Learning Platform for Brief Motivational Interviewing by Nurses in Cardiovascular Care: A Pilot Study

Guillaume Fontaine[1,2], RN, MSc; Sylvie Cossette[1,2], RN, PhD; Sonia Heppell[2,3], NP, MSc; Louise Boyer[2,4], RN, PhD; Tanya Mailhot[1,2], RN, BSc; Marie-Josée Simard[3], RN, MSc; Jean-Francois Tanguay[1,3,5], MD, FRCPC, FACC, FAHA, FESC

[1]Montreal Heart Institute Research Center, Montréal, QC, Canada
[2]Faculty of Nursing, Université de Montréal, Montréal, QC, Canada
[3]Montreal Heart Institute, Montréal, QC, Canada
[4]Center for Innovation in Nursing Education, Montréal, QC, Canada
[5]Faculty of Medicine, Université de Montréal, Montréal, QC, Canada

Corresponding Author:
Guillaume Fontaine, RN, MSc
Montreal Heart Institute Research Center
5000 Belanger Street
Room S-2490
Montréal, QC, H1T 1C8
Canada
Phone: 1 514 376 3330 ext 3069
Fax: 1 514 593 7441
Email: guillaume.fontaine@umontreal.ca

Abstract

Background: Brief motivational interviewing (MI) can contribute to reductions in morbidity and mortality related to coronary artery disease, through health behavior change. Brief MI, unlike more intensive interventions, was proposed to meet the needs of clinicians with little spare time. While the provision of face-to-face brief MI training on a large scale is complicated, Web-based e-learning is promising because of the flexibility it offers.

Objective: The primary objective of this pilot study was to examine the feasibility and acceptability of a Web-based e-learning platform for brief MI (MOTIV@CŒUR), which was evaluated by nurses in cardiovascular care. The secondary objective was to assess the preliminary effect of the training on nurses' perceived brief MI skills and self-reported clinical use of brief MI.

Methods: We conducted a single-group, pre-post pilot study involving nurses working in a coronary care unit to evaluate MOTIV@CŒUR, which is a Web-based e-learning platform for brief MI, consisting of two sessions lasting 30 and 20 minutes. MOTIV@CŒUR covers 4 real-life clinical situations through role-modeling videos showing nurse-client interactions. A brief introduction to MI is followed by role playing, during which a nurse practitioner evaluates clients' motivation to change and intervenes according to the principles of brief MI. The clinical situations target smoking, medication adherence, physical activity, and diet. Nurses were asked to complete both Web-based training sessions asynchronously within 20 days, which allowed assessment of the feasibility of the intervention. Data regarding acceptability and preliminary effects (perceived skills in brief MI, and self-reported clinical use of conviction and confidence interventions) were self-assessed through Web-based questionnaires 30 days (±5 days) after the first session.

Results: We enrolled 27 women and 4 men (mean age 37, SD 9 years) in March 2016. Of the 31 participants, 24 (77%, 95% CI 63%–91%) completed both sessions in ≤20 days. At 30 days, 28 of the 31 participants (90%) had completed at least one session. The training was rated as highly acceptable, with the highest scores observed for information quality (mean 6.26, SD 0.60; scale 0–7), perceived ease of use (mean 6.16, SD 0.78; scale 0–7), and system quality (mean 6.15, SD 0.58; scale 0–7). Posttraining scores for self-reported clinical use of confidence interventions were higher than pretraining scores (mean 34.72, SD 6.29 vs mean 31.48, SD 6.75, respectively; P=.03; scale 10–50). Other results were nonsignificant.

Conclusions: Brief MI training using a Web-based e-learning platform including role-modeling videos is both feasible and acceptable according to cardiovascular care nurses. Further research is required to evaluate the e-learning platform in a randomized controlled trial.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 16510888; http://www.isrctn.com/ISRCTN16510888 (Archived by WebCite at http://www.webcitation.org/6jf7dr7bx)

(J Med Internet Res 2016;18(8):e224) doi:10.2196/jmir.6298

KEYWORDS

motivational interviewing; cardiovascular diseases; coronary artery disease; health behavior change; Web-based learning; e-learning; pilot study

Introduction

Background

Coronary artery disease contributes significantly to worldwide morbidity and mortality [1]. According to clinical practice guidelines, the reduction of cardiovascular risk factors through health behavior change plays a critical role in treatment for coronary artery disease [2]. Smoking cessation, medication adherence, physical activity, and diet are often cited as examples of health behaviors that are amenable to change and allow risk factor reduction [3,4]. Health behavior change is determined mainly by *conviction* (ie, knowledge and understanding of the disease, personal meaning, and the relevance of that knowledge) and *confidence* in one's ability to change successfully [5,6]. These determinants of health behavior change can be addressed via client-centered interventions that target individuals' beliefs, values, and motivation [7-10].

Brief motivational interviewing (MI) is a client-centered approach designed to guide individuals through collaborative conversational style and to solidify their motivation and commitment to health behavior change [7,11]. Unlike longer interventions, brief MI was proposed to meet the needs of clinicians who have little time to use the full range of MI techniques in practice [12-14]. The scientific literature generally supports the efficacy of brief MI in various health care settings, including those involving smoking cessation, medication adherence, physical activity, and diet [7,14-17]. While brief MI is promising, health care practitioners often lack time, basic training, or continuous education opportunities to update their knowledge and skills regarding increasing clients' motivation for change [18]. A systematic review [18] evaluated 10 studies involving health care practitioners' use of brief to intensive MI training methods. The duration of the training ranged from 20 minutes to 24 hours, while the format varied from face-to-face sessions to short video modules presented in a classroom setting. Results of the review were generally favorable, suggesting that MI training generates an increase in knowledge, skills, and clinical use. In contrast, very few studies have thus far examined MI training delivered via Web-based e-learning. In fact, of 36 studies included in 3 different systematic reviews concerning MI training, none evaluated Web-based MI training [18-20].

E-learning, defined as instruction delivered on a digital device [21], has been shown to be effective for health care practitioners, with knowledge acquisition and clinical skill development equal or superior to those observed with face-to-face instruction [22-27]. Web-based e-learning can reduce the cost and time involved in providing continuing education, as it offers flexibility with respect to learning times and locations and can reach an unlimited number of clinicians [28]. Web-based e-learning can therefore enhance health care practitioners' knowledge and skills, as a prerequisite for effective use of health behavior interventions such as brief MI [25,29].

User acceptance of Web-based e-learning for specific sociodemographic groups of health care practitioners, such as nurses, is a topic of great interest [30-32]. However, the literature concerning the subject is scarce. According to the unified theory of acceptance and use of technology, various factors influence user acceptance of technology, which in turn influences technology use [33]. Careful attention must then be paid to learners' perceptions and attitudes toward workplace e-learning in order to optimize the knowledge, skills, and clinical use of brief MI [30,34]. The integration of interactivity measures and audiovisual media in e-learning may positively affect learners' perceptions and attitudes [29].

Video-based e-learning showcasing clinical simulation has attracted strong interest from clinicians and researchers [23,35-37]. Videos can facilitate knowledge acquisition and clinical skill development through pedagogical material that matches the reality of clinical settings [36,38-40]. Video-based e-learning has the potential to "enliven abstract concepts, demonstrate real-world applications of complex principles, motivate the learner, organize thoughts and actions of highly cognitive processes, and heighten learner attention and interest" [36]. This is particularly interesting, because MI is usually learned through observation of role models in face-to-face or, most recently, videotaped clinical simulations [11,18].

However, little is known about the educational effectiveness of brief MI training via a Web-based e-learning platform. To our knowledge, cardiovascular nurses' MI-related skill development and clinical use of brief MI have not been evaluated. Therefore, in this study, we developed and pilot tested a Web-based e-learning platform for brief MI, which included videos in which nurses could observe brief MI in a real-life clinical context.

Study Objectives

The primary objective of this pilot study was to examine the feasibility and acceptability of a Web-based e-learning platform for brief MI (MOTIV@CŒUR), which was evaluated by nurses in cardiovascular care. The primary end point of the pilot study was the proportion of nurses who had completed both training sessions 20 days after the initiation of the training session.

The secondary objective was to assess the preliminary effect of MOTIV@CŒUR on nurses' perceived skill in, and self-reported clinical use of, brief MI.

Methods

Study Design and Setting

We conducted a single-group, pre-post pilot study involving cardiovascular nurses to assess MOTIV@CŒUR. We conducted the study at the coronary care unit (CCU) at a tertiary care hospital center in Montreal, Canada. The pilot study was registered (ISRCTN16510888), as well as being approved by the Scientific and Ethics Committee of the Montreal Heart Institute Research Center (reference number: 2015-1948). Our study is reported in accordance with the CONSORT-EHEALTH checklist version 1.6.1 [41] (see Multimedia Appendix 1). No content or methodological modifications were made after study commencement.

Participants

We recruited a convenience sample of nurses employed at the CCU. Nurses were eligible for participation if they were working at the CCU during the study period. The inclusion criteria were employment in a temporary replacement or permanent position at the CCU and basic computer skills. The exclusion criterion was completion of MI training in the preceding year.

Procedure

Enrollment and follow-up occurred between March and May 2016 (see Table 1) [28,42-44]. We recruited nurses through individual face-to-face encounters at the CCU. Participants were informed that they would need to complete the training and study requirements on their personal time without financial compensation. However, it was stated that they would receive a certificate attesting to 1 hour of continuing education after completing the training. After receiving an explanation regarding the study and providing written consent, participants completed a paper-based sociodemographic questionnaire. An individual identification number, username, and password were then provided to participants, to allow them to log in to the e-learning platform throughout the study. They also received a training information sheet that explained MOTIV@CŒUR using screen captures and colorful textual content. During the 15-day period following enrollment, an initial email containing the URL for the Web-based e-learning platform was sent to each participant.

Table 1. Schedule of enrollment, interventions, and assessments for MOTIV@CŒUR.[a]

Participant timeline	Study period and time points				
	Enrollment	Experimentation			Closeout
	t_0 days –20 to 0	t_{1a} day 1	t_{1b} day 1	t_2 day 15 (±5 days)	t_3 day 30 (±5 days)
Enrollment					
Eligibility screen and informed consent	×				
Intervention encounters					
Training sessions in brief MI			Session 1 (30 min)	Session 2 (20 min)	
Assessments					
Sociodemographic questionnaire	×				
Primary objectives					
Feasibility of the Web-based e-learning platform for brief MI	×	×	×	×	×
Acceptability of the Web-based e-learning platform for brief MI					×
Secondary objectives					
Perceived skill in brief MI		×			×
Self-reported clinical use of brief MI		×			×

[a]Template adapted from the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines [42].

[b]MI: motivational interviewing.

[c]Measured throughout the study with indicators from Feeley and Cossette [43].

[d]Measured with Cheng's tool [28].

[e]Measured with the adapted tool of Paradis et al [44].

The Web-Based E-Learning Platform for Brief MI: MOTIV@CŒUR

MOTIV@CŒUR (in French, which translates as MOTIV@HEART in English) is a Web-based e-learning platform for brief MI, which includes role-modeling videos. The intervention content is based on the work of key authors in brief MI [11,44-47].

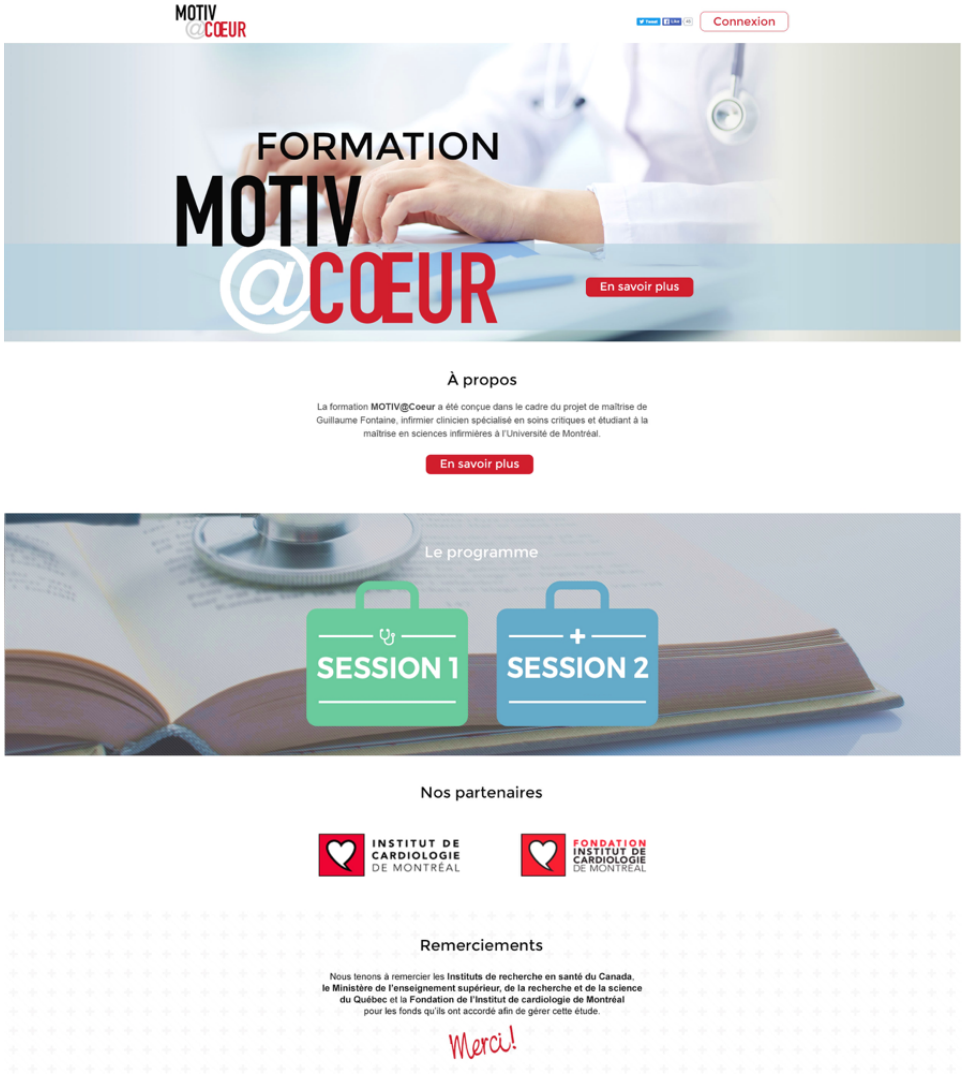
Development Process

The MOTIV@CŒUR Web-based platform was developed by an independent consulting firm in Montreal, Canada. We chose the firm because it designs interactive websites whose format is adaptive to computers, tablets, and smartphones. MOTIV@CŒUR is based on the open-source learning platform Moodle 3.0 (Moodle Pty Ltd, Perth, Australia). The MOTIV@CŒUR homepage (see Figure 1) was designed to create an appealing first impression, using visual material and dynamic components.

Subsequent adaptation of the content of the e-learning platform for brief MI is possible via access to the Moodle course management system provided to the research team. However, changes in the design of the website require the involvement of the consulting firm.

To ensure the preservation of data related to the implementation MOTIV@CŒUR and the usage statistics for nurses, the website and data regarding its use were hosted on secure computer servers at the research setting for the duration of the study.

Figure 1. MOTIV@COEUR homepage (in French).



MOTIV@CŒUR Access

MOTIV@CŒUR can be accessed via a fixed URL. The availability of the website was restricted to the study period. Participants could log in to the e-learning platform from home or work via the device of their choice, using their personal log-in credentials, which were provided during the face-to-face encounter at the CCU. We suggested that participants change their passwords after the initial log-in. Passwords could also be reset via their personal email accounts if forgotten.

MOTIV@CŒUR Content

The content of the brief MI on the Web-based e-learning platform was developed by the project team, which included 1 MI expert and an experienced cardiology researcher, who supervised the development of the intervention, led by GF. In addition, 2 MI experts validated the content. The intervention

was pretested with 5 nurses who were not part of the sample of nurse participants in this study. We adjusted MOTIV@CŒUR content according to the nurses' comments before we recruited the study sample.

MOTIV@CŒUR was designed around 4 scenarios, each presenting a clinical case involving a client with a given level of conviction (low or high) and confidence (low or high) regarding change in a health behavior (see Figure 2) [47]. Each motivation profile was associated with one of the following health behaviors: smoking cessation, medication adherence, physical activity, or diet. For instance, clinical case #1 presents the association between low conviction and confidence levels for smoking cessation. We chose 4 different associations between motivation levels and health behaviors as examples that could be extrapolated to other health behaviors for individuals with any motivation profile. The team developed scenarios for each clinical case, based on real-life experience. During the scenarios, the nurse introduced herself, targeted the health behavior in each clinical case, assessed the level of conviction and confidence regarding change, and engaged the client in a brief MI conversation. Following each scenario, a second video showed the cardiology nurse practitioner (CNP) explaining why each intervention was retained in response to the client's motivation profile.

The content of brief MI for the 4 scenarios was based on the model developed by Bédard [47], who adapted the work of Miller and Rollnick [11]. In this model, *conviction* represents the extent to which each individual perceives practical and emotional benefits to the change of a health behavior. *Confidence* represents the extent to which the individual is confident of being able to achieve change [47]. After assessing the client's motivation, the practitioner provides tailored brief MI to increase conviction and confidence regarding health behavior change. Through videotaped role modeling, nurses could observe the CNP involving brief MI in a real-life clinical context. Videos were recorded at the research setting in a real patient room, to represent the real-life context (see Multimedia Appendix 2 for screenshots), with 4 volunteers (2 men and 2 women) representing different ages.

MOTIV@CŒUR was conceptualized to ensure that participants would observe real-life examples, allowing them to (1) familiarize themselves with the spirit of brief MI, (2) acquire basic skills in brief MI (eg, open-ended questions, validation, and reformulation), (3) recognize and reinforce the change discourse, specifically that involving conviction and confidence, (4) learn to create and strengthen the change discourse, (5) learn to accept resistance to avoid confrontation, (6) understand how to develop a plan, and (7) understand how to help clients to initiate change [11,48].

MOTIV@CŒUR Structure

MOTIV@CŒUR consists of 2 training sessions including 13 video modules (see Figure 3). The planned durations were 30 minutes for the first session (S1) and 20 minutes for the second session (S2). Following an introduction and statement of objectives, each session was initiated with a video containing a theoretical introduction to brief MI. Each clinical case was then separated into 3 sections: (1) a textual presentation of the clinical case on the screen, (2) a video of brief MI, in which the CNP interacted with each client, and (3) a video in which the CNP explained why each intervention was retained in response to the client's motivation profile. Both sessions concluded with a reminder of the key concepts and tips for real-world use of brief MI.

Figure 2. The 4 clinical cases and motivation profiles presented in MOTIV@CŒUR.

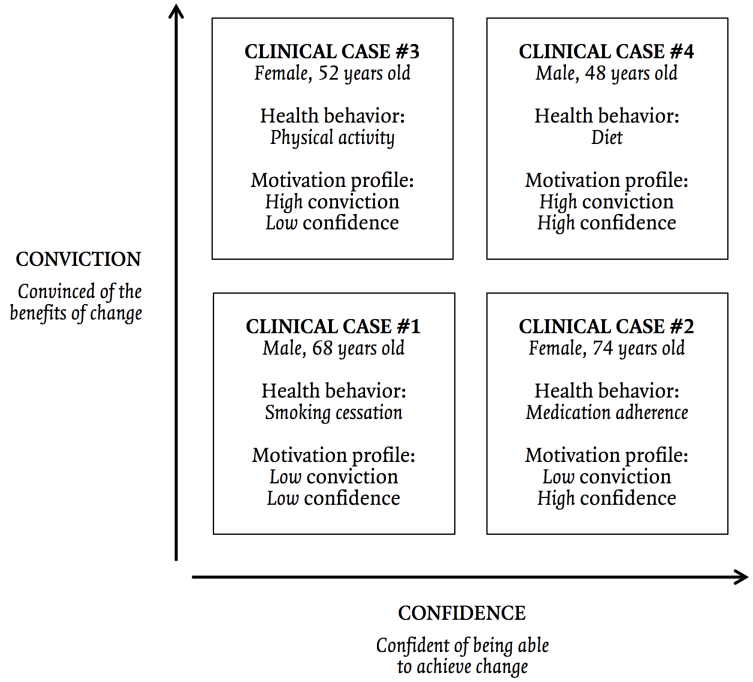
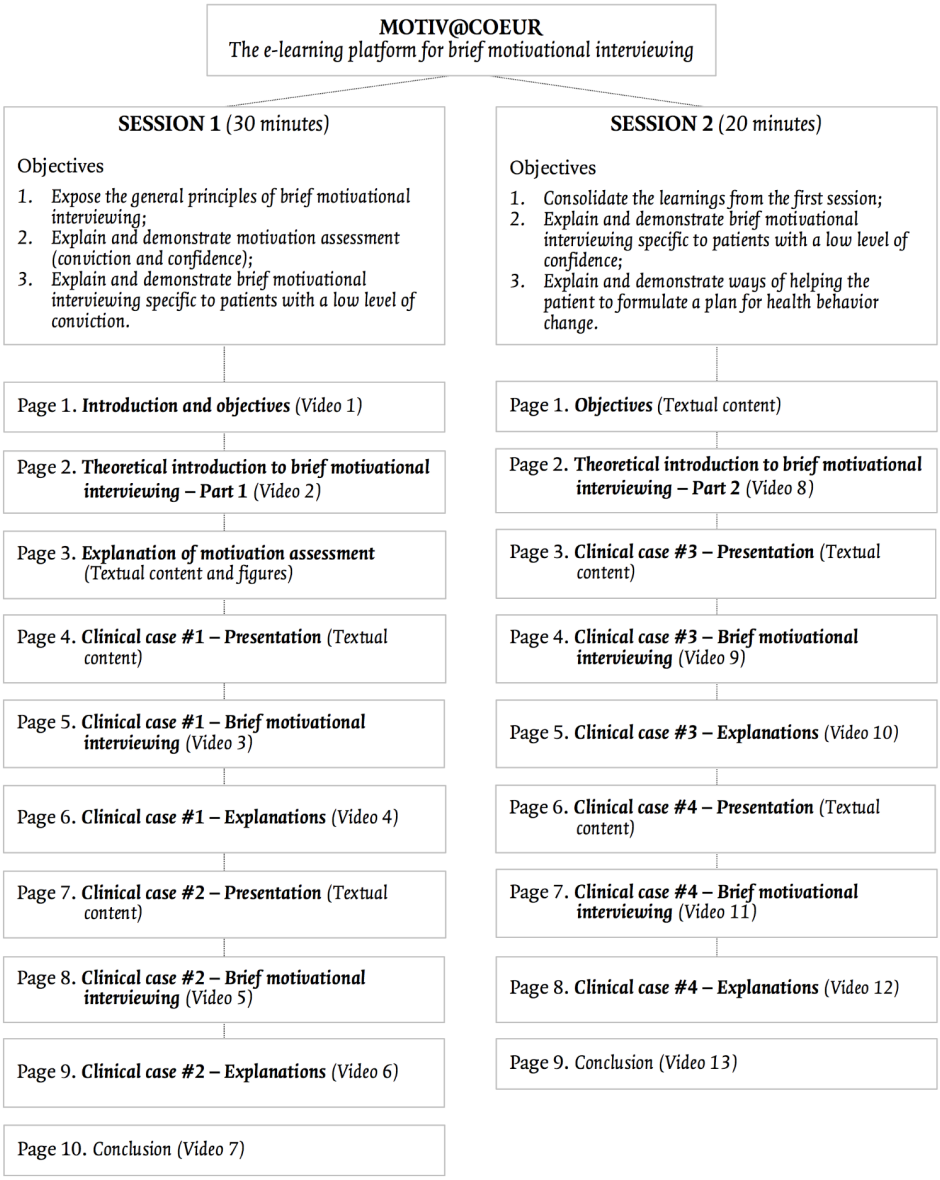


Figure 3. Structure of MOTIV@CŒUR, a web-based e-learning platform for brief motivational interviewing.



Use Parameters

We asked participants to complete S1 within 5 days of receiving the initial email sent after enrollment. Two options were provided for completing the first MOTIV@CŒUR Web-based training session. Participants could complete the Web-based training individually in a dedicated room equipped with computers at the study hospital, during a scheduled session in which a facilitator would explain the project and procedure for accessing MOTIV@CŒUR to each participant. Participants could also complete the training at home with remote support (eg, by email or telephone). In addition, they were encouraged to practice brief MI techniques observed in the video in their regular clinical practice, if appropriate.

We asked nurses to complete S2 either at home or at the hospital, 2 weeks after S1. There was no computer constraint limiting completion of S2 earlier or later than this. However, participants were required to complete both sessions within 20 days.

Reminders, Level of Human Involvement, and Co-interventions

We planned a maximum of 3 email or telephone reminders at 3-day intervals for each of the 3 time points in the study (S1, S2, and outcome measures). A maximum of 9 emails or telephone reminders could be sent throughout the study period.

The intervention was completely asynchronous. The research team was available at all times, to provide technical support in person or via mail or telephone. Apart from the technical support provided when necessary (access to the website, log-in, and password), we offered no other intervention, such as that involving information and content explanation regarding the brief MI.

Outcome Measures

The primary feasibility outcome was the completion of both training sessions 20 days after initiation of S1. We also assessed additional feasibility outcomes regarding recruitment and study completion.

Secondary outcomes included the acceptability of MOTIV@CŒUR according to the cardiovascular nurses, skills perceived in brief MI, and self-reported clinical use of brief MI with coronary patients. These outcomes were self-assessed via Web-based questionnaires.

MOTIV@CŒUR Feasibility

We measured the feasibility of the Web-based e-learning platform for brief MI from recruitment to closeout using indicators collected throughout the study period, based on pilot study evaluation criteria established by Feeley and Cossette [43].

Feasibility indicators collected by the research team included the proportion of enrolled nurses in the eligible population, recruitment duration, and completion of outcome measures.

Feasibility indicators were also extracted from the Moodle platform. Moodle collects information about each user with an exact time stamp for each action (ie, change of a page in a module or completion of a module). We interpreted the interval between 2 actions as engagement with the site or absence from the site. Since an inappropriately long latency period between the user's actions would overestimate the time spent on each session, we defined a maximum latency period, fixed at 15 minutes. When a latency period exceeded this threshold, we deducted it from the time spent on MOTIV@CŒUR. Feasibility indicators extracted from the Moodle platform for each user included the duration of each session, number of sessions completed, and time elapsed between the completion of S1 and S2.

We also recorded types, numbers, and timestamps for reminders sent to participants in an Excel file, version 15.16 (Microsoft).

MOTIV@CŒUR Acceptability

We used the model of information systems quality antecedents on nurses' acceptance of e-learning, developed by Cheng [28], to assess posttraining acceptability of the Web-based e-learning platform for brief MI. The tool evaluates nurses' perception of the e-learning system, using 27 items grouped into 2 main dimensions: global system quality and technology acceptance. These dimensions are based on DeLone and McLean's [49] work in information systems quality and van der Heijden's [50] technology acceptance model. The model is subdivided into 8 subdimensions, of which 4 are related to global system quality (system, information, service, and user interface design quality), and 4 are related to technology acceptance (perceived usefulness, perceived ease of use, perceived enjoyment, and intention to use). The items in each subdimension were described the original paper [28]. Responses are provided using a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), with 4 representing neutral responses. The score for each subdimension is calculated by summing the scores for the responses to the items therein and dividing the result by the number of items in the subdimension. A higher total score indicates greater acceptability (possible range: 0–7). Cronbach alphas for the scale were between .70 and .96 in previous studies [28]. The tool was translated into French using the back-translation method defined by the World Health Organization [51]. The content validity of the translated items was then determined by an expert, who provided adjustments to the language and structure of the items. A pretest was performed and included nurses who were not involved in the project.

Nurses could also provide suggestions and comments regarding MOTIV@CŒUR at the end of the acceptability questionnaire.

Preliminary Efficacy of MOTIV@CŒUR

We adapted the tool of nursing interventions specific to conviction and confidence levels and stages of change developed by Paradis et al [44], to assess perceived skill in brief MI and the self-reported clinical use of brief MI before and after training. We reduced the number of interventions from 55 to 26, retaining only those that targeted conviction and confidence, as this was the primary focus of the brief MI training in the study. The content was then validated by 2 MI experts.

The scale contained 26 intervention items grouped under 2 motivational intervention dimensions: conviction (16 items) and confidence (10 items). Based on the work of Cossette et al [52], 2 questions were asked for each intervention item to assess outcomes. The first question ("How comfortable do you feel doing it?") assessed nurses' perceived skill in performing each intervention. The second question ("How often do you do it?") assessed nurses' self-reported clinical use of each intervention. Each question was used to calculate a total score and 2 subdimension scores for conviction and confidence. The response scale for each question ranged from 1 (not at all) to 5 (extremely) and provided 2 total scores ranging from 26 to 130. A higher score for the first question indicated higher perceived skill in brief MI, and a higher score for the second question indicated higher clinical use of brief MI. To calculate total scores, we recoded a maximum of 3 missing values per participant in the mode for each item.

Other Measures

A self-administered sociodemographic paper questionnaire was completed at enrollment to collect data regarding nurses' general profiles concerning sex, age, language, educational level, year of entry to the hospital, employment status, duration of experience in nursing and cardiovascular acute care, shift, and type of position held at the CCU. We also asked participants whether they had previously completed Web-based training.

Sample Size

To examine the primary feasibility outcome, we defined success as the completion of both training sessions by 80% of participants within 20 days. We expected this rate to be 80%; therefore, we targeted a sample of 30 participants to allow estimation with accuracy of ±14.3% and a confidence level of 95%.

Statistical Analysis

With respect to sociodemographic, acceptability, and preliminary effect variables, we calculated means and SDs for continuous variables, and counts and percentages for categorical variables. We also used descriptive statistics for the feasibility criteria, as follows: (1) proportion of enrolled nurses in the eligible target population (expected: 50%), (2) time required to complete recruitment (expected: 60 ± 30 days), (3) proportion

of nurses who completed both sessions in ≤20 days (expected: 80%), (4) proportion of nurses who completed both sessions in 60 ± 10 min (expected: 80%), and (5) proportion of nurses who completed outcome measures (expected: 80%).

We compared changes between pre- and posttraining measures for perceived skill in brief MI and self-reported clinical use of brief MI using Student *t* test for paired samples, with a 2-sided significance level of .05. We also performed Student *t* test for paired samples, with the same parameters used, for the 4 subdimension scores for pre- and posttraining conviction and confidence. All statistical tests were 2-sided and performed using IBM SPSS version 23.0 (IBM Corporation). We verified basic assumptions, such as normal distribution, before analysis.

Results

Participant Characteristics

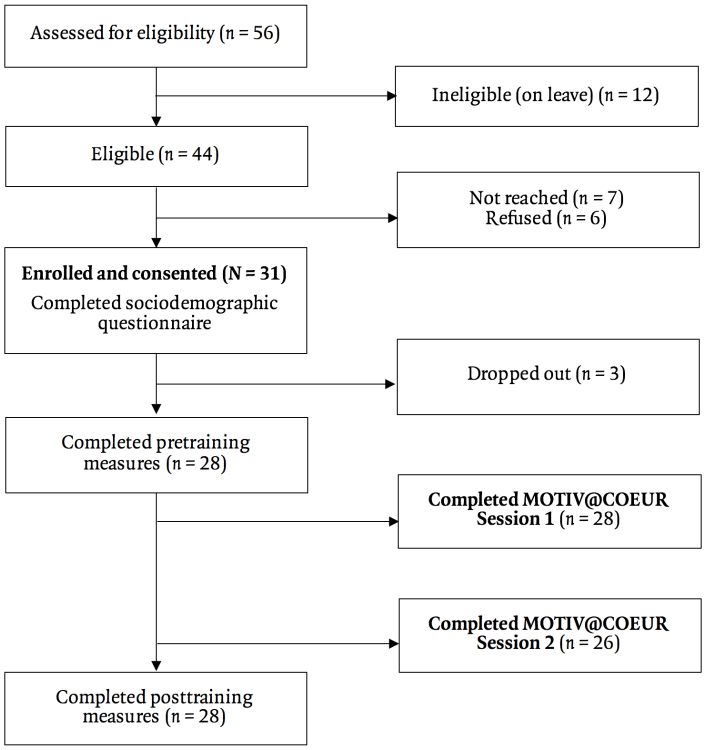
Most participants were women, and participants' mean age was 37 years (see Table 2). The majority had completed university-level education and worked full-time as bedside nurses. The duration of participants' experience as critical cardiovascular care nurses ranged from 1 month to 37 years, with a mean of 11 years. Nurses were almost evenly distributed across all work shifts, with 5 working rotating shifts. More than three-quarters of participants had previously completed Web-based training, but none had undertaken MI in the preceding year.

Table 2. Nurses' baseline sociodemographic data (N=31).

Characteristic	Mean (SD) or n (%)
Sex (female), n (%)	27 (87%)
Age, in years, mean (SD)	37 (9)
Education (Bachelor's degree or higher), n (%)	18 (58%)
Position (full-time), n (%)	18 (58%)
Position in coronary care unit (bedside nurse), n (%)	27 (87%)
Experience in acute care, in years, mean (SD)	11 (10)
Shift, n (%)	
Day	9 (29%)
Evening	9 (29%)
Night	8 (26%)
Rotation	5 (16%)
Previously completed Web-based training (yes)	24 (89%)[a]

[a]n=27.

Figure 4. MOTIV@CŒUR study participation flowchart.



Feasibility Results

Recruitment

The feasibility criteria for participant recruitment were all met. Of the 56 nurses employed at the CCU, 44 were eligible for study participation, and 31 (70%) were enrolled in the study between March and May 2016 (see Figure 4). This exceeded the target proportion of 50%. Moreover, recruitment was completed within 11 days, which was a significantly shorter period than the expected period of 30–90 days.

Training Nurses via the Web-Based E-Learning Platform for Brief MI

With regard to the primary feasibility outcome, 24 of the 31 recruited participants (77%, 95% CI 63%–92%) completed both training sessions within 20 days following initiation of S1 (see Table 3). This was close to the criterion for determining success (ie, 80%). Another 2 nurses completed S2 within 26 and 30 days of S1. In addition, 28 participants had completed S1 and 26 had completed S2 at 30 days. A total of 3 participants dropped out before beginning the training, resulting in 28 participants completing pretraining and posttraining measures.

Table 3. Feasibility of MOTIV@CŒUR (N=31).

Feasibility or outcome variable	No. or n (%)
Feasibility criteria	
1. Nurses in the eligible target population (expected: 50%), n (%)	31 (70%)
2. Duration of recruitment (expected: 30–90 days), no. of days	11
3. Completed both sessions within ≤20 days (expected: 80%), n (%)	24 (77%)
Completed the first session at 30 days, n (%)	28 (90%)
Completed both sessions at 30 days, n (%)	26 (84%)
4. Completed both sessions within 60 ± 10 min (expected: 80%), n (%)[a]	19 (73%)
5. Completed posttraining measures (expected: 80%), n (%)	28 (90%)
Completion of pretraining measures and first session, n (%)	
Completed before a reminder was sent	10 (32%)
Completed after 1 reminder was sent	10 (32%)
Completed after 2 reminders were sent	7 (23%)
Completed after 3 reminders were sent	1 (3%)
Never completed	2 (6%)
Completion of second session, n (%) [b]	
Completed before a reminder was sent	14 (50%)
Completed after 1 reminder was sent	10 (36%)
Completed after 2 reminders were sent	0
Completed after 3 reminders were sent	2 (7%)
Never completed	2 (7%)
Completion of posttraining measures, n (%) [b]	
Completed before a reminder was sent	13 (46%)
Completed after 1 reminder was sent	8 (29%)
Completed after 2 reminders were sent	6 (21%)
Completed after 3 reminders were sent	1 (4%)
Never completed	0

[a]n=26.

[b]n=28.

The results showed that 25 participants completed S1 and 22 completed S2 during a single connection. The mean durations were 31 (SD 6) minutes for S1 and 19 (SD 6) minutes for S2. The mean total training duration was 50 (SD 11) minutes, which was consistent with the expected duration for MOTIV@CŒUR. The mean period between the completion of S1 and S2 was 13 (SD 7) days, which was close to the recommended time of 2 weeks.

Of the 31 participants, 10 (32%) completed the baseline measures and S1 without requiring a reminder after the initial email providing instructions regarding accessing the Web-based e-learning platform. This proportion was higher in the rest of the study: 14 of 28 participants (50%) completed S2 without a reminder, and 13 of 28 participants (46%) completed the outcome measures without a reminder. Across the 3 time points, the first email reminder was more effective than the second and third reminders and doubled the number of participants who fulfilled the requirements.

In total, 80 emails and telephone reminders were sent throughout the study period. More specifically, 44 email, 16 telephone, and 20 voicemail reminders were sent. Of these, the email reminders were the most effective. Of the 44 email reminders, 27 (61%) resulted in the completion of requirements at each time point (S1, S2, and outcome measures), while 9 of 16 (56%) telephone reminders and 9 of 20 (45%) voicemail reminders were effective throughout the study period.

Acceptability Outcomes

The Web-based e-learning platform for brief MI was considered highly acceptable by cardiovascular nurses across all 8 dimensions of Cheng's [28] model (see Table 4).

Table 4. Posttraining acceptability of MOTIV@CŒUR (n=28).

Outcome variable	No. of items	Possible range	Mean (SD) score
Global system quality	15	0–7	5.95 (0.48)
System quality	5	0–7	6.15 (0.58)
Information quality	4	0–7	6.26 (0.60)
Service quality	3	0–7	5.28 (0.96)
User interface design quality	3	0–7	6.12 (0.69)
Technology acceptance	12	0–7	5.90 (0.75)
Perceived usefulness	3	0–7	5.64 (0.81)
Perceived ease of use	3	0–7	6.16 (0.78)
Perceived enjoyment	3	0–7	5.80 (1.01)
Intention to use	3	0–7	6.01 (0.84)

The 4 dimensions concerning system quality were evaluated favorably, and each received a mean score of >5 on the 7-point Likert scale. The 3 items that received the highest scores for system quality were the flexibility of MOTIV@CŒUR regarding learning time and location, presentation of course materials in a readable multimedia format, and the delivery schedule for the learning content. The 3 items that received the lowest scores were related to the quality of support services, as 11 nurses did not use them and provided neutral scores, which were below the observed scores of >5 for other items. The information quality subdimension received the highest score in the dimension related to global system quality.

The 4 dimensions concerning technology acceptance were evaluated very favorably by all participants and received scores of >5 on the 7-point Likert scale. The 3 items that received the highest scores were related to the ease of use of MOTIV@CŒUR, the usefulness of MOTIV@CŒUR for learning, and the opinion that MOTIV@CŒUR should be available to other nurses and professionals. While most participants agreed or strongly agreed that MOTIV@CŒUR was useful in their learning, they appeared less convinced of the superiority of e-learning relative to traditional face-to-face methods. Indeed, the 3 items that received the lowest scores but still scored >5 were related to enhanced learning effectiveness compared with other training methods, increased learning efficiency, and enjoyment while training with MOTIV@CŒUR. Finally, the overwhelming majority of participants agreed or strongly agreed that they would use the e-learning platform again if it were made available with more content and resources.

Comments of participants at the end of the acceptability questionnaire underlined the simplicity, clarity, and dynamism of the e-learning platform. One participant suggested developing a checklist on the training content to be made available to nurses in the clinical setting. Another participant proposed conducting practical workshops to implement the learning acquired during the Web-based training. Overall, the feedback from participants was positive and indicated significant interest in the Web-based e-learning platform for brief MI.

Preliminary Efficacy Outcomes

Regarding the preliminary efficacy of MOTIV@CŒUR with respect to perceived skill in brief MI, posttraining scores for all dimensions were higher than pretraining scores. However, the raw differences were small and nonsignificant (see Table 5).

Table 5. Preliminary effect of MOTIV@CŒUR on perceived skill in brief motivational interviewing (MI) and self-reported clinical use of brief MI.

Outcome variable	No. of items	Possible range	Mean (SD) scores		*P* value
			Pretraining	Posttraining	
Perceived skill in brief MI [a,b]	26	26–130	95.19 (16.37)	97.50 (15.38)	.40
Conviction interventions	16	16–80	60.62 (9.79)	61.53 (8.87)	.54
Confidence interventions	10	10–50	34.59 (7.01)	35.93 (6.98)	.30
Self-reported clinical use of brief MI [a,c]	26	26–130	89.60 (15.74)	94.28 (13.64)	.13
Conviction interventions	16	16–80	58.12 (9.44)	59.56 (7.97)	.41
Confidence interventions	10	10–50	31.48 (6.75)	34.72 (6.29)	.03

[a]Higher is better.

[b]n=26.

[c]n=25.

In addition, regarding the preliminary efficacy of MOTIV@CŒUR with respect to self-reported clinical use of brief MI, posttraining scores for all dimensions were higher than pretraining scores. A significant effect was observed for

self-reported clinical use of brief MI to increase clients' confidence in change (P=.03). Other results were nonsignificant.

Discussion

This study involved the design, implementation, and evaluation of a Web-based e-learning platform for brief MI, which included role-modeling videos for nurses in cardiovascular care. We demonstrated the feasibility, acceptability, and preliminary efficacy of the intervention. In addition, preliminary posttraining results regarding perceived skill and clinical use of brief MI were all more favorable than those observed in the pretraining assessment. Overall, the feedback received from participants was positive.

While some previous studies examined Web-based MI training with health care practitioners [53-56], to our knowledge, this study was the first to examine cardiovascular nurses' evaluation of an asynchronous Web-based e-learning platform for brief MI. We were successful in recruiting 31 participants within 11 days, of whom 28 completed posttraining measures. This demonstrates cardiovascular care nurses' significant interest in Web-based e-learning and interventions targeting health behavior change. The strong participation in the study could reflect the applicability and credibility of the use of brief MI in acute care settings. Brief MI demonstrated in the MOTIV@CŒUR videos lasted 3–4 minutes. This duration is more likely to be feasible in clinical settings than in longer motivational interventions [12-14].

Previous studies suggested that technical difficulties, such as a lack of Internet access, could impede the ease with which information and communication technology could be used by health care practitioners [22,57-59]; however, this was not the case in our study, as we did not experience problems with computers. We informed nurses that they were required to be at ease with basic computer use, prior to enrollment, and the research team was available for prompt technical support via email. The e-learning training progressed very well without significant technical difficulties. Participants asked occasional questions (eg, regarding a malfunctioning URL link), but no one experienced difficulty in using the Web-based e-learning platform. This could suggest that nurses in acute care settings are familiar with the use of information and communication technology for clinical and pedagogical purposes. Of the 27 participants who completed the acceptability measures in this study, 24 (89%) had previously completed Web-based training for other topics. The streamlining and improvement of the user interface design in Web-based training platforms could also have affected the ease with which participants used the system [60].

Participant reminders are often overlooked but crucial to asynchronous Web-based e-learning. Literature concerning the subject is scarce; only a few studies have been conducted, and they reported incomplete data regarding frequency, content, numbers, and mode of delivery (eg, telephone or email) for reminders sent to participants. For example, one study [61] proposed up to 3 automated email reminders for incomplete modules. Other studies included 2 automated emails sent 7 days apart, with an additional personalized email and telephone call if required [62], weekly reminders [63], and 2 reminders after 2 weeks [64]. This heterogeneity shows a lack of consensus regarding best practice with respect to the reminders sent to participants. In this study, we decided to send a maximum of 3 telephone or email reminders 3 days apart, at each time point to avoid oversoliciting participants. Two reminders ensured that approximately 90% of participants completed the sessions and measures. Relative to telephone and voicemail reminders, email reminders were more effective in ensuring the completion of requirements at each time point. This finding could inform future studies.

Our study's high acceptability scores suggested that the Web-based e-learning platform for brief MI, based on Moodle, could be ready for inclusion in a larger study. However, some participants asked for further details and interactivity measures, which could be included in future iterations of the platform. The positive aspects of e-learning observed in this study, such as flexibility and control regarding the learning time and location, are consistent with those reported in the literature [22,25,26]. This could be explained by the adaptive format of MOTIV@CŒUR, which can be used anywhere via smartphones, tablets, and computers; however, we did not collect this information. Moreover, participants appreciated the presentation of MOTIV@CŒUR course materials in a multimedia format, as they all reported acceptability scores of >6 for this item in the posttraining assessment. This extends existing literature concerning the feasibility and acceptability of illustrating complex clinical processes, such as brief MI, in video modules [18,38,56].

The next step of this project is to optimize the tailoring, structure, and content of brief MI in the Web-based e-learning platform. Moreover, we intend to evaluate this platform in a randomized controlled trial, to assess its efficacy in comparison with alternative instructional methods such as face-to-face training and reading. Assessment of participant knowledge based on training content is an outcome we will explore in our future research. Moreover, objective measures are required for clinical skills and motivational interventions provided in health care settings. We also intend to assess the effect of brief MI, provided by health care practitioners, on health behavior change in coronary clients.

Future research should assess tailored, interactive, Web-based e-learning platforms for brief MI, as this was not the focus of our study, and the scientific literature has demonstrated the efficacy of such features [29]. In addition to tailoring the platform to health care practitioners' knowledge and experience, researchers should develop an algorithm that accounts for each participant's characteristics and specific needs (for instance, some participants asked for additional content, while others were satisfied with what was provided in MOTIV@CŒUR). In doing so, they could ensure that every participant follows an individualized path that could lead to enhanced knowledge and clinical skills [25,29]. The efficacy of interactivity measures in e-learning has been demonstrated in the scientific literature [25,29]. Web-based e-learning platforms for brief MI could benefit from the inclusion of self-assessment questions, interactive models and figures, and thought-stimulating activities

[29]. When combined with videos, these elements are potentially valuable for scientific, pedagogical, and clinical purposes.

Future research should explore how to assess the effects of Web-based e-learning for brief MI on objective results in clinical settings. Indeed, despite the progress that has been made in recent years, evaluating the effects of e-learning on real clinical behavior and client outcomes remains a challenge [60,65]. With regard to clinical behavior, researchers should assess the effect of new skills acquired via Web-based e-learning for brief MI on practice using methods other than those involving self-report measures. Supervised clinical simulations of brief MI in parallel with Web-based training could be an interesting means of assessing changes in clinical practice.

Regarding the clinical implications of the study, the results regarding feasibility and acceptability were encouraging and showed that cardiovascular nurses were willing and able to use a Web-based e-learning platform for brief MI to develop skills related to health behavior change. This suggests that Web-based training covering a larger scope of clinical situations and levels of motivation could be designed to assist health care practitioners in providing health behavior change interventions. These interventions could target a larger spectrum of risk factors other than those related to coronary artery disease.

Strengths and Limitations of the Study

The strengths of the study include adherence to the study protocol, the prospective registration of the study, and encouraging feasibility and acceptability results. In addition, no MOTIV@CŒUR-related technical problems occurred during the study period.

The study demonstrated the potential of Web-based e-learning training for brief MI, but it was subject to some limitations. First, as it was a pilot study, it was not designed for adequate power. Second, the Web-based, self-administrated questionnaires used in the study are not objective measures of real clinical use of brief MI. Third, the single-group, pre-post study design did not allow for causal inferences.

Most participants had experimented with Web-based training prior to entering the study. This could provide a partial explanation as to why the Web-based e-learning platform showed such high levels of acceptability. A study with a more diverse population of nurses and other health care practitioners could be interesting and allow researchers to determine whether sociodemographic variables increase acceptability scores and affect knowledge acquisition and clinical outcomes. However, this proved difficult in this pilot study, as the small sample size did not allow for enough power.

Not all participants enrolled in the study ultimately used MOTIV@CŒUR for training in brief MI, as 3 individuals dropped out before beginning the training. However, the global participation rate in the study was superior to those observed in similar studies. Indeed, 28 of the 31 participants (90%) used MOTIV@CŒUR, and this proportion ranged from 82% to 89% in other studies [66-69]. While our study included cardiovascular nurses, it is possible that other health care practitioners could benefit from the training.

Conclusion

Information and communication technology is instrumental in the future of health care practitioners' education. Indeed, technology is ubiquitous in clinical, professional, and academic settings. Researchers should consider a wide variety of factors, to provide rich, interactive, tailored Web-based e-learning and enhance health care practitioners' knowledge, skills, and clinical interventions. The optimization of factors related to system quality and technology acceptance could contribute to the way in which care is learned, planned, and provided in health care settings for years to come. Further research is required to improve understanding of health care practitioners' interactions and technology use in learning, and the impact of Web-based e-learning on patient care. Our results showed that the Web-based e-learning platform for brief MI was feasible and acceptable according to nurses in cardiovascular care. Moreover, the preliminary posttraining results regarding perceived skill and clinical use of brief MI were all more favorable than those observed in the pretraining assessment. MOTIV@CŒUR, which includes role-modeling videos, could introduce nurses to brief MI for the reduction of cardiovascular risk and exert an impact on their skills regarding motivational interventions.

Future research should focus on tailoring Web-based e-learning platforms to health care practitioners' existing knowledge and experience, to provide individualized paths and fulfill specific learning needs. Further, such training would benefit from the inclusion of additional interactivity measures to enhance the learning experience.

Acknowledgments

The authors wish to thank the multidisciplinary team responsible for the conceptualization and operationalization of MOTIV@CŒUR. We are grateful to the 4 volunteers who portrayed clients in the video modules. We are also grateful to Sophie Langlois and John Kayser, who validated the MI content. Finally, we would like to thank Eric Maillet, who validated the acceptability questionnaire, and the 5 nurses who pretested MOTIV@CŒUR.

The authors would like to thank the Canadian Institutes for Health Research; the Quebec Ministry of Higher Education; the Montreal Heart Institute (MHI) Research Center; the MHI Foundation; and the Faculty of Nursing at the University of Montreal for their financial support.

Authors' Contributions

All authors contributed to the study design and protocol. GF: study conception and design, data collection and conception of the analytical plan, manuscript writing, and final approval of the manuscript. SC: study conception and design, data collection and conception of the analytical plan, manuscript writing, and final approval of the manuscript. SH: study conception and design of the brief MI content of the Web-based e-learning platform, manuscript writing, and final approval of the manuscript. LB: study conception and instructional design of the Web-based e-learning platform, manuscript writing, and final approval of the manuscript. TM: study conception and design, manuscript writing, and final approval of the manuscript. MJS: study conception and design, manuscript revision, and final approval of the manuscript. JFT: study conception and design, manuscript revision, and final approval of the manuscript.

Conflicts of Interest

The authors of this study own MOTIV@CŒUR.

Multimedia Appendix 1

CONSORT-EHEALTH Checklist V 1.6.1.

[PDF File (Adobe PDF File), 762KB - jmir_v18i8e224_app1.pdf]

Multimedia Appendix 2

Screenshots.

[PPTX File, 14MB - jmir_v18i8e224_app2.pptx]

References

1. Finegold JA, Asaria P, Francis DP. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. Int J Cardiol 2013 Sep 30;168(2):934-945 [FREE Full text] [doi: 10.1016/j.ijcard.2012.10.046] [Medline: 23218570]
2. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren M, European Association for Cardiovascular Prevention & Rehabilitation (EACPR), ESC Committee for Practice Guidelines (CPG). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J 2012 Jul;33(13):1635-1701 [FREE Full text] [doi: 10.1093/eurheartj/ehs092] [Medline: 22555213]
3. Blokstra A, van Dis I, Verschuren WM. Efficacy of multifactorial lifestyle interventions in patients with established cardiovascular diseases and high risk groups. Eur J Cardiovasc Nurs 2012 Mar;11(1):97-104. [doi: 10.1016/j.ejcnurse.2010.10.005] [Medline: 21130687]
4. Janssen V, De Gucht V, Dusseldorp E, Maes S. Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. Eur J Prev Cardiol 2013 Aug;20(4):620-640. [doi: 10.1177/2047487312462824] [Medline: 23022703]
5. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [Medline: 3378902]
6. Strecher VJ, Rosenstock IM. The health belief model. In: Baum A, editor. Handbook of Psychology, Health and Medicine. Cambridge, UK: Cambridge University Press; 1997:113-116.
7. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
8. Britt E, Hudson SM, Blampied NM. Motivational interviewing in health settings: a review. Patient Educ Couns 2004 May;53(2):147-155. [doi: 10.1016/S0738-3991(03)00141-1] [Medline: 15140454]
9. Chow CK, Jolly S, Rao-Melacini P, Fox KA, Anand SS, Yusuf S. Association of diet, exercise, and smoking modification with risk of early cardiovascular events after acute coronary syndromes. Circulation 2010 Feb 16;121(6):750-758 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.891523] [Medline: 20124123]
10. de Waure C, Lauret G, Ricciardi W, Ferket B, Teijink J, Spronk S, et al. Lifestyle interventions in patients with coronary heart disease: a systematic review. Am J Prev Med 2013 Aug;45(2):207-216. [doi: 10.1016/j.amepre.2013.03.020] [Medline: 23867029]
11. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd edition. New York, NY: Guilford Press; 2012.
12. McGrady A, Burkes R, Badenhop D, McGinnis R. Effects of a brief intervention on retention of patients in a cardiac rehabilitation program. Appl Psychophysiol Biofeedback 2014 Dec;39(3-4):163-170. [doi: 10.1007/s10484-014-9252-y] [Medline: 25150038]

13. Pietrabissa G, Sorgente A, Castelnuovo G. Integrating motivational interviewing with brief strategic therapy for heart patients. Procedia Soc Behav Sci 2015 Jan;165:136-143. [doi: 10.1016/j.sbspro.2014.12.615]
14. Ski CF, Thompson DR. Motivational interviewing as a brief intervention to improve cardiovascular health. Eur J Cardiovasc Nurs 2013 Jun;12(3):226-229. [doi: 10.1177/1474515112472271] [Medline: 23303765]
15. Ren Y, Browning C, Yang H, Thomas S. Motivational interviewing and its application in the management of coronary heart disease. Family Med Commun Hlth 2013 Sep 01;1(3):48-54. [doi: 10.15212/fmch.2013.0309]
16. Heckman C, Egleston B, Hofmann M. Efficacy of motivational interviewing for smoking cessation: a systematic review and meta-analysis. Tob Control 2010;19(5):410-416. [doi: 10.1136/tc.2009.033175] [Medline: 20675688]
17. Knight K, McGowan L, Dickens C, Bundy C. A systematic review of motivational interviewing in physical health care settings. Br J Health Psychol 2006;11(2):319-332. [doi: 10.1348/135910705X52516] [Medline: 16643702]
18. Soderlund L, Madson M, Rubak S, Nilsen P. A systematic review of motivational interviewing training for general health care practitioners. Patient Educ Couns 2011;84(1):16-26. [doi: 10.1016/j.pec.2010.06.025] [Medline: 20667432]
19. Madson M, Loignon A, Lane C. Training in motivational interviewing: a systematic review. J Subst Abuse Treat 2009;36(1):101-109. [doi: 10.1016/j.jsat.2008.05.005] [Medline: 18657936]
20. Barwick MA, Bennett LM, Johnson SN, McGowan J, Moore JE. Training health and mental health professionals in motivational interviewing: a systematic review. Child Youth Serv Rev 2012 Sep;34(9):1786-1795. [doi: 10.1016/j.childyouth.2012.05.012]
21. Clark R, Mayer R. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. 4th edition. Hoboken, NJ: John Wiley & Sons; 2011.
22. Bembridge E, Levett-Jones T, Jeong S. The transferability of information and communication technology skills from university to the workplace: a qualitative descriptive study. Nurse Educ Today 2011;31(3):245-252. [doi: 10.1016/j.nedt.2010.10.020] [Medline: 21093125]
23. Button D, Harrington A, Belan I. E-learning & information communication technology (ICT) in nursing education: A review of the literature. Nurse Educ Today 2014 Oct;34(10):1311-1323. [doi: 10.1016/j.nedt.2013.05.002] [Medline: 23786869]
24. Kala S, Isaramalai S, Pohthong A. Electronic learning and constructivism: a model for nursing education. Nurse Educ Today 2010 Jan;30(1):61-66. [doi: 10.1016/j.nedt.2009.06.002] [Medline: 19573956]
25. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
26. Lahti M, Hätönen H, Välimäki M. Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: a systematic review and meta-analysis. Int J Nurs Stud 2014 Jan;51(1):136-149. [doi: 10.1016/j.ijnurstu.2012.12.017] [Medline: 23384695]
27. Means B, Toyama Y, Murphy R, Bakia M, Jones K. Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies. Washington, DC: US Department of Education; 2009. URL: https://www.sri.com/sites/default/files/publications/imports/EvaluationEvidenceBasedPracticeOnlineLearning.pdf [accessed 2016-08-15] [WebCite Cache ID 6jmdGWiw9]
28. Cheng Y. The effects of information systems quality on nurses' acceptance of the electronic learning system. J Nurs Res 2012 Mar;20(1):19-30. [doi: 10.1097/JNR.0b013e31824777aa] [Medline: 22333963]
29. 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]
30. Cheng B, Wang M, Yang SJ, Kinshuk M, Peng J. Acceptance of competency-based workplace e-learning systems: effects of individual and peer learning support. Comput Educ 2011 Aug;57(1):1317-1333. [doi: 10.1016/j.compedu.2011.01.018]
31. Nesterowicz K, Librowski T, Edelbring S. Validating e-learning in continuing pharmacy education: user acceptance and knowledge change. BMC Med Educ 2014 Feb;14(33):1-7 [FREE Full text] [doi: 10.1186/1472-6920-14-33] [Medline: 24528547]
32. Chow M, Herold DK, Choo T, Chan K. Extending the technology acceptance model to explore the intention to use Second Life for enhancing healthcare education. Comput Educ 2012 Dec;59(4):1136-1144. [doi: 10.1016/j.compedu.2012.05.011]
33. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478.
34. Admiraal W, Lockhorst D. E-Learning in small and medium-sized enterprises across Europe: Attitudes towards technology, learning and training. Int Small Bus J 2009 Dec 22;27(6):743-767. [doi: 10.1177/0266242609344244]
35. Bloomfield JG, Jones A. Using e-learning to support clinical skills acquisition: exploring the experiences and perceptions of graduate first-year pre-registration nursing students - a mixed method study. Nurse Educ Today 2013 Dec;33(12):1605-1611. [doi: 10.1016/j.nedt.2013.01.024] [Medline: 23473860]
36. Hartland W, Biddle C, Fallacaro M. Audiovisual facilitation of clinical knowledge: a paradigm for dispersed student education based on Paivio's Dual Coding Theory. AANA J 2008 Jun;76(3):194-198. [Medline: 18567323]
37. Zhang D, Zhou L, Briggs RO, Nunamaker JF. Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness. Inform Manage 2006 Jan;43(1):15-27. [doi: 10.1016/j.im.2005.01.004]

38. Kelly M, Lyng C, McGrath M, Cannon G. A multi-method study to determine the effectiveness of, and student attitudes to, online instructional videos for teaching clinical nursing skills. Nurse Educ Today 2009 Apr;29(3):292-300. [doi: 10.1016/j.nedt.2008.09.004] [Medline: 19121875]
39. Sharpnack PA, Goliat L, Baker JR, Rogers K, Shockey P. Thinking like a nurse: using video simulation to rehearse for professional practice. Clin Simulation Nurs 2013 Dec;9(12):e571-e577. [doi: 10.1016/j.ecns.2013.05.004]
40. Cardoso AF, Moreli L, Braga FT, Vasques CI, Santos CB, Carvalho EC. Effect of a video on developing skills in undergraduate nursing students for the management of totally implantable central venous access ports. Nurse Educ Today 2012 Aug;32(6):709-713. [doi: 10.1016/j.nedt.2011.09.012] [Medline: 22036273]
41. 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]
42. Chan A, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. Lancet 2013 Jan 12;381(9861):91-92. [doi: 10.1016/S0140-6736(12)62160-6] [Medline: 23305999]
43. Feeley N, Cossette S. Pilot studies for randomized clinical trials. In: Henly S, editor. Routledge International Handbook of Advanced Quantitative Methods in Nursing Research. New York, NY: Routledge; 2015.
44. Paradis V, Cossette S, Frasure-Smith N, Heppell S, Guertin M. The efficacy of a motivational nursing intervention based on the stages of change on self-care in heart failure patients. J Cardiovasc Nurs 2010;25(2):130-141. [doi: 10.1097/JCN.0b013e3181c52497] [Medline: 20168193]
45. Miller WR, Moyers TB. Eight stages in learning motivational interviewing. J Teaching Addict 2006 Jan;5(1):3-17. [doi: 10.1300/J188v05n01_02]
46. Paradis V. Évaluation d'une Intervention Infirmière Motivationnelle Selon les Stades de Changement (MSSC) sur des Comportements d'Auto-soins Chez des Patients Insuffisants Cardiaques [master's thesis]. Montreal, QC: Université de Montréal; 2009.
47. Bédard J. 3-Minute Empowerment.: Jacques Bédard, MD; 2016. URL: http://www.jbedardmd.com/files/3_Minutes_Empowerment.pdf [accessed 2016-08-13] [WebCite Cache ID 6jjbSCowU]
48. Schwalbe CS, Oh HY, Zweben A. Sustaining motivational interviewing: a meta-analysis of training studies. Addiction 2014 Aug;109(8):1287-1294. [doi: 10.1111/add.12558] [Medline: 24661345]
49. DeLone W, McLean E. The DeLone and McLean model of information systems success: a ten-year update. J Manage Inform Syst 2003;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
50. van der Heijden H. User acceptance of hedonic information systems. MIS Q 2004;28(4):695-704.
51. World Health Organization. Process of Translation and Adaptation of Instruments. Geneva: WHO; 2016. URL: http://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2016-08-13] [WebCite Cache ID 6jjbXisyt]
52. Cossette S, Cara C, Ricard N, Pépin J. Development and psychometric evaluation of a scale to assess patient-nurse interactions with a caring worldview. Int J Nurs Stud 2005;42:673-686.
53. Mitchell S, Heyden R, Heyden N, Schroy P, Andrew S, Sadikova E, et al. A pilot study of motivational interviewing training in a virtual world. J Med Internet Res 2011;13(3):e77 [FREE Full text] [doi: 10.2196/jmir.1825] [Medline: 21946183]
54. Butler CC, Simpson SA, Hood K, Cohen D, Pickles T, Spanou C, et al. Training practitioners to deliver opportunistic multiple behaviour change counselling in primary care: a cluster randomised trial. BMJ 2013;346:1-25 [FREE Full text] [Medline: 23512758]
55. Welch J. Building a foundation for brief motivational interviewing: communication to promote health literacy and behavior change. J Contin Educ Nurs 2014 Dec;45(12):566-572. [doi: 10.3928/00220124-20141120-03] [Medline: 25401341]
56. Kaltman S, WinklerPrins V, Serrano A, Talisman N. Enhancing motivational interviewing training in a family medicine clerkship. Teach Learn Med 2015;27(1):80-84. [doi: 10.1080/10401334.2014.979179] [Medline: 25584475]
57. Eley R, Fallon T, Soar J, Buikstra E, Hegney D. The status of training and education in information and computer technology of Australian nurses: a national survey. J Clin Nurs 2008 Oct;17(20):2758-2767. [doi: 10.1111/j.1365-2702.2008.02285.x] [Medline: 18808645]
58. Maag MM. Nursing students' attitudes toward technology. Nurse Educ 2006;31(3):112-118. [doi: 10.1097/00006223-200605000-00007]
59. Moule P, Ward R, Lockyer L. Nursing and healthcare students' experiences and use of e-learning in higher education. J Adv Nurs 2010 Dec;66(12):2785-2795. [doi: 10.1111/j.1365-2648.2010.05453.x] [Medline: 20946565]
60. Hung J. Trends of e-learning research from 2000 to 2008: use of text mining and bibliometrics. Br J Educ Technol 2012;43(1):5-16. [doi: 10.1111/j.1467-8535.2010.01144.x]
61. Chang TP, Pham PK, Sobolewski B, Doughty CB, Jamal N, Kwan KY, et al. Pediatric emergency medicine asynchronous e-learning: a multicenter randomized controlled Solomon four-group study. Acad Emerg Med 2014 Aug;21(8):912-919 [FREE Full text] [doi: 10.1111/acem.12434] [Medline: 25154469]
62. Stansfeld SA, Kerry S, Chandola T, Russell J, Berney L, Hounsome N, et al. Pilot study of a cluster randomised trial of a guided e-learning health promotion intervention for managers based on management standards for the improvement of employee well-being and reduction of sickness absence: GEM Study. BMJ Open 2015;5(10):e007981 [FREE Full text] [doi: 10.1136/bmjopen-2015-007981] [Medline: 26503383]

63. Nicastro E, Lo Vecchio A, Liguoro I, Chmielewska A, De Bruyn C, Dolinsek J, et al. The impact of e-learning on adherence to guidelines for acute gastroenteritis: a single-arm intervention study. PLoS One 2015;10(7):e0132213 [FREE Full text] [doi: 10.1371/journal.pone.0132213] [Medline: 26148301]
64. Lahti ME, Kontio RM, Välimäki M. Impact of an e-learning course on clinical oractice in osychiatric hospitals: Nurse managers' views. Perspect Psychiatr Care 2016 Jan;52(1):40-48. [doi: 10.1111/ppc.12100] [Medline: 25624098]
65. Sinclair PM, Kable A, Levett-Jones T, Booth D. The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: A systematic review. Int J Nurs Stud 2016 May;57:70-81. [doi: 10.1016/j.ijnurstu.2016.01.011] [Medline: 27045566]
66. van de Steeg L, IJkema R, Wagner C, Langelaan M. The effect of an e-learning course on nursing staff's knowledge of delirium: a before-and-after study. BMC Med Educ 2015;15:12 [FREE Full text] [doi: 10.1186/s12909-015-0289-2] [Medline: 25653115]
67. Van Lancker A, Baldewijns K, Verhaeghe R, Robays H, Buyle F, Colman R, et al. The effectiveness of an e-learning course on medication calculation in nursing students: a clustered quasi-experimental study. J Adv Nurs 2016 Sep;72(9):2054-2064. [doi: 10.1111/jan.12967] [Medline: 27060466]
68. van de Steeg L, IJkema R, Langelaan M, Wagner C. Can an e-learning course improve nursing care for older people at risk of delirium: a stepped wedge cluster randomised trial. BMC Geriatr 2014;14:69 [FREE Full text] [doi: 10.1186/1471-2318-14-69] [Medline: 24884739]
69. Ebben RH, van Grunsven PM, Moors ML, Aldenhoven P, de Vaan VJ, van Hout R, et al. A tailored e-learning program to improve handover in the chain of emergency care: a pre-test post-test study. Scand J Trauma Resusc Emerg Med 2015;23(1):1-11 [FREE Full text] [doi: 10.1186/s13049-015-0113-3] [Medline: 25887239]

Abbreviations

CCU: coronary care unit
CNP: cardiology nurse practitioner
MI: motivational interviewing
S1: first session of MOTIV@CŒUR
S2: second session of MOTIV@CŒUR

Edited by G Eysenbach; submitted 30.06.16; peer-reviewed by J Welch, J Apolinário-Hagen; comments to author 27.07.16; revised version received 03.08.16; accepted 08.08.16; published 18.08.16.

Please cite as:
Fontaine G, Cossette S, Heppell S, Boyer L, Mailhot T, Simard MJ, Tanguay JF
Evaluation of a Web-Based E-Learning Platform for Brief Motivational Interviewing by Nurses in Cardiovascular Care: A Pilot Study
J Med Internet Res 2016;18(8):e224
URL: http://www.jmir.org/2016/8/e224/
doi:10.2196/jmir.6298
PMID:27539960

Original Paper

A Multirelational Social Network Analysis of an Online Health Community for Smoking Cessation

Kang Zhao[1], PhD; Xi Wang[2], MSc; Sarah Cha[3], MSPH; Amy M Cohn[3,4], PhD; George D Papandonatos[5], PhD; Michael S Amato[3], PhD; Jennifer L Pearson[3,6], PhD; Amanda L Graham[3,4], PhD

[1]Department of Management Sciences, Tippie College of Business, The University of Iowa, Iowa City, IA, United States

[2]Interdisciplinary Graduate Program in Informatics, The University of Iowa, Iowa City, IA, United States

[3]Schroeder Institute for Tobacco Research and Policy Studies, Truth Initiative, Washington, DC, United States

[4]Department of Oncology, Georgetown University Medical Center / Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, United States

[5]Center for Statistical Sciences, Brown University, Providence, RI, United States

[6]Department of Health, Behavior, and Society, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Kang Zhao, PhD
Department of Management Sciences
Tippie College of Business
The University of Iowa
S224 PBB
Iowa City, IA, 52242
United States
Phone: 1 3193353831
Fax: 1 319 335 0297
Email: kang-zhao@uiowa.edu

Abstract

Background: Online health communities (OHCs) provide a convenient and commonly used way for people to connect around shared health experiences, exchange information, and receive social support. Users often interact with peers via multiple communication methods, forming a multirelational social network. Use of OHCs is common among smokers, but to date, there have been no studies on users' online interactions via different means of online communications and how such interactions are related to smoking cessation. Such information can be retrieved in multirelational social networks and could be useful in the design and management of OHCs.

Objective: To examine the social network structure of an OHC for smoking cessation using a multirelational approach, and to explore links between subnetwork position (ie, centrality) and smoking abstinence.

Methods: We used NetworkX to construct 4 subnetworks based on users' interactions via blogs, group discussions, message boards, and private messages. We illustrated topological properties of each subnetwork, including its degree distribution, density, and connectedness, and compared similarities among these subnetworks by correlating node centrality and measuring edge overlap. We also investigated coevolution dynamics of this multirelational network by analyzing tie formation sequences across subnetworks. In a subset of users who participated in a randomized, smoking cessation treatment trial, we conducted user profiling based on users' centralities in the 4 subnetworks and identified user groups using clustering techniques. We further examined 30-day smoking abstinence at 3 months postenrollment in relation to users' centralities in the 4 subnetworks.

Results: The 4 subnetworks have different topological characteristics, with message board having the most nodes (36,536) and group discussion having the highest network density (4.35×10^{-3}). Blog and message board subnetworks had the most similar structures with an in-degree correlation of .45, out-degree correlation of .55, and Jaccard coefficient of .23 for edge overlap. A new tie in the group discussion subnetwork had the lowest probability of triggering subsequent ties among the same two users in other subnetworks: 6.33% (54,142/855,893) for 2-tie sequences and 2.13% (18,207/855,893) for 3-tie sequences. Users' centralities varied across the 4 subnetworks. Among a subset of users enrolled in a randomized trial, those with higher centralities across subnetworks generally had higher abstinence rates, although high centrality in the group discussion subnetwork was not associated with higher abstinence rates.

Conclusions: A multirelational approach revealed insights that could not be obtained by analyzing the aggregated network alone, such as the ineffectiveness of group discussions in triggering social ties of other types, the advantage of blogs, message boards, and private messages in leading to subsequent social ties of other types, and the weak connection between one's centrality in the group discussion subnetwork and smoking abstinence. These insights have implications for the design and management of online social networks for smoking cessation.

(J Med Internet Res 2016;18(8):e233) doi:10.2196/jmir.5985

KEYWORDS

social networks; smoking cessation; community networks

Introduction

Over the past decade, many people have turned to the Internet to find health-related information and support. According to the Pew Research Center, 72% of adult Internet users in the United States use the Internet for health-related purposes. Of those, 26% have read or watched someone else's experience about health or medical issues in the last 12 months and 16% have used the Internet to find others who might share the same health concerns in the last year [1]. Interactions with peers who share similar health problems are facilitated by online health communities (OHCs), which are Internet-based online groups or websites specifically designed for both patients and caregivers to learn about an illness, seek and offer support, and connect with others in similar circumstances [2]. Online health communities enable individuals to connect via forums, discussion boards, private messages, and other forms of synchronous and asynchronous social interaction. In addition to their popularity, the physical and psychological benefits of participation in OHCs have been well documented in numerous studies (eg, [3-6]). Given the proliferation and popularity of OHCs, it is important to understand the experiences and behaviors of users in these network contexts so that the design and management of OHCs can be improved or optimized.

Various aspects of OHCs have been studied, such as topics of online discussions [7-11], the nature and exchange of various types of social support [12-15], users' participation patterns [16-19], and the psychological mechanisms through which participation affects health outcomes [20-22]. Less studied have been the social network structures of OHCs and the role of network characteristics in understanding individual user patterns and outcomes. Social network analyses can help identify community structures at the network level (ie, considering the entire network), as well as individual behaviors and positions at the individual level (ie, considering individuals and their ties with peers). To date, social network analyses of OHCs have largely focused on social networks based on a single type of communication (eg, posting comments to threaded discussions [23,24]) or have aggregated different types of communications into one network [25]. However, most social networks—both online and offline—are multirelational (also called multiplex or multidimensional networks), composed of myriad social relationships with family members, neighbors, classmates, colleagues, etc [26]. In OHCs, users' communications via different channels foster different types of social relations or ties. For example, private message ties may be more intimate and influential than ties formed based on the exchange of messages in a group discussion. Multirelational analyses of social networks have provided important new insights into information flows, individual centralities, growth models, link prediction, and community discoveries. For example, social ties based on one type of relationship can predict the formation of ties based on another type of relationship [26]. Differentiating social ties based on different relationships can also contribute to the prediction of individuals' preferences [27,28].

This study adopted a multirelational perspective in examining the network structure and dynamics of a popular OHC for smoking cessation. We examined the structure of the social network, as well as the coevolution of different types of subnetworks. Numerous publications highlight the importance of social influences on a range of smoking behaviors, including initiation, cessation, and relapse, in offline settings. Thus, we also illustrated how users' behavior patterns in different subnetworks were related to their smoking status, using outcome data available for a subset of OHC members enrolled in a randomized trial. Our primary goals were to characterize multirelational social networks in an OHC for smoking cessation, identify dynamic coevolution of multirelational networks, and explore potential links between users' online social network engagement and health behavior using a multirelational approach. To our knowledge, this study is the first to analyze large-scale multirelational social networks among OHC users of a Web-based, smoking cessation program. Furthermore, while previous studies have enumerated social networks based only on users' posting behaviors, our multirelational social network incorporated private behaviors as well by considering both posting and reading behaviors of users. This study lays the foundation for an ongoing series of analyses aimed at understanding and optimizing the multirelational behaviors of a large OHC for smoking cessation.

Methods

Intervention

We conducted these analyses using longitudinal data from BecomeAnEX, a Web-based smoking cessation program developed and managed by Truth Initiative (formerly American Legacy Foundation). Launched in 2008, BecomeAnEX was developed in accordance with the Clinical Practice Guidelines for Treating Tobacco Use and Dependence [29]. Through an interactive, multimedia experience, BecomeAnEX assists users in setting a quit date, understanding their smoking habits and preparing to quit, selecting and using Food and Drug Administration–approved medications, and connecting with others for social support in the BecomeAnEX community. A

national mass media campaign [30] and ongoing Web-based advertising have resulted in more than 700,000 registered users since its inception.

The BecomeAnEX community is composed of thousands of current and former smokers who interact via 4 primary communication channels. Users can exchange private messages via the site; users who have opted-in to receive email notifications are informed when they have received a new message. Message board posts are public communications made on a member's profile page. All users have a community profile that can be customized with photos and personal information. Group discussions are threaded discussions among users with similar experiences or interests (eg, "March Quit Dates," "Over 50 BecomeAnEXs"). Blogs are single entries made by users about their experiences, which appear in reverse chronological order on the site. Users can comment on others' blog posts, creating threaded discussions similar to group discussions. Communication between and communication among members via blogs (and comments), message boards, and group discussions are all public communications that can be accessed by all BecomeAnEX users. Private messages occur only between two users. Blogs and group discussions elicit many-to-many communications, whereas posts on message boards and private messages are one-to-one communications. A community administrator addresses technical issues and spammers, but otherwise the community is largely unmoderated.

All user actions are date and time stamped and stored in a relational database. Before analysis, users' identifiers were converted into alphanumeric strings using cryptographic hash functions, which makes this conversion infeasible to invert. The content of private messages was not included in the dataset to protect privacy.

Multirelational Social Network Analyses

The Python programming package NetworkX (v1.11) was used to construct and analyze social networks. The multirelational network consists of 4 subnetworks: private messages (PM), message boards (MB), group discussions (GD), and blogs (BL). In each subnetwork, a node represents an individual user, while a directed tie pointing from user A to user B means that B accessed information contributed by A or, in other words, information from A reached B. Taking the blog subnetwork as an example, if B posted a comment to one of A's blogs then we assume B read (or at least skimmed) the original blog post, and so there is a tie pointing from A to B (A→B) indicating that A's contribution has reached B. Similarly, if A's clickstream (ie, the logs of clicking URLs) suggests that he or she has read that comment from B, then we add a B→A tie to reciprocate the earlier A→B tie. In such a directed network, a node's in-degree refers to the number of other nodes that have ties pointing to it (ie, the number of people who may have influenced that user). Conversely, a node's out-degree is the number of its outgoing ties (ie, the number of people that user has potentially influenced). A node's total degree is the total number of its network neighbors irrespective of tie direction. By incorporating both posting (outgoing ties if a post was read by others) and reading (incoming ties) behaviors, our subnetworks can better capture how information flows among OHC users via each means of communication. When combining all nodes and ties in the 4 subnetworks, an aggregated network emerges, where a tie means two users have had some type of interaction in the community.

Our analysis proceeded in 4 steps. First, we conducted topological analysis to illustrate the characteristics of the 4 subnetworks. We examined the number of nodes with total degree greater than zero, the number of edges, density (defined as the number of actual ties divided by the number of possible ties), and the proportion of ties that were reciprocated. To compare the connectedness of the subnetworks, we identified the largest strongly connected component (LSCC). A strongly connected component is a subset of a network, in which there is a directed path between every pair of nodes. The LSCC is the one with the most nodes among all strongly connected components of a network. For each subnetwork, we also calculated the average shortest path among nodes in its LSCC. In general, the larger the LSCC and the shorter the average path length within the LSCC, the more connected the network.

Second, we measured structural similarities among the subnetworks using 2 metrics: centrality correlations at the individual level and tie overlap at the network level. At the individual level, one's centrality can be captured by in- and out-degrees. Higher degrees usually mean higher centralities. We correlated each node's rank by in- and out-degrees in one subnetwork with the same node's rank by in- and out-degrees in the other 3 subnetworks. A high correlation coefficient between two subnetworks suggests that individuals with high centrality in one subnetwork tend to have high centrality in another. At the network level, the tie overlap between two subnetworks was calculated with Jaccard coefficients [31]. A high Jaccard coefficient between two subnetworks signals that if there is a tie from node *i* to node *j* in one subnetwork, there is a high probability that a tie also exists from *i* to *j* in another subnetwork.

Third, coevolution analysis was used to demonstrate tie formation dynamics across subnetworks. Building on analyses of the static characteristics (ie, topology) and structural similarities of the subnetworks, we also investigated coevolution dynamics between the 4 subnetworks. We were specifically interested in how the formation of a tie between two users in one subnetwork triggered the formation of ties between the same two users in other subnetworks. For each subnetwork, we calculated the probability that this subnetwork hosts the first tie among all pairs of nodes that were connected in any of the 4 subnetworks. We also investigated whether the same pair of nodes that formed their first tie in one of the subnetworks would form new ties in other subnetworks. To answer this question, we analyzed the temporal sequence of tie formations, and calculated the probabilities to form subsequent ties in the second and third subnetworks given the subnetwork in which the first tie was formed, along with the most common tie sequences.

Finally, user profiling was used to identify whether centralities in different subnetworks had different implications for abstinence rates. We used Gaussian mixture models (GMMs), an unsupervised clustering technique, to divide users into groups based on their centralities in the 4 subnetworks so that those

with similar centralities across subnetworks were placed in the same group. As the input for the profiling process, each user is represented by a vector with 8 elements, each one being the user's in- and out-degree in the 4 subnetworks. To determine the number of user groups (K), we tried different K values (from 2 to 10) for GMM and selected the value that represented the best fit with our data as determined by log-likelihood.

The user profiling analysis was based on a subsample of N=1337 BecomeAnEX users who participated in a randomized smoking cessation trial (NCT01544153) and were assigned to the control arm (BecomeAnEX alone). The trial has been described in detail elsewhere [32]. All participants were current smokers at baseline; 30-day point prevalence abstinence was assessed at 3 months after enrollment ("In the past 30 days, have you smoked any cigarettes at all, even a puff?"). The overall response rate for the trial at 3 months was 58.41% (781/1337). Users who did not complete the follow-up survey were conservatively counted as smoking under the intent-to-treat principle. Of the 1337 BecomeAnEX users in this sample, 12.27% (164/1337) reported 30-day point prevalence abstinence at 3 months. Differences in abstinence rates between the user groups identified in the GMMs described above were examined using analysis of variance.

The study protocol was approved by Chesapeake Institutional Review Board (protocol #CR00040526).

Results

Description of Dataset

The dataset used in this study spanned the period from January 1, 2010, to May 31, 2015, and included records of both posting and reading behaviors of N=71,251 users who accessed content of the community on BecomeAnEX by clicking and reading a post (eg, a blog, a message board post, or a group discussion thread) or a private message. The community was migrated from a different platform before this period, which resulted in a slightly different user experience. Our analyses focus on this time frame given the stability of the social network feature set.

Topological Analysis

Figure 1 shows distributions of total degrees in the aggregated network (part A), in-degrees (part B), and out-degrees (part C) for the 4 subnetworks. The distribution of total degrees in the aggregated network was similar to the power-law degree distribution that is typical for a scale-free network. In a power-law degree distribution, the probability that a node has degree k follows $P(k)=c \times k^{-r}$, where c and r are network-specific constants. In a log-log plot, a power-law degree distribution features a downward-sloping straight line that is similar to Figure 1, part A. However, the in- and out-degree distributions of the 4 subnetworks suggested that each subnetwork, in fact, had different topological characteristics. The private message subnetwork featured power-law distributions for both in- and out-degrees, but the other 3 were hardly scale-free networks as their curves in log-log plots were nonlinear. For example, the blog and group discussion subnetworks had relatively flat distributions for low in- and out-degrees. On the one hand, blog and group discussion distributions conformed to the generally observed pattern among scale-free networks that nodes with higher degrees appear less frequently. On the other hand, the message board and the group discussion subnetworks featured sudden increases in the number of nodes with in-degree around 10 and 18, respectively. Additionally, there were more users with zero out-degrees than those with zero in-degrees, because many users only read community content without contributing and thus had no outgoing ties.

Descriptive statistics of the aggregated network and each of the 4 subnetworks are presented in Table 1. Among the 4 subnetworks, the message board subnetwork had the most nodes, followed by the private message subnetwork. However, the private message subnetwork also had the lowest density. The high number of nodes with nonzero degree in the private message subnetwork was attributable to many nodes with in-degree of 1. The presence of welcome messages was also reflected by the low reciprocity of the private message subnetwork: only 8.94% (4970/55,585) of the ties were reciprocated. By contrast, even though its number of nodes ranked only third among the 4, the blog subnetwork had the shortest average path length in the LSCC, the second highest density, and the second highest reciprocity rate, indicating a well-connected network in which people actively interacted with each other. Among the 4 subnetworks, the private message subnetwork was the least connected with the smallest LSCC (6.87% (2404/34,996) of nodes in the LSCC) and the longest average shortest path in the LSCC (3.74).

Structural Similarity

The topological analysis described in the previous section treated each subnetwork as independent. However, two individuals may be connected in more than one subnetwork in the online community. We computed how many pairs of nodes were connected in different subnetworks. As shown in Table 2, although many pairs of nodes were connected in only 1 subnetwork, there were still more than 370,000 pairs of nodes that were connected in 2 or more subnetworks.

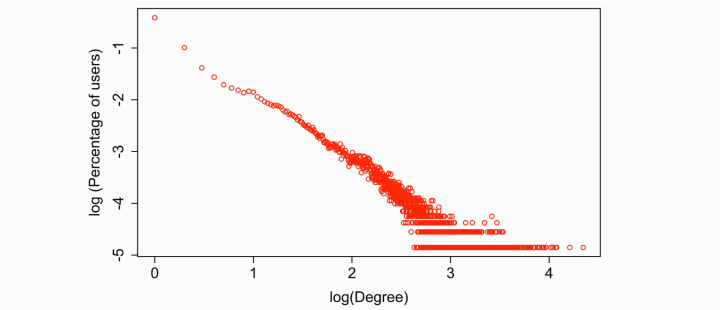
As shown in Tables 3 and 4, the blog and message board subnetworks had the most similar topologies. They had the highest correlation in node centralities (ρ=.45), as well as the top Jaccard coefficient (.23) that was at least 4 times higher than the others. Meanwhile, the private message subnetwork was quite different from the other 3 subnetworks. Although those with high out-degree in the private message subnetwork also tended to have high out-degree in other subnetworks (with moderate correlations), in-degree in the private message subnetwork was negatively correlated with in-degrees in other subnetworks. Active contributors in the other 3 subnetworks tended to send messages to more people, but those who received messages from more people did not necessarily read more posts from others.

Table 1. Descriptive statistics of the aggregated network and the 4 subnetworks.

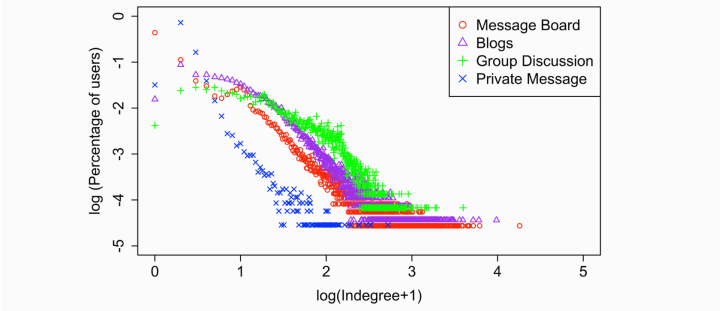
Characteristics	Aggregated	Blog	Message board	Group discussion	Private message
Number of nodes with degree >0	71,251	27,461	36,536	14,827	34,996
Number of edges	2,578,659	1,065,514	1,027,694	956,506	60,555
Density	5.08×10^{-4}	1.41×10^{-3}	7.70×10^{-4}	4.35×10^{-3}	4.94×10^{-5}
% Of reciprocated ties	18.22 (397,339/2,181,320)	23.61 (203,485/862,029)	29.62 (234,873/792,821)	3.57 (32,928/923,578)	8.94 (4970/55,585)
% Of nodes in the LSCC[a]	35.64 (25,395/71,251)	35.00 (9611/27,461)	41.61 (15,203/36,536)	26.48 (3926/14,827)	6.87 (2404/34,996)
Average shortest path length in LSCC	2.86	2.29	2.68	2.40	3.74

[a]LSCC: largest strongly connected component.

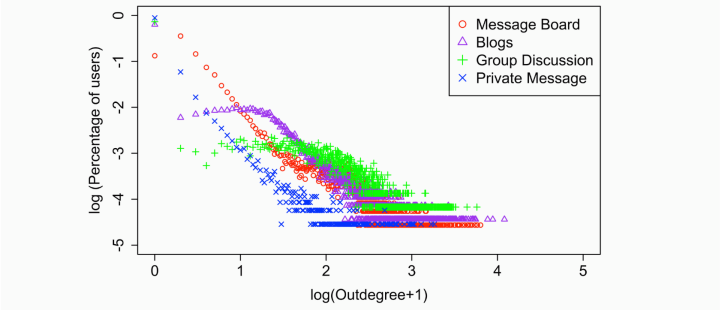
Figure 1. Network degree distributions for the aggregated network and 4 subnetworks.



(a) Distribution of total degrees in the aggregated network.



(b) Distribution of in-degrees for the four sub-networks.



(c) Distributions of out-degrees for the four sub-networks.

Table 2. The number of node pairs with ties in different networks.

Node pairs	Number of node pairs
Pairs connected in 1 subnetwork only	1,807,720
Pairs connected in 2 subnetworks	300,758
Pairs connected in 3 subnetworks	66,591
Pairs connected in 4 subnetworks	6251

Table 3. Spearman rank correlation coefficients between individual nodes' in-degree (above the diagonal) and out-degree (below the diagonal) across the 4 subnetworks.

Network	Blog	Message board	Group discussion	Private message
Blog	—	.45[a]	.35[a]	−.23[a]
Message board	.55[a]	—	.40[a]	−.10[a]
Group discussion	.33[a]	.32[a]	—	−.10[a]
Private message	.43[a]	.35[a]	.35[a]	—

[a]$P<.001$.

Table 4. Tie overlap measured by Jaccard coefficients between the 4 subnetworks.

Subnetwork	Blog	Message board	Group discussion	Private message
Blog	—	0.23	0.05	0.02
Message board	—	—	0.05	0.02
Group discussion	—	—	—	0.01

The Coevolution of Multirelational Networks

As shown in Table 5, the largest proportion of first ties (39.24% (855,893/2,181,320)) occurred in the group discussion subnetwork, 33.67% (734,559/2,181,320) occurred in the blog subnetwork, 28.30% (617,287/2,181,320) occurred in the message board subnetwork, and only 1.87% (40,728/2,181,320) occurred in the private message subnetwork, which had the fewest edges. Among those who formed their first ties in the blog subnetwork, 27.22% (199,913/734,559) formed their second ties in another subnetwork, most commonly the message board subnetwork as BL→MB is the most frequent 2-tie sequence. Also, for pairs that were first connected in the blog subnetwork, only 4.37% (32,126/734,559) were eventually connected via a third tie, with BL→MB→GD being the most frequent 3-tie sequence. Comparing the conditional probabilities of forming the second tie given the first tie in each subnetwork, we found that a first tie in the blog and message board subnetworks had similar probabilities of leading to a second and third tie in other subnetworks. By contrast, first ties in the group discussion subnetwork had the lowest probability of developing into subsequent ties.

Table 5. Probabilities (P) of subnetworks to host the first tie between two nodes, conditional probabilities of subsequent ties in other subnetworks, and top tie sequences.

Subnetwork hosting the 1st tie	P(hosting the 1st tie), %	P(forming 2nd ties in other subnetwork | 1st tie), %	Top 2-tie sequence by P(sequence | 1st tie), %	P(forming 3rd ties in other subnetwork | 1st tie), %	Top 3-tie sequence by P(sequence | 1st tie), %
Blog	33.67 (734,559/2,181,320)	27.22 (199,913/734,559)	BL[a]→MB[b] 86.03 (135,245/157,212)	4.37 (32,126/734,559)	BL→MB→GD[c] 9.23 (14,517/157,212)
Message board	28.30 (617,287/2,181,320)	28.52 (176,031/617,287)	MB→BL 79.52 (96,418/121,244)	5.10 (31,492/617,287)	MB→BL→GD 11.47 (13,909/121,244)
Group discussion	39.24 (855,893/2,181,320)	6.33 (54,142/855,893)	GD→MB 58.43 (27,795/45,573)	2.13 (18,207/855,893)	GD→BL→MB 14.47 (6,882/47,573)
Private message	1.87 (40,728/2,181,320)	26.05 (10,611/40,728)	PM[d]→BL 44.2 (378/855)	3.65 (1487/40,728)	PM→MB→BL 5.1 (44/855)

[a]BL: blog.

[b]MB: message board.

[c]GD: group discussion.

[d]PM: private message.

User Profiling and Abstinence

Gaussian mixture model with K=7 generated user groups that fit our data the best—the log-likelihood reached a plateau when K=7. Adding more clusters only increased the likelihood by 0.4%-4% (K=8, 9, and 10), but lower K values (K=2 to 6) reduced the likelihood by 16%-61%. Table 6 lists the average centrality (in- and out-degrees in the 4 subnetworks) of each of the 7 groups, along with the number of users and 30-day point prevalence abstinence (ppa) rates at 3 months for each user group. Groups are sorted from the highest to the lowest abstinence rates.

Table 6. User groups and their average in- and out-degrees in 4 subnetworks.

User group	MB[a] In	MB Out	BL[b] In	BL Out	GD[c] In	GD Out	PM[d] In	PM Out	No. of users	30-Day ppa at 3 months[e], %
1. Super users	118.8	150.1	176.9	183.7	118.9	30.2	6.3	6.3	18	55.6 (10/18)
2. Regular contributors	8.8	4.5	17.8	9.4	51.5	95.0	0.5	0.0	13	38.5 (5/13)
3. Regular contributors	11.1	19.1	24.8	25.8	17.8	0.0	0.7	0.3	88	30.7 (27/88)
4. Lurkers	3.9	1.0	12.6	0.0	56.4	0.0	0.4	0.0	68	14.7 (10/68)
5. Lurkers	3.0	0.7	14.4	0.0	0.0	0.0	0.4	0.0	118	14.4 (17/118)
6. Inactive users	0.0	0.9	0.0	0.0	0.0	0.0	1.2	0.0	210	9.5 (20/210)
7. Inactive users	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	822	9.1 (75/822)

[a]MB: message board.

[b]BL: blog.

[c]GD: group discussion.

[d]PM: private message.

[e]The 30-day point prevalence abstinence (ppa) at 3 months calculated under intent-to-treat principle with nonresponders counted as smokers.

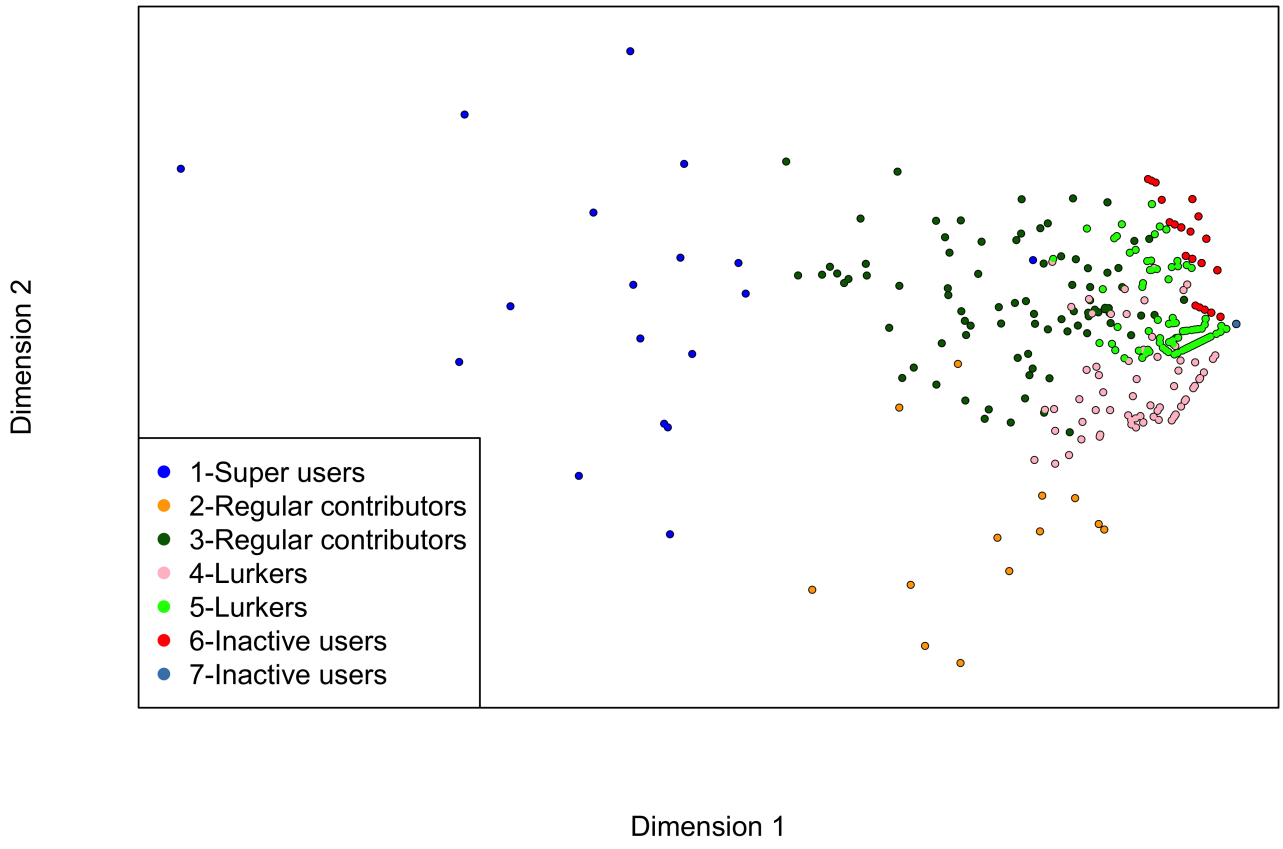
Users in group 1 were highly connected users with many incoming and outgoing ties across the 4 subnetworks. Groups 2 and 3 represented regular contributors who not only read what others posted, but also contributed content that was read by others, although they were less connected than those in group 1. Groups 4 and 5 were “lurkers” who mainly read posts from others but contributed little or no content of their own. The largest 2 groups (groups 6 and 7) consisted of trial participants who never visited the BecomeAnEX community (but may have used other smoking cessation features or content on the website), although those in group 6 received private messages and visits to their message boards from an average of about 1 other user. Figure 2 shows the differences among the 7 groups of users after using multidimensional scaling to map the 8-dimensional data into 2-dimensional space.

The overall comparison between user groups found that high degree centralities were associated with high abstinence rates. For example, well-connected users in group 1 had significantly higher abstinence rates than regular contributors in group 3 ($F_{1,104}$=4.15, P=.04), lurkers in group 4 ($F_{1,84}$=15.38, P<.001) and group 5 ($F_{1,134}$=18.66, P<.001), and isolated users in group 6 ($F_{1,226}$=35.22, P<.001) and group 7 ($F_{1,838}$=43.83, P<.001). Regular contributors in groups 2 and 3 also had significantly higher abstinence rates than lurkers (group 2 vs group 4: $F_{1,79}$=4.19, P=.04; group 2 vs group 5: $F_{1,129}$=4.96, P=.03; group 3 vs group 4: $F_{1,154}$=5.53, P=.02; group 3 vs group 5: $F_{1,204}$=8.19, P=.005) and inactive users (group 2 vs group 6: $F_{1,221}$=10.70, P<.001; group 2 vs group 7: $F_{1,833}$=12.88, P<.001; group 3 vs group 6: $F_{1,296}$=22.32, P<.001; group 3 vs group 7: $F_{1,908}$=38.61, P<.001). The robustness of these findings is supported by additional analyses that examined abstinence rates under a less-conservative, responder-only approach. Abstinence rates for groups 1 to 7 were 71.4% (10/14), 55.6% (5/9), 39.1% (27/69), 24.4% (10/41), 21.5% (17/79), 16.4% (20/122), and 16.8% (75/447), respectively. The rank order of the 7 groups is largely consistent, with the exception of group 7, which has a slightly high abstinence rate than group 6 under this analytic approach.

The multirelational network approach enabled the discovery of meaningful subgroups of participants, using information that would have been lost in an aggregated network analysis. For example, users in group 3 and group 4 had similar total degrees in the aggregated network (73.8 and 71.7, respectively). However, Table 6 reveals that members of group 3 were active across all subnetworks, whereas members of group 4 were active almost exclusively in the group discussion subnetwork. These patterns were significantly associated with abstinence as we showed in the previous paragraph ($F_{1,154}$=5.53, P=.02), suggesting a weak relationship of centrality in the group discussion subnetwork with abstinence.

In addition, the specific subnetwork in which users gained their centralities resulted in varying abstinence rates. For instance, having high in- and out-degrees in the group discussion subnetwork alone did not necessarily suggest high abstinence rates. Lurkers in group 4 had the second highest average in-degree in the group discussion subnetwork, but the abstinence rate in group 4 was not significantly different from that of otherwise similarly connected lurkers in group 5 ($F_{1,184}$=0.003, P=.96), or of isolated users in group 6 ($F_{1,276}$=1.43, P=.23) or group 7 ($F_{1,888}$=2.27, P=.13). Similarly, group 2 had the highest average out-degree in the group discussion subnetwork, yet its abstinence rate was not significantly different from users in group 3 ($F_{1,99}$=0.31, P=.58), who had much lower centralities in the group discussion subnetwork but higher centralities in the blog and message board subnetworks.

Figure 2. Multidimensional scaling of the 7 user groups.



Discussion

Principal Findings

To our knowledge, this study is the first to analyze a smoking cessation OHC from the perspective of a multirelational social network. We constructed 4 subnetworks based on users' interactions via 4 communication channels and illustrated the value of a multirelational approach through topological analysis, coevolution analysis, and user profiling analysis. We found that the subnetworks based on different types of relationships had different topological characteristics. Specifically, the blog subnetwork was the most connected. The blog and message board subnetworks were topologically similar, whereas the private message subnetwork was topologically distinct from others.

Coevolution analyses of subnetwork tie formation dynamics found that although the group discussion subnetwork was the most common subnetwork for the initial formation of ties between users, ties formed there also had the lowest probability of leading to additional ties in another subnetwork. This may have been because the many-to-many group-based interactions did not encourage relationship building at the dyadic (ie, one-to-one) level. By contrast, roughly a quarter of users who formed their first ties in one of the other subnetworks, including in the private message subnetwork, went on to form additional ties in a second subnetwork. When two BecomeAnEX users are first connected via private messages, it is likely to be via a welcome message from one member to another. Even though many such messages may be a mere formality, they do seem to encourage users to build more ties in other social networks, notably the blog and the message board subnetworks. However, because we did not use the content of private messages to protect users' privacy, we cannot directly validate whether these messages were indeed welcome messages.

User profiling based on users' centralities across the 4 subnetworks showed that users can have different centralities in different subnetworks. This further highlights the importance of examining subnetworks within OHCs. For example, although users with high centralities across all 4 subnetworks had high abstinence rates, aggregating these subnetworks into one network would have lost valuable information about users' online and offline behaviors. In other words, having high total degrees, high in-degrees, or high out-degrees in the aggregated network was not necessarily related to abstinence. Instead, our multirelational approach revealed that the subnetwork in which a user gained his or her centralities mattered.

Analyzing centrality with a multirelational approach is likely to be particularly useful for researchers and website designers interested in improving the effectiveness of OHCs as health interventions. This approach is capable of identifying which communication channels are facilitative of desired outcomes and which channels are not. We found that high centrality in the blog and message board subnetworks was positively associated with abstinence, whereas high centrality in the group discussion subnetwork was not. Recall that the group discussion subnetwork has the lowest reciprocity rate of 3.57% (32,928/923,578). Having a high degree in this subnetwork does not necessarily mean the user interacted or bonded with more peers in the community. These findings suggest that the group discussion feature may not be contributing to the health behavior change goals of the OHC and may be a candidate for revision to serve a more useful function or removal so as to avoid distracting new users from more active and/or effective communication channels. These insights would have been obscured with an aggregated network analysis; the multirelational approach allowed the signal of blog and message board centrality to be distinguished from the noise of group discussion centrality.

Implications

These findings shed light on users' online behaviors in a multirelational social network in an OHC for smoking cessation and inform community design or redesign, management, and interventions for smoking cessation and other health-risk behaviors using Web-based platforms for behavior change. For example, because the blog and the message board subnetworks were similar in structure and often triggered the formation of subsequent ties in each other, better integration of blogs and message boards may help users connect with each other more easily. Private messages can be a good way to welcome new users and encourage them to build more ties with peers using other means of communication, such as visiting message boards. Conversely, group discussions had the lowest probabilities of triggering subsequent ties in other subnetworks.

Comparison With Prior Work

Our observation that users with higher centralities had higher abstinence rates is consistent with previous research on the role of online social networks in smoking cessation. Two recent studies [33,34] demonstrated that smokers who participated in an online community—even just browsing or "lurking" the posts made by others—were more likely to be abstinent than those who did not participate at all in the community. These studies used statistical methods to account for the possibility of selection bias (ie, more active users of an OHC may be more motivated to make changes to their behavior), lending credence to causal links between online community engagement and smoking outcomes. Given the observational nature of the analyses in this paper, however, we cannot conclude that social network position, per se, is causally related to abstinence. Nevertheless, understanding more about behaviors within a social network highlights factors associated with positive outcomes. These factors could be harnessed in future interventions to improve longer-term cessation rates. Other studies have also identified the existence of key established members who have different roles within a smoking cessation network [19,35,36], but these studies have primarily focused on user behaviors or content of posts and have yet to link these behaviors to abstinence outcomes.

Although previous social network research has adopted the multirelational approach to study online social networks, the focus was mainly on traditional network analysis tasks, such as node ranking, link prediction, network evolution, and community discoveries [25,26,29-31,37,38]. Few have explored individual behaviors in the context of multirelational social networks, especially offline behaviors.

Limitations

This research has a few limitations. First, we showed that users with different roles based on their centralities in subnetworks can have different abstinence rates, but we cannot make causal statements regarding the links between centralities in certain subnetworks and abstinence. Second, the user profiling analysis was based only on a group of users who enrolled in a randomized trial. Third, we considered only the social network among users and did not incorporate the textual content of their interactions. This would be an interesting direction for future work to better understand what users shared and talked about in OHCs. Finally, we did not assess or examine other social influences that could affect smoking behaviors, such as family, friends, health care providers, and social media channels. It is important to determine whether and how these offline sources of social support interact with network dynamics that occur within OHCs for smoking cessation.

Future Research Directions

Directions for future work include investigating how information flows between nodes via different channels of communication. Topic modeling techniques can be used to capture what people talked about in each communication channel to model and predict the coevolution of multirelational social networks. The outcome of topic modeling also has the potential to reveal the evolution of users into specific self-assigned roles within an online community (eg, “Elder,” “Conflict Resolver”). Future work with this network will seek to identify content, communication strategies, and network connections that improve abstinence outcomes.

Conclusions

This study represents one of the first efforts to study the structure and dynamics of a large-scale OHC for smoking cessation. Specifically, user behavior patterns in the subnetworks were found to be differentially associated with important outcomes, including formation of subsequent ties to the network as well as abstinence from smoking. Whereas blogs, message boards, and private messages are effective in triggering subsequent social ties in other subnetworks, group discussions are not. Centralities in the group discussion subnetwork are not indicative of smoking outcome either. The results highlight the value of the multirelational approach in analyzing large-scale online social networks among OHC users. Our research also contributes to multirelational social network analysis by showing that multirelational network analysis of online ties can provide valuable insights for understanding individual health behaviors.

Acknowledgments

This work is supported by the National Cancer Institute of the National Institutes of Health (#R01 CA192345).

Conflicts of Interest

SC, AMC, MSA, JLP, and ALG are employees of Truth Initiative, which runs the BecomeAnEX smoking cessation website.

References

1. Fox S, Duggan M. pewinternet. 2013 Jan. Peer-to-Peer Health Care URL: http://www.pewinternet.org/2013/01/15/peer-to-peer-health-care/ [accessed 2016-05-08] [WebCite Cache ID 6hMEK6MAP]
2. Neal L, Oakley K, Lindgaard G, Kaufman D, Leimeister J, Selker T. Online health communities. In: Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2007 Presented at: Computer/Human Interaction 2007 Conference; April 28, 2007 - May 3, 2007; New York, NY p. 2129-2132.
3. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
4. Idriss SZ, Kvedar JC, Watson AJ. The role of online support communities: benefits of expanded social networks to patients with psoriasis. Arch Dermatol 2009 Jan;145(1):46-51. [doi: 10.1001/archdermatol.2008.529] [Medline: 19153342]
5. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
6. McKechnie V, Barker C, Stott J. The effectiveness of an Internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
7. Attard A, Coulson N. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Hum Behav 2012 Mar;28(2):500-506. [doi: 10.1016/j.chb.2011.10.022]
8. Burri M, Baujard V, Etter J. A qualitative analysis of an internet discussion forum for recent ex-smokers. Nicotine Tob Res 2006 Dec;8 Suppl 1:S13-S19. [Medline: 17491166]
9. Mo PK, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374. [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
10. Portier K, Greer G, Rokach L, Ofek N, Wang Y, Biyani P, et al. Understanding Topics and Sentiment in an Online Cancer Survivor Community. JNCI Monographs 2014 Jan 06;2013(47):195-198. [doi: 10.1093/jncimonographs/lgt025] [Medline: 24395991]

11. Selby P, van Mierlo T, Voci S, Parent D, Cunningham J. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
12. Bambina A. Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication. Youngstown, NY: Cambria Press; 2007.
13. Batenburg A, Das E. Emotional coping differences among breast cancer patients from an online support group: a cross-sectional study. J Med Internet Res 2014;16(2):e28 [FREE Full text] [doi: 10.2196/jmir.2831] [Medline: 24499687]
14. Bar-Lev S. "We are here to give you emotional support": performing emotions in an online HIV/AIDS support group. Qual Health Res 2008 Apr;18(4):509-521. [doi: 10.1177/1049732307311680] [Medline: 18192435]
15. Chuang K, Yang C. Informational support exchanges using different computer-mediated communication formats in a social media alcoholism community. J Assn Inf Sci Tec 2013 Oct 23;65(1):37-52. [doi: 10.1002/asi.22960]
16. 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]
17. Van't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
18. Wang X, Zhao K, Street WN. Social Support and User Engagement in Online Health Communities. : Springer; 2014 Presented at: Smart Health: International Conference, IC SH 2014; July 10-11, 2014; Beijing, China p. 97-110.
19. Healey B, Hoek J, Edwards R. Posting behaviour patterns in an online smoking cessation social network: implications for intervention design and development. PLoS One 2014;9(9):e106603 [FREE Full text] [doi: 10.1371/journal.pone.0106603] [Medline: 25192174]
20. Mo PK, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
21. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
22. Vambheim SM, Wangberg SC, Johnsen JK, Wynn R. Language use in an internet support group for smoking cessation: development of sense of community. Inform Health Soc Care 2013 Jan;38(1):67-78. [doi: 10.3109/17538157.2012.710685] [Medline: 22958060]
23. Yang CC, Tang X. Estimating User Influence in the MedHelp Social Network. IEEE Intell Syst 2012 Sep;27(5):44-50. [doi: 10.1109/MIS.2010.113]
24. Zhao K, Greer G, Yen J, Mitra P, Portier K. Leader identification in an online health community for cancer survivors: a social network-based classification approach. Inf Syst E-Bus Manage 2014 Sep 16;13(4):629-645. [doi: 10.1007/s10257-014-0260-5]
25. Cobb N, Graham A, Abrams D. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
26. Zhao K, Yen J, Ngamassi LM, Maitland C, Tapia A. Simulating inter-organizational collaboration network: a multi-relational and event-based approach. Simulation 2011 Sep 19;88(5):617-633. [doi: 10.1177/0037549711421942]
27. Boccaletti S, Bianconi G, Criado R, del Genio C, Gómez-Gardeñes J, Romance M, et al. The structure and dynamics of multilayer networks. Physics Reports 2014 Nov;544(1):1-122. [doi: 10.1016/j.physrep.2014.07.001]
28. Tang L, Liu H. Toward Predicting Collective Behavior via Social Dimension Extraction. IEEE Intell. Syst 2010 Jul;25(4):19-25. [doi: 10.1109/MIS.2010.36]
29. Fiore MC, Jaén CR, Baker TB, Clinical Practice Guideline Treating Tobacco Use and Dependence Panel. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
30. McCausland K, Curry L, Mushro A, Carothers S, Xiao H, Vallone D. Promoting a web-based smoking cessation intervention: implications for practice. Cases Public Health Commun Mark. 2011. p. 3-26 URL: http://publichealth.gwu.edu/departments/pch/phcm/casesjournal/volume5summer/peer-reviewed/V5_Case1PR.pdf [WebCite Cache ID 6jytRGo0r]
31. Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. Proceedings of the National Academy of Sciences of the United States of America 2010 Aug 03;107(31):13636-13641. [doi: 10.1073/pnas.1004008107] [Medline: 20643965]
32. Graham AL, Cha S, Papandonatos GD, Cobb NK, Mushro A, Fang Y, et al. Improving adherence to web-based cessation programs: a randomized controlled trial study protocol. Trials 2013;14:48 [FREE Full text] [doi: 10.1186/1745-6215-14-48] [Medline: 23414086]
33. Papandonatos GD, Erar B, Stanton CA, Graham AL. Online Community Use Predicts Abstinence in Combined Internet/Phone Intervention for Smoking Cessation. J Consult Clin Psychol 2016;84(7):633-644. [doi: 10.1037/ccp0000099] [Medline: 27100127]

34. Graham A, Papandonatos G, Erar B, Stanton C. Use of an online smoking cessation community promotes abstinence: Results of propensity score weighting. Health Psychol (Suppl)? 2015;34:1295. [doi: 10.1037/hea0000278] [Medline: 26651470]
35. Myneni S, Cobb NK, Cohen T. Finding meaning in social media: content-based social network analysis of QuitNet to identify new opportunities for health promotion. Stud Health Technol Inform 2013;192:807-811. [Medline: 23920669]
36. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
37. Yang Y, Chawla N, Sun Y, Hani J. Predicting Links in Multi-relational and Heterogeneous Networks. In: 2012 IEEE 12th International Conference on Data Mining. 2012 Presented at: 2012 IEEE 12th International Conference on Data Mining; 10-13 Dec 2012; Brussels p. 755-764. [doi: 10.1109/ICDM.2012.144]
38. Ng M, Li X, Ye Y. MultiRank: co-ranking for objects and relations in multi-relational data. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY: ACM; 2011 Presented at: KDD-2011; August 21-24, 2011; San Diego, CA p. 1217-1225. [doi: 10.1145/2020408.2020594]
39. Blankers M, Smit ES, van der Pol Peggy, de VH, Hoving C, van LM. The Missing=Smoking Assumption: A Fallacy in Internet-Based Smoking Cessation Trials? Nicotine Tob Res 2016 Jan;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]

Abbreviations

GMM: Gaussian mixture model
LSCC: largest strongly connected component
OHC: online health community

Edited by G Eysenbach; submitted 17.05.16; peer-reviewed by N Cobb, S Lee, N De Cock; comments to author 23.06.16; revised version received 25.07.16; accepted 12.08.16; published 25.08.16.

Please cite as:
Zhao K, Wang X, Cha S, Cohn AM, Papandonatos GD, Amato MS, Pearson JL, Graham AL
A Multirelational Social Network Analysis of an Online Health Community for Smoking Cessation
J Med Internet Res 2016;18(8):e233
URL: http://www.jmir.org/2016/8/e233/
doi:10.2196/jmir.5985
PMID:27562640

Original Paper

Making Quality Health Websites a National Public Health Priority: Toward Quality Standards

Theresa Devine[1*], MPH; Jordan Broderick[2*], MA; Linda M Harris[1*], PhD; Huijuan Wu[3*], MSc, MEng, PhD; Sandra Williams Hilfiker[3*], MA

[1]U.S. Department of Health and Human Services (HHS), Office of Disease Prevention and Promotion (ODPHP), Rockville, MD, United States

[2]U.S. Department of Health and Human Services (HHS), National Vaccine Program Office, Washington, DC, United States

[3]CommunicateHealth, Inc., Northampton, MA, United States

[*]all authors contributed equally

Corresponding Author:
Huijuan Wu, MSc, MEng, PhD
CommunicateHealth, Inc.
26 Market St.
Northampton, MA, 01060
United States
Phone: 1 413 582 0425
Fax: 1 413 517 0545
Email: huijuan@communicatehealth.com

Abstract

Background: Most US adults have limited health literacy skills. They struggle to understand complex health information and services and to make informed health decisions. The Internet has quickly become one of the most popular places for people to search for information about their health, thereby making access to quality information on the Web a priority. However, there are no standardized criteria for evaluating Web-based health information. Every 10 years, the US Department of Health and Human Services' Office of Disease Prevention and Health Promotion (ODPHP) develops a set of measurable objectives for improving the health of the nation over the coming decade, known as Healthy People. There are two objectives in Healthy People 2020 related to website quality. The first is objective Health Communication and Health Information Technology (HC/HIT) 8.1: increase the proportion of health-related websites that meet 3 or more evaluation criteria for disclosing information that can be used to assess information reliability. The second is objective HC/HIT-8.2: increase the proportion of health-related websites that follow established usability principles.

Objective: The ODPHP conducted a nationwide assessment of the quality of Web-based health information using the Healthy People 2020 objectives. The ODPHP aimed to establish (1) a standardized approach to defining and measuring the quality of health websites; (2) benchmarks for measurement; (3) baseline data points to capture the current status of website quality; and (4) targets to drive improvement.

Methods: The ODPHP developed the National Quality Health Website Survey instrument to assess the quality of health-related websites. The ODPHP used this survey to review 100 top-ranked health-related websites in order to set baseline data points for these two objectives. The ODPHP then set targets to drive improvement by 2020.

Results: This study reviewed 100 health-related websites. For objective HC/HIT-8.1, a total of 58 out of 100 (58.0%) websites met 3 or more out of 6 reliability criteria. For objective HC/HIT-8.2, a total of 42 out of 100 (42.0%) websites followed 10 or more out of 19 established usability principles. On the basis of these baseline data points, ODPHP set targets for the year 2020 that meet the minimal statistical significance—increasing objective HC/HIT-8.1 data point to 70.5% and objective HC/HIT-8.2 data point to 55.7%.

Conclusions: This research is a critical first step in evaluating the quality of Web-based health information. The criteria proposed by ODPHP provide methods to assess website quality for professionals designing, developing, and managing health-related websites. The criteria, baseline data, and targets are valuable tools for driving quality improvement.

(J Med Internet Res 2016;18(8):e211) doi:10.2196/jmir.5999

KEYWORDS

online health information; health literacy; reliability; usability; measurement

Introduction

Most US adults (90%) have limited health literacy skills [1], which means that many struggle to make sense of the complex information and services intended to help prevent disease and promote our health [2-5]. One promising path to improve the health literacy of the adult population is to increase the availability of evidence-based, understandable, easy-to-find sources of health information [6-8].

While medical experts continue to play a vital role in the health decision–making process, the Internet has quickly become one of the most popular places for people to search for information about their health [9,10]. Research indicates that not only are adults of all generations searching online for health information [11,12], but they are using what they find to make health care decisions, either on behalf of themselves or a loved one [13]. Thus, improving the quality of health-related websites has the potential to improve the health literacy—and the health—of the population [14,15].

Increasing access to quality online information is a shared priority among national and international policy makers. President Obama's Digital Government Strategy calls for new and better ways to deliver digital information and services [16] and the World Health Organization has called for greater transparency, privacy and security, codes of conduct, and individual choice and control of health-related websites [17].

Yet, in spite of the critical role of quality in health-related websites, there are no standardized criteria for assessing it [6,8,18]. Such criteria have been elusive since researchers called for operationalized definitions of quality in their 2002 meta-analysis of empirical website assessments [7]. Therefore, the first steps in achieving real gains in health-related website quality are to establish (1) a standardized approach to defining and measuring the quality of health websites; (2) benchmarks for measurement; (3) baseline data points to capture the current status of website quality; and (4) targets to drive improvement. This paper describes the efforts of the US Department of Health and Human Services' Office of Disease Prevention and Health Promotion (ODPHP) to establish these four items.

Methods

Defining and Measuring Quality Health Websites for Healthy People

Every 10 years, ODPHP manages the development of a set of measurable objectives for improving the health of the nation over the coming decade, known as Healthy People. Two objectives (Health Communication and Health Information Technology, HC/HIT-8.1 and 8.2, see Textbox 1) in Healthy People 2020 relate to website quality: one calls for improved information reliability; the other calls for improved website usability. Both were devised and operationalized by experts in health communication and technology. Together, these objectives provide a working definition of website quality and a promising path toward overall quality improvement of Web-based health information.

Textbox 1. Healthy People objectives Health Communication and Health Information Technology (HC/HIT) 8.1 and 8.2.

HC/HIT-8.1: Increase the proportion of health-related websites that meet three or more evaluation criteria for disclosing information that can be used to assess information reliability

HC/HIT-8.2: Increase the proportion of health-related websites that follow established usability principles

Information Reliability

Information reliability refers to the accuracy and credibility of website content as well as transparency in the purpose and ownership of the site [7,19]. This information can help users discern the origin and quality of Web-based content [20]. It is one of the most commonly identified indicators of website quality and has been widely referenced by public, private, and nonprofit organizations committed to improving the quality of Web-based health information such as the Medical Library Association, Health on the Net Foundation, Consumers Union, and the Agency for Healthcare Research and Quality [21-24].

Healthy People objective HC/HIT-8.1 was first introduced in Healthy People 2010 as a developmental objective. In 2005, ODPHP convened a Technical Expert Workgroup to identify reliability criteria based on established Web standards [25,26]. The ODPHP then developed the Website Information Reliability Evaluation Instrument (Multimedia Appendix 1), which includes specific reliability requirements for each of the 6 criteria identified by the Workgroup (see Table 1).

Table 1. Objective HC/HIT-8.1 criteria and reliability requirements.

Criteria	Reliability requirements
Identity	Name of person or organization responsible for website
	Street address for person or organization responsible for website
	Identified source of funding for website
Purpose	Statement of purpose or mission for website
	Uses and limitations of services provided
	Association with commercial products or services
Content development	Differentiation of advertising from nonadvertising content
	Medical, editorial, or quality review practices or policies
	Authorship of health content (per page of health content)
Privacy	Privacy policy
	How personal information is protected
User feedback	Feedback form or mechanism
	How information from users is used[a]
Content updating	Date content created (per page of health content)
	Date content reviewed, updated, modified, or revised (per page of health content)
	Copyright date[a]

[a] Optional requirements.

Website Usability

Usability standards tend to fall into three major categories [27]. The first category focuses on how the information is organized, commonly referred to as *information architecture*. The second category looks at how users navigate the information on a website, known as *site design*. The third category emphasizes how users interact with content on the website, referred to as *content design*. These three elements are commonly addressed in both federal and private guidelines on website usability [27-32].

Usability is an important component of website quality, affecting a user's ability to access and understand information online [33]. In fact, the design of a website is one of the most important indicators of website credibility and quality for users [34-37].

In 2012, an expert panel was used to establish an empirical definition of usability for Healthy People 2020. Panel members were selected from academic, private, and government sectors based on their expertise in website usability and health communication. With input from the panel, ODPHP developed the Website Usability Evaluation Instrument (see Multimedia Appendix 2) to measure progress toward Healthy People objective HC/HIT-8.2. The instrument assesses the three aforementioned website usability categories, using 19 established usability principles across 59 task-based usability measures (see Table 2). The Site Design category includes 9 composites that assess basic design elements of the site, including how the site looks, how the site functions, and how a user can interact with the site. The Information Architecture category includes 7 composites that assess how the site content is organized, including navigation, grouping, and labeling. Lastly, the Content Design category includes 3 composites that assess how the content is written and formatted, and includes plain language principles. Each of the 59 measures is rated on a scale of 1 to 4 based on the level of difficulty of performing the task on the website (1 being "task failure" and 4 being "minimal problems"). An average rating score is calculated for each usability principle. The benchmark was set to require an average score of 3.5 for 10 or more of the 19 usability principles.

Table 2. Objective HC/HIT-8.2 established usability principles and measures.

Categories	Established usability principles
Site Design	1. Use conventional interaction elements
	2. Make it obvious what is clickable and what is not
	3. Minimize vertical scrolling
	4. Ensure that the Back button behaves predictably
	5. Provide clear feedback signals for actions
	6. Ensure site is accessible for users with disabilities and uses elements of 508 compliance
	7. Provide a simplified user experience
	8. Incorporate multimedia
	9. Offer a functional home page
Information Architecture	10. Present a clear visual hierarchy
	11. Provide easy search functionality
	12. Clearly label content categories
	13. Make pages easy to skim or scan
	14. Make elements on the page easy to read
	15. Visually group related topics
	16. Make sure text and background colors contrast
Content Design	17. Focus the writing on audience and purpose
	18. Use the users' language; minimize jargon and technical terms
	19. Allow for interaction with the content

With Healthy People objectives HC/HIT-8.1 and 8.2, ODPHP established standardized criteria to define and measure the quality of health-related websites and set benchmarks for measurement. Next, the criteria were applied to a sample of health-related websites in order to identify baseline data points, which are required for all Healthy People measurable objectives. The reliability (objective HC/HIT-8.1) and usability (objective HC/HIT-8.2) instruments were combined into a single instrument: the National Quality Health Website Survey. Targets were set to drive improvement by the year 2020.

Sampling

The ODPHP identified the 100 top-ranked health-related websites (see Multimedia Appendix 3) for 3 months (August-October 2014) from the Alexa Top Sites pool—health category [38]. The data were collected from Alexa on October 14, 2014. Alexa's traffic ranks are based on the traffic data provided by users in Alexa's global data panel over a rolling 3-month period. A site's ranking is based on a combined measure of unique visitors and page views [39]. Websites were considered health related if they had at least three items of health information as it is broadly defined by the e-Health Code of Ethics (see Textbox 2) [40]. Duplicated websites were consolidated with the exception of microsites within the National Institutes of Health such as PubMed and MedlinePlus. A number of websites were excluded from the sample based on the exclusion criteria in Textbox 3.

Textbox 2. Criteria for selecting health websites (from e-Health Code of Ethics).

- Health information includes information for staying well, preventing and managing disease, and making other decisions related to health and health care.
- It includes information for making decisions about health products and health services.
- It may be in the form of data, text, audio, and/or video.
- It may involve enhancements through programming and interactivity.
- Health products include drugs, medical devices, and other goods used to diagnose and treat illnesses or injuries or to maintain health. Health products include both drugs and medical devices subject to regulatory approval by agencies such as the US Food and Drug Administration or UK Medicines Control Agency and vitamin, herbal, or other nutritional supplements and other products not subject to such regulatory oversight.
- Health services include specific, personal medical care or advice; management of medical records; communication between health care providers and/or patients and health plans or insurers or health care facilities regarding treatment decisions, claims, billing for services, and so on; and other services provided to support health care.
- Health services also include listservs, bulletin boards, chat rooms, and other online venues for the exchange of health information.
- Like health information, health services may be in the form of data, text, audio, and/or video and may involve enhancements through programming and interactivity.

Textbox 3. Exclusion criteria.

- Sites that are not about human beings
- Sites that are owned or maintained in a foreign country
- Sites that are specifically for health industry professional development, listing job postings for health professionals or research grants available for health researchers
- Sites that are designed only to introduce, sell, or support specific medical commercial products or technology solutions for the health or medical industry
- Sites that provide platforms for laboratory services
- Sites accessible only to members or paying subscribers who must enter an identifying log-in name and password
- Sites about beauty or cosmetic products or hairstyles
- Sites about health or medical education programs
- Sites about fitness industry professional development or gym memberships
- Sites providing pharmacy price comparison information
- Online forums or groups, or other social media platforms for informal discussions regarding health

The final sample included websites sponsored by three types of organizations: 48 out of 100 websites (48.0%) were for profit, 36 out of 100 (36.0%) were nonprofit, and 16 out of 100 (16.0%) were government.

Interrater Reliability

A senior usability researcher and a research associate reviewed the sample websites, following a reviewer training process to reach a certain level of interrater reliability (IRR). The team used Altman's Kappa Benchmark Scale (see Table 3) [41].

Table 3. Altman's Kappa Benchmark Scale.

Kappa statistic	Strength of agreement
<.20	Poor
.21-.40	Fair
.41-.60	Moderate
.61-.80	Good
.81-1.00	Very good

Objective HC/HIT-8.1 criteria are primarily composed of yes or no questions. Cohen's kappa was used to measure the agreement for nominal scales. A benchmark kappa score of .80 (a score generally accepted as demonstrating a sufficient degree of IRR) was used for objective HC/HIT-8.1 [41]. The criteria for objective HC/HIT-8.2 are scored on a 4-point scale, making it more difficult to reach a perfect IRR. A benchmark kappa score of .61 was used for objective HC/HIT-8.2. Interclass correlation (ICC) was used for assessing ordinal and interval scales. To ensure IRR, both reviewers assessed the same 6 websites during the initial training process. The IRR was calculated for each assessment; discrepancies were identified and resolved between the 2 reviewers until the IRR scores met the benchmark kappa scores.

After the initial training process, the reviewers randomly selected 4 additional websites from the sample to measure IRR scores. The IRR scores for the Website Information Reliability Evaluation Instrument (kappa .83) and the Website Usability Evaluation Instrument (ICC .76) both met the benchmarks.

Website Review Process

After ensuring IRR, the reviewers divided the remaining 90 websites into 2 equal groups. Each reviewer assessed an equal number of websites.

To score each item in the Website Information Reliability Evaluation Instrument, reviewers began by randomly selecting 3 pages from different sections of the site. Each of the items in the instrument was scored based on a review of at least 3 different pages and no individual page was reviewed more than once. For items in the instrument that refer to a specific page or feature (eg, home page–related items and search function items), the review also included those specific pages or features. As each review progressed, the reviewer revised the scores for previously scored items as needed. In general, reviewers examined about 150 pages on each website.

Results

Baseline data points for 2015 were calculated for both Healthy People 2020 website quality objectives.

Information Reliability Baseline

For Healthy People objective HC/HIT-8.1, a total of 58 out of 100 health-related websites (58.0%) met 3 or more of the 6 information reliability criteria. Only 2 out of 100 websites (2.0%) met all the criteria, and 1 out of 100 websites (1.0%) met none of the 6 criteria. See Figure 1.

Figure 2 shows the percentage of health-related websites in compliance with specific criteria. User Feedback (90.0%) had the highest percentage in compliance, followed by Privacy (83.0%). Only 4 out of 100 websites (4.0%) met the criterion for Content Updating.

Table 4 presents the percentage of websites in compliance by criterion and by information reliability requirements associated with each criterion.

Table 4. Estimated percentages of websites in compliance, by information reliability criterion and required disclosure elements.

Criterion and required disclosure elements	N	Count	Percent (%)	SE[a] (%)	Lower bound 95% CI (%)	Upper bound 95% CI (%)
Identity	100	37	37.0	4.83	27.5	46.5
Name	100	93	93.0	2.55	88.0	98.0
Street address	100	83	83.0	3.76	75.6	90.4
Funding sources	100	44	44.0	4.96	34.3	53.7
Purpose	100	52	52.0	5.00	42.2	61.8
Purpose or mission	100	79	79.0	4.07	71.0	87.0
Uses and limitations	100	82	82.0	3.84	74.5	89.5
Association with commercial products	100	71	71.0	4.54	62.1	79.9
Content Development	100	15	15.0	3.57	8.0	22.0
Identify advertising content	60	28	46.7	6.44	34.0	59.3
Describe editorial policy	100	39	39.0	4.88	29.4	48.6
Authorship	100	38	38.0	4.85	28.5	47.5
Privacy	100	83	83.0	3.76	75.6	90.4
Privacy policy	100	96	96.0	1.96	92.2	99.8
Describe protection of personal information	100	83	83.0	3.76	75.6	90.4
User Feedback	100	90	90.0	3.00	84.1	95.9
Feedback mechanism	100	90	90.0	3.00	84.1	95.9
Content Updating	100	4	4.0	1.96	0.2	7.8
Display date created	100	25	25.0	4.33	16.5	33.5
Display date reviewed or updated	100	28	28.0	4.49	19.2	36.8

[a] SE: standard error.

Figure 1. Percentage of websites in compliance, with the number of reliability criteria.

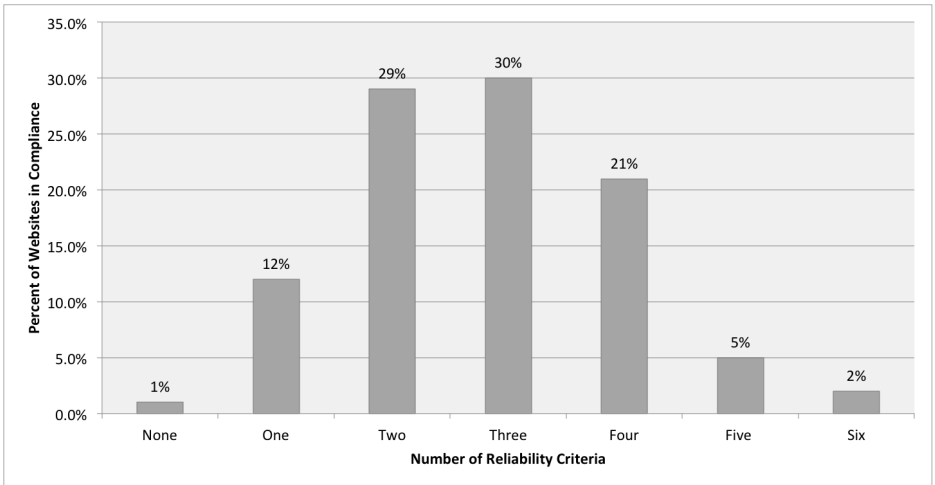


Figure 2. Percentage of websites in compliance, by specific reliability criteria.

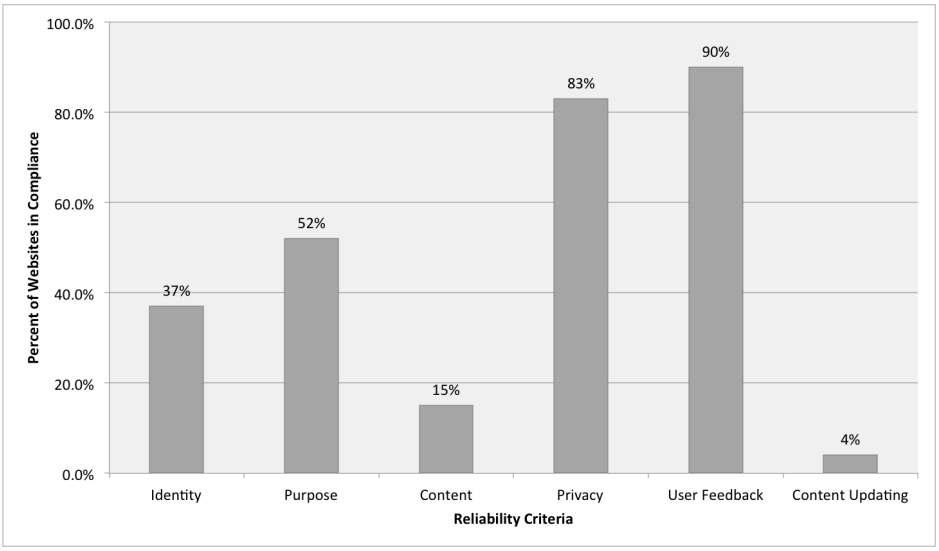
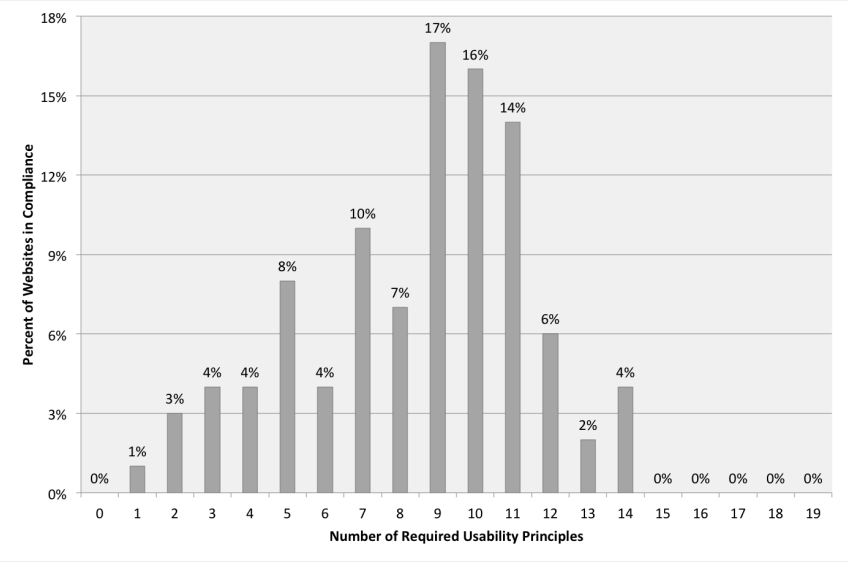


Table 5. Estimated percentages of websites in compliance, by criterion and established usability principles.

Criterion and established usability principles	Number (n=100)	Percent (%)	SE[a] (%)	Lower bound 95% CI (%)	Upper bound 95% CI (%)
Site Design					
1. Use conventional interaction elements	78	78.0	4.14	69.9	86.1
2. Make it obvious what is clickable and what is not	68	68.0	4.66	58.9	77.1
3. Minimize vertical scrolling	24	24.0	4.27	15.6	32.4
4. Ensure that the Back button behaves predictably	100	100.0	0.00	100.0	100.0
5. Provide clear feedback signals for actions	39	39.4	4.91	29.8	49.0
6. Ensure site is accessible for users with disabilities and uses elements of 508 compliance	6	6.0	2.37	1.3	10.7
7. Provide a simplified user experience	30	30.0	4.58	21.0	39.0
8. Incorporate multimedia	70	70.0	4.58	61.0	79.0
9. Offer a functional home page	30	30.0	4.58	21.0	39.0
Information Architecture					
10. Present a clear visual hierarchy	42	42.0	4.94	32.3	51.7
11. Provide easy search functionality	17	17.0	3.76	9.6	24.4
12. Clearly label content categories	25	25.0	4.33	16.5	33.5
13. Make pages easy to skim or scan	45	45.0	4.97	35.2	54.8
14. Make elements on the page easy to read	72	72.0	4.49	63.2	80.8
15. Visually group related topics	45	45.0	4.97	35.2	54.8
16. Make sure text and background colors contrast	74	74.0	4.39	65.4	82.6
Content Design					
17. Focus the writing on audience and purpose	30	30.0	4.58	21.0	39.0
18. Use the users' language; minimize jargon and technical terms	32	32.0	4.66	22.9	41.1
19. Allow for interaction with the content	19	19.0	3.92	11.3	26.7

[a] SE: standard error.

Figure 3. Percentage of websites in compliance, by the number of usability principles.



Website Usability Baseline

For Healthy People objective HC/HIT-8.2, a total of 42 out of 100 health-related websites (42.0%) met 10 or more out of 19 established usability principles. Figure 3 shows the distribution of website compliance by the number of usability principles met. All websites met at least one usability principle. Almost half of the websites (47.0%) met between 9 and 11 principles. None of the websites met 15 or more of the established usability principles. See Table 5 for the percentage of websites in compliance broken down by the 19 usability principles.

Discussion

Baseline Data Points

The 2015 review of health-related websites identified 2 baseline data points. For Healthy People 2020 objective HC/HIT-8.1, a total of 58 out of 100 health-related websites (58.0%) met 3 or more out of 6 reliability criteria. For Healthy People 2020 objective HC/HIT-8.2, a total of 42 out of 100 health-related websites (42.0%) followed 10 or more out of 19 established usability principles.

This research revealed significant shortcomings in the quality of today's Web-based health information landscape, particularly in disclosing sources of funding and authorship, clearly differentiating between advertisements and original content, complying with universal accessibility guidelines, providing simple search and print functionality, and minimizing scientific and technical jargon.

2020 Targets

The ODPHP set the following targets for 2020 that meet the minimal statistical significance. For Healthy People 2020 objective HC/HIT-8.1, 70.5% of health-related websites will meet 3 or more out of 6 reliability criteria. For Healthy People 2020 objective HC/HIT-8.2, 55.7% of health-related websites will follow 10 or more out of 19 established usability principles.

Improving the quality of health-related websites is critical to national efforts to promote health literacy and shared decision making. Until now, there has been no standardized approach to defining and measuring the quality of Web-based health information. With Healthy People 2020 objectives HC/HIT-8.1 and 8.2, ODPHP has developed and validated such an approach, and established baseline data points and national benchmarks to track progress over time.

The 2015 study confirmed that there is, indeed, room for improvement. The ODPHP's research revealed significant shortcomings in the quality of today's Web-based health information landscape, particularly in the following areas:

- Disclosing sources of funding and authorship
- Clearly differentiating between advertisements and original content
- Complying with universal accessibility guidelines (eg, Section 508 of the Amendment to the Rehabilitation Act of 1973)
- Providing simple search and print functionality
- Minimizing scientific and technical jargon

Limitations

This study has several limitations. First, the samples of this study were selected from Alexa Top Sites—health category. The research team had no control over the quality of the website rankings performed by Alexa.

Additionally, some of the usability principles might change over time. For example, in this study, only 24 out of 100 websites (24.0%) followed the principle of minimizing vertical scrolling. However, with the proliferation of mobile and responsive design technology, users are becoming more accustomed to navigating websites by scrolling. Minimizing vertical scrolling may not remain a usability principle in the future.

Finally, the survey instrument is somewhat subjective, especially for objective HC/HIT-8.2. Although the research team controlled the reliability by measuring the IRR for several websites in the sample, there was still variation across reviewers that may affect the results.

Conclusions

The quality and accessibility of Web-based health information is a key factor in improving access to health services and facilitating informed health decision making. The criteria proposed by ODPHP provide methods to assess the quality of health-related websites and provide baseline data and targets to drive quality improvement. In addition to having implications for website developers and policy makers, this work also points to the need for consumer education related to the quality of Web-based health information. Such education efforts are critical in a time when nearly 3 in 4 Internet users are looking for health information online [42].

To promote increased quality of health-related websites, ODPHP updated and published Health Literacy Online: A Guide to Simplifying the User Experience, second edition [43]. Health Literacy Online is based on literature related to cognitive processing and online behavior and on usability research with more than 800 participants. It features actionable information that website owners, content writers, designers, and developers can use to create quality health websites. The recommendations in Health Literacy Online provide a clear road map for achieving the Healthy People 2020 objectives to increase the proportion of quality health-related websites (Objective HC/HIT-8). The ODPHP is working with other federal agencies to adopt the principles in Health Literacy Online and is presenting these strategies for improving health websites at national conferences.

Acknowledgments

This research was funded by the US Department of Health and Human Services' Office of Disease Prevention and Health Promotion (ODPHP). The ODPHP, the Centers for Disease Control and Prevention's Office of the Associate Director for Communication, and the Office of the National Coordinator for Health Information Technology co-lead the Healthy People 2020 Health Communication and Health Information Technology topic area. The ODPHP contracted with CommunicateHealth, Inc., to conduct the data collection and analysis discussed in this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Website Information Reliability Evaluation Instrument.

[PDF File (Adobe PDF File), 83KB - jmir_v18i8e211_app1.pdf]

Multimedia Appendix 2

Website Usability Evaluation Instrument.

[PDF File (Adobe PDF File), 74KB - jmir_v18i8e211_app2.pdf]

Multimedia Appendix 3

Sample Websites.

[PDF File (Adobe PDF File), 103KB - jmir_v18i8e211_app3.pdf]

References

1. Kutner M, Greenberg E, Jin Y, Paulsen C. NCES 2006-483. Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2006. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy URL: http://files.eric.ed.gov/fulltext/ED493284.pdf [accessed 2016-07-18] [WebCite Cache ID 6j5iikXBc]
2. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: The state of the art. Health Educ Res 2001;16(6):671-692 [FREE Full text] [Medline: 11780707]
3. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
4. 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]
5. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 May 20:- Epub ahead of print. [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
6. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the Internet: let's Google it. J Pediatr 2012;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]

7. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [Medline: 12020305]
8. Neumark Y, Flum L, Lopez-Quintero C, Shtarkshall R. Quality of online health information about oral contraceptives from Hebrew-language websites. Isr J Health Policy Res 2012;1(1):38 [FREE Full text] [doi: 10.1186/2045-4015-1-38] [Medline: 23006798]
9. Fox S, Rainie L. Pew Research Center. 2002. Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick URL: http://www.pewinternet.org/files/old-media/Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf [accessed 2016-04-21] [WebCite Cache ID 6gwCQqT2V]
10. Fox S, Jones S. Pew Research Center. 2009. The social life of health information URL: http://www.pewinternet.org/2009/06/11/the-social-life-of-health-information/ [accessed 2016-04-21] [WebCite Cache ID 6gwGpOpGT]
11. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
12. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
13. Fox S, Duggan M. Pew Research Center. 2013. Health online URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2016-04-21] [WebCite Cache ID 6gwH4m8OG]
14. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. J Med Internet Res 2014;16(10):e224 [FREE Full text] [doi: 10.2196/jmir.3341] [Medline: 25348028]
15. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. BMJ 2007;335(7609):24-27 [FREE Full text] [doi: 10.1136/bmj.39246.581169.80] [Medline: 17615222]
16. The White House. Whitehouse.gov. 2012. Presidential Memorandum -- Building a 21st Century Digital Government URL: https://www.whitehouse.gov/the-press-office/2012/05/23/presidential-memorandum-building-21st-century-digital-government [accessed 2016-04-21] [WebCite Cache ID 6gwI68EyU]
17. WHO. 2014. The health Internet: governing principles for a .health top-level domain URL: http://www.who.int/ehealth/programmes/governance/en/index3.html [accessed 2016-04-21] [WebCite Cache ID 6gwHCCTtF]
18. Kitchens B, Harle CA, Li S. Quality of health-related online search results. Decision Support Systems 2014;57:454-462. [doi: 10.1016/j.dss.2012.10.050]
19. Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. BMJ 1999;318(7184):647-649 [FREE Full text] [Medline: 10066209]
20. Van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
21. Medical Library Association. Find Good Health Information URL: http://www.mlanet.org/resources/userguide.html [accessed 2016-04-21] [WebCite Cache ID 6gwI1yEDJ]
22. Health on the Net Foundation. HON. 2014. The commitment to reliable health and medical information on the internet URL: http://www.hon.ch/HONcode/Patients/Visitor/visitor.html [accessed 2016-04-21] [WebCite Cache ID 6gwHGhsE8]
23. Consumers Union. Mission URL: http://consumersunion.org/about/mission/ [accessed 2016-04-21] [WebCite Cache ID 6gwHK2YK2]
24. The Agency for Healthcare Research and Quality. AHRQ. 2006. Guidelines for Ensuring the Quality of Information Disseminated to the Public URL: http://www.ahrq.gov/research/publications/ahrq-info-quality-gdlns.html [accessed 2016-04-21] [WebCite Cache ID 6gwITjffN]
25. URAC. Tspwebs. 2001. Health Web Site Standards, Version 1.0 URL: http://webs.tspwebs.net/URAC%20Standards.pdf [accessed 2016-04-27] [WebCite Cache ID 6h5HG6pSD]
26. Health Improvement Institute. 2005. Health Website Ratings Instrument, Version 2.0 URL: https://dl.dropboxusercontent.com/content_link/XcTWVhlqdUN8rYGqTRDc4uLTG2NOHy54swzmXbVvrgzm619tqqYNeQSgfsgMMFGi/file [accessed 2016-04-27] [WebCite Cache ID 6h5NYfdew]
27. The U.S. Department of Health Human Services. Research-Based Web Design & Usability Guidelines (Enlarge/E). Washington, DC: U.S. Government Printing Office; 2006.
28. Travis D. Userfocus. 2014. 247 web usability guidelines URL: http://www.userfocus.co.uk/resources/guidelines.html [accessed 2016-04-21] [WebCite Cache ID 6gwJkQTV3]
29. UsabilityNet. 2006. Design guidelines for the Web URL: http://www.usabilitynet.org/tools/webdesign.htm [accessed 2016-04-21] [WebCite Cache ID 6gwJohNql]
30. Sharpened Productions. 2015. 10 Rules of Web Design URL: http://sharpened.com/web_design_rules [accessed 2016-04-21] [WebCite Cache ID 6gwJteSD7]
31. Krug S. Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability. 3rd edition. Berkeley, CA: New Riders; 2014.

32. Redish J. Letting Go of the Words: Writing Web Content that Works (Interactive Technologies). 2nd edition. Burlington, MA: Morgan Kaufmann; 2012.
33. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of "Dr Google": A Survey on Health Information-Seeking Behaviors and Navigational Needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
34. Weymann N, Härter M, Dirmaier J. Quality of online information on type 2 diabetes: a cross-sectional study. Health Promot Int 2015;30(4):821-831. [doi: 10.1093/heapro/dau019] [Medline: 24688114]
35. Stanford J, Tauber ER, Fogg BJ, Marable L. Ebusiness Forum. 2002. Experts vs. Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites URL: http://www.ebusinessforum.gr/old/content/downloads/comparativeCredibilityStudy.pdf [accessed 2016-04-15] [WebCite Cache ID 6gwKAeTk6]
36. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
37. 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]
38. Alexa. 2014. Top Sites -- health category URL: http://www.alexa.com/topsites/category/Top/Health [accessed 2014-10-15] [WebCite Cache ID 6i9KIek4q]
39. Alexa. 2016. How are Alexa's traffic rankings determined? URL: https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined- [accessed 2016-06-10] [WebCite Cache ID 6i9Kj8Hvo]
40. Rippen H, Risk A. e-Health Code of Ethics (May 24). J Med Internet Res 2000;2(2):E9 [FREE Full text] [doi: 10.2196/jmir.2.2.e9] [Medline: 11720928]
41. Altman D. Practical Statistics for Medical Research. Boca Raton, FL: CRC Press; 1990.
42. Pew Research Center. 2013. Health Fact Sheet URL: http://www.pewinternet.org/fact-sheets/health-fact-sheet/ [accessed 2016-04-21] [WebCite Cache ID 6gwHc9ajM]
43. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Health.gov. 2015. Health Literacy Online: 2nd Edition URL: http://health.gov/healthliteracyonline/ [accessed 2016-06-10] [WebCite Cache ID 6iAbZO9Wl]

Abbreviations

HC/HIT: Health Communication and Health Information Technology
ICC: interclass correlation
IRR: interrater reliability
ODPHP: Office of Disease Prevention and Health Promotion

Edited by G Eysenbach; submitted 19.05.16; peer-reviewed by K Summers, G Reddish; comments to author 08.06.16; revised version received 22.06.16; accepted 23.06.16; published 02.08.16.

Please cite as:
Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW
Making Quality Health Websites a National Public Health Priority: Toward Quality Standards
J Med Internet Res 2016;18(8):e211
URL: http://www.jmir.org/2016/8/e211/
doi:10.2196/jmir.5999
PMID:27485512

Original Paper

Exploring Patients' Views Toward Giving Web-Based Feedback and Ratings to General Practitioners in England: A Qualitative Descriptive Study

Salma Patel[1], BSc (Hons); Rebecca Cain[1], BA (Hons), PhD; Kevin Neailey[1], BSc (Hons), PhD; Lucy Hooberman[1], MA (Hons) Oxon, FRSA

WMG, Coventry, United Kingdom

Corresponding Author:
Salma Patel, BSc (Hons)
WMG
University of Warwick
Coventry, CV47AL
United Kingdom
Phone: 44 24 7657 5951
Fax: 44 24 7657 5951
Email: salma.patel@warwick.ac.uk

Abstract

Background: Patient feedback websites or doctor rating websites are increasingly being used by patients to give feedback about their health care experiences. There is little known about why patients in England may give Web-based feedback and what may motivate or dissuade them from giving Web-based feedback.

Objective: The aim of this study was to explore patients' views toward giving Web-based feedback and ratings to general practitioners (GPs), within the context of other feedback methods available in primary care in England, and in particular, paper-based feedback cards.

Methods: A descriptive exploratory qualitative approach using face-to-face semistructured interviews was used in this study. Purposive sampling was used to recruit 18 participants from different age groups in London and Coventry. Interviews were transcribed verbatim and analyzed using applied thematic analysis.

Results: Half of the participants in this study were not aware of the opportunity to leave feedback for GPs, and there was limited awareness about the methods available to leave feedback for a GP. The majority of participants were not convinced that formal patient feedback was needed by GPs or would be used by GPs for improvement, regardless of whether they gave it via a website or on paper. Some participants said or suggested that they may leave feedback on a website rather than on a paper-based feedback card for several reasons: because of the ability and ease of giving it remotely; because it would be shared with the public; and because it would be taken more seriously by GPs. Others, however, suggested that they would not use a website to leave feedback for the opposite reasons: because of accessibility issues; privacy and security concerns; and because they felt feedback left on a website may be ignored.

Conclusions: Patient feedback and rating websites as they currently are will not replace other mechanisms for patients in England to leave feedback for a GP. Rather, they may motivate a small number of patients who have more altruistic motives or wish to place collective pressure on a GP to give Web-based feedback. If the National Health Service or GP practices want more patients to leave Web-based feedback, we suggest they first make patients aware that they can leave anonymous feedback securely on a website for a GP. They can then convince them that their feedback is needed and wanted by GPs for improvement, and that the reviews they leave on the website will be of benefit to other patients to decide which GP to see or which GP practice to join.

(J Med Internet Res 2016;18(8):e217) doi:10.2196/jmir.5865

KEYWORDS

Web-based reviews; physician quality; primary care; Internet; quality patient empowerment; quality transparency; public reporting

Introduction

In England, patients and carers can leave feedback about their experience of a consultation with a general practitioner (GP) using a multitude of methods [1]. These include in-house surveys, the National General Practice Patient Survey, suggestion boxes, surveys for revalidation, the Friends and Family Test, Care Quality Commission ratings, and the NHS Choices and other feedback websites. The number of people leaving ratings and reviews on the Web for products and services in other sectors has exponentially increased [2-4]. A similar type of growth can also be seen, although not to the same magnitude, in the number of patients and carers in England leaving feedback on the Web about their health care experience [5-9].

In 2007, the National Health Service (NHS) introduced the NHS Choices website. Part of the intention and part of the site was designed to encourage patients to provide feedback on health care services [10]. Consequently, on this website, for primary care, patients and carers can (1) view feedback and ratings left by other patients and carers and (2) leave feedback, reviews, or ratings of their health care experience under the GP practice's name [11]. The former is part of the "choice" agenda that aims to give patients the tools to choose which GP practice to join [12-14]. The latter, the NHS in England states, gives patients a "voice" to air their feedback and concerns independently in the public domain, which they argue will not only increase transparency but also bring improvement and help empower patients [8,10,15]. However, there is little evidence to date to suggest that this has happened.

Research into doctor rating and patient feedback websites is increasing (studies in the United Kingdom [6-8,10,14,16-19], the United States [20-22], Germany [23-27], the Netherlands [28], and Australia [29]). There is some evidence, not always consistent, to suggest that there is an association between Web-based ratings and quality of care [5,6,21,30,31]. In England, although there was some evidence to support a moderate association between patient experience about primary care narrated on a website and via conventional patient surveys, the association with clinical quality of primary care was found to be weak [7].

Studies conducted outside England [25,32-35] have explored what type of patients use patient rating and feedback websites. Two studies conducted in England [36,37] explored patients' awareness and consideration of their future use of doctor rating websites, as well as some of the demographic predictors for people willing to leave feedback on doctor rating websites. However, none of these studies explored patients' own views toward patient feedback websites, such as whether they perceive any benefits or risks in relation to leaving feedback on a website, or what may motivate or dissuade them to leave feedback on a website [9]. There is also little understanding of how these attitudes and preferences differ from attitudes and preferences toward other feedback methods. Therefore, the aim of this study was to explore patients' views toward giving Web-based feedback and ratings to GPs in England, within the context of other feedback methods available in primary care, in particular paper-based feedback cards. The intention is to use the findings from this study to create a questionnaire that could be used across England to explore nationwide public views and understanding toward giving feedback on a website about GPs.

Methods

Defining the Context of the Study

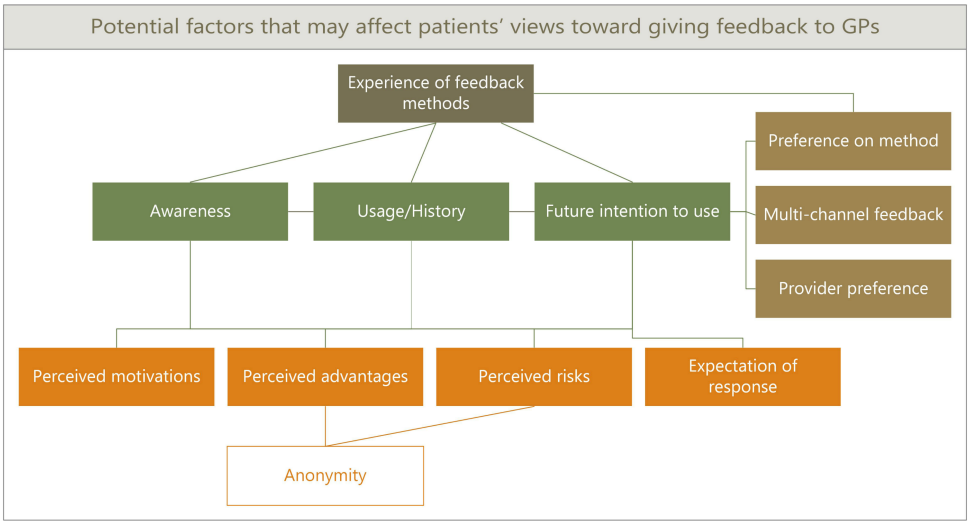
The nature of patient feedback websites appears to be evolving quickly. Therefore, to ensure that the research questions developed for this study were up-to-date, and the context within which this study was developed could be understood, before starting this study (in April 2015), the following were outlined: the key stakeholders involved with patient feedback websites (see Multimedia Appendix 1), the characteristics of the Web-based patient feedback platforms available in England (see Multimedia Appendix 2), and the different pathways that a patient may take to use patient feedback websites in England (see Multimedia Appendix 3).

Data Collection

This study was exploratory and descriptive in design because there was very little known about patients' views toward patient feedback websites. Qualitative semistructured interviews were used because this gave the depth required and allowed probing of participants [38]. A deductive conceptual framework was created (see Figure 1) based on existing literature and knowledge gaps, and this, as well as guidance suggested by Bryman [39] and Matthews and Ross [40], was used to design the topic guide (see Multimedia Appendix 4 for a copy of the topic guide). The topic guide was pilot-tested on 2 members of the public before use in the interviews.

Two materials were used in the interviews to provide information to participants. The first was the NHS Friends and Family Test card and the second material contained a screenshot of a GP practice page on the NHS Choices website. Two card sorting exercises were also used to help participants explain which methods they would most prefer to use to leave feedback for GPs. The methods selected were based on feedback methods mentioned by Brown et al [41], Silva [42], and Coulter [43] in patient feedback literature, and are listed in the topic guide (see Multimedia Appendix 4).

Figure 1. A deductive conceptual framework used to design the topic guide. GP: general practitioner.



Sampling and Recruitment

Purposive sampling was used to recruit participants so that the sample would represent 3 patients from each age group between the ages of 20 and 80 years. Participants were screened before recruitment to ensure they had at least one consultation with a GP in the past year. A total of 18 participants (10 female; 8 male) were recruited from 4 locations in England: East London, North London, South London, and Coventry. A total of 15 participants were interviewed initially, after which the data were analyzed because the data appeared to have reached close to thematic saturation. Then, 3 further interviews were conducted and analyzed, and the themes that emerged validated and supported the existing themes found. The data were now believed to have reached thematic saturation [44], and therefore no further interviews were conducted.

Study Interviews

Participants were sent an invitation letter and information sheet beforehand and were interviewed using the topic guide in a private meeting room or at the participant's home. Informed consent was obtained from all participants. Each interview was on average 30 minutes long and was recorded digitally. The study had ethical approval from the Biomedical and Scientific Research Ethics Committee at the University of Warwick (ref REGO-2015-1472; May 2015).

Data Preparation and Analysis

The interviews were transcribed verbatim and double-checked for inaccuracies. Transcripts were then transferred to NVivo (QSR International) where they were analyzed using applied thematic analysis [45]. This is a form of inductive (data-driven) thematic analysis that has a pragmatic focus and allows the use of tools appropriate for the analytical process, such as structural coding, quantification, word searches, and deviant case analyses. A structural coding framework consisting of 25 sections was created, which was applied to the first 15 transcripts. Data were then collated relevant to each new code, and a codebook was created in Microsoft Word for each section. As the codebook developed, codes were refined, combined, and deleted from both NVivo and the codebook. Themes were then generated from the codes and reviewed. The 3 new interview transcripts were then added at this stage and went through the aforementioned steps. The themes that emerged supported and validated the existing themes found. The analysis was conducted by the first author (SP), and the codebook was checked for accuracy, internal homogeneity, and external heterogeneity by the second author (RC).

Results

Overview

Participants were asked about their views toward giving feedback to GPs, with a focus in particular on patient feedback websites and on paper-based feedback cards. Participants discussed their awareness and past usage of the Web-based and offline modes of feedback to leave feedback for a GP, as well as their attitudes, motivations, and consideration for future use of both websites and paper-based feedback cards. The interviews focused mainly on the NHS Choices website as the Web-based patient feedback mode and the NHS Friends and Family Test feedback card as the offline mode to leave feedback, both of which are available in general practice in England and are generally unsolicited forms of feedback.

This paper presents only the major themes that emerged from the data. The first 4 themes (1-4) were not specific to a method or mode of feedback; rather, they were found in relation to both paper-based feedback and patient feedback websites. The final 3 themes (5-7) were unique to patient feedback websites, and they allude to the additional considerations that patients need to give when considering using websites to leave feedback for a GP.

Theme 1: Limited Awareness About Methods to Leave Feedback for GPs, Especially on a Website

In this study, 5 participants had given feedback about a GP in the past using non-Web-based methods, and the remaining 13 had not. Interestingly, however, almost half of the participants (n=8) did not know that they could leave feedback about a GP using any method:

> *I haven't seen this [NHS Friends and Family Test card] before, probably haven't looked* [P18]

Similarly, the majority of participants (n=16) were not aware of the existence of patient feedback websites, and only 1 female participant aged 47 years had experience of giving feedback on a website for a GP. However, more than half of participants (n=12) said they would happily leave feedback about a GP if they were asked to by the GP or the practice, and 13 participants said they may consider giving feedback on a website or on paper in the future.

Theme 2: Preference for Mode of Feedback Depends on the Nature of Feedback

The majority of participants preferred to give positive feedback directly face-to-face to the GP, and almost half also preferred to give negative feedback directly face-to-face to the GP:

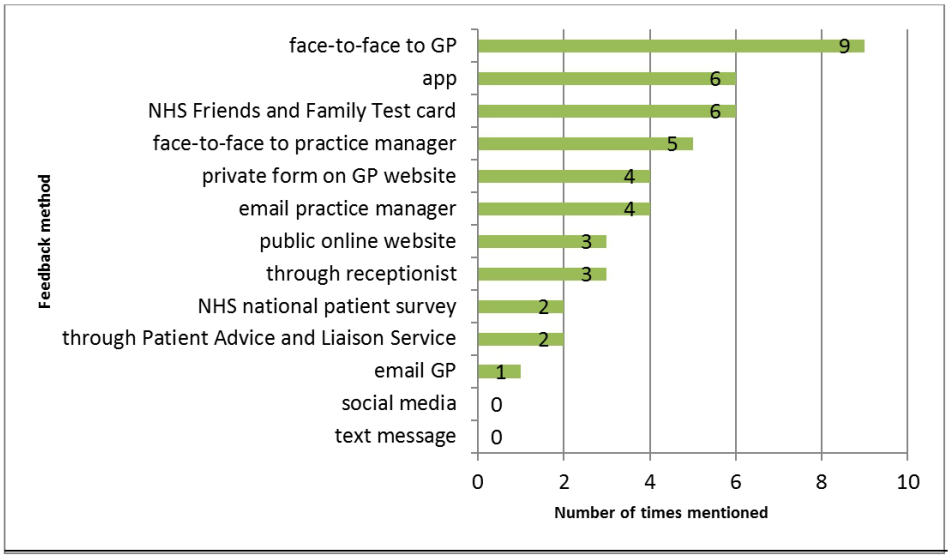
> *If I was unhappy with my GP I would make an appointment and tell her that I was unhappy. I wouldn't mess about* [P16]

The other methods by which participants most preferred to give feedback were through an app, filling in the NHS Friends and Family Test card, giving the feedback to the practice manager, and leaving the feedback on a private form on the GP website (see Figure 2).

Participants in this study were not keen on using social media (such as Facebook or Twitter) to leave feedback for a GP, emailing or texting the feedback, or using the national patient survey. Among the digital methods the least popular with participants was social media, followed by emailing the GP directly and text messaging. A total of 3 participants (aged between 35 and 55 years) mentioned the website in their top 3 preferred ways to leave feedback about a GP. However, almost all participants added a caveat and said that their preference of which method to use to give feedback to a GP would actually differ depending on the nature of the feedback, that is, whether the feedback was positive or negative:

> *It depends [on] what feedback you are giving [positive or negative].* [P1]

Figure 2. A chart showing the number of times the feedback method was selected (through the card sorting exercise) in participants' 3 most preferred ways to leave feedback for a general practitioner (GP). NHS: National Health Service.



Theme 3: Extreme Experience Is the Primary Driver to Leave Feedback, Both on Paper and on a Website

Almost all of the participants (n=17) described their past experience with a GP as satisfactory or good, and therefore they felt that there was no need for them to leave feedback about a GP:

> *I suppose it feels a bit silly to make the effort to go to give feedback to say "yeah everything was fine.* [P1]

However, all participants (n=18) agreed that if they experienced an extreme experience in the future, they would leave feedback for a GP:

> *If I felt that the level of service [was] exceptionally good or exceptionally poor, I'd be inclined to leave feedback* [P15]

Furthermore, the majority of participants said they were more likely to leave feedback (on a website or using any other method) when they had experienced an extreme negative experience, rather than an extreme positive experience.

Theme 4: Patients Need Convincing That Feedback (Both Paper and on a Website) Is Needed and Will Be Used for Improvement

Many participants questioned whether the feedback they leave for a GP would lead to any kind of improvement. Half of the participants (n=9) believed that giving feedback to a GP would not make a difference to the GP's behavior or practice. A total of 5 participants were unsure whether the GP would even see the patient feedback or respond to it. Furthermore, more than half of participants (n=13) believed that GPs do not want patient feedback, otherwise GPs would ask them to leave feedback for them:

> *If I was given a card every time I went and they said "can you tick it" then I would tick it and pop it in the box on the way out, but it seems sort of an odd thing*

> *to do if I don't know they [GPs] particularly want it.* [P1]

A total of 4 participants said that a GP does not need feedback, and a quarter of participants (n=6) said that GPs could utilize their time better by treating patients instead of using their time to read patient feedback. However, more than half of participants (n=12) said they would happily leave feedback about a GP if they were asked to leave feedback by the GP or the practice:

> *If the paper [to give feedback] was given to me, I would definitely leave feedback.* [P7]

Furthermore, more than half of participants (n=13) explained that if they were to leave feedback for a GP in the future, their reason for doing so would be to highlight good and bad practice and identify opportunities for improvement:

> *I think it's good to highlight good practice, where things go well...but equally with regards to whether [sic] things don't go as well* [P4]

Theme 5: Transparency of Patient Rating and Feedback Websites

A total of 12 participants believed that patient feedback being on a website and in the public domain is advantageous for patients. Among them, 7 participants explained that this was because the public and other organizations could benefit from such feedback because they could evaluate from patient experiences how well GPs and GP practices were performing:

> *Because it is public isn't it, and shows the whole world [sic] can see how well the practice is doing.* [P6]

Other participants (n=4) explained that because the feedback is in the public domain, GPs and the GP practice would take Web-based patient feedback much more seriously than feedback left using other methods, because they would feel more accountable:

> *With online, because it is in the open, once it is there, it is pretty much like a branding for them, so it's almost like they have to take it more seriously* [P10]

Furthermore, 2 participants believed that patient feedback being on a website was advantageous to the GP practice too, because feedback would be easier to collate and there would be less room for error when transferring that feedback to GPs. However, 3 participants believed that these types of feedback websites could be a breeding ground for false complaints, negativity, and abuse:

> *If you put things that are negative online, it just creates a breeding ground for more...and then becomes a slating of the surgery* [P2]

One participant said that because of this the GP practice may actually view the feedback left on these websites with skepticism, which, according to her, defeats the purpose of leaving feedback to bring about change or improvement. Similarly, 5 participants (most older than 60 years) questioned the value of patient feedback websites by arguing that these types of feedback websites are not useful to them or to the public:

> *What's the value in people scrolling down and reading I've had a particularly good or bad experience? [P9]*

Theme 6: Concerns About Privacy, Security, and Anonymity of Patient Feedback Websites

More than half of participants (n=10) from all age groups had privacy concerns about leaving feedback on a website and were worried that their identity could be traced, even when leaving feedback anonymously. In contrast, only 2 participants felt that their identity could be revealed if they left feedback for a GP using the NHS Friends and Family Test card. One participant was worried that disclosing her identity when leaving negative feedback on a website or using any other method could risk damaging her relationship with her GP. However, when participants were asked specifically whether they believed leaving negative feedback about a GP would have an impact on their relationship with a GP, most participants believed that GPs were professional, and therefore leaving negative feedback for a GP would not have an impact on their relationship with a GP.

Leaving Their Real Name on a Website

A total of 6 participants said they were happy to leave their real name on a website when they left feedback about a GP on a patient feedback website, because they believed their feedback would be more effective with their name on it, because GPs could then use the feedback for improvement:

> *I always think it is important to [leave one's name], because if you don't, then that person can't get back to you to say how can we improve? Because I always believe it should always be solution focused, so you can't just sit and moan without thinking how it could be improved, both sides really.* [P5]

Furthermore, 1 participant mentioned that he would be happy to leave his real name on negative feedback on a website because he could always see another GP in the practice. However, 7 participants from all age groups were not happy to leave their real name on a website because of privacy concerns, because of their need to remain anonymous, and because they were worried that they could be identified by a GP.

Leaving Information About Their Diagnoses on a Website

More than half of the participants said that if their diagnosis was a commonly occurring diagnosis they would not mind leaving it on a website. However, if it was quite specific or an embarrassing ailment, they would hesitate to leave it:

> *I've had both my hips replaced, I don't mind people knowing that...if it was a very personal issue than probably not. Same with online.* [P4]

Naming a General Practitioner When Leaving Feedback on a Website

The majority of participants were happy to name a GP when they left positive feedback about a GP on a website. However, when leaving negative feedback on a website, participants disagreed as to whether a GP should be named. A total of 4 participants said that feedback would be more useful if a GP is

named. One of the reasons given was that if the patient wants improvement, the GP needs to be named in the feedback, especially if the GP is part of a larger practice:

> *There are 18-20 [GPs] working on the same day [in my GP practice], it's hard to know which doctor you are talking about* [P3]

However, 7 participants felt that it was unfair to name GPs, because the feedback left on a website could damage the GP's reputation or personal confidentiality, and it could just be that the GP was having a bad day:

> *I think they deserve privacy. I live in the public world and I know how that feels, and if I fail I don't necessarily need it everywhere, and same with them* [P6]

Theme 7: Accessibility of Patient Rating and Feedback Websites

Almost half of participants (n=7), all younger than 50 years, believed that a website is more accessible because it is available all the time and can be used from anywhere, and therefore it is also easier to use:

> *Yeah, cos you can do it [give feedback] any time. You know you don't have to do it there and then. Or you don't have to go home and come back to collect something paper-based, you can do it at home, 2am in the morning* [P2]

Furthermore, 1 participant, who was younger than 30 years, suggested that giving feedback on a website would make it easier for her to be critical of her GP:

> *I think I would feel more comfortable typing it [i.e., critical feedback] (laugh), it's just, I don't know, I think it's just psychological, I just feel like if I put it down myself [on paper], I wrote it, then it'd be, yeah, I wouldn't feel as comfortable being as expressive that [sic] I'd like to be. Is that weird?* [P3]

However, more than half of the participants (n=11) expressed that a website is less accessible. A total of 4 participants (who were all older than 60 years) said this was because they do not have a computer or they do not know how to navigate a website, whereas others who were familiar with the Internet and used the Internet felt they did not want to go on a website for nonwork purposes.

Discussion

Principal Findings

In this study, patients as a group are divided about their attitudes toward using feedback and rating websites in the future to leave feedback for a GP. Some patients do not want to leave feedback or do not feel the need to leave feedback in the future (regardless of the method of feedback offered to them), whereas others, who may be willing to leave feedback for a GP, are for or against leaving feedback on a patient feedback website.

The results suggest that some patients may be motivated in the future to leave Web-based feedback rather than paper-based feedback because (1) they can give feedback anytime from anywhere, (2) it allows them to share their experience with the public so others can see what went right or wrong, or (3) they believe that the GP will take Web-based feedback more seriously. On the other hand, however, others suggested that they would not use a website to leave feedback because (1) they cannot use a personal computer or website (mentioned only by participants older than 60 years), (2) they have privacy concerns about leaving feedback on a website, or (3) they believe that feedback left on a website will not be taken seriously by the GP or the practice, because other patients may be abusing the website or using it as a negative breeding ground. These findings can be used by the NHS and patient feedback website providers to effectively target marketing material and address these patient concerns about patient feedback websites that have emerged from this study.

Furthermore, although participants younger than 50 years appeared to perceive giving feedback on a website easier than giving it on paper, this does not mean that they were convinced of the value of giving feedback about a GP on a public feedback website. Privacy and security were important to all of the participants in this study regardless of age, and this suggests that if patients feel a website is not secure enough or will not preserve their anonymity, they will be reluctant to use such a website to leave feedback about GPs. The NHS and other patient feedback website providers need to reassure patients that their websites are secure and will maintain patient privacy.

Comparison With Prior Work

Since 1978, patient and public involvement has been part of NHS policy, and there has been increasing emphasis on collecting patient experience narratives and feedback both in the NHS and outside it [46]. It was surprising, therefore, that half of the participants in this study were not aware they could leave feedback for or about a GP. In addition, the majority of participants were also not aware of the existence of patient feedback websites. However, the latter is in line with findings from a study by Galizzi et al [36] who found that only 15% of a sample of Londoners were aware of doctor rating websites. This is in contrast to the United States and Germany, where recent studies found that approximately a quarter of respondents had used a physician rating website [25,34]. However, Patel et al [47] suggest this may be partly because of the higher usage of private health care in the United States and Germany.

One of the criticisms of the NHS Choices website in England is that its user-driven content is biased and it contains very few numbers of reviews and ratings, which are not representative of a GP or GP practice's performance [47]. This was supported by a study in England, which found that less than 1% of all GP consultations had been reviewed on the Web [7], and studies from the United States [30,48-50], Germany [35,51,52], and Australia [29] all indicated that less than 30% of doctors had been rated on the Web. General practitioners in England also suggested that their patients are not aware about the existence of patient feedback websites [47]. The findings from this study appear to support this. However, they also suggest that the lack of awareness and usage among patients is not limited to patient feedback websites; rather, patients appear to have limited awareness of other feedback methods that are present in GP

practices too. More positively, however, participants also suggested that this could be reversed if the GP or the practice actively asked them to leave feedback about a GP (rather than by just providing tokenistic methods, such as leaving forms at the reception desk), and this may convince them that their feedback, even if it is mediocre feedback, is of some value to the GP or the GP practice for improvement.

Despite the phenomenal increase in Internet usage and ownership of computers in UK households, there is still a digital divide present in society, where 11% of adults in the United Kingdom in 2015 have never used the Internet [53], and 37% of adults aged 65-74 years and 65% of those older than 75 years do not have access to the Internet at home [54]. This was reflected in our study too, where almost all of the participants older than 60 years said they did not have access to a computer or the knowledge to use such websites. This suggests that some parts of society, mainly the elderly, may be excluded from patient feedback websites, and Trigg [8] proposed that this may be a type of social exclusion for those who most need health care and access to such feedback websites. Interestingly though, even among those who did have access to the Internet in this study and who were familiar with the Internet, a few just did not want to use the Internet for purposes outside of work.

Patients in this study felt that their primary motivation to leave feedback for a GP (irrespective of whether it is on a website or on paper) was to help improve GPs' professional practice, and this may explain why many in this study preferred to leave feedback directly with the GP or practice, because they believed the GP could then make the necessary changes. This type of motivation is described as "helping the organization" by researchers in the field of consumer behavior, who explore what motivates people to communicate positive and negative sentiments through word-of-mouth about consumer products [55,56]. However, the difference is that this type of motivation was attributed to positive feedback only, whereas in this study, patients attributed it to negative feedback too. This also appears to dismiss the concerns raised in the literature [7,57-59], and by GPs in a previous study [47], that some patients have malicious intentions when they leave feedback on a website.

Two additional perceived patient motivations for leaving feedback on a website were found in this study, and these were exclusive to leaving feedback on a website for GPs. The first of these perceived motivations would fall under the term "altruism" described in the field of e-consumer behavior [60]; this was the ability to benefit other patients and organizations by sharing feedback in the public domain, so that (1) it ensures that others do not share the same negative experience and (2) other patients can use the reviews to decide which GP to see or which GP practice to join. The latter has been part of the "patient choice" agenda in the NHS [11,28], and the NHS argues that this type of "choice" will drive improvement and empower patients [61]. More than half of the participants in this study spoke positively of this advantage; however, there has been considerable criticism of the choice agenda in the literature [12,62].

The second perceived patient motivation to leave Web-based patient feedback mentioned in this study was its collective power to force improvement. This exercising of power over an organization has also been described by Yoo and Gretzel [63] as a motivator for people leaving Web-based travel reviews. Similarly, Ben Bradshaw, a former British Minister for Health, argued that Web-based patient feedback will force doctors to improve their performance and bedside manner out of fear that patients may post on the Web about them [64,65]. However, the majority of GPs in a previous study [47] disagreed that this would bring about a positive change; rather, they believed it would just force GPs to practice more defensively. Davidson et al [66] also found that just because stories about the quality of services appeared in the public domain and affected an organization's reputation, this did not mean that they would automatically become drivers for improvement in the NHS. Furthermore, 1 participant in this study highlighted that leaving feedback on a website, she believes, will not be taken more seriously by the GP as the feedback may be looked at with skepticism, because patient feedback websites may be seen as negative breeding grounds by GPs. This appears to be supported by some GPs in a previous study [47] who saw little value in Web-based patient feedback and had concerns about them.

Patients' views about leaving their name on feedback that they would leave on a website in the future were found to be mixed. On the one hand, some patients had concerns about privacy, whereas others suggested the feedback would be more useful to GPs if they as patients left their name on it; 7 GPs in a previous study [47] also believed the same. Similarly, views were mixed about whether GPs should be named in the feedback provided, and a previous study [47] found that GPs preferred to receive practice-based feedback, where they as GPs would not be named by the patient on the feedback left on a website. However, 4 participants in this study believed that feedback would be more useful if the GP is named, because there is no other way to identify the GP, especially if the GP is part of a larger practice. In a study by Patel et al [47], GPs similarly questioned the usefulness of a piece of feedback if it was anonymous to GP and the patient, and remarked that it was difficult to work out who the comment was for and about and therefore could not be used for improvement.

Findings from this study suggest that there is no single most preferred method for patients to give feedback about a GP, and Entwistle et al [67] also found the same in their study with Scottish patients. However, in this study, giving feedback directly to the GP and the practice was the most preferred way for the majority of the participants to leave feedback. This is significant, because it appears to suggest that some patients do not feel the need to formalize the feedback they give about a GP. The results also appear to suggest that if patients feel heard within the practice, they may be less likely to seek out other external ways to leave feedback.

The results from this study also suggest that patients will change their method of giving feedback based on the type of feedback they want to leave (negative or positive) and the type of experience they have. This is significant because it suggests that patient feedback left on a website for a GP—that other patients can then use to make a "choice" of provider—may very well be biased, because it appears to be that patients pick and choose which type of feedback they give on a website and

which, for example, they directly tell their GP after a consultation.

All of the participants in this study said that they would consider leaving feedback for a GP (Web-based or using another method) in the future when they had experienced an extreme experience, mainly an extreme negative experience. This appears to support the argument made by GPs in a previous study [47] as well as physician representatives [68,69] that the majority of Web-based patient feedback is extreme negative opinion. This is usually counteracted in the literature with the statement that studies (including [7,10,20,23,30,35,48,59,70-72]) in and out of the United Kingdom have found that the majority of feedback left on physician review websites is positive [73]. The findings from this study appear to contradict that and further suggest that regardless of whether patient feedback is given on a website or not, patients are much more likely to leave feedback when they have experienced an extreme negative experience.

Most participants in this study felt quite comfortable giving negative feedback directly to the GP, and they did not believe leaving negative feedback for a GP would have an impact on their relationship with a GP. This contradicts Dorr and Lipkin's [74] stance that the doctor-patient relationship is "sacred" and therefore patients would not risk jeopardizing that relationship. However, it appears to support the argument by Kaba and Sooriakumaran [75] that the one-sided power in a doctor-patient relationship is swiftly shifting in the United Kingdom, and the push for patient-centered care means that both parties are now more likely to be involved in decision-making processes.

Conclusions and Recommendations

The findings of this study appear to suggest that the current low usage of patient feedback websites in England may be partially because many patients do not know that they can leave feedback at all about GPs, Web-based or otherwise, and within the group that does know about leaving feedback for GPs, some do not want to leave feedback, regardless of which method of feedback is offered to them. This is in part because they are not convinced that GPs want patient feedback or need patient feedback. However, the findings also suggest that those patients who do want to leave feedback about a GP would choose the method based on the following: the type of feedback they want to give, whether that particular method of giving feedback was convenient for them, whether they believed the feedback method was secure and appropriate to use, and whether they believed that the feedback would reach the GP using that method and would be used for improvement. These generic factors (found in this study) associated with preference of feedback method may be used by the NHS and other health providers to evaluate whether proposed new methods to collect patient feedback are appropriate and will be effective.

The findings also suggest that patient feedback websites as they currently are will not replace other mechanisms for patients to give feedback to a GP, but they may motivate a small number of patients who have more altruistic motives or wish to place collective pressure on a GP to give feedback on the website. If the NHS or GPs want more patients to leave feedback on the website, the findings suggest they first make patients aware that they can leave anonymous feedback securely on a website for a GP. They could then convince them actively that their feedback is needed and wanted by GPs for improvement and that the reviews they leave on a website will be of benefit to other patients to decide which GP to see or GP practice to join. The findings also suggest that some patients may prefer to give feedback using a Web-based method because it is easier and more accessible, but at the same time they may want their feedback to remain private for the GP or GP practice to view only. Future research will explore this and examine whether the other findings from this study can be found at a population level in England.

Limitations

Findings from this study provide valuable insight into patients' views and motivations toward Web-based patient feedback in the context of primary care. However, the findings need to be used with some caution because, even though the data appeared to reach thematic saturation, the sample size for this study was small (n=18), and participants were recruited from 4 locations only. Therefore, it is difficult to conclude to what extent findings can be found in the general population of patients. However, the findings are useful for scoping further research, and future research will examine to what extent findings from this study can be found at a population level in England.

Acknowledgments

We would like to thank all the participants who took part in this study. The research was funded by the Engineering and Physical Sciences Research Council under the Participation in Healthcare Environment Engineering program (#EP/H022031/1).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Key stakeholders of Web-based patient feedback for general practitioners in England (as of April 2015).

[PDF File (Adobe PDF File), 40KB - jmir_v18i8e217_app1.pdf]

Multimedia Appendix 2

The different patient feedback websites in England as of April 2015.

[PDF File (Adobe PDF File), 46KB - jmir_v18i8e217_app2.pdf]

Multimedia Appendix 3

Characteristics of websites in England (as of April 2015) where patients can leave feedback for general practitioners publicly.

[PDF File (Adobe PDF File), 31KB - jmir_v18i8e217_app3.pdf]

Multimedia Appendix 4

Topic guide used for the interviews.

[PDF File (Adobe PDF File), 62KB - jmir_v18i8e217_app4.pdf]

References

1. Gillam S, Newbould J. Patient participation groups in general practice: what are they for, where are they going? BMJ 2016;352:i673. [doi: 10.1136/bmj.i673] [Medline: 26893458]
2. Schuckert M, Liu X, Law R. Insights into Suspicious Online Ratings: Direct Evidence from TripAdvisor. Asia Pacific Journal of Tourism Research 2015 Apr 21;21(3):259-272 [FREE Full text] [doi: 10.1080/10941665.2015.1029954]
3. Li H, Ye Q, Law R. Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. Asia Pacific Journal of Tourism Research 2013 Oct;18(7):784-802. [doi: 10.1080/10941665.2012.708351]
4. Filieri R. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. Journal of Business Research 2015 Jun;68(6):1261-1270. [doi: 10.1016/j.jbusres.2014.11.006]
5. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
6. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter R, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
7. Greaves F, Pape U, Lee H, Smith D, Darzi A, Majeed A. Patients' Ratings of Family Physician Practices on the Internet: Usage and Associations With Conventional Measures of Quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280]
8. Trigg L. Patients' opinions of health care providers for supporting choice and quality improvement. J Health Serv Res Policy 2011 Apr;16(2):102-107. [doi: 10.1258/jhsrp.2010.010010] [Medline: 20870690]
9. Powell J, Boylan A, Greaves F. Harnessing patient feedback data: A challenge for policy and service improvement. Digital Health 2015 Nov 30;1:1-3 [FREE Full text] [doi: 10.1177/2055207615617910]
10. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. Jt Comm J Qual Patient Saf 2013 Jan;39(1):7-15. [Medline: 23367647]
11. NHS Choices. Comments policy - Terms and conditions - NHS Choices Internet URL: http://www.nhs.uk/aboutNHSChoices/aboutnhschoices/termsandconditions/Pages/commentspolicy.aspx [accessed 2015-01-30] [WebCite Cache ID 6jRN2TXp6]
12. Fotaki M. Centre for the Health and Public Interest. UK; 2014. What market-based patient choice can't do for the NHS?: The theory and evidence of how choice works in health care URL: http://chpi.org.uk/wp-content/uploads/2014/03/What-market-based-patient-choice-cant-do-for-the-NHS-CHPI.pdf [accessed 2015-11-07] [WebCite Cache ID 6crZO3XgL]
13. Powell T. UK briefing papers. 2015. General practice in England URL: http://www.ukbriefingpapers.co.uk/briefingpaper/SN06906 [accessed 2016-08-01] [WebCite Cache ID 6jRNQT8jK]
14. King D, Zaman S, Zaman S, Kahlon G, Naik A, Jessel A, et al. Identifying Quality Indicators Used by Patients to Choose Secondary Health Care Providers: A Mixed Methods Approach. JMIR Mhealth Uhealth 2015;3(2):e65 [FREE Full text] [doi: 10.2196/mhealth.3808] [Medline: 26048441]
15. Greaves F, Millett C, Nuki P. England's Experience incorporating "anecdotal" reports from consumers into their national reporting system: lessons for the United States of what to do or not to do? Med Care Res Rev 2014 Oct;71(5 Suppl):65S-80S. [doi: 10.1177/1077558714535470] [Medline: 24836765]
16. Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]
17. Greaves F, Laverty A, Ramirez-Cano D, Pulman S, Moilanen K, Darzi A, et al. Analysis of patients' comments about hospitals in the English NHS via Twitter, and comparison with patient surveys. BMJ Qual Saf 2014 Mar 17;23(4):348 [FREE Full text] [doi: 10.1136/bmjqs-2014-002893.5]

18. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Machine learning and sentiment analysis of unstructured free-text information about patient experience online. The Lancet 2012 Nov;380:S10 [FREE Full text] [doi: 10.1016/S0140-6736(13)60366-9]
19. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]
20. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
21. Bardach N, Asteria-Peñaloza R, Boscardin W, Dudley R. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
22. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015;15(1):416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
23. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
24. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
25. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
26. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
27. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
28. Adams S. POST-PANOPTIC SURVEILLANCE THROUGH HEALTHCARE RATING SITES. Information, Communication & Society 2013 Mar;16(2):215-235 [FREE Full text] [doi: 10.1080/1369118X.2012.701657]
29. Atkinson S. Current status of online rating of Australian doctors. Aust J Prim Health 2014;20(3):222-223. [doi: 10.1071/PY14056] [Medline: 24852125]
30. Gao G, McCullough J, Agarwal R, Jha A. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
31. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
32. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]
33. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
34. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
35. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
36. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):- Page no. not available [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
37. The Tavistock Institute. Potential benefits of web-based feedback for improving health service quality: Lessons emerging from the evaluation of Patient Opinion website Internet URL: http://www.tavinstitute.org/projects/potential-benefits-of-web-based-feedback-for-improving-health-service-quality-lessons-emerging-from-the-evaluation-of-patient-opinion-website/ [accessed 2016-08-02] [WebCite Cache ID 6jReWzk47]
38. Tracy S. Qualitative research methodscollecting evidence, crafting analysis, communicating impact. UK: Wiley-Blackwell; 2012.
39. Bryman A. Social research methods. Oxford: Oxford University Press; 2008.
40. Matthews B, Ross L. Research Methods: A Practical Guide for the Social Sciences. UK: Longman; 2010.
41. Brown H, Davidson D, Ellins J. University of Birmingham. 2009. Real-time Patient Feedback Project URL: http://www.birmingham.ac.uk/Documents/college-social-sciences/social-policy/HSMC/publications/2011/real-time-patient-feedback.pdf [accessed 2016-08-02] [WebCite Cache ID 6jReq0ju6]
42. Silva D. The Health Foundation. 2013. Measuring the patient experience URL: http://www.health.org.uk/publication/measuring-patient-experience [accessed 2016-08-01] [WebCite Cache ID 6jRf7XEjE]

43. Coulter A. Picker Institute Europe. Oxford: Picker Institute; 2006. Engaging patients in their healthcare URL: http://www.pickereurope.org/wp-content/uploads/2014/10/Engaging-patients-in-their-healthcare-how-is-the-UK-doing....pdf [accessed 2016-08-02] [WebCite Cache ID 6jRfHWYXu]
44. O'Reilly M, Parker N. 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qualitative Research 2012 May 17;13(2):190-197 [FREE Full text] [doi: 10.1177/1468794112446106]
45. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. USA: SAGE Publications, Inc; 2012.
46. Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. Patient 2014;7(3):235-241. [doi: 10.1007/s40271-014-0060-5] [Medline: 24831941]
47. Patel S, Cain R, Neailey K, Hooberman L. General Practitioners' Concerns About Online Patient Feedback: Findings From a Descriptive Exploratory Qualitative Study in England. J Med Internet Res 2015;17(12):e276 [FREE Full text] [doi: 10.2196/jmir.4989] [Medline: 26681299]
48. Lagu T, Hannon N, Rothberg M, Lindenauer P. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
49. Mostaghimi A, Crotty B, Landon B. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]
50. Ellimoottil C, Hart A, Greco K, Quek M, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
51. Emmert M, Maryschok M, Eisenreich S, Schöffski O. [Websites to assess quality of care--appropriate to identify good physicians?] [Article in German]. Gesundheitswesen 2009 Apr;71(4):e18-e27. [doi: 10.1055/s-0028-1103288] [Medline: 19221986]
52. Strech D, Reimann S. Der Status quo ihrer Bewertungskriterien, Bewertungstendenzen und Nutzung: German Language Physician Rating Sites: The Status Quo of Evaluation Criteria, Evaluation Tendencies and Utilization. Gesundheitswesen 2012;74:502-503 [FREE Full text] [doi: 10.1186/1472-6963-10-332]
53. Office for National Statistics. UK: ONS; 2015. Internet Users, 2015 URL: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171778_404497.pdf [accessed 2016-08-02] [WebCite Cache ID 6jRgRutSL]
54. Ofcom. 2015. Adults' media use and attitudes: Report 2015 Internet URL: http://stakeholders.ofcom.org.uk/market-data-research/other/research-publications/adults/media-lit-10years/ [accessed 2016-08-02] [WebCite Cache ID 6jRjMTWML]
55. Sundaram D, Mitra K, Webster C. Word-of-Mouth Communications: A Motivational Analysis. Advances in Consumer Research 1998;527:532.
56. Trigg L. LSE Health and Social Care. 2012. Using Online Reviews in Social Care URL: http://eprints.lse.ac.uk/46483/1/__libfile_REPOSITORY_Content_PSSRU_Discussion%20Papers_DP2836.pdf [accessed 2016-08-02] [WebCite Cache ID 6jRgpEzoP]
57. Segal J. The role of the Internet in doctor performance rating. Pain Physician 2009;12(3):659-664 [FREE Full text] [Medline: 19461833]
58. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. JAMA 2010 Oct 20;304(15):1711-1712. [doi: 10.1001/jama.2010.1499] [Medline: 20959582]
59. Kadry B, Chu L, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
60. Parikh A, Behnke C, Vorvoreanu M, Almanza B, Nelson D. Motives for reading and articulating user-generated restaurant reviews on Yelp.com. JHTT 2014 Aug 12;5(2):160-176 [FREE Full text] [doi: 10.1108/JHTT-04-2013-0011]
61. Department of Health, UK. London: Crown Publications; 2003. Building on the Best: Choice, Responsiveness and Equity in the NHS URL: https://www.gov.uk/government/publications/building-on-the-best-choice-responsiveness-and-equity-in-the-nhs [accessed 2016-08-02] [WebCite Cache ID 6jRjFTSN7]
62. Dixon A, Robertson R, Appleby J, Burge P, Devlin N, Magee H. Kingsfund. London: The King's Fund; 2010. How patients choose and how providers respond Internet URL: http://www.kingsfund.org.uk/publications/patient-choice [accessed 2016-08-02] [WebCite Cache ID 6jRhWj3on]
63. Yoo K, Gretzel U. What Motivates Consumers to Write Online Travel Reviews? Information Technology & Tourism 2008 Dec 01;10(4):283-295. [doi: 10.3727/109830508788403114]
64. Symons J. The Sun. 2008 Dec 31. Top of the docs NHS ratings URL: http://www.thesun.co.uk/sol/homepage/woman/health/health/article2083788.ece [accessed 2012-04-19] [WebCite Cache ID 6h4pEsz81]
65. Carvel J. Guardian. 2008 Dec 30. Patients to rate and review their GPs on NHS website URL: http://www.theguardian.com/society/2008/dec/30/doctors-rating-website-nhs [accessed 2016-04-27] [WebCite Cache ID 6h4pHvQpq]
66. Davidson D, Joyner O, Drabble D, Cullen J, Hills D. The University of Birmingham. Independence, Engagement, Relationship and Partnering: Essential dimensions in models of web-based feedback mechanisms for quality improvement URL: http://www.birmingham.ac.uk/Documents/college-social-sciences/social-policy/HSMC/research/patient-opinion-final-report.pdf [accessed 2016-08-02] [WebCite Cache ID 6jRiwKmXN]

67. Entwistle VA, Andrew JE, Emslie MJ, Walker KA, Dorrian C, Angus VC, et al. Public opinion on systems for feeding back views to the National Health Service. Qual Saf Health Care 2003 Dec;12(6):435-442 [FREE Full text] [Medline: 14645759]
68. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009 Mar 17;338:b1033 [FREE Full text] [doi: 10.1136/bmj.b1033]
69. Boffey D. The Guardian. London: The Guardian; 2011. GPs' fury over poison pen online tirades URL: http://www.theguardian.com/society/2011/dec/03/gps-poison-pen-online [accessed 2015-11-07] [WebCite Cache ID 6crZfvo3C]
70. Black E, Thompson L, Saliba H, Dawson K, Black N. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253. [Medline: 20359403]
71. Detz A, López A, Sarkar U. Long-term doctor-patient relationships: patient perspective from online reviews. J Med Internet Res 2013;15(7):e131 [FREE Full text] [doi: 10.2196/jmir.2552] [Medline: 23819959]
72. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4] [Medline: 22207014]
73. Verhoef L, Van de Belt T, Engelen L, Schoonhoven L, Kool R. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
74. Dorr GS, Lipkin M. The doctor-patient relationship: challenges, opportunities, and strategies. J Gen Intern Med 1999 Jan;14 Suppl 1:S26-S33 [FREE Full text] [Medline: 9933492]
75. Kaba R, Sooriakumaran P. The evolution of the doctor-patient relationship. Int J Surg 2007 Feb;5(1):57-65 [FREE Full text] [doi: 10.1016/j.ijsu.2006.01.005] [Medline: 17386916]

Abbreviations

GP: general practitioner
NHS: National Health Service

Edited by G Eysenbach; submitted 28.04.16; peer-reviewed by F Alemi, T Lagu; comments to author 19.05.16; revised version received 23.06.16; accepted 11.07.16; published 05.08.16.

Please cite as:
Patel S, Cain R, Neailey K, Hooberman L
Exploring Patients' Views Toward Giving Web-Based Feedback and Ratings to General Practitioners in England: A Qualitative Descriptive Study
J Med Internet Res 2016;18(8):e217
URL: http://www.jmir.org/2016/8/e217/
doi:10.2196/jmir.5865
PMID:27496366

Letter to the Editor

The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors

Jon-Patrick Allem[1*], MA, PhD; Emilio Ferrara[2*], PhD

[1]Keck School of Medicine, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States

[2]Information Sciences Institute, Department of Computer Science, University of Southern California, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Keck School of Medicine
Department of Preventive Medicine
University of Southern California
2001 N. Soto Street, 3rd Floor Mail
Los Angeles, CA, 90032
United States
Phone: 1 8586030812
Fax: 1 3234428201
Email: allem@usc.edu

Related Articles:

Comment on: http://www.jmir.org/2016/2/e41/

Comment in: http://www.jmir.org/2017/6/e165/

(J Med Internet Res 2016;18(8):e219) doi:10.2196/jmir.6185

KEYWORDS

Internet; surveillance; electronic cigarettes; Twitter; social media

In a recent issue of *JMIR*, Kim and colleagues described a framework for data collection, quality assessment, and reporting standards for social media data used in health research [1]. The authors' framework was based on two principles: retrieval precision or "how much of retrieved data is relevant" and retrieval recall or "how much of the relevant data is retrieved." With an in-depth knowledge of the subject matter under investigation, and refinement of the keywords to develop reliable search filters, the authors suggested that irrelevant content could be weeded out and high-quality data collection could be assured. Using the topic of electronic cigarettes (e-cigarettes), discussed on Twitter, as a case study to showcase their framework, the authors demonstrated how reporting standards could be made systematic and transparent. While the authors cogently argued for better reporting standards in social media data used in health research, and their principles regarding retrieval precision and retrieval recall were thoughtfully laid out, they overlooked the importance of identifying the sources of the content being captured during data collection. For example, Twitter has quickly become subject to third party manipulation where automated accounts are created by industry groups and private companies that aim to influence discussions and promote specific ideas or products [2]. This fact is absent from the framework of Kim and colleagues [1] and according to their principle of retrieval precision, researchers could classify tweets about e-cigarettes as high-quality data regardless of its origin.

Recent research has suggested that between 70% and 80% of tweets mentioning e-cigarettes stem from automated accounts [3]. Studies using tweets and that aimed at gaining insights to individual-level attitudes and behaviors are now faced with data with substantial bias and noise. Any results drawn upon this data and not preprocessed with de-noising techniques lose validity and significance. To ignore this bias in Twitter data would be akin to a public health researcher ignoring the bias from having a sample of participants, in a survey-based study on tobacco-related attitudes, where 700 of the 1000 participants happened to be gainfully employed by a tobacco company. The survey researcher would be forced to rethink their sampling frame, and the same dilemma applies to the social media researcher relying on Twitter as their data source. We propose herein that appropriate analyses be implemented to obtain valid data sets that remove sources of bias and noise before applying the framework of Kim and colleagues.

Twitter screen names responsible for each tweet collected in a data set should be obtained and each account's recent history, interactions, and metadata should be analyzed to determine whether the account is a social bot, a computer algorithm designed to automatically produce content and engage with humans on Twitter [2]. These social bots are meant to appear to be individuals operating Twitter accounts that are complete with metadata (name, location, pithy quote) and a photo or an image. Tweets from these accounts pollute social and health research data sets and need to be identified and removed. Programs like "Bot Or Not?" [2] use a classification system that groups each Twitter account's features into 6 main classes: Network (diffusion patterns), User (metadata), Friends (account's contacts), Temporal (tweet rate), and Sentiment (content of message). This classification system ultimately generates a score that falls on a spectrum that can then be used to determine the likelihood of any one account being a social bot. If an account is identified as a social bot then that account and any tweets produced from that account should be removed from the dataset. This platform is freely available, easy to use, and has shown to be successful in reducing bias and noise in datasets from earlier studies led by computer scientists [2].

Using Twitter to examine e-cigarette-related discussion is a novel approach; however, the signal-to-noise ratio has become increasingly low [3]. In other words, the ratio of information representative of individuals' perceptions, sentiments, and behavior is low as compared with the content from social bots. Prior studies have attempted to increase the signal-to-noise ratio by employing crude techniques (eg, removing any tweet that is accompanied by a URL [4]. However, this approach and other blunt approaches (eg, methods solely relying on community detection or methods solely relying on innocent by association paradigms—an account interacting with a human user is considered human) result in misclassification (eg, the removal of a valid tweet from the data set simply because it was accompanied by a URL or keeping an invalid tweet because a human interacted with the account it originated from) [5]. The debiasing techniques available to social media researchers proposed herein can be used to overcome earlier limitations.

Social bots are only one source of bias in studies of Twitter posts. For example, the population of Twitter users over represents young people and ethnic minority groups, when compared to the general population in the United States. This source of bias cannot be easily resolved by machine algorithms and correcting such biases should be a focus of future research. The use of social bots are not confined to discussions of e-cigarettes but have been found to infiltrate political discourse, manipulate the stock market, acquire personal information, and disseminate misinformation [5]. "Bot or Not?" is not a perfect system for bot detection, however, it scores a detection accuracy above 95% suggesting biases from inappropriate removal of legitimate accounts is minimal especially when compared with earlier approaches [5]. Researchers need to take advantage of the resources designed to reliably identify and remove third party accounts responsible for the noise in social media data. Once debiasing techniques have been exploited, frameworks for data collection, quality assessment, and reporting standards for social media data used in health research should be employed.

Acknowledgments

Research reported in this publication was supported by Grant # P50CA180905 from the National Cancer Institute and the FDA Center for Tobacco Products (CTP). The NIH or FDA had no role in study design, collection, analysis, and interpretation of data, writing the report, and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or FDA.

Conflicts of Interest

None declared.

References

1. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. J Med Internet Res 2016;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
2. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. Botornot: A system to evaluate social bots. Presented at: The 25th International Conference Companion on World Wide Web; 2016; Montreal, Canada p. 273-274.
3. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter. PLoS One 2016;11(7):e0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]
4. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
5. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun. ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]

Edited by P Bamidis; submitted 09.06.16; peer-reviewed by A Benton, L Fernandez-Luque; comments to author 15.07.16; accepted 27.07.16; published 09.08.16.

Please cite as:
Allem JP, Ferrara E
The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors
J Med Internet Res 2016;18(8):e219
URL: http://www.jmir.org/2016/8/e219/
doi:10.2196/jmir.6185
PMID:27507563

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org