Journal of Medical Internet Research

Contents

Original Papers

Corrigenda and Addenda

Original Paper

Program Use and Outcome Change in a Web-Based Trauma Intervention: Individual and Social Factors

Zhiyun Wang[1,2], PhD; Jianping Wang[3], PhD, MD; Andreas Maercker[2], PhD, MD

[1]School of Philosophy, Department of Psychology, Wuhan University, Wuhan, China

[2]Department of Psychology, University of Zurich, Zurich, Switzerland

[3]School of Psychology, Beijing Normal University, Beijing, China

Corresponding Author:
Jianping Wang, PhD, MD
School of Psychology
Beijing Normal University
No. 19, XinJieKouWai St., HaiDian District
Beijing, 100875
China
Phone: 86 10 5880 9004
Fax: 86 10 5880 0519
Email: wjphh@bnu.edu.cn

Abstract

Background: Insight into user adherence to Web-based intervention programs and into its relationship to intervention effect is needed.

Objective: The objective of this study was to examine use of a Web-based self-help intervention program, the Chinese version of My Trauma Recovery (CMTR), among Chinese traumatized individuals, and to investigate the relationship between program use and user characteristics before the intervention and change in outcomes after the intervention and at 3-months' follow-up.

Methods: The sample consisted of 56 urban survivors of different trauma types and 90 rural survivors of the 2008 Sichuan earthquake, who used the CMTR in 1 month on their own or guided by volunteers in a counseling center. Predictors were demographics (sex, age, highest education, marital status, and annual family income), health problems (trauma duration, posttraumatic symptoms, and depression), psychological factors (coping self-efficacy), and social factors (social functioning impairment and social support). Program use was assessed by general program usage (eg, number of visiting days) and program adherence (eg, webpages completed in modules). Outcome measures were the Posttraumatic Diagnostic Scale (PDS), Symptom Checklist 90-Depression (SCL-D), Trauma Coping Self-Efficacy scale (CSE), Crisis Support Scale (CSS), and Social Functioning Impairment questionnaire (SFI) adopted from the CMTR.

Results: (1) Program use: rural participants had a larger total number of visiting days ($F_{1,144}=40.50$, $P<.001$) and visited more program modules in 1 month ($\chi^2_3=73.67$, $P<.001$) than urban participants. (2) Predictors and program use: total number of visiting days was correlated with CSS at pretest ($r=.22$, $P=.009$), and total number of completed webpages was associated with SFI at pretest ($r=.19$, $P=.02$). Number of webpages completed in modules was correlated with all demographic, disease severity, psychological, and social factors at pretest. (3) Program use and outcomes change: in general, use of the triggers and self-talk modules showed a consistent positive association with improvement in PDS, SCL-D, SFI, and CSE. The relaxation module was associated with positive change in PDS, but with negative change in CSS and SFI. The professional help module was associated with positive change in SCL-D, but its use on the first day was associated with negative change in CSS and CSE. The unhelpful coping module was associated with negative change in SFI. The mastery tools module showed a consistent association with negative change in PDS and SCL-D.

Conclusions: These findings suggest that both individual (eg, demographic, health problems, psychological) and social factors (eg, social functioning, social support) should be considered when delivering Web-based interventions, particularly in collectivist cultures. Specific program adherence indicators (eg, webpages completed in each module, activity types completed), rather than general program usage indicators (eg, total number or time of visiting), should be developed to examine the effectiveness of various program modules or elements.

Introduction

In recent years, the Internet has been adopted as a valuable tool to deliver physical and mental health services to large populations [1]. Research has revealed significant treatment effects of Web-based intervention programs for a variety of mental disorders, such as depression, anxiety, and posttraumatic stress disorder (PTSD) [2,3]. However, increasing the effectiveness of Web-based intervention programs faces challenges, such as high dropout and poor user adherence reported in previous studies [4,5]. According to Christensen and colleagues [6], dropout refers to a participant not completing the research trial protocol or trial assessments associated with a Web-based intervention; and adherence refers to the extent to which participants experience the content of the Web-based intervention, which is the focus of this study. Further, adherence can be examined by the indicators of general program usage (eg, number of log-ins or time spent on intervention programs) and program adherence (ie, the extent to which intervention programs are used in accordance to recommendations, such as content modules completed or activities completed) [7].

The relationship between user exposure to intervention programs and the effect of Web-based interventions can be complicated [8] and may be a dose-response relationship [9]. Donkin et al [10] found, however, that compared with low program users, medium users showed little additional benefit from Web-based intervention, and suggested that "concentrated use of the program (eg, completing multiple modules per log-in) or passive exposure to material (as measured by modules completed) may not be as useful as regular shorter periods of use with higher levels of activity in each of these log-ins." Thus, deeper insight into the process of Web-based program use and into its relationship to intervention effect is needed.

To improve Web-based interventions for target groups with various physical and mental health problems, previous studies have investigated potential predictors of program use, with mixed findings. One group of predictors is user characteristics, including (1) demographic variables, such as age, sex, education, marital status, and socioeconomic status [7,11], (2) health problems, such as severity of the target disease, duration of the target disease, and subjective health status [12,13], (3) psychological factors, such as illness attitudes and beliefs, expectations, motivation, and self-efficacy [6,14], and (4) social factors, such as family characteristics (eg, parenting practices and styles, socioeconomic status of the family), and support from partners and friends [12,15]. Another group of predictors includes the characteristics of Web-based interventions, such as feedback, interactive elements, email or telephone contact, and reminders [16,17].

Based on these findings, this study investigated the role of demographics, health problems, and psychological and social factors in Chinese traumatized persons' use of a Web-based self-guided intervention program, the Chinese version of My Trauma Recovery (CMTR). The program showed preliminary short-term treatment effects on PTSD and depressive symptoms in a randomized controlled trial in 2 Chinese populations [18]. The trial showed a high dropout rate, however, with 40.8% of participants not completing the research protocol [19-21]. Further analysis revealed that participant dropout was associated with such social factors as needs in trauma disclosure and perceived social acknowledgment or disapproval by family and extended social environments. This indicated a necessity to examine social factors together with other factors for better understanding of program use across cultures.

Note that, although it has been argued that Web-based interventions are potentially beneficial for rural residents to receive mental health help, because they have little access to face-to-face mental health resources [22], few empirical studies have evaluated the use and efficacy of such programs among rural users. More precisely, previous studies primarily focused on Internet users, who might come from rural as well as urban areas, and found that more highly educated, older, and female users were more likely to adhere to Web-based interventions [23]. However, for this study we recruited non-Internet users from rural areas who met the required literacy level for Web-based interventions. These people tend to be older and less educated, to have a lower income, and typically have much less Internet use experience. Volunteers provided them with (minimal) support with Internet service problems in a counseling center so that they could complete and benefit from the CMTR program [18]. The literature suggests that social support and contact would increase the use of Web-based intervention programs, particularly the number and duration of visits [16]. This study thus examined the difference in program use between rural and urban users. It offers implications for future application of Web-based interventions, for example, in psychological and mental health aid for extensive rural populations who are in need of mental health service but have little access even to online mental health resources.

In sum, this study aimed to investigate (1) how urban and rural participants used the CMTR program, including general program usage and program adherence, (2) how program use was related to demographics (ie, sex, age, highest level of education attained, marital status, and annual family income), health problems (ie, PTSD and depressive symptom severity, trauma duration), psychological factors (ie, coping self-efficacy), and social factors (ie, social functioning impairment and social support after trauma) before the intervention, and (3) how program use was

associated with change in outcomes after the treatment and at 3-months' follow-up.

Methods

Materials

The CMTR program was translated, with minimal cultural adaptation of pictures, audio and video segments, and professional hotlines, from the English My Trauma Recovery program (previously referred to as Journey to Trauma Recovery), which aims to improve trauma recovery by increasing individual coping self-efficacy and coping skills [24]. CMTR contains 6 recovery modules offering education and exercises for 6 trauma recovery-related topics: professional help, relaxation, self-talk, social support, triggers, and unhelpful coping [24]. There are 2 other interactive sections: self-test and mastery tools. Specifically, users are encouraged to take the self-test on posttraumatic stress reactions and coping self-efficacy after log-in and once a week during the treatment period. The mastery tools section offers hyperlinks to all exercises in the 6 recovery modules, so that users can easily access them in future. Because few users took the self-test during the treatment period, we report only the data on the 6 recovery modules and the exercise section.

Sample and Procedure

We used a sample of 56 urban and 90 rural participants, which was part of the sample (ie, 90 urban and 93 rural participants who completed the pretest) reported earlier [18]. Specifically, of the 93 rural participants, we excluded 3 for not using the 6 recovery modules and the mastery tools section after log-in. Of the 90 urban participants, we excluded 29 for not logging in and 5 for not using the 7 modules after log-in. Based on chi-square analysis results, the 34 urban dropouts (ie, 29 plus 5) did not differ significantly from the 56 participants in sex, age, marital status, highest education level attained, and family income. Analyses of variance (ANOVAs) revealed significant difference in the Crisis Support Scale (CSS) (dropouts: mean score 1.93, SD 0.85; users: mean score 1.58, SD 0.72; $F_{1,88}$=4.23, P=.04), but no significant differences in the Posttraumatic Diagnostic Scale (PDS), Symptom Checklist 90-Depression (SCL-D), Social Functioning Impairment questionnaire (SFI), or Trauma Coping Self-Efficacy scale (CSE) ($F_{1,88}$ range 0.11–1.28, all P>.26) at pretest.

Urban and rural participants used the program in some different way. The urban participants had experienced a variety of traumatic events (eg, physical assault, unexpected death of someone close, or serious accidents), reported at least two PTSD symptoms in a trauma screening questionnaire, and had sufficient Internet access time (≥360 minutes in 4 weeks). They had Internet access at home or work and thus decided themselves when, where, and how often to visit the program during the 1-month treatment period. Research assistants did not contact them or give reminders until posttest. The rural participants were survivors of the 2008 earthquake in Beichuan county in Sichuan province, and reported at least two PTSD symptoms in the trauma screening questionnaire. Due to lack of personal Internet access, they used the program 5 times (at least 30

minutes per time) in a counseling center's computer room. Some participants used the program more than 5 times because they broke off one or more intervention sessions for personal reasons and then made them up later. Note that research assistants provided support for all urban and rural participants only for technical problems with CMTR. When participants asked for support with their distress or the program content, research assistants first evaluated the participants' needs and, if these were not acute, sent the participants a brief reply that CMTR was a self-help program and they would get further support, if needed, after the follow-up test. Otherwise, the research assistants would interrupt the session for other treatment, which did not happen in this study.

Measures

Posttraumatic Diagnostic Scale

This scale includes 17 PTSD symptom items assessing the frequency of trauma-related symptoms in the past month on a 4-point scale (0=not at all or only 1 time, 3=5 or more times a week or almost always) [25]. The internal consistency of the scale in this study was alpha=.92.

Symptom Checklist 90-Depression

We used the 13-item depression subscale of the Symptom Checklist-90 [26] to measure to what extent participants had been bothered by depressive symptoms in the past month on a 5-point scale, ranging from 0 (not at all) to 4 (extremely). The internal consistency of the scale in this study was alpha=.94.

Social Functioning Impairment

We adopted 4 questions from the My Trauma Recovery program to examine individual functional impairment after trauma experiences. An example question is "To what extent have your reactions to what has happened reduced your ability to complete your normal responsibilities (eg, job, school, home, childcare duties)?" Participants answered the questions on a 5-point scale (0=not at all, 4=extremely). The internal consistency of the questionnaire in this study was alpha=.88.

Crisis Support Scale

This 7-item scale measures received practical support, received sympathy, being able to find someone to talk to about the traumatic experience, and overall satisfaction with received social support after trauma [27]. Example statements are "I can talk to someone who has had a similar experience" and "People are being supportive and empathetic of me." Participants responded to the statements on a 5-point scale (0=not at all, 4=extremely). The internal consistency of the scale in this study was alpha=.86.

Trauma Coping Self-Efficacy Scale

This 10-item scale is a short version of the Coping Self-Efficacy Scale for Trauma [28]. It measures to what extent participants felt capable of coping with PTSD reactions at different assessment points. The 5-point scale ranges from 0 (not at all) to 4 (extremely). The internal consistency of the scale in this study was alpha=.83.

Program Use

Total Number of Days Visiting CMTR

After the first log-in, most participants left CMTR without logging out and then visited the program without logging in next time. It was thus difficult to calculate the number of log-ins or the amount of time spent on CMTR. Thus, we counted as 1 day when a participant visited the program within one 24-hour day. The maximum could be 30 days during the 1-month treatment period.

Total Number of CMTR Webpages Completed

We tracked how many CMTR webpages participants completed during the 1-month treatment period. Because participants were encouraged to use the program as many times as they wanted, webpages were counted repeatedly when repeated use occurred. Thus, the total number of webpages completed could be larger than the number of webpages contained in CMTR.

Number of Modules Visited

The maximum could be 7 modules in this study; that is, 6 recovery modules and the mastery tools module. Note that participants could have visited 1 module but not have completed all the webpages in the module.

Number and Proportion of Webpages Completed in Each Module

We calculated the proportion of each module completed during the treatment period. The number of webpages completed could be larger than the actual total number of webpages in each module due to repeated use.

Number of Modules Visited per Day

We tracked how many modules participants visited on different days through the treatment course. The maximum could be 7 modules.

Number and Proportion of Webpages Completed in Each Module on the First Day

We tracked the proportion of each module that was completed on the first day. The number of webpages completed in each module could be larger than the actual total number of webpages due to repeated use.

Data Analysis

We analyzed the data using IBM SPSS version 22 (IBM Corporation). Chi-square analysis tested urban and rural subsample differences in demographic characteristics. We used 1-way ANOVA to test subsample differences in the pretest scores on PDS, SCL-D, SFI, CSS, and CSE, and in program use indicators. We conducted 1-way ANOVA and correlation analyses to explore the relationship between program use and demographics, health problems, psychological factors, and social factors at pretest. Finally, we conducted linear regression analysis for PDS, SCL-D, SFI, CSS, and CSE posttreatment incremental difference scores (ie, posttest values minus pretest values, then divided by the variance of pretest values). In the linear regression analysis, we entered trauma duration and 6 dummy variables as independent variables in the first block (stepwise method; these were sex female or not, age 26–40 years old or not, family income US $0–4000 or not, marital status of married or not, highest education high middle school/bachelor's degree or not, and sample urban or not) and the indicators of program use in the second block (enter method). Similarly, we conducted linear regression analysis for PDS, SCL-D, SFI, CSS, and CSE 3-month follow-up incremental difference scores (ie, follow-up values minus pretest values, then divided by the variance of pretest values).

Results

Overall Program Use

Table 1 presents the demographic statistics of the 56 urban (38.4%) and 90 rural participants (61.6%). The rural subsample consisted of more female, older, and married participants than did the urban subsample. The rural subsample also reported a lower annual family income and level of education. Table 2 shows subsample means (SD) and correlations of variables at pretest. The urban participants reported higher levels of depression and social functioning impairment, but a lower level of social support, than the rural participants. In addition, the rural participants had a longer trauma duration than the urban participants. Correlation analysis showed that a longer trauma duration was associated with participants' lower level of depression. While PTSD symptoms, depression, and social functioning impairment correlated positively with each other, social support correlated negatively with depression and social functioning and positively with coping self-efficacy.

XSL•FO

RenderX

Table 1. Demographic characteristics of participants in the Chinese version of My Trauma Recovery intervention (N=146; urban: n=56; rural: n=90).

Characteristic	Urban n (%)	Rural n (%)	χ^2	df	P value
Sex					
Female	38 (67.9)	74 (82.2)	3.99	1	.046
Male	18 (32.1)	16 (17.8)			
Age range (years)					
16–25	29 (51.8)	1 (1.1)	63.98	2	<.001
26–40	23 (41.1)	41 (45.6)			
41–70	4 (7.1)	48 (53.3)			
Annual family income ($US)					
0–4000	17 (30.4)	81 (90)	54.94	2	<.001
4001–10,000	22 (39.3)	5 (5.6)			
≥10,001	13 (23.2)	2 (2.2)			
Missing	4 (7.1)	2 (2.2)			
Marital status					
Single	43 (76.8)	5 (5.6)	79.37	1	<.001
Married	13 (23.2)	85 (94.4)			
Education					
Junior middle school/lower	1 (1.8)	64 (71.1)	94.36	2	<.001
High middle school	7 (12.5)	19 (21.1)			
Bachelor's degree/higher	48 (85.7)	7 (7.8)			

Table 2. Subsample difference and correlations at pretest in the Chinese version of My Trauma Recovery intervention (N=146; urban: n=56; rural: n=90).

Test	Subsample scores, mean (SD)		$F_{1,144}$	P value	PDS[a]	P value	SCL-D[b]	P value	SFI[c]	P value	CSS[d]	P value	CSE[e]	P value
	Urban	Rural												
PDS	1.70 (0.57)	1.70 (0.51)	0.00	.99	1									
SCL-D	2.53 (0.83)	2.13 (0.81)	7.91	.006	.68	<.001	1							
SFI	2.65 (0.95)	2.11 (0.96)	10.96	.001	.64	<.001	.62	<.001	1					
CSS	1.63 (0.75)	2.26 (0.55)	34.40	<.001	−.13	.12	−.31	<.001	−.24	.004	1			
CSE	1.99 (0.67)	1.91 (0.53)	0.75	.39	−.08	.31	−.06	.50	.04	.62	.32	<.001	1	
DUR[f]	30.16 (45.71)	49.66 (4.24)	16.22	<.001	−.14	.09	−.21	.01	−.06	.45	.15	.07	−.06	.47

[a]PDS: Posttraumatic Diagnostic Scale.

[b]SCL-D: Symptom Checklist 90-Depression scale.

[c]SFI: Social Functioning Impairment.

[d]CSS: Crisis Support Scale.

[e]CSE: Trauma Coping Self-Efficacy scale.

[f]DUR: Trauma duration (in months).

XSL•FO
RenderX

On average, the total number of days visiting CMTR was 3.50 (SD 2.82; minimum 1, maximum 12) among urban participants, of whom 80% (45/56) visited the modules for ≤4 days. The total number of days was 5.51 (SD 0.81; minimum 4, maximum 8) among rural participants, of whom 87% (78/90) used the modules for 5 or 6 days. The subsample difference reached significance ($F_{1,144}$=40.50, P<.001). Urban and rural participants did not differ in the total number of CMTR webpages completed (urban: mean 88.05, SD 76.86; rural: mean 100.13, SD 14.14; $F_{1,144}$=2.12, P=.15).

Module Use

On average, the number of modules visited was 4.63 (SD 2.17) for urban participants and 5.48 (SD 0.66) for rural participants (see Table 3). Chi-square analysis, with categories 1 to 4 combined, revealed a significant subsample difference (χ^2_3=73.67, P<.001).

Table 4 shows the number of webpages completed in each module during the treatment period (top half). The rural participants completed significantly more webpages in the relaxation, unhelpful coping, professional help, and social support modules, but fewer webpages in the trauma triggers and mastery tools modules than the urban participants.

Table 3. Number of modules visited among urban and rural participants in the Chinese version of My Trauma Recovery intervention.

Number of modules	Urban (n=56)		Rural (n=90)	
	Frequency	Proportion	Frequency	Proportion
1	4	7.1%		
2	10	17.9%		
3	7	12.5%		
4	4	7.1%	3	3.3%
5	7	12.5%	46	51.1%
6	5	8.9%	36	40.0%
7	19	33.9%	5	5.6%

Table 4. Number and proportion of completed webpages in each module in the Chinese version of My Trauma Recovery intervention.

Module[a]	Urban (n=56)			Rural (n=90)			$F_{1,144}$	P value
	Mean	SD	%	Mean	SD	%		
In 1 month								
Relaxation	4.18	4.99	104.5	6.02	1.69	150.6	10.40	.002
Triggers	7.41	9.80	74.1	2.70	6.19	27.0	12.69	<.001
Coping	8.86	11.83	55.4	17.62	6.79	110.1	32.38	<.001
Help	4.34	6.57	62.0	10.84	4.47	154.9	50.68	<.001
Self-talk	24.34	23.10	93.6	27.42	4.33	105.5	1.52	.22
Support	13.13	14.69	77.2	17.89	7.64	105.2	6.61	.01
Tools	4.86	4.97	69.4	0.60	1.26	8.6	60.12	<.001
On the first day								
Relaxation	1.23	1.94	30.8	1.58	2.74	39.4	0.68	.41
Triggers	1.38	4.30	13.8	0.23	1.48	2.3	5.35	.02
Coping	2.29	5.15	14.3	0.29	2.74	1.8	9.32	.003
Help	0.88	2.85	12.5	3.78	4.65	54.0	17.65	<.001
Self-talk	7.64	14.12	29.4	3.06	8.75	11.8	5.88	.02
Support	4.70	8.15	27.6	1.90	5.74	11.2	5.90	.02
Tools	1.84	2.43	26.3	0.19	0.93	2.7	33.57	<.001

[a]The total number of webpages in the 7 modules (from top to bottom) is 4, 10, 16, 7, 26, 17 and 7.

Program Use per Day

As Figure 1 shows, the number of modules visited by urban participants decreased through the treatment course, while rural participants' module use was relatively stable. Table 4 presents the number of webpages completed in each module on the first day (bottom half). While the rural participants completed significantly more webpages in the professional help module than the urban participants, the urban participants completed more webpages in other modules except the relaxation module on the first day.

Figure 1. Number of modules visited per day by urban (n=56) and rural (n=90) participants in the Chinese version of My Trauma Recovery intervention.
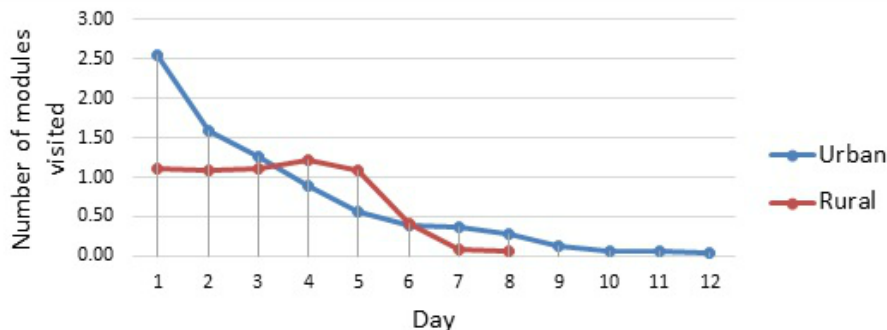
Program Use and User Characteristics at Pretest

A 1-way ANOVA among urban and rural participants revealed no significant sex, age, education, marital status, and family income differences in the total number of visiting days or the total number of CMTR webpages completed (all $P>.12$). As for the number of completed webpages in each module, the urban participants aged 26–40 years completed significantly more webpages in the mastery tools module than those aged 16–25 years ($F_{2,53}=3.39$, $P=.04$). Among rural participants, female participants visited more webpages in the professional help module than did male participants ($F_{1,88}=5.02$, $P=.03$), and married participants visited more webpages in the relaxation module ($F_{1,88}=5.11$, $P=.03$). In addition, correlation analysis showed no significant relationship between trauma duration and the 3 program use indicators in each subsample.

Table 5 presents the correlations between PDS, SCL-D, SFI, CSS, and CSE pretest scores and the program use indicators.

The total number of days was positively correlated with CSS score, and total number of webpages completed was positively correlated with SFI score. The number of webpages completed in the modules during the treatment period showed a positive association with PDS, SCL-D, SFI, and CSE scores, but a negative correlation with the CSS score. In general, these variables showed more significant correlations with the number of webpages completed in the modules on the first day than the number completed in 1 month.

Program Use and Outcomes Change at Posttest and Follow-Up

Table 6 shows the results of regression analysis of PDS, SCL-D, SFI, CSS, and CSE posttreatment and follow-up incremental difference scores on the number of webpages completed in each module in 1 month. Table 7 showed the results of regression analysis when the number of webpages completed in each module on the first day was entered in the second block as an independent variable.

Table 5. Correlations of program use and variable scores at pretest in the Chinese version of My Trauma Recovery intervention (N=146).

Program use indicator	PDS[a]	P value	SCL-D[b]	P value	SFI[c]	P value	CSS[d]	P value	CSE[e]	P value
Total number of days	−.04	.65	−.06	.47	.02	.82	.22	.009	.06	.51
Total number of webpages completed	.11	.19	.08	.33	.19	.02	−.03	.76	.13	.11
Number of pages completed in each module in 1 month										
Relaxation	.15	.07	.11	.19	.11	.18	.10	.25	.13	.11
Triggers	.04	.62	.15	.08	.18	.03	−.07	.38	.07	.43
Coping	.08	.37	−.09	.29	−.01	.92	.12	.15	.05	.53
Help	.20	.02	.10	.23	.18	.03	.15	.08	.04	.66
Self-talk	.08	.33	.03	.72	.15	.08	.01	.89	.06	.48
Support	.03	.69	.03	.71	.07	.39	−.07	.39	.16	.06
Tools	.10	.24	.20	.02	.32	<.001	−.29	<.001	.17	.04
Number of pages completed in each module on the first day										
Relaxation	.02	.78	−.03	.69	−.02	.86	.13	.11	.01	.86
Triggers	.18	.03	.25	.002	.24	.004	−.13	.13	−.05	.55
Coping	.19	.02	.23	.006	.20	.02	−.21	.01	−.01	.92
Help	.15	.08	.06	.51	−.01	.86	.22	.009	.05	.55
Self-talk	.08	.33	.17	.04	.15	.07	−.22	.006	−.23	.006
Support	.15	.07	.23	.005	.12	.16	−.19	.02	.07	.38
Tools	.15	.08	.24	.004	.22	.009	−.29	<.001	.02	.81

[a]PDS: Posttraumatic Diagnostic Scale.

[b]SCL-D: Symptom Checklist 90-Depression.

[c]SFI: Social Functioning Impairment.

[d]CSS: Crisis Support Scale.

[e]CSE: Trauma Coping Self-Efficacy scale.

Table 6. Regressions on the number of webpages completed in each module in 1 month by participants in the Chinese version of My Trauma Recovery intervention (N=146).

Module	PDS[a]			SCL-D[b]			SFI[c]			CSS[d]			CSE[e]		
	b	Beta	*P* value	b	Beta	*P* value	b	Beta	*P* value	b	Beta	*P* value	b	Beta	*P* value
Posttreatment incremental difference (n=122)															
Relaxation	–0.01	–0.02	.89	–0.03	–0.08	.47	1.34	0.22	.04	–0.01	–0.04	.75	–0.06	–0.16	.17
Triggers	0.01	0.04	.71	0.01	0.04	.71	–0.16	–0.06	.55	–0.01	–0.05	.67	0.01	0.07	.55
Coping	0.01	0.10	.40	0.02	0.14	.22	0.50	0.23	.04	0.00	0.002	.98	–0.02	–0.13	.29
Help	–0.01	–0.05	.67	–0.05	–0.25	.04	0.11	0.03	.78	–0.003	–0.02	.90	0.03	0.14	.26
Self-talk	–0.02	–0.26	.047	–0.02	–0.23	.06	–0.48	–0.33	.005	0.01	0.08	.54	0.01	0.13	.31
Support	0.01	0.10	.39	–0.01	–0.11	.31	0.01	0.004	.97	0.00	–0.003	.98	–0.00	–0.04	.73
Tools	–0.02	–0.05	.65	0.09	0.28	.046	0.001	0.00	.999	0.04	0.10	.50	–0.03	–0.08	.60
R^2	.05			.17			.21			.09			.15		
Follow-up incremental difference (n=113)															
Relaxation	–0.04	–0.10	.38	–0.02	–0.05	.62	1.35	0.17	.09	–0.08	–0.21	.07	–0.07	–0.16	.19
Triggers	0.02	0.12	.28	0.02	0.10	.35	0.18	0.05	.58	–0.01	–0.03	.78	0.05	0.24	.04
Coping	–0.01	–0.09	.48	0.00	0.03	.79	0.57	0.21	.049	–0.01	–0.05	.69	–0.01	–0.04	.75
Help	0.03	0.13	.32	–0.01	–0.06	.63	–0.42	–0.10	.37	–0.01	–0.05	.71	0.00	0.003	.98
Self-talk	–0.01	–0.07	.59	–0.02	–0.17	.15	–0.79	–0.39	<.001	0.01	0.13	.29	0.01	0.08	.52
Support	–0.01	–0.10	.39	–0.01	–0.05	.63	–0.22	–0.09	.36	0.01	0.10	.36	–0.01	–0.07	.54
Tools	0.08	0.26	.07	0.10	0.31	.02	0.17	0.02	.81	–0.03	–0.08	.55	–0.02	–0.06	.61
R^2	.17			.26			.28			.27			.07		

[a]PDS: Posttraumatic Diagnostic Scale.

[b]SCL-D: Symptom Checklist 90-Depression.

[c]SFI: Social Functioning Impairment.

[d]CSS: Crisis Support Scale.

[e]CSE: Trauma Coping Self-Efficacy scale.

XSL•FO
RenderX

Table 7. Regressions on the number of webpages completed in each module on the first day by participants in the Chinese version of My Trauma Recovery intervention (N=146).

Module	PDS[a]			SCL-D[b]			SFI[c]			CSS[d]			CSE[e]		
	b	Beta	P value	b	Beta	P value	b	Beta	P value	b	Beta	P value	b	Beta	P value
Posttreatment incremental difference (n=122)															
Relaxation	–0.06	–0.13	.18	0.00	0.01	.93	0.58	0.07	.42	–0.10	–0.20	.04	–0.02	–0.03	.73
Triggers	–0.04	–0.10	.49	–0.04	–0.10	.46	–1.53	–0.21	.07	–0.00	–0.01	.97	0.04	0.09	.47
Coping	0.01	0.04	.77	–0.001	–0.002	.99	1.35	0.27	.014	0.03	0.10	.41	–0.01	–0.04	.76
Help	–0.02	–0.09	.36	–0.05	–0.23	.03	–0.04	–0.01	.93	–0.06	–0.23	.03	–0.05	–0.17	.09
Self-talk	–0.01	–0.05	.62	–0.02	–0.17	.09	–0.33	–0.18	.06	0.01	0.06	.56	0.02	0.21	.04
Support	–0.00	–0.01	.93	–0.01	–0.05	.63	0.11	0.04	.71	–0.02	–0.12	.26	–0.01	–0.06	.54
Tools	–0.02	–0.02	.86	0.09	0.13	.28	1.92	0.16	.15	0.04	0.05	.71	–0.11	–0.15	.21
R^2	.04			.15			.22			.14			.18		
Follow-up incremental difference (n=113)															
Relaxation	–0.10	–0.21	.06	–0.03	–0.07	.50	1.22	0.12	.21	–0.07	–0.16	.12	–0.08	–0.14	.19
Triggers	0.04	0.04	.68	0.07	0.08	.35	–0.10	–0.01	.95	–0.08	–0.09	.31	–0.05	–0.04	.65
Coping	0.02	0.05	.58	–0.02	–0.04	.68	1.22	0.16	.08	0.01	0.01	.89	0.01	0.02	.88
Help	–0.04	–0.16	.18	–0.01	–0.03	.76	0.24	0.04	.69	–0.04	–0.14	.21	–0.07	–0.21	.045
Self-talk	–0.01	–0.11	.28	–0.02	–0.18	.06	–0.08	–0.03	.74	0.01	0.11	.24	0.02	0.14	.17
Support	–0.01	–0.04	.70	–0.02	–0.09	.36	0.07	0.01	.89	0.02	0.11	.27	0.01	0.04	.73
Tools	0.09	0.11	.37	0.07	0.09	.45	0.47	0.03	.78	–0.10	–0.12	.29	0.01	0.01	.91
R^2	.15			.23			.19			.28			.08		

[a]PDS: Post-traumatic Diagnostic Scale.

[b]SCL-D: Symptom Checklist 90-Depression.

[c]SFI: Social Functioning Impairment.

[d]CSS: Crisis Support Scale.

[e]CSE: Trauma Coping Self-Efficacy scale.

As shown, after controlling for demographic variables, subsample, and trauma duration, use of the relaxation (on the first day), professional help (in 1 month, on the first day), and self-talk (in 1 month, on the first day) modules was associated with score reductions in the PDS and SCL-D. However, use of the mastery tools module in 1 month was unexpectedly associated with score increases in PDS and SCL-D.

In addition, use of the triggers (in 1 month) and self-talk (on the first day) modules was associated with score increases in CSE, but use of the professional help module on the first day was associated with a score reduction in CSE.

As for social variables, use of the triggers (on the first day) and self-talk (in 1 month, on the first day) modules was associated with score reductions in SFI. However, use of the relaxation (in 1 month) and unhelpful coping (in 1 month, on the first day) modules was associated with score increases in SFI. Also, use of the relaxation (in 1 month, on the first day) and professional help (on the first day) modules was associated with score reductions in CSS.

Discussion

Main Findings

Program Use

The urban and rural participants used the CMTR program for 1 month in different ways and showed different patterns of program use. In general, the rural participants used the program for more days, visited more modules, and completed more webpages in most modules than the urban participants in 1 month. However, the rural participants may actually have been passive users compared with the urban participants. First, they completed a much smaller proportion of the mastery tools module, which contains all of the tools and skills exercises. Second, while the urban participants quickly decreased the number of modules they visited from the second day, which is consistent with previous findings [29], the rural participants visited on average 1 module in each session. This might have been due to the prescribed intervention procedure among these rural users. In this sense, the program adherence indicator of number of webpages completed in each module is more informative than other general program use indicators.

It is important to note that the relatively passive program use among rural participants might have been due to their lack of Internet use experience (eg, being less skilled at module switching) rather than lack of motivation. The data revealed that the rural participants completed a larger proportion of the professional help module on the first day, as well as during the treatment period, which may indicate a motivation to receive interventions. If this is the case, future studies could effectively increase the program use in rural populations by improving their Internet use skills before applying interventions.

User Characteristics and Program Use

The total number of visiting days and completed webpages showed a positive association only with social factors. That is, users with higher social support and social functioning impairment tended to use the program more. As for the number of webpages completed in the modules, demographic, disease severity, and psychological and social factors all showed some significant correlations. Consistent with previous findings [23], the urban participants aged 26–40 years completed more webpages in the mastery tools module than those aged 16–25 years. Among rural participants, female and married participants visited more webpages in different modules than did male and single participants.

PTSD symptoms and depression before treatment were positively related to module use, particularly to that on the first day. This suggests that disease severity has a motivational role in program use, which is consistent with previous findings [6]. Moreover, the better correlation on the first day supports the CMTR program's original intention: that users will be directed to those modules that may be most useful for them based on their self-test results. These findings thus might help us to learn about users' needs and then improve the intervention efficacy even with limited program use. Given that a large number of individuals dropped out after a few sessions but still benefited from Web-based interventions [18,30], it may be valuable in future research to design short-term as well as long-term use patterns of such interventions in order to benefit more users.

While social functioning impairment was positively related to module use, similar to disease severity variables, our findings on the other social factor (ie, social support) were more complex. Social support at pretest was positively associated with the total number of days visiting CMTR, but negatively related to the use of most modules, except for the professional help module, which introduces what happens in face-to-face professional counseling. These findings indicated that, on the one hand, individuals with less social support might be less interested in seeking face-to-face professional help, and be motivated to use Web-based intervention programs. On the other hand, more social support might provide resources for individuals to insist on using a self-help intervention program.

The motivational role and resource role of social support were also reported in previous studies [31]. For example, Yli-Uotila et al [32] found that patients with cancer used the Internet as a source of social support due to the need for emotional and informational support, a lack of support outside the Internet, the ease of communication, and the negative experiences caused by the illness. As for the resource role, previous studies revealed

that support from partners and friends was associated with individual adherence to Web-based interventions [12]. Thus, further research should pay attention to the various roles of social factors to better understand who chooses Web-based interventions and why, and how they use the program through the intervention course, particularly in collectivist cultures.

Finally, participants' coping self-efficacy was negatively related to the use of the self-talk module on the first day, but positively related to use of the mastery tools module in 1 month. Thus, coping self-efficacy might also play a motivational role in pushing individuals to use the program content, and resources could be offered for individuals particularly to complete interactive tasks (eg, exercises and activities in a program), providing further support for the role of self-efficacy in improving Web-based intervention program use [12,13].

Program Use and Outcomes Change

Among the 7 modules that we examined in this study, the self-talk and triggers modules showed a consistent association with improvement in 4 outcomes (ie, PTSD symptoms, depression, coping self-efficacy, and social functioning impairment). The relaxation module was associated with positive change in PTSD symptoms, but with negative change in 2 social variables (ie, social functioning impairment and social support). The professional help module was associated with a positive change in depression, but its use on the first day was related to a negative change in coping self-efficacy and social support. The unhelpful coping module showed an association with negative change in social functioning impairment. The mastery tools module use in 1 month showed a consistent association with a negative change in disease severity variables (ie, PTSD symptoms and depression).

These findings provide a better understanding of how the CMTR program worked to help the participants in dealing with their PTSD symptoms and depression after trauma [18]. They also showed that the CMTR modules did not play a unified role in treatment. Thus, the effectiveness of various elements in Web-based intervention programs should be examined (eg, education, self-monitoring, feedback or tailored information, self-management training, and personal exercise program) [33]. Also, future studies need to examine different delivery forms of these elements, for example, providing exercises outside of or within its educational context.

Moreover, it is surprising that the CMTR modules seem to have had much less of an effect in improving users' social support and social functioning impairment. Given that individuals turn to the Internet for social support, as well as for health information, the perceived low effectiveness in social support improvement may also cause user dropout and low adherence to Web-based intervention programs [34]. It is thus essential to consider social factors in Web-based interventions, for example, providing special social support-related content (eg, the social support module in CMTR) for social skills training, or using some form of service delivery (eg, contact with therapists or other users, forum posts, and blog posts; [35]) to offer social experiences for users. This is also one goal in future research on the CMTR program. Other goals are updating contents based

on the latest research advances and making more cultural adaptations.

Limitations

First, this study had limitation in sampling. Due to the small size and self-selected sample, our findings cannot be generalized to populations from hospitals or outpatient clinics. Future studies need to examine CMTR program use in a larger representative sample. In particular, attention should be paid to the difference in Web-based intervention program use between urban and rural users. As a potential target group of Web-based interventions, rural residents may benefit from (minimal) guided self-help interventions, although lack of Internet use experience could hinder their full use of these interventions. Future research should investigate more directly the impact of this group's other characteristics (eg, Internet use experience, lay beliefs about health problems) on their program use and explore effective methods to improve their program use in a practical way, such as offering short-term computer training before interventions.

A second limitation was a lack of control over the impact of program factors (eg, different number of webpages in modules, and the types of webpages in modules) on module use. Such an impact might be complex. For example, the self-talk, social support, and unhelpful coping modules all contain a relatively large number of webpages (ie, 16–26 pages), but the (urban) participants used them in very different ways. Thus, more specific indicators should be adopted to assess the active use of these program elements rather than passive exposure to the general program.

Finally, the findings on first-day and 1-month program use have implications for further research. Finding out user and program characteristics that promote initial program use and continuous program use will be helpful. Understanding the quicker and slower effectiveness of various program elements would also be helpful.

Conflicts of Interest

None declared.

References

1. Klein B, Mitchell J, Gilson K, Shandley K, Austin D, Kiropoulos L, et al. A therapist-assisted Internet-based CBT intervention for posttraumatic stress disorder: preliminary results. Cogn Behav Ther 2009;38(2):121-131. [doi: 10.1080/16506070902803483] [Medline: 20183691]

2. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]

3. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]

4. 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]

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

6. 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]

7. 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]

8. 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]

9. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]

10. 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 Oct;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]

XSL•FO

RenderX

11. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]

12. Davis MJ, Addis ME. Predictors of attrition from behavioral medicine treatments. Ann Behav Med 1999;21(4):339-349. [Medline: 10721442]

13. Nock MK, Ferriter C. Parent management of attendance and adherence in child and adolescent therapy: a conceptual and empirical review. Clin Child Fam Psychol Rev 2005 Jun;8(2):149-166. [Medline: 15984084]

14. Bucks RS, Hawkins K, Skinner TC, Horn S, Seddon P, Horne R. Adherence to treatment in adolescents with cystic fibrosis: the role of illness perceptions and treatment beliefs. J Pediatr Psychol 2009 Sep;34(8):893-902 [FREE Full text] [doi: 10.1093/jpepsy/jsn135] [Medline: 19196850]

15. Mâsse LC, Watts AW, Barr SI, Tu AW, Panagiotopoulos C, Geller J, et al. Individual and household predictors of adolescents' adherence to a web-based intervention. Ann Behav Med 2015 Jun;49(3):371-383. [doi: 10.1007/s12160-014-9658-z] [Medline: 25270826]

16. Brouwer W, Kroeze W, Crutzen R, de Mooijer 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]

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

18. Wang Z, Wang J, Maercker A. Chinese My Trauma Recovery, a Web-based intervention for traumatized persons in two parallel samples: randomized controlled trial. J Med Internet Res 2013;15(9):e213 [FREE Full text] [doi: 10.2196/jmir.2690] [Medline: 24080137]

19. Wang Z, Küffer A, Wang J, Maercker A. Nutzung eines webbasierten Selbsthilfeprogramms für Traumaopfer: Auswertung von Dropout-Raten und deren Prädiktoren [Adherence to a web-based self-help program for trauma survivors: analyses of drop-out rates and predictors]. Verhaltenstherapie 2014;24(1):6-14. [doi: 10.1159/000358472]

20. Wang Z, Wang J, Maercker A. An Internet-based self-help intervention program application for traumatized persons [in Chinese]. Chin Ment Health J 2013;27(8):583-590. [doi: 10.3969/j.issn.1000-6729.2013.08.006]

21. Xu W, Wang J, Wang Z, Li Y, Yu W, Xie Q, et al. Web-based intervention improves social acknowledgement and disclosure of trauma, leading to a reduction in posttraumatic stress disorder symptoms. J Health Psychol 2015 May 1:1-14. [doi: 10.1177/1359105315583371] [Medline: 25934590]

22. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]

23. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]

24. Benight CC, Ruzek JI, Waldrep E. Internet interventions for traumatic stress: a review and theoretically based example. J Trauma Stress 2008 Dec;21(6):513-520. [doi: 10.1002/jts.20371] [Medline: 19107724]

25. Foa EB, Cashman L, Jaycox L, Perry K. The validation of a self-report measure of posttraumatic stress disorder: the Posttraumatic Diagnostic Scale. Psychol Assess 1997;9(4):445-451. [doi: 10.1037/1040-3590.9.4.445]

26. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull 1973 Jan;9(1):13-28. [Medline: 4682398]

27. Joseph S, Williams R, Yule W. Crisis support, attributional style, coping style, and post-traumatic symptoms. Pers Individual Differences 1992 Nov;13(11):1249-1251. [doi: 10.1016/0191-8869(92)90262-N]

28. Steinmetz SE, Benight CC, Bishop SL, James LE. My Disaster Recovery: a pilot randomized controlled trial of an Internet intervention. Anxiety Stress Coping 2012;25(5):593-600 [FREE Full text] [doi: 10.1080/10615806.2011.604869] [Medline: 21834731]

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

30. Merrill KA, Tolbert VE, Wade WA. Effectiveness of cognitive therapy for depression in a community mental health center: a benchmarking study. J Consult Clin Psychol 2003 Apr;71(2):404-409. [Medline: 12699035]

31. 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]

32. Yli-Uotila T, Rantanen A, Suominen T. Motives of cancer patients for using the Internet to seek social support. Eur J Cancer Care (Engl) 2013 Mar;22(2):261-271. [doi: 10.1111/ecc.12025] [Medline: 23320398]

33. 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]

34.   Wang Y, Kraut RE, Levine JM. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. J Med Internet Res 2015;17(4):e99 [FREE Full text] [doi: 10.2196/jmir.3558] [Medline: 25896033]
35.   Graham AL, Papandonatos GD, Kang H, Moreno JL, Abrams DB. Development and validation of the online social support for smokers scale. J Med Internet Res 2011;13(3):e69 [FREE Full text] [doi: 10.2196/jmir.1801] [Medline: 21955465]

Abbreviations

ANOVA: analysis of variance
CMTR: Chinese version of My Trauma Recovery
CSE: Trauma Coping Self-Efficacy scale
CSS: Crisis Support Scale
PDS: Posttraumatic Diagnostic Scale
PTSD: posttraumatic stress disorder
SCL-D: Symptom Checklist 90-Depression
SFI: Social Functioning Impairment

Original Paper

Post-9/11 Veterans and Their Partners Improve Mental Health Outcomes with a Self-directed Mobile and Web-based Wellness Training Program: A Randomized Controlled Trial

Janet R Kahn[1], LMT, PhD; William Collinge[2], PhD; Robert Soltysik[3], MS (Statistics)

[1]College of Medicine, Department of Psychiatry, University of Vermont, Burlington, VT, United States

[2]Collinge and Associates, Inc., Eugene, OR, United States

[3]Optimal Data Analysis, LLC, Glen Burnie, MD, United States

Corresponding Author:
Janet R Kahn, LMT, PhD
College of Medicine
Department of Psychiatry
University of Vermont
240 Maple Street
Burlington, VT, 05401
United States
Phone: 1 802 578 2990
Fax: 1 802 658 8525
Email: Janet.Kahn@uvm.edu

Abstract

Background: Veterans with history of deployment in the Global War on Terror face significant and ongoing challenges with high prevalences of adverse psychological, physical, spiritual, and family impacts. Together, these challenges contribute to an emerging public health crisis likely to extend well into the future. Innovative approaches are needed that reach veterans and their family members with strategies they can employ over time in their daily lives to promote improved adjustment and well-being.

Objective: The objective of this study was to evaluate effects of use of a Web-based, self-directed program of instruction in mind- and body-based wellness skills to be employed by Global War on Terror veterans and their significant relationship partners on mental health and wellness outcomes associated with postdeployment readjustment.

Methods: We recruited 160 veteran-partner dyads in 4 regions of the United States (San Diego, CA; Dallas, TX; Fayetteville, NC; and New York, NY) through publicity by the Iraq and Afghanistan Veterans of America to its membership. Dyads were randomly allocated to 1 of 4 study arms: Mission Reconnect (MR) program alone, MR plus the Prevention and Relationship Enhancement Program (PREP) for Strong Bonds weekend program for military couples, PREP alone, and waitlist control. We administered a battery of standardized and investigator-generated instruments assessing mental health outcomes at baseline, 8 weeks, and 16 weeks. Dyads in the MR arms were provided Web-based and mobile app video and audio instruction in a set of mindfulness-related stress reduction and contemplative practices, as well as partner massage for reciprocal use. All participants provided weekly reports on frequency and duration of self-care practices for the first 8 weeks, and at 16 weeks.

Results: During the first 8-week reporting period, veterans and partners assigned to MR arms used some aspect of the program a mean of 20 times per week, totaling nearly 2.5 hours per week, with only modest declines in use at 16 weeks. Significant improvements were seen at 8 and 16 weeks in measures of posttraumatic stress disorder, depression, sleep quality, perceived stress, resilience, self-compassion, and pain for participants assigned to MR arms. In addition, significant reductions in self-reported levels of pain, tension, irritability, anxiety, and depression were associated with use of partner massage.

Conclusions: Both veterans and partners were able to learn and make sustained use of a range of wellness practices taught in the MR program. Home-based, self-directed interventions may be of particular service to veterans who are distant from, averse to, or prohibited by schedule from using professional services. Leveraging the partner relationship may enhance sustained use of self-directed interventions for this population. Use of the MR program appears to be an accessible, low-cost approach that supports well-being and reduces multiple symptoms among post-9/11 veterans and their partners.

Trial Registration: Clinicaltrials.gov NCT01680419; https://clinicaltrials.gov/ct2/show/NCT01680419 (Archived by WebCite at http://www.webcitation.org/6jJuadfzj)

XSL•FO
RenderX

Introduction

Since 2001, roughly 2.3 million US military personnel have been deployed to the Global War on Terror, many more than once. These veterans have displayed high rates of comorbidity of chronic pain, posttraumatic stress disorder (PTSD), mild traumatic brain injury, and other conditions, creating an urgent need for innovative, accessible interventions and multimodal treatment approaches [1,2]. PTSD rates are estimated at ≤30% [3]. A review of 29 studies found "prevalence rates of adult men and women previously deployed ranging from 5% to 20% for those who do not seek treatment and around 50% for those who do seek treatment" [4].

In addition, up to 81.5% of Global War on Terror veterans have acute or chronic pain [5-7]. In fact, many veterans live with *complex pain* resulting from multiple physical injuries, the pain from which may be exacerbated by high rates of emotional distress and mental problems resulting from traumatic brain injury [8]. Suicidality also remains a concern. A recent analysis of the military suicide prevention provisions mandated by a presidential executive order in 2012 concludes they have not been fully and effectively implemented and the goal of reducing military suicide "remains elusive" [9].

Postdeployment screening has suggested that many returning veterans may have problems that warrant treatment, but the majority may not receive treatment [10,11]. The RAND Center for Military Health Policy Research found that barriers deterring veterans from seeking help include concerns about negative career repercussions, belief that treatment won't be effective, the prospect of long wait times, limited availability of providers, and the potential side effects of medications. They concluded that continued research is needed to develop more effective treatment options [12]. Together, the complexity of many veterans' situations and the multiple interferences with receiving treatment contribute to an emerging public health crisis that is likely to extend well into the future [13].

While veterans face several risk factors for long-term mental health problems, higher interpersonal support is protective [14]. The critical role of such support was recognized by a joint work group of researchers from the US National Institute of Mental Health, the US Department of Veterans Affairs (VA), and the US Department of Defense (DoD) that singled out couple-focused interventions as among the needed directions for new research [15]. Indeed, significant attention has been focused on the impact of deployment on the spouse, as 55% of returning soldiers are married [16]. Evidence indicates that spouses may experience greater levels of emotional stress than soldiers, with soldiers' combat exposure reflected in higher spousal stress levels [17]. A review of 14 studies found that longer deployments, deployment extensions, and PTSD in military personnel were associated with psychological problems for the spouse [18]. Finally, there is evidence that the stresses of deployment may adversely affect marital satisfaction in military couples well after a return [19]. Thus, for a primary relationship to serve the much-needed support function for veterans, the impact of deployment on the *veteran*, the *partner*, and the *relationship* all need to be recognized and addressed.

The DoD's most widely used effort to engage the relationship dyad in reintegration has been the Prevention and Relationship Enhancement Program (PREP) for Strong Bonds (PREP Inc, Greenwood Village, CO, USA), a standardized program administered by military chaplains [20]. However, due to limited resources, the Strong Bonds program has the modest goal to target only 18% of postdeployment Army personnel and their families (Chaplain (Maj) J Bartels, Strong Bonds Program Operations Manager, Chaplaincy Headquarters, US Department of the Army, oral communication, June 8, 2016). Additional smaller efforts exist, such as Families OverComing Under Stress [21], and most recently the VA system has begun offering the Warrior to Soulmate program at a few VA hospitals [22]. Unfortunately, such programs have inherent barriers limiting their reach into the full population in need. These include geographic distance, required time away from work and children, a group-based experience, which is aversive to some, and the requirement of qualified professional leadership, which is costly and not always readily available.

Alternative approaches are needed that are both accessible and acceptable to veterans and their partners, to help mitigate the long-term impacts of deployment on their well-being and relationship stability. One such approach is the use of Internet-based multimedia instruction in both individual and collaborative self-care strategies. This paper reports the results of a study of an integrated program of mind- and body-based therapies delivered in multimedia format by the Internet and mobile app. Entitled Mission Reconnect (MR), the program was designed as a dyadic intervention for post-9/11 veterans and their partners to use individually and together, teaching selected self-care strategies aimed at addressing short- and long-term impacts of deployment and promoting well-being.

A phase I feasibility study of MR found high compliance and significant improvements in measures of perceived stress, depression, PTSD, and self-compassion for both members of the dyad. In addition, significant reductions were reported in pain, tension, irritability, anxiety, and depression for veterans following partner-delivered massage [23]. These results indicated that meaningful improvements in well-being are possible with this form of intervention. The objective of this subsequent study is to evaluate effects of use of MR on mental health outcomes associated with postdeployment readjustment in a randomized controlled trial with an active comparator (ClinicalTrials trial registration number NCT01680419).

Methods

The MR Program

The MR program is grounded in the biopsychosocial model of health [24]. Just as the risks and threats to well-being in the target population are multidimensional and affect multiple systems (psychological, social, spiritual and physical), multidimensional intervention addressing these multiple systems simultaneously may be beneficial. VA staff treating this comorbid population have coined the term postdeployment multisymptom disorder and noted that treatments focusing on only a single condition or symptom produce "suboptimal outcomes" [25]. Thus, we composed an integrated program of practices supporting psychological (PTSD, stress, depression, resilience, self-compassion), social (mutual support, relationship satisfaction, collaborative participation, compassion), and physical outcomes (physical pain, tension, sleep quality).

The specific methods used are grounded in the evidence bases of mindfulness-based therapies [26,27], massage therapy [28,29], positive emotions [30-32], and caregiver education [33,34]. MR is designed for autonomous, self-directed use without dependence on professional instruction. This makes it a resource for users not accessing formal services, whether due to geographic obstacles, lack of suitable services, or personal preference. It is not designed as a substitute for formal clinical services but can be used to complement or enhance them.

The program delivers the instruction by techniques grouped into three categories— *connecting with yourself*, *connecting with quiet*, and *connecting with your partner* —together comprising the 11 individual activities described in Table 1. Instruction is provided via videos, guided audio exercises, and written materials, all accessible on the program website [35] and mobile device apps (iOS, Android, and Windows Phone versions). The study website was a static design but has been a made responsive design since completion of the study. The information offered on the website and the mobile app is redundant except for the print materials (see the Supporting Materials subsection below), which can be downloaded from the website only.

Video Content

We obtained instructional footage by filming 2 workshops teaching MR to 11 Iraq and Afghanistan veterans and their partners. The Program Overview video (54 minutes) introduces the program and provides instructional sequences for the individual exercises, accompanied by commentary and discussion by workshop participants and Dr Wayne Jonas, a former military physician and expert in the field of integrative health care. Footage of massage instruction, also with veterans and partners, is presented in the program's Massage Instruction video (37 minutes). A Massage Video Supplement (3 minutes) addresses logistics of using home furniture. The video menu, as participants saw it, can be viewed in Figure 1.

Audio Content

The 9 audio exercises range in length from 1 to 22 minutes. Users are encouraged to listen, learn the practices, and then use each technique with or without guided instruction as they wish. As Figure 2 indicates, users could also view reasons why to use each practice. Clicking on Why for any practice would produce 3 to 5 brief statements, based on research, indicating potential benefit from that practice. For instance, for centering, they would see "Helps calm and relax the mind and body;" "Helps you feel grounded and present;" "Helps you be less reactive to thoughts and feelings;" "Gives you more choice about how to respond to events and feelings;" and "Helps you feel peaceful more easily."

Supporting Materials

A massage instruction booklet and a 1-page illustrated massage reminder handout are downloadable from the website [35]. A *What if?* feature on the website and the app enables users to access advice on how to apply program techniques in specific challenging situations such as problems with sleep, focus, and concentration. Clicking on any of the arrows next to a *What If?* topic, as seen in Figure 3, would open to a page displaying both written suggestions and active audio links. Users can submit questions through the *What if?* interface or suggest future content to enhance the program. *Optional Audios* include the guided audio exercises using the alternative gender voice to that used in the main program (see Multimedia Appendix 1 for a video introduction to MR offering an overview of the elements of the program).

Table 1. Content of the Mission Reconnect Program.

| Component	Description	Run time	Media
Program overview	Introduction to the structure and components of the program; instructional sequences; motivational interviews with experts and participants.	54:29	Video
Connecting with yourself			
Morning gratitude	A brief practice of starting each day with a moment of spoken gratitude to encourage positive mood, empathy, forgiveness, focus, and sleep quality.	2:49	Audio
Mirror greeting	A brief practice of greeting oneself in a mirror with positive self-regard to encourage self-confidence and resilience, and to decrease tension.	1:18	Audio
Loosening and relaxing	A brief practice of standing and rhythmically moving to relax, release physical tension, bring awareness into the body, and feel happier.	5:50	Video & audio
Waking up the body	A brief practice of rhythmically patting and stimulating meridians to increase energy and blood flow, and one's sense of aliveness.	9:26	Video & audio
Reset and refresh	Evoking a series of yawns to release tension, interrupt stressful patterns of thought or feeling, prompt oxytocin production, and return attention to the present moment.	2:42	Video & audio
Connecting with quiet			
Centering	A guided exercise of basic mindfulness instruction focused on using the breath as an anchor to the present moment and to reduce reactivity.	9:41	Video & audio
Movement into stillness	A brief practice of gentle swaying movement to bring body and mind into alignment in the present moment. Effects are similar to centering; method is ideal for those who cannot sit still.	4:08	Video & audio
Deep relaxation	A guided exercise lying down or sitting and combining aspects of traditional progressive relaxation and yoga nidra meditation, for sleep enhancement and relaxation.	23:12	Video & audio
Connecting with your partner			
Seeing each other	A guided contemplative exercise focused on eliciting appreciation, compassion, and forgiveness for one's partner and oneself, done alone or together.	8:00	Video & audio
Giving massage Receiving massage	Practicing being present with one's partner by providing comfort and relaxation through simple massage techniques, and by receiving massage.	37:35 main, 3:00 supplement, booklet 34 pp, reminder handout 2 pp.	Video & print

Figure 1. Mission Reconnect Video Menu.
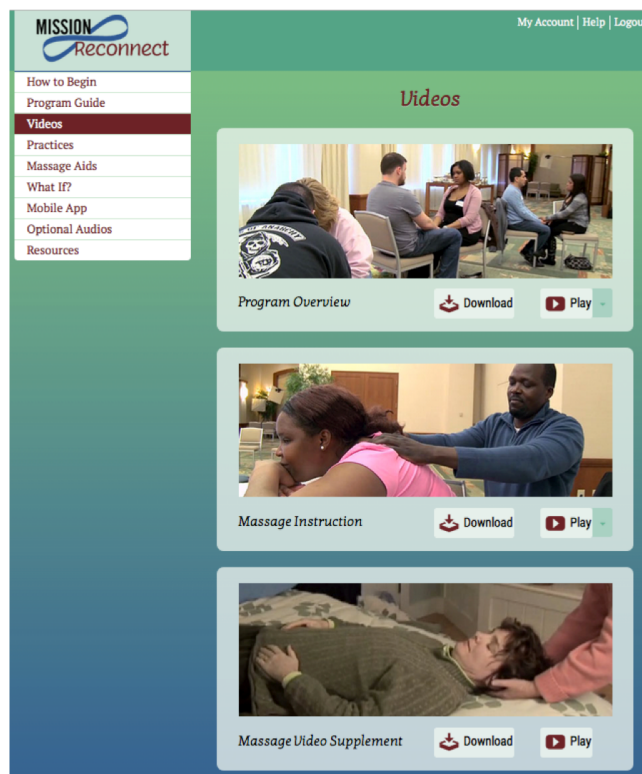
Figure 2. Mission Reconnect Practice Menu.
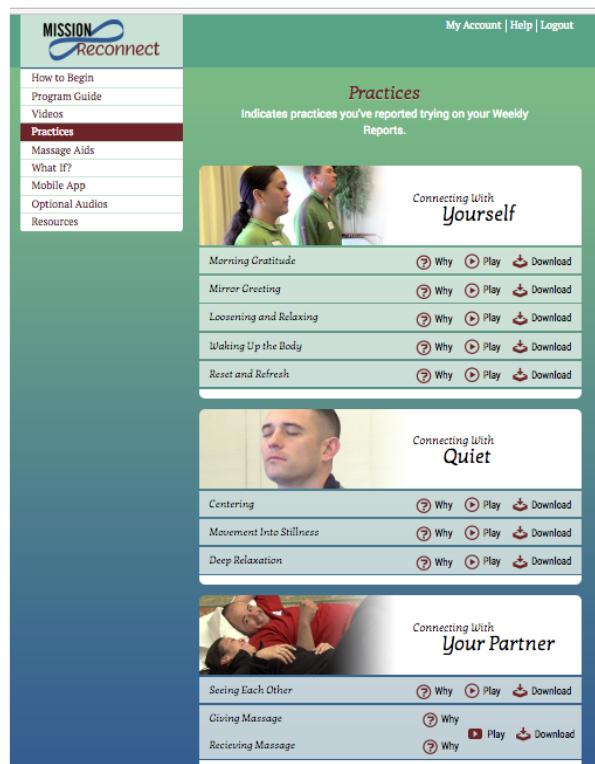
Figure 3. Mission Reconnect What If? Menu.

Aims and Hypotheses

Aim 1 was to evaluate the impact of MR on mental health outcomes associated with postdeployment readjustment (primary outcomes). Hypothesis 1 was that veterans and partners assigned to MR would report significantly greater improvements in stress, depression, PTSD symptoms, self-compassion, sleep quality, resilience, social support, and relationship satisfaction compared with those in the non-MR comparison groups.

Aim 2 was to evaluate program use and satisfaction among participants in the MR program (secondary outcomes). Hypothesis 2 was that participants assigned to MR would comply with recommendations to use the nonmassage techniques of the program weekly and would average 30 minutes per week of nonmassage practice. Hypothesis 3 was that participants assigned to MR would comply with recommendations to exchange massages weekly and would receive an average of at least one massage per week. Hypothesis 4 was that users of MR receiving massage would report significant positive massage effects.

Aim 3 was to collect exploratory data on change in pain levels associated with the use of MR. There were no specific hypotheses associated with this aim.

Aim 4 was to collect exploratory data on the presence of moral injury. There were no specific hypotheses associated with this aim.

Study Design

To evaluate outcomes of MR, we planned a 4-arm randomized controlled trial using a standard-of-care comparator to evaluate both comparative and additive effects over a 16-week period. The comparator was the PREP for Strong Bonds program used widely by the military for relationship enhancement and postdeployment reintegration [36]. This is a standardized, evidence-based program conducted during weekend residential retreats with experiential exercises, facilitated by Army Chaplain Corps-trained chaplains. The multidimensional program uses methods from cognitive behavioral therapy and communication-oriented marital enhancement programs developed by Markman et al, Denver University Center for Marital and Family Studies [37]. Thus, the 4 arms to be studied were (1) MR alone, (2) MR+PREP, (3) PREP alone, and (3) a waitlist control. Sample size was determined by power analysis applied to the phase I feasibility data [23]. We planned a sample of 160 dyads, with 40 dyads allocated to each arm.

Participants

Eligibility

Eligible applicants were veteran-partner dyads in which the veteran had a history of deployment in a post-9/11 combat operation. This includes what the US DoD has labeled Operation Iraqi Freedom, Operation Enduring Freedom, and Operation New Dawn. As proof of deployment, veterans typically provided their Form DD214, a discharge or separation from service document provided by the military that includes the veteran's military service record. This verification of deployment had to be provided before random allocation. The partner could be a spouse, life partner, fiancé or fiancée, girlfriend or boyfriend, or in other significant relationship identified by the veteran as fulfilling the role of "partner" for the purposes of mutual support. While most dyads had partners with no military

experience, some dual-veteran dyads also applied and were accepted. No queries were made regarding prior use of meditation or other complementary and alternative techniques.

Publicity and Recruitment

Iraq and Afghanistan Veterans of America provided publicity for recruitment through their website, social media channels, and geographically targeted emails to their membership announcing the study [38]. To achieve geographic diversity, we recruited study cohorts for all 4 study arms in each of 4 metropolitan areas: San Diego, CA; Dallas, TX; Fayetteville, NC; and New York, NY, USA.

Application and Consent

The New England Independent Review Board (Needham, MA) provided institutional review board oversight. Prospective participants completed a Web-based application (PsychData, State College, PA, USA), which included the consent form (downloadable as a pdf file). Applicants gave electronic consent by marking a box attesting they had read the consent form when they submitted the application. Applicants were interviewed by phone by a research assistant to determine candidacy—that is, that the dyad met all the eligibility requirements and was available for the planned meeting date in their city if they were selected. Launch meetings were held in each city for each arm (content is listed below in the Protocols subsection). All meetings were held in hotel conference rooms, except in New York, where the MR-only and waitlist meetings were held in a conference room at the offices of Iraq and Afghanistan Veterans of America, the organization that had helped in recruitment.

Random Allocation

From each city's eligible candidate pool, we selected, based on our ethnic and racial diversity criteria, 40 candidate dyads to be randomly allocated to the 4 arms in that city (random allocation was generated through Randomization.com [39]). Candidates not selected were placed on an alternates list. After random allocation and before any data collection, we notified selected dyads of their assignment and required date of attendance at a launch meeting to begin the study for their arm. Those who declined or dropped out before the launch meeting were replaced by a randomly selected dyad from that city's alternates list, to fill the vacated slot so as to preserve the original randomization. If a city had insufficient candidates to fill a slot, we moved the empty slot to its same arm in another city and used the new city's alternates list to randomly select a dyad for the vacant slot, to satisfy the overall study randomization plan of 40 dyads per arm (this was done twice).

Protocols

MR-Only

Dyads attended a 90-minute launch meeting designed to ensure that they understood their responsibilities to the study, and the logistics of both data collection and participant compensation. Each MR-only launch meeting included all dyads in that city that had been randomly allocated to the MR-only arm. The investigators introduced themselves, attendees introduced themselves, and the investigators went over the logistics of participation. The 10-minute Introduction chapter of the Program

Overview video was shown followed by explanation of data collection procedures materials. Participants' responsibilities were described, including trying all MR techniques at least once during the first few weeks, sharing weekly massage with one's partner, and providing Web-based data through 8 weekly reports and 3 surveys. Data collection procedures were explained, as were the mechanics of compensation linked to data collection, and each participant received a debit card that had been preloaded with compensation for the baseline survey they had completed prior to the launch meeting. The importance to the study of the MR-only arm was explained, participants were thanked for their willingness to participate in the study, and the meeting was concluded. No actual instruction or practice took place in the launch meeting. For the duration of the study, MR dyads received a weekly e-newsletter from the investigators focused on general reiteration of instructions and encouragement to try all the practices.

MR+PREP

Dyads attended a standardized weekend residential PREP retreat led by a PREP-trained army chaplain. A total of 3 chaplains were used across the 4 cities. A teleconference among the chaplains, investigators, and a trainer from PREP Inc headquarters was conducted to establish fidelity with current PREP content, and a standard program with 12 hours of content was agreed upon before the study. PREP weekends lasted from Friday evening to noon Sunday. Instruction focused on communication and relationship building, problem solving, stress and relaxation, intimacy, forgiveness, and commitment. After lunch Sunday, the dyads attended an MR launch meeting with the same protocol as described for the MR-only participants above.

PREP-Only

Dyads attended a standardized PREP weekend as described above. At the program's conclusion, the chaplain (scripted by the investigators to be the same as for waitlist participants, described below) provided the launch meeting content for non-MR participants. As with other arms, attendees were given instructions for data collection on the project website for the remainder of the study. The participants' compensation was explained and attendees received their debit cards that had been preloaded with compensation for the baseline survey they had completed prior to the PREP weekend. The importance of their arm to the integrity and value of the study was explained, and participants were thanked for their willingness to participate in the study. Participants were instructed to continue their usual behavior regarding self-care or wellness-related activity. At the end of the study, they would be given access to the MR program.

Waitlist

Dyads attended a 90-minute launch meeting with the investigators to receive their instructions for data collection on the project website. As with other arms, they were introduced to the purpose of the project, the design of the study, the mechanism for their compensation, and the importance of the waitlist control arm to the study. They were instructed to continue with their usual behavior regarding self-care or wellness-related activity. They were thanked for their participation and given their debit cards preloaded with

compensation for the baseline survey they had completed prior to the meeting. At the end of the study, they would be given access to the MR program.

Human Contact at Launch

We recognized in designing the study that the quantity of human contact at launch for each arm was an important consideration; hence, we planned to assure that the MR-only and waitlist arms had an equal duration (90 minutes) of contact during launch. The 2 PREP arms of course received a full weekend intervention in person to allow the comparisons between self-directed and in-person intervention sought in the study. Between the 2 PREP arms, there was some difference in total contact over the full weekend, with the PREP-only group receiving less time dedicated to launch information because there was no need for the MR portion. We deemed this difference of negligible importance given participants' exposure to a full weekend of contact; it proved most practical to incorporate their launch content at the end of their workshop rather than making them stay an extra 90 minutes.

Prompting

Each participant received email notifications of data collection tasks due (weekly reports, surveys). A personal project calendar on the website and app displayed her or his data collection due dates (non-MR participants could not access intervention content).

Data Collection

All data collection was Web based. The initial application was submitted on PsychData.com, and surveys and weekly reports during the study period were completed on the project website.

Survey Data

A survey package was administered 3 times: at baseline (T1, prior to the launch meeting), at 8 weeks (T2), and at 16 weeks (T3, end of study). We used the following standardized instruments: the Perceived Stress Scale-10 item (PSS) [40], Beck Depression Inventory (BDI) [41], PTSD Checklist-Civilian version (PCL-C) [42], Self-Compassion Scale (SCStotal) [43], Response to Stressful Experiences Scale (RSES) [44], Multidimensional Scale of Perceived Social Support (MSPSStotal) [45], Pittsburgh Sleep Quality Index (PSQItotal) [46], and Revised Dyadic Adjustment Scale (RDAS) [47]. In addition, 2 investigator-generated Likert-scaled (0–10 points) questions asked respondents to rate their usual pain level over the past week (PainUsual) and their best pain level over the past week (PainBest).

Veterans were asked 5 investigator-generated items exploring the concept of moral injury associated with military service. The results of this exploratory moral injury inquiry are not reported here, as addressing moral injury directly was not an aim of the MR program. We will report those data in a subsequent publication.

Weekly Reports

During the initial 8 weeks, all participants completed a Web-based weekly report on their use of wellness activities as follows (see Multimedia Appendix 2 and 3 for weekly report forms used by MR and non-MR participants).

For the MR arms, the weekly report recorded frequency and duration of use of each MR program activity. In addition, participants were to designate one 20-minute massage per week as their massage reporting session, for which they completed a massage session card (hard copies were provided at the launch meeting). On the card, they rated symptoms of pain, tension, being on edge or irritable, anxiety or worry, and depression, on a 0–10 Likert scale, both before and again 15 minutes after the massage (data on massage session effects). They were then to upload the responses recorded on the card to their weekly report.

For participants in the non-MR arms, the weekly Web-based report assessed the number and types of activities used during the past week to (1) relax, reduce stress, or support general well-being, (2) ease physical pain or tension, (3) support one's partner or strengthen the relationship, and (4) improve sleep quality.

No weekly reports were collected during weeks 9–15 of the study. In the final survey package (16 weeks), a weekly report was included for the last week of activity.

Compensation

Each participant was given a debit card at their launch meeting for payment for data collection. Cards were automatically funded via the data collection website when reports and surveys were submitted: US $40 after each survey and US $20 after each weekly report.

Statistical Methods

We used intent-to-treat analysis. For the survey data, paired-sample t tests were performed for each pairwise contrast of times T1 (baseline), T2 (8 weeks), and T3 (16 weeks). Descriptive statistics were used to report the weekly report data for MR arms and non-MR arms. Pre-post massage session effects were evaluated using the Wilcoxon signed rank test for each weekly pairwise contrast. In addition, Kendall tau-b was evaluated across the 8-week set of samples of premassage effects. The 4 study arms were compared with each other for each of the survey outcomes at each time point using 2-sample t tests. The MR and MR+PREP groups were combined to define a common metric concerning frequency and duration of MR activities. This metric was correlated to the survey data and analyzed by Pearson correlation. Improvement in survey scores at time T2 was also predicted from the frequency and use of MR activities with Kendall tau-b. Linear discriminant analysis was performed at baseline to investigate differences in survey outcomes at T2 and T3 between the waitlist control and each of the other 3 comparison groups.

We used a sequentially rejective Sidak Bonferroni-type multiple comparisons procedure to ensure the desired experimentwise type I error rate (P<.05). This procedure has been demonstrated to more efficiently ensure the desired experimentwise P value when compared with Dunn's procedure [48].

For dual-veteran couples, we performed subanalyses for veterans as a class and nonveteran partners as a class. In dyads where

both members met the criteria for veteran, we included both in analyses of veteran data, and not in partner data.

Results

Sample

A total of 238 dyads (476 individuals) that met eligibility criteria applied and were assessed for candidacy for 160 slots. This process included submission of a copy of a DD214 form or other official military document that verified their deployment

status. Figure 4 shows the reasons prompting exclusion from selection and the flow of participants through the project (see Multimedia Appendix 4 [49] for the CONSORT eHealth checklist). All 320 participants completed their baseline surveys, with 313 completing the 8-week follow-up and 311 also completing the 16-week follow-up. Within the sample of 160 dyads were 21 dyads in which both members were veterans. Thus, the sample comprised 181 veterans and 139 nonveteran partners (hence the differences in numbers in the tables below separately reporting veteran and partner outcomes).

Figure 4. CONSORT Flow Diagram.
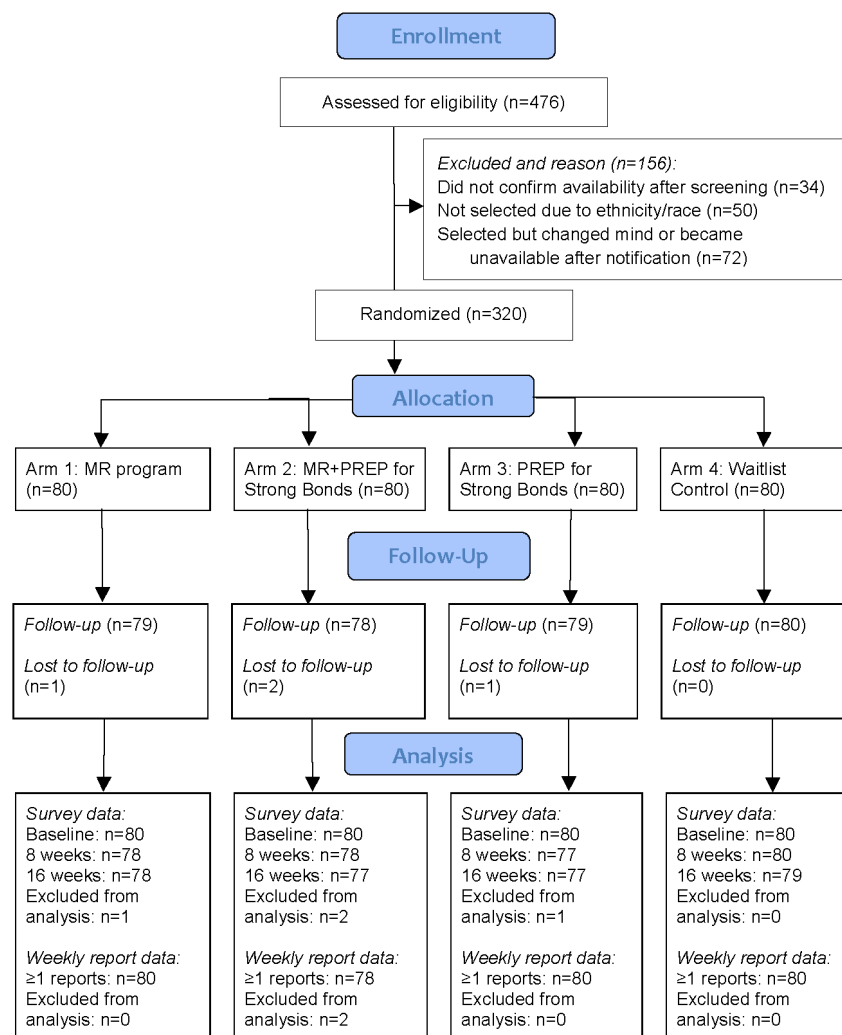
Demographics

Most dyads (151/160, 94.4%) were married or living together as life partners. Mean relationship duration was 7.2 years with a range of 1–32 years. The sample included 3 same-sex couples and 21 couples in which both members met the study "veteran" criteria. As Multimedia Appendix 5 indicates, over half of the participants identified as white, followed by Hispanic, African American, Asian, Native American, and Hawaiian/Pacific Islander. Regarding education, just over 10% of the sample had a high school diploma/general equivalency diploma or less; the largest group had attended some college or trade school; others had completed an Associate or other 2-year degree. The

remainder had completed a Bachelor degree or more, with a noticeable number of participants having either received a graduate degree or completed some graduate work. The vast majority of veterans were male (147/181, 81.2%) and the vast majority of partners (129/139, 92.8%) were female.

Service History

The majority of veterans in this study served in the Army (102/181, 56.4%), followed by the Marine Corps (43/181, 23.8%), the Navy (23/181, 12.7%), and the Air Force (13/181, 7.2%). At the time the study launched, the bulk of veterans had retired or separated from the service (123/181, 68.0%) and the remainder were still serving (58/181, 32.0%). On average, these

veterans had served 2 deployments. The most frequent year of first deployment was 2005, and of last return, 2008.

Fidelity (Use of MR)

There were no absolute requirements for use of MR practices other than to both give and receive one 20-minute massage per week for the first 8 weeks. We wanted participants in the MR arms to use the range of nonmassage techniques offered in the way they felt was most efficacious for them, knowing that this could vary widely. We were interested to learn which practices participants chose to use most frequently. Thus, we requested that all participants in both MR arms try every practice at least once early on, then use whatever practices they felt helped them most, as often as they wished.

With a complex and self-directed program such as MR, use, or fidelity, is a critical question. If we build it, will they use it? And will those who use it more derive more benefit? With these questions in mind, we tracked use through their weekly reports.

As Table 2 indicates, during weeks 1–8, on average veterans and partners used some aspect of MR 20 times per week, totaling nearly 2.5 hours per week. With an average of more than 17 "uses" of nonmassage practices per week by both veterans and partners, and an average of 1.4 massages per week received by both veterans and partners, the usage data met and surpassed our hypothesized use of the program (hypotheses 2 and 3). The most frequently used practice cluster was *connecting with yourself*, with morning gratitude being most popular, followed by the mirror greeting. Since this entire cluster involves relatively brief practices, frequency does not correlate with most time spent. Nearly twice as much time was spent in *connecting with quiet* and even more than that in *connecting with your partner*. The 2 massages each averaged roughly 20 minutes and the contemplative exercise of seeing each other with fresh eyes averaged nearly 15 minutes. The large standard deviations in this table indicate widely differing usage patterns. Cumulative totals for each cluster indicate that both veterans and partners spent over half the MR time per week in exercises to enhance connection with their partner. (See Multimedia Appendix 6 for MR usage reports for weeks 2, 8, and 16.)

Weeks 9–16 required no weekly reports, but the final survey at T3 included the weekly report questions answered just for week 16, to collect data on any longitudinal pattern changes. Frequency of activity use had decreased somewhat in all 3 MR clusters for veterans, as had time spent, shrinking a bit, yet still remaining above 90 minutes of weekly use on average. Partners, on the other hand, increased their mean total time per week by over 10 minutes, due entirely to increased use of the *connecting with yourself* and *connecting with quiet* practices.

Table 2. Weekly Mission Reconnect use over the 8-week monitoring period (n=79 dyads of veterans and their partners).

Component	Veterans (692 reports)		Partners (559 reports)	
	Frequency, mean (SD)	Minutes, mean (SD)	Frequency, mean (SD)	Minutes, mean (SD)
Connecting with yourself				
Loosening and relaxing	1.9 (1.9)	12.8 (23.3)	1.8 (2.2)	11.8 (19.9)
Waking up the body	1.9 (2.1)	12.0 (23.5)	1.7 (2.2)	12.3 (25.6)
Reset and refresh	2.2 (3.1)	N/A[a]	2.6 (3.4)	N/A
Morning gratitude	2.6 (2.2)	N/A	3.1 (2.4)	N/A
Mirror greeting	2.5 (2.9)	N/A	2.5 (2.5)	N/A
Cumulative subtotals	11.1 (9.6)	24.8 (44.8)	11.7 (10.2)	24.2 (43.6)
Connecting with quiet				
Centering	2.1 (2.3)	16.3 (27.1)	2.3 (2.8)	17.7 (27.4)
Movement into stillness	1.3 (1.6)	8.5 (15.5)	1.3 (2.1)	11.7 (26.6)
Deep relaxation	1.7 (1.9)	22.5 (39.0)	1.5 (2.1)	17.3 (30.1)
Cumulative subtotals	5.0 (4.9)	47.3 (65.7)	5.2 (6.0)	46.7 (76.1)
Connecting with your partner				
Seeing each other	1.2 (1.7)	14.1 (35.0)	1.4 (2.0)	15.8 (34.4)
Giving massage to your partner	1.4 (1.8)	29.0 (38.3)	1.4 (2.3)	29.0 (88.5)
Receiving massage from your partner	1.4 (1.8)	23.9 (53.6)	1.4 (2.3)	28.8 (88.3)
Cumulative subtotals	4.1 (4.4)	72.5 (103.3)	4.0 (5.7)	73.2 (186.3)
Totals for all activities	20.2 (15.1)	144.6 (164.8)	20.5 (18.3)	142.2 (244.0)

[a]N/A: not applicable or not available. Since these practices take very little time, often less than one minute, we asked subjects to report frequency only, not minutes spent.

Non-MR Participants' Wellness Strategies

Participants in the PREP-alone and waitlist arms employed a range of techniques to help themselves relax and to support their general well-being, ease pain or tension, support their partner and strengthen their relationship, or improve their sleep. These included prescription and nonprescription drugs, sex, meditation, exercise, baths, yoga, and reading. Both veterans and partners reported most frequently engaging in activities for relaxation or well-being, followed by support for the relationship or partner, and easing pain or tension, and least frequently for improving sleep. Veterans averaged 3.6 total activities per week and partners 3.9 (see Multimedia Appendix 7 for a description of activities reported).

Survey Data: Veterans

While we assigned 40 dyads to each arm, for dual-veteran dyads we analyzed both participants as veterans, leading to sample sizes larger than 40. Table 3 presents veterans' scores by arm at all 3 testing points. Baseline differences between arms were few and relatively modest except a PainBest difference between arms 1 and 3. *P* values shown reflect significant differences for T2 and T3 relative to baseline scores.

The most striking differences were between the MR-only and waitlist control arms. The MR-only veterans had improvements at 8 weeks in a broad array of mental health dimensions, indeed, everything other than pain scores, dyadic adjustment scores, and perceived social support. All improvements were sustained at 16 weeks with the exception of sleep improvement

(PSQItotal). Perceived ability to respond to stressful events (RSES) had further improved.

The PREP-only veterans (arm 3) showed a modest gain in three important dimensions at T2 (PSS, BDI, and PCL-C), which became stronger by T3. The MR+PREP arm was included to see whether there would be added value by combining the programs. These results indicate that adding MR to PREP produces improvement in more domains than PREP alone (adding improvements in self-compassion and capacity to respond to stress). However, the combination yielded a lower magnitude of improvement in those arenas than resulted from MR alone. Thus, aim 1, our hypothesis that veterans and partners assigned to MR would report significantly greater improvements in a wide range of mental health outcomes, was partially met in that veterans in the MR-only arm reported significant improvements for a broader array of mental health outcomes than veterans in the other arms.

PTSD Subanalysis

The data indicate significant reductions, of varying degrees, in PCL-C scores for veterans in all 3 intervention arms. To assess outcomes for veterans entering with high levels of PTSD, we additionally analyzed those with baseline scores ≥50, which is commonly considered a clinical diagnostic criterion. This subset included 53/181 (29.3%) veterans in the sample. As seen in Table 4, high scorers in the MR-only arm had clinically significant reductions, dropping them to below the diagnostic threshold at both follow-ups. Their peers in the MR+PREP and PREP-alone arms also showed significant reductions, though of lesser magnitude.

Table 3. Veterans' within-group changes from baseline to 8- and 16-week follow-ups (paired *t* tests).

Scale	Arm 1: MR[a]			Arm 2: MR+PREP[b]			Arm 3: PREP			Arm 4: waitlist		
	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value
Baseline (T1)												
PSS[c]	45	20.0 (6.7)		44	19.2 (6.3)		43	21.1 (6.4)		48	20.5 (6.5)	
BDI[d]	45	14.5 (10.8)		44	16.1 (12.5)		43	19.5 (12.0)		48	16.8 (12.7)	
PCL-C[e]	45	38.4 (16.5)		44	41.7 (18.3)		43	42.1 (16.1)		48	41.1 (15.8)	
SCStotal[f]	45	73.9 (20.8)		44	72.6 (21.9)		44	73.2 (22.0)		48	72.3 (18.8)	
RSES[g]	45	57.8 (14.1)		44	55.3 (19.7)		44	50.9 (20.5)		48	54.7 (15.3)	
MSPSStotal[h]	45	60.2 (18.3)		44	62.1 (15.2)		44	59.8 (17.1)		48	62.2 (15.6)	
PSQItotal[i]	45	9.2 (4.2)		44	11.2 (4.4)		44	10.7 (4.0)		48	10.5 (4.1)	
RDAS[j]	45	46.1 (10.2)		44	47.0 (9.6)		44	42.7 (11.5)		48	43.1 (9.9)	
PainUsual[k]	45	3.1 (2.6)		44	4.5 (2.7)		44	4.5 (2.9)		48	4.2 (2.6)	
PainBest[l]	45	1.6 (1.8)		44	2.5 (2.3)		44	2.7 (2.1)		48	2.1 (2.0)	
8 weeks (T2)												
PSS	44	15.5 (7.2)	.0001*	42	16.2 (7.3)	.007	43	19.2(6.6)	.03	47	18.4 (7.2)	.002*
BDI	43	9.4 (12.3)	.0004*	42	10.9 (10.5)	.002*	40	14.8 (12.4)	.004*	45	15.0 (13.1)	.43
PCL-C	44	32.3 (15.8)	.0002*	42	35.0 (16.2)	.006	42	37.9 (16.7)	.02	48	39.6 (15.8)	.15
SCStotal	44	85.6 (22.5)	.0001*	43	77.5 (18.8)	.09	39	77.2 (21.3)	.37	45	74.8 (18.4)	.32
RSES	44	64.6 (15.6)	.002*	43	57.7 (18.7)	.31	42	53.0 (22.3)	.43	47	55.1 (16.6)	.57
MSPSStotal	43	62.5 (17.7)	.44	42	65.9 (14.4)	.17	42	61.0 (17.0)	.64	47	61.9 (15.7)	.77
PSQItotal	44	7.6 (4.1)	.003*	42	10.2 (4.7)	.07	42	10.8 (4.3)	.62	48	10.0 (4.5)	.25
RDAS	43	47.1 (11.0)	.64	40	50.7 (9.8)	.03	42	42.8 (11.9)	.98	46	43.0 (10.5)	.56
PainUsual	44	3.4 (2.8)	.56	43	4.0 (2.8)	.25	41	4.0 (2.7)	.25	48	4.2 (2.6)	.95
PainBest	44	1.8 (2.3)	.55	43	2.3 (2.2)	.72	41	2.2 (2.1)	.09	48	2.2 (2.0)	.57
16 weeks (T3)												
PSS	44	15.0 (7.3)	.0001*	43	16.2 (7.8)	.009	42	18.2 (5.6)	.0001*	48	19.5 (7.2)	.27
BDI	44	8.7 (12.5)	.0003*	42	12.4 (14.7)	.04	41	13.9 (11.8)	.0009*	46	14.7 (13.1)	.99
PCL-C	44	31.3 (15.7)	.0002*	42	36.1 (18.7)	.01	42	35.1 (14.1)	.0004*	46	39.4 (17.3)	.77
SCStotal	43	87.7 (24.4)	.0001*	43	82.3 (2.2)	.004*	42	76.7 (22.3)	.12	47	73.4 (2.2)	.63
RSES	43	65.3 (17.6)	.0006*	42	62.0 (2.1)	.007	42	54.5 (18.0)	.06	48	56.4 (17.3)	.59
MSPSStotal	43	64.9 (16.9)	.08	43	63.0 (18.1)	.99	42	62.5 (15.2)	.32	47	62.0 (18.2)	.49
PSQItotal	44	8.2 (5.0)	.08	41	10.7 (4.8)	.45	41	10.0 (4.5)	.21	47	10.7 (4.5)	.635
RDAS	44	47.9 (1.9)	.21	39	49.6 (11.3)	.27	42	42.5 (11.6)	.67	46	43.9 (9.4)	.857
PainUsual	44	2.7 (2.4)	.11	42	4.3 (3.1)	.66	41	4.2 (3.0)	.44	46	4.5 (2.7)	.426
PainBest	44	1.4 (1.9)	.37	42	2.5 (2.5)	.47	41	2.6 (2.2)	.63	46	2.5 (2.3)	.362

[a]MR: Mission Reconnect.

[b]PREP: Prevention and Relationship Enhancement Program.

[c]PSS: Perceived Stress Scale-10 item.

[d]BDI: Beck Depression Inventory.

[e]PCL-C: PTSD Checklist-Civilian version.

XSL•FO

RenderX

[f]SCStotal: Self-Compassion Scale.

[g]RSES: Response to Stressful Experiences Scale.

[h]MSPSStotal: Multidimensional Scale of Perceived Social Support.

[i]PSQItotal: Pittsburgh Sleep Quality Index.

[j]RDAS: Revised Dyadic Adjustment Scale.

[k]PainUsual: rating of usual pain level over the past week.

[l]PainBest: rating of best pain level over the past week.

*$P<.05$ after adjustment for experimentwise error rate (true P values before adjustment displayed).

Table 4. Changes at 8 (T2) and 16 weeks (T3) in PTSD Checklist-Civilian version (PCL-C) scores for veterans with baseline (T1) score ≥50 (paired t tests)

Arm	n	T1 scores	T1 vs T2 scores				T1 vs T3 scores			
		Mean (SD)	Mean change	t	df	P value	Mean change	t	df	P value
MR[a]	12	61.2 (10.0)	–12.2	–4.6	11	.001*	–13.5	–3.9	11	.003*
MR+PREP[b]	14	63.9 (9.7)	–10.6	–2.2	13	.048	–5.6	–1.4	13	.18
PREP	13	62.8 (8.3)	–7.9	–2.3	12	.046*	–12.1	–3.5	12	.006*
Waitlist	14	60.6 (7.0)	–3.7	–1.9	13	.08	–1.2	–.5	13	.66

[a]MR: Mission Reconnect.

[b]PREP: Prevention and Relationship Enhancement Program.

*$P<.05$ after adjustment for experimentwise error rate (true P values before adjustment displayed).

Group Comparisons

Table 5 presents results of 2-sample t tests at 8 weeks, indicating no significant differences between the 2 non-MR arms at that point. However, there were some notable differences between MR-only and each of the other arms, particularly regarding the capacity to respond to stress and quality of sleep during the first 8 weeks. Table 6 shows results at T3, indicating that contrasts had strengthened, with more dimensions reaching significant differences by 16 weeks. These tables indicate that the number of mental health outcomes for which difference in magnitude of improvement was significant for MR compared with the other groups was limited and did not match our expectations as stated in hypothesis 1.

Table 5. Veterans' between-group differences at 8 weeks (T2) (2-sample *t* tests).

Scale	Arm comparisons (*P* values)					
	MR[a] vs MR+PREP[b]	MR vs PREP	MR vs waitlist	MR+PREP vs PREP	MR+PREP vs waitlist	PREP vs waitlist
PSS[c]	.66	.02	.07	.05	.17	.56
BDI[d]	.56	.054	.04	.13	.12	.92
PCL-C[e]	.44	.12	.03	.43	.18	.61
SCStotal[f]	.07	.09	.02	.95	.49	.57
RSES[g]	.06	.007*	.006*	.30	.49	.62
MSPSStotal[h]	.34	.69	.87	.16	.22	.79
PSQItotal[i]	.008*	.001*	.01*	.58	.82	.42
RDAS[j]	.12	.09	.08	.002*	.001*	.94
PainUsual[k]	.31	.27	.16	.94	.74	.80
PainBest[l]	.35	.41	.32	.90	.99	.90

[a]MR: Mission Reconnect.

[b]PREP: Prevention and Relationship Enhancement Program.

[c]PSS: Perceived Stress Scale-10 item.

[d]BDI: Beck Depression Inventory.

[e]PCL-C: PTSD Checklist-Civilian version.

[f]SCStotal: Self-Compassion Scale.

[g]RSES: Response to Stressful Experiences Scale.

[h]MSPSStotal: Multidimensional Scale of Perceived Social Support.

[i]PSQItotal: Pittsburgh Sleep Quality Index.

[j]RDAS: Revised Dyadic Adjustment Scale.

[k]PainUsual: rating of usual pain level over the past week.

[l]PainBest: rating of best pain level over the past week.

*$P<.05$ after adjustment for experimentwise error rate (true *P* values before adjustment displayed).

XSL•FO

RenderX

Table 6. Veterans' between-group differences at 16 weeks (T3) (2-sample *t* tests).

Scale	Arm comparisons (*P* values)					
	MR[a] vs MR+PREP[b]	MR vs PREP	MR vs waitlist	MR+PREP vs PREP	MR+PREP vs waitlist	PREP vs waitlist
PSS[c]	.44	.03	.004*	.19	.04	.33
BDI[d]	.22	.054	.03	.61	.43	.76
PCL-C[e]	.20	.24	.02	.78	.40	.21
SCStotal[f]	.27	.03	.003*	.23	.04	.48
RSES[g]	.42	.006*	.02	.07	.16	.60
MSPSStotal[h]	.61	.49	.43	.90	.80	.89
PSQItotal[i]	.02	.09	.01	.48	.99	.45
RDAS[j]	.49	.03	.07	.007*	.01	.51
PainUsual[k]	.01*	.01	.002*	.89	.76	.64
PainBest[l]	.02	.008	.02	.87	.96	.82

[a]MR: Mission Reconnect.

[b]PREP: Prevention and Relationship Enhancement Program.

[c]PSS: Perceived Stress Scale-10 item.

[d]BDI: Beck Depression Inventory.

[e]PCL-C: PTSD Checklist-Civilian version.

[f]SCStotal: Self-Compassion Scale.

[g]RSES: Response to Stressful Experiences Scale.

[h]MSPSStotal: Multidimensional Scale of Perceived Social Support.

[i]PSQItotal: Pittsburgh Sleep Quality Index.

[j]RDAS: Revised Dyadic Adjustment Scale.

[k]PainUsual: rating of usual pain level over the past week.

[l]PainBest: rating of best pain level over the past week.

*$P<.05$ after adjustment for experimentwise error rate (true *P* values before adjustment displayed).

Survey Data: Partners

The 4 partner arms were also quite comparable at baseline (Table 7). Differences were that arm 1 had higher perceived stress than arm 4, and arm 3 differed from arms 2 and 4 on the PainBest and PainUsual scores. However, partners differed from veterans in terms of program outcomes.

Examining change within each arm from baseline to T2 and T3, we see no significant changes for either of the non-MR arms. The MR-only arm shows change in the same arenas for partners as for veterans, but a bit less powerfully at T2 and somewhat diminished at T3. For MR+PREP, the changes for partners strengthen from T2 to T3, which may reflect partners' increased use of the program in weeks 8–16, when veterans had decreased both duration and frequency of use. Despite this within-group change for the MR arms, there were no statistically significant differences between the 4 arms as time progressed, other than for pain.

Group differences for partners at T2 and T3 were not noteworthy and are not reported here (available upon request).

Table 7. Partners' within-group changes from baseline to 8- and 16-week follow-ups (paired *t* tests), as measured by standardized instruments.

Scale	Arm 1: MR[a]			Arm 2: MR+PREP[b]			Arm 3: PREP			Arm 4: waitlist		
	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value	n	Mean (SD)	*P* value
Baseline (T1)												
PSS[c]	35	20.2 (5.7)		36	18.3 (6.5)		36	17.5 (7.2)		32	17.0 (5.7)	
BDI[d]	35	13.6 (9.7)		36	13.3 (1.4)		36	10.9 (1.9)		32	9.9 (8.1)	
PCL-C[e]	35	33.7 (12.6)		36	33.4 (15.0)		36	29.2 (11.7)		32	29.1 (9.0)	
SCStotal[f]	35	81.7 (18.5)		36	77.0 (2.5)		36	85.1 (23.1)		32	84.3 (16.4)	
RSES[g]	35	62.9 (14.6)		36	59.0 (19.5)		36	63.0 (17.8)		32	62.8 (16.0)	
MSPSStotal[h]	35	63.6 (15.2)		36	65.4 (18.2)		36	66.2 (13.7)		32	67.3 (11.3)	
PSQItotal[i]	35	9.1 (4.5)		36	9.0 (4.2)		36	7.8 (4.3)		32	8.0 (4.3)	
RDAS[j]	35	45.4 (12.5)		36	46.4 (11.7)		36	44.8 (1.5)		32	44.7 (8.7)	
PainUsual[k]	35	2.9 (2.8)		36	3.9 (2.0)		36	2.5 (2.2)		32	3.4 (2.5)	
PainBest[l]	35	1.4 (1.8)		36	1.8 (1.9)		36	.9 (1.2)		32	1.6 (1.7)	
8 Weeks (T2)												
PSS	34	16.3 (6.2)	.004*	35	16.5 (7.0)	.02	35	16.2 (6.9)	.17	31	18.1 (5.8)	.27
BDI	34	7.6 (9.9)	.0001*	35	10.0 (11.2)	.02	36	10.2 (11.1)	.70	31	10.3 (1.2)	.99
PCL-C	32	29.1 (12.7)	.0005*	33	30.7 (14.3)	.04	36	30.3 (14.5)	.67	32	29.6 (1.2)	.77
SCStotal	33	90.2 (19.6)	.01	33	83.5 (21.6)	.006	35	87.4 (22.3)	.25	31	84.7 (18.9)	.63
RSES	34	65.7 (13.7)	.35	35	61.2 (2.5)	.23	34	61.4 (19.0)	.65	32	61.7 (17.8)	.59
MSPSStotal	34	68.2 (12.7)	.01	33	65.4 (2.6)	.96	36	68.1 (15.2)	.25	31	66.3 (11.9)	.49
PSQItotal	34	6.4 (4.2)	.0002*	34	7.9 (4.6)	.17	36	7.4 (4.1)	.44	31	8.4 (5.2)	.64
RDAS	34	46.8 (13.2)	.31	35	47.7 (12.3)	.39	36	44.5 (11.5)	.88	32	44.6 (1.3)	.86
PainUsual	34	3.1 (2.4)	.64	34	3.1 (2.8)	.04	36	2.4 (2.4)	.78	32	3.7 (2.6)	.43
PainBest	34	1.2 (1.8)	.67	34	1.6 (2.4)	.80	36	.9 (1.3)	.92	32	1.9 (2.2)	.36
16 Weeks (T3)												
PSS	33	15.6 (7.0)	.002*	35	15.1 (7.1)	.002*	35	16.0 (7.7)	.13	32	17.6 (7.1)	.48
BDI	33	9.3 (11.1)	.02	32	8.2 (1.0)	.002*	36	10.3 (1.3)	.68	32	10.8 (9.1)	.51
PCL-C	34	30.3 (15.1)	.09	34	28.7 (14.4)	.007*	35	29.1 (14.6)	.66	32	29.9 (12.0)	.65
SCStotal	33	91.1 (21.2)	.003*	32	87.5 (2.8)	.001*	33	88.0 (22.3)	.21	32	86.6 (2.4)	.22
RSES	31	65.2 (13.1)	.17	34	62.5 (2.9)	.104	34	62.1 (18.9)	.88	32	61.4 (16.6)	.40
MSPSStotal	34	65.0 (16.3)	.63	32	62.1 (22.9)	.33	34	67.9 (11.3)	.48	32	64.3 (14.3)	.22
PSQItotal	33	7.4 (4.5)	.01	34	7.3 (4.4)	.003*	34	7.9 (4.5)	.28	32	8.8 (4.4)	.24
RDAS	33	47.3 (12.3)	.19	33	47.8 (13.5)	.77	34	44.6 (1.7)	.64	32	44.2 (1.8)	.60
PainUsual	32	2.5 (2.6)	.20	34	2.7 (2.8)	.03	34	2.3 (2.2)	.73	32	3.7 (3.1)	.53
PainBest	32	1.2 (2.0)	.54	34	1.3 (2.0)	.09	34	1.4 (1.9)	.05	32	2.1 (2.2)	.29

[a]MR: Mission Reconnect.

[b]PREP: Prevention and Relationship Enhancement Program.

[c]PSS: Perceived Stress Scale-10 item.

[d]BDI: Beck Depression Inventory.

[e]PCL-C: PTSD Checklist-Civilian version.

XSL·FO

RenderX

[f]SCStotal: Self-Compassion Scale.

[g]RSES: Response to Stressful Experiences Scale.

[h]MSPSStotal: Multidimensional Scale of Perceived Social Support.

[i]PSQItotal: Pittsburgh Sleep Quality Index.

[j]RDAS: Revised Dyadic Adjustment Scale.

[k]PainUsual: rating of usual pain level over the past week.

[l]PainBest: rating of best pain level over the past week.

*P<.05 after adjustment for experimentwise error rate (true P values before adjustment displayed).

Massage Effects

We included partner massage as an important element of MR both to give couples another way to communicate their care and because our own prior studies found that brief massage by family members significantly reduced troublesome symptoms [23,34]. Both veterans and partners reported highly significant reductions in all assessed symptoms of physical pain, tension, irritability, anxiety or worry, and depression shortly after receiving massage (Table 8), thus supporting hypothesis 4. In addition, both veterans and partners reported significant reductions in most premassage symptoms over the 8-week period (Table 9). While the short-term improvements recorded in the pre- and postmassage scores can reasonably be attributed to massage effects, the premassage changes over time cannot. They may be influenced, at least in part, by the MR program use overall, but we did not directly test that.

Table 8. Massage session effects on veteran-partner dyads: changes in symptom ratings (Wilcoxon signed rank tests).

Symptoms	Before mean (SD)	After mean (SD)	S[a]	P value
Veterans (n=453)				
Physical pain	3.6 (2.6)	2.1 (2.0)	26555	.0001*
Physical tension	4.2 (2.3)	1.9 (1.8)	40798.5	.0001*
On edge/irritable	3.8 (2.7)	1.6 (1.9)	34376	.0001*
Anxiety/worry	3.7 (2.8)	1.7 (2.1)	31278	.0001*
Depression	2.3 (2.7)	1.2 (1.9)	13169.5	.0001*
Partners (n=371)				
Physical pain	3.6 (2.6)	2.1 (2.1)	20518	.0001*
Physical tension	4.5 (2.5)	2.0 (1.9)	28796	.0001*
On edge/irritable	4.0 (2.7)	1.6 (1.9)	26398	.0001*
Anxiety/worry	4.1 (2.7)	1.9 (2.1)	23646.5	.0001*
Depression	2.3 (2.7)	1.4 (2.1)	6310.5	.0001*

[a]Signed rank statistic.

*P<.05 after adjustment for experimentwise error rate (true P values before adjustment displayed).

Table 9. Changes in premassage symptom ratings over 8 weeks in veteran-partner dyads (Kendall tau-b results).

Symptoms	τ	P value
Veterans (453 session reports)		
Physical pain	−.003	.92
Physical tension	−.116	.001*
On edge/irritable	−.113	.002*
Anxiety/worry	−.122	.0005*
Depression	−.055	.13
Partners (371 session reports)		
Physical pain	−.019	.63
Physical tension	−.096	.01
On Edge/irritable	−.089	.02
Anxiety/worry	−.112	.004*
Depression	−.088	.03

*$P < .05$ after adjustment for experimentwise error rate (true P values before adjustment displayed).

Power

In view of the encouraging phase I feasibility results [23], after adding a margin of conservatism, we chose a moderately large effect size, corresponding to a value of 0.7 for Cohen *d*. The analysis of veterans' within-group change (Table 3), assuming n=40 per arm, and type I error probability alpha=.05, obtained a power of 1–beta=99%. Between-group changes (Table 5, Table 6) yielded a power of 87%. Partners' within-group changes (Table 7), with n=32 per arm, obtained a power of 96%. Massage session effects (Table 8) yielded a power >99% for both veterans and partners.

User Satisfaction

On the final survey (T3) we asked MR participants how likely they would be to recommend the program to a friend. On a 0–10 scale, veterans' mean score was 8.7, and partners' mean was 9.1, indicating high user satisfaction with the program.

Discussion

Key Findings

Our primary intention in designing MR was to offer veterans and their partners a flexible form of instruction in simple ways to improve their own well-being. The range of outcomes in which users had significant improvement can be seen as confirmation that MR teaches skills that improved their well-being. The fact that improvements, including reduced PTSD symptoms, and increased self-compassion, were sustained at the 16-week follow-up is particularly promising.

Hypothesis 1 was generally supported by the findings that MR participants had significant improvement on far more mental health outcomes than did participants in other arms of the study. The exceptions in this case were sleep quality (partners using MR had more stable improvements than veterans), and lack of movement in dyadic adjustment and perceived social support. We speculate that the lack of change in relationship variables

may have been due to high baseline levels, which may have led to the decision to participate, but this is conjecture and warrants further study. Hypothesis 1 was not fully supported in that, while the 2-sample *t* tests indicated significantly stronger improvement among MR participants on many mental health outcomes, only a portion of these remained following adjustment for potential experimentwise error. The remaining hypotheses were strongly supported.

It is notable that both veterans and partners in the MR arms used this self-directed program for over 2 hours per week during the initial 8 weeks, surpassing our hypothesized use, and averaged well over an hour per week throughout the 16-week data collection period. We suspect the control each user had over which program elements to use and when contributed to the high use level, underscoring the value of user preference in long-term adherence to, and ensuing effectiveness of, self-care approaches for this population.

Sustained use may also have been enhanced through the inherent support and encouragement of compliance by coparticipation with a significant relationship partner. We note, however, that the majority of MR practices (except for massage) can be used either alone or with others. This adds to the convenience for young parents, allowing one partner to practice while the other attends to their children.

In comparing the trial's 4 arms, the contrast for both veterans and partners is greatest between MR-only and waitlist control arms, a predictable finding based on phase I results. The contrast in outcomes for the MR-only versus PREP arms, especially at 8 weeks, is notable. While PREP has been found in multiple studies to support significant improvement in variables related to relationship dynamics and intimacy for couples, we chose outcome measures to assess other dimensions of mental health. We selected some MR components based on evidence of health-promoting neurological and neurochemical effects; potential easing of symptoms related to PTSD, stress, and depression; and likelihood of enhancing self-compassion and

forgiveness. We expected the physical exercises to amplify energy, improve mood, and reduce susceptibility to the fight-flight-freeze response. While PREP offers some content aimed at relaxation, stress reduction, and forgiveness, it devotes more time to relationship skills per se than does MR.

Thus, one interpretation of the outcomes is that MR performed well at its goals and PREP does less well at the things MR is designed to do. As to whether combining the programs adds value, adding MR to PREP appears to amplify improvement in self-compassion and response to stressful events. Adding PREP did not enhance other results for MR-only.

The program showed strong benefit for both men and women veterans and for their partners, although benefits were strongest for veterans. Based on these findings, which included veterans from all branches of service, and participants both on active duty and retired or separated from service, it appears that the program provides a safe, low-cost self-care intervention that enhances overall well-being. A modified version could be useful for single veterans.

Limitations

This study evaluated MR in a community-based sample with no inclusion or exclusion criteria related to specific mental or physical health parameters. Thus, the program's impact in clinically defined populations remains to be assessed. Other limitations are that the follow-up period was limited to 16 weeks and that the program was tested only with dyads, not single veterans. In addition, all participants were required to attend an in-person launch meeting, and it is possible that this excluded potential applicants for whom such attendance was impossible for reasons of time, geography, or something else. Finally, the instructional program as tested did not include video closed captioning or verbatim transcripts of the audio instruction, and thus would not have accommodated users with hearing limitations. These enhancements will be considered in future upgrading.

Suggested Research

Four directions for future research on MR for veterans present themselves. First is to test whether adherence and benefit are still present at the 1-year mark or even further with a partnered sample such as we used. Second is to offer MR, minus the partner massage, to nonpartnered veterans to see whether lack of a partner affects use or outcomes. Third, it is important to follow this study with research on diagnostically defined

samples (eg, high PTSD) to see what benefit might be provided for a clinically based sample. Fourth, we also recommend that health services research be done to determine how best to offer MR. This study demonstrates that veteran-partner dyads can learn a range of physical and contemplative self-care practices on their own from a media program accessed via mobile app or webstream, and that they can derive great benefit across several dimensions of well-being. Still, it is possible that, especially for a clinically based or nonpartnered sample, use of the program would be enhanced by introduction through in-person group instruction, for example, with 1 to 4 structured sessions. While this would add cost to delivery of the program, it is possible that added benefit could justify the cost. Related to this, one could retest MR, forgoing the launch meetings and substituting Web-based instruction in the provision of weekly report and survey data.

In addition, we recognize that many nonveterans have PTSD, physical pain, lack of self-compassion, etc. MR, as it now exists, or with minor changes, could and should be tested with other samples. This could include others who have experienced severe trauma, such as refugees, youth and adults exposed to mass shootings such as that in an Orlando, FL nightclub in 2016, as well as people with chronic pain and more generalized anxiety.

Conclusions

This study adds to the growing literature on the power of brief, repeated mind- and body-based practices to address physical, psychological, and spiritual well-being [50-52]. The results indicate that MR is a widely accessible, low-cost approach that supports well-being and reduces multiple symptoms among post-9/11 veterans and their partners. Both veterans and partners were able to learn and make sustained use of a range of wellness practices from a media-only source. While the launch meetings offered 90 minutes of human contact, they offered no instruction in any of the practices. These usage findings contrast with others' findings of high dropout rates for clinic-based programs such as meditation instruction [53]. Home-based, self-directed interventions may be of particular service to veterans who are distant from, averse to, or prohibited by schedule from using professional services. The partner relationship may enhance sustained use of self-directed interventions for this population. Finally, these data suggest MR to be superior to waitlist and PREP for Strong Bonds in both within-group and between-group assessments. Notably, this distinction appears to increase over time.

Conflicts of Interest

The Mission Reconnect program is owned and distributed by Mission Reconnect LLC, which was founded by Drs Kahn and Collinge to make this program widely available.

Multimedia Appendix 1

Video Introduction to Mission Reconnect.

[MP4 File (MP4 Video), 78MB - jmir_v18i9e255_app1.mp4]

Multimedia Appendix 2

Weekly report cards for Mission Reconnect participants (MR Only and MR+PREP) during weeks 1-8.

[PDF File (Adobe PDF File), 97KB - jmir_v18i9e255_app2.pdf]

Multimedia Appendix 3

Weekly report cards for non-Mission Reconnect participants (PREP+Waitlist) during weeks 1-8.

[PDF File (Adobe PDF File), 90KB - jmir_v18i9e255_app3.pdf]

Multimedia Appendix 4

CONSORT eHealth checklist.

[PDF File (Adobe PDF File), 7MB - jmir_v18i9e255_app4.pdf]

Multimedia Appendix 5

Table 2. Demographic Information.

[PDF File (Adobe PDF File), 28KB - jmir_v18i9e255_app5.pdf]

Multimedia Appendix 6

Mission Reconnect program utilization at weeks 2, 8, and 16.

[PNG File, 193KB - jmir_v18i9e255_app6.png]

Multimedia Appendix 7

Wellness strategies of non-Mission Reconnect participants (PREP + Waitlist).

[PNG File, 104KB - jmir_v18i9e255_app7.png]

References

1. Otis J, McGlinchey R, Vasterling J, Kerns R. Complicating factors associated with mild traumatic brain injury: impact on pain and posttraumatic stress disorder treatment. J Clin Psychol Med Settings 2011;18(2):145-154. [Medline: 21626354]
2. Sharpless B, Barber J. A clinician's guide to PTSD treatments for returning veterans. Prof Psychol Res Pr 2011;42(1):8-15. [Medline: 21475611]
3. Outcalt SD, Ang DC, Wu J, Sargent C, Yu Z, Bair MJ. Pain experience of Iraq and Afghanistan Veterans with comorbid chronic pain and posttraumatic stress. J Rehabil Res Dev 2014;51(4):559-570 [FREE Full text] [doi: 10.1682/JRRD.2013.06.0134] [Medline: 25144169]
4. Ramchand R, Schell TL, Karney BR, Osilla KC, Burns RM, Caldarone LB. Disparate prevalence estimates of PTSD among service members who served in Iraq and Afghanistan: possible explanations. J Trauma Stress 2010 Feb;23(1):59-68 [FREE Full text] [doi: 10.1002/jts.20486] [Medline: 20135699]
5. Matthias MS, McGuire AB, Kukla M, Daggy J, Myers LJ, Bair MJ. A brief peer support intervention for veterans with chronic musculoskeletal pain: a pilot study of feasibility and effectiveness. Pain Med 2015 Jan;16(1):81-87 [FREE Full text] [doi: 10.1111/pme.12571] [Medline: 25312858]
6. Kerns RD, Otis J, Rosenberg R, Reid MC. Veterans' reports of pain and associations with ratings of health, health-risk behaviors, affective distress, and use of the healthcare system. J Rehabil Res Dev 2003;40(5):371-379 [FREE Full text] [Medline: 15080222]
7. Haskell SG, Ning Y, Krebs E, Goulet J, Mattocks K, Kerns R, et al. Prevalence of painful musculoskeletal conditions in female and male veterans in 7 years after return from deployment in Operation Enduring Freedom/Operation Iraqi Freedom. Clin J Pain 2012 Feb;28(2):163-167. [doi: 10.1097/AJP.0b013e318223d951] [Medline: 21677563]
8. Alschuler KN, Otis JD. An examination of the impact of clinically significant levels of posttraumatic stress disorder symptomatology on the classification of pain as mild, moderate, or severe in a sample of veterans with chronic pain. Psychol Serv 2014 Aug;11(3):273-280. [doi: 10.1037/a0036201] [Medline: 24635041]

9.   Schuman D. A value-critical choice analysis of a policy to prevent suicide in veterans and service members. Soc Work Public Health 2016:2-12. [Medline: 27254385]

10.   Milliken CS, Auchterlonie JL, Hoge CW. Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. JAMA 2007 Nov 14;298(18):2141-2148. [doi: 10.1001/jama.298.18.2141] [Medline: 18000197]

11.   Tanielian T, Jaycox L. Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery. Santa Monica, CA: Rand Center for Military Health Policy Research; 2008.

12.   Tanielian T. Bridging the Gaps in Treating Veterans With Post-Deployment Mental Health Problems.: Rand Corporation; 2011 Dec 5. URL: http://www.rand.org/commentary/2011/12/05/AH.html [accessed 2016-09-09] [WebCite Cache ID 6kOchCZKM]

13.   Baker MS. Casualties of the Global War on Terror and their future impact on health care and society: a looming public health crisis. Mil Med 2014 Apr;179(4):348-355. [doi: 10.7205/MILMED-D-13-00471] [Medline: 24690957]

14.   Hourani L, Bender R, Weimer B, Peeler R, Bradshaw M, Lane M, et al. Longitudinal study of resilience and mental health in Marines leaving military service. J Affect Disord 2012 Jul;139(2):154-165. [doi: 10.1016/j.jad.2012.01.008] [Medline: 22381952]

15.   Department of Veterans Affairs Office of Research and Development, National Institute of Mental Health, United States Army Medical Research and Materiel Command. Mapping the Landscape of Deployment Related Adjustment and Mental Disorders: A Meeting Summary of a Working Group to Inform Research. 2006. URL: http://www.nimh.nih.gov/research-priorities/scientific-meetings/2006/mapping-the-landscape-of-deployment-related-adjustment-and-mental-disorders_34850.pdf [accessed 2016-09-14] [WebCite Cache ID 6kWFK11W9]

16.   The White House. Strengthening Our Military Families: Meeting America's Commitment. Presidential Study Directive/PSD-9. 2011 Jan. URL: http://archive.defense.gov/home/features/2011/0111_initiative/strengthening_our_military_january_2011.pdf [accessed 2016-09-14] [WebCite Cache ID 6kWFOGrMG]

17.   Allen ES, Rhoades GK, Stanley SM, Markman HJ. On the home front: stress for recently deployed Army couples. Fam Process 2011 Jun;50(2):235-247 [FREE Full text] [doi: 10.1111/j.1545-5300.2011.01357.x] [Medline: 21564063]

18.   de Burgh HT, White CJ, Fear NT, Iversen AC. The impact of deployment to Iraq or Afghanistan on partners and wives of military personnel. Int Rev Psychiatry 2011 Apr;23(2):192-200. [doi: 10.3109/09540261.2011.560144] [Medline: 21521089]

19.   McLeland KC, Sutton GW, Schumm WR. Marital satisfaction before and after deployments associated with the Global War on Terror. Psychol Rep 2008 Dec;103(3):836-844. [doi: 10.2466/pr0.103.3.836-844] [Medline: 19320220]

20.   Allen ES, Rhoades GK, Stanley SM, Loew B, Markman HJ. The effects of marriage education for army couples with a history of infidelity. J Fam Psychol 2012 Feb;26(1):26-35 [FREE Full text] [doi: 10.1037/a0026742] [Medline: 22229880]

21.   Beardslee W, Lester P, Klosinski L, Saltzman W, Woodward K, Nash W, et al. Family-centered preventive intervention for military families: implications for implementation science. Prev Sci 2011 Dec;12(4):339-348 [FREE Full text] [doi: 10.1007/s11121-011-0234-5] [Medline: 21761154]

22.   VA San Diego Healthcare System. From Warrior to Soul Mate Weekend Couples Retreats: Reclaim your Relationship after Military Service. Washington, DC: US Department of Veterans Affairs; 2016 Jun 30. URL: http://www.sandiego.va.gov/services/From_Warrior_to_Soul_Mate.asp [accessed 2016-09-09] [WebCite Cache ID 6kOfHv8Lq]

23.   Collinge W, Kahn J, Soltysik R. Promoting reintegration of National Guard veterans and their partners using a self-directed program of integrative therapies: a pilot study. Mil Med 2012 Dec;177(12):1477-1485 [FREE Full text] [Medline: 23397692]

24.   Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 8;196(4286):129-136. [Medline: 847460]

25.   Walker RL, Clark ME, Sanders SH. The "Postdeployment multi-symptom disorder": an emerging syndrome in need of a new treatment paradigm. Psychol Serv 2010;7(3):136-147. [doi: 10.1037/a0019684]

26.   Hempel S, Taylor S, Marshall N, Miake-Lye I, Beroes J, Shanman R, et al. Evidence map of mindfulness. QUERI 2014 Oct:1-36. [Medline: 25577939]

27.   Canadian Agency for Drugs and Technologies in Health. Mindfulness interventions for the treatment of post-traumatic stress disorder, generalized anxiety Disor der, Depression, and substance use disorders: a review of the clinical effectiveness and guidelines. CADTH Rapid Response Reports 2015 Jun:1-43. [Medline: 26180855]

28.   Field T. Massage therapy research review. Complement Ther Clin Pract 2014 Nov;20(4):224-229. [doi: 10.1016/j.ctcp.2014.07.002] [Medline: 25172313]

29.   Crawford C, Boyd C, Paat CF, Price A, Xenakis L, Yang E, Evidence for Massage Therapy (EMT) Working Group. The impact of massage therapy on function in pain populations: a systematic review and meta-analysis of randomized controlled trials: part I, patients experiencing pain in the general population. Pain Med 2016 May 10:pii. [doi: 10.1093/pm/pnw099] [Medline: 27165971]

30.   Fredrickson B. Cultivating positive emotions to optimize health and well-being. Prev Treatment 2000 Mar 07;3(1):1-25 [FREE Full text] [doi: 10.1037/1522-3736.3.1.31a]

31.   Caputo A. The relationship between gratitude and loneliness: the potential benefits of gratitude for promoting social bonds. Eur J Psychol 2015 May;11(2):323-334 [FREE Full text] [doi: 10.5964/ejop.v11i2.826] [Medline: 27247660]

32.   Tugade MM, Fredrickson BL. Resilient individuals use positive emotions to bounce back from negative emotional experiences. J Pers Soc Psychol 2004 Feb;86(2):320-333 [FREE Full text] [doi: 10.1037/0022-3514.86.2.320] [Medline: 14769087]

33.   Kozak L, Vig E, Simons C, Eugenio E, Collinge W, Chapko M. A feasibility study of caregiver-provided massage as supportive care for Veterans with cancer. J Support Oncol 2013 Sep;11(3):133-143. [Medline: 24400393]

34.   Collinge W, Kahn J, Walton T, Kozak L, Bauer-Wu S, Fletcher K, et al. Touch, Caring, and Cancer: randomized controlled trial of a multimedia caregiver education program. Support Care Cancer 2013 May;21(5):1405-1414 [FREE Full text] [doi: 10.1007/s00520-012-1682-6] [Medline: 23262808]

35.   Collinge W, Kahn J. Mission Reconnect. 2014. URL: https://missionreconnect.com/ [accessed 2016-09-09] [WebCite Cache ID 6kOjcf4OZ]

36.   Stanley SM, Allen ES, Markman HJ, Rhoades GK, Prentice DL. Decreasing divorce in army couples: results from a randomized controlled trial using PREP for Strong Bonds. J Couple Relatsh Ther 2010 Apr;9(2):149-160 [FREE Full text] [doi: 10.1080/15332691003694901] [Medline: 20634994]

37.   Markman HJ, Stanley SM, Blumberg SL. Fighting for Your Marriage: A Deluxe Revised Edition of the Classic Best-Seller for Enhancing Marriage and Preventing Divorce. San Francisco, CA: Jossey-Bass; 2010.

38.   Iran and Afghanistan Veterans of America. Be Part of the Movement. New York, NY: IAVA URL: http://iava.org/ [accessed 2016-09-09] [WebCite Cache ID 6kOk7Vrad]

39.   Dallal GE. Randomization.com. 2013 Mar 29. URL: http://www.randomization.com/ [accessed 2016-09-09] [WebCite Cache ID 6kOkHwcDR]

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

41.   Beck A, Steer R. Beck Depression Inventory Manual. San Antonio, TX: Harcourt Brace; 1996:1-38.

42.   Weathers F, Huska J, Keane T. PCL-C for DSM-IV. Boston, MA: National Center for PTSD Behavioral Science Division; 1991. URL: http://www.mirecc.va.gov/docs/visn6/3_PTSD_CheckList_and_Scoring.pdf [accessed 2016-03-23] [WebCite Cache ID 6gE03udXx]

43.   Neff K. The development and validation of a scale to measure self-compassion. Self Identity 2003;2:223-250.

44.   Johnson DC, Polusny MA, Erbes CR, King D, King L, Litz BT, et al. Development and initial validation of the Response to Stressful Experiences Scale. Mil Med 2011 Feb;176(2):161-169. [Medline: 21366078]

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

46.   Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

47.   Busby DM, Christensen C, Crane DR, Larson JH. A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: construct hierarchy and multidimensional scales. J Marital Fam Ther 1995;32(3):289-308. [doi: 10.1111/jmft.12095]

48.   Yarnold P, Soltysik R. Optimal Data Analysis: A Guidebook with Software for Windows. Washington, DC: APA Books; 2004.

49.   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]

50.   Fredrickson BL, Cohn MA, Coffey KA, Pek J, Finkel SM. Open hearts build lives: positive emotions, induced through loving-kindness meditation, build consequential personal resources. J Pers Soc Psychol 2008 Nov;95(5):1045-1062 [FREE Full text] [doi: 10.1037/a0013262] [Medline: 18954193]

51.   Garland EL, Fredrickson B, Kring AM, Johnson DP, Meyer PS, Penn DL. Upward spirals of positive emotions counter downward spirals of negativity: insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. Clin Psychol Rev 2010 Nov;30(7):849-864 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.002] [Medline: 20363063]

52.   Seppälä EM, Nitschke JB, Tudorascu DL, Hayes A, Goldstein MR, Nguyen DTH, et al. Breathing-based meditation decreases posttraumatic stress disorder symptoms in U.S. military veterans: a randomized controlled longitudinal study. J Trauma Stress 2014 Aug;27(4):397-405 [FREE Full text] [doi: 10.1002/jts.21936] [Medline: 25158633]

53.   Williams J, Gierisch J, McDuffie J, Strauss J, Nagi A. An Overview of Complementary and Alternative Medicine Therapies for Anxiety and Depressive Disorders: Supplement to Efficacy of Complementary and Alternative Medicine Therapies for Posttraumatic Stress Disorder. Evidence-based Synthesis Program. Washington, DC: Department of Veterans Affairs; 2011:1-23.

Abbreviations

BDI: Beck Depression Inventory

XSL•FO

RenderX

DoD: Department of Defense
MR: Mission Reconnect
MSPSStotal: Multidimensional Scale of Perceived Social Support
PainBest: rating of best pain level over the past week
PainUsual: rating of usual pain level over the past week
PCL-C: PTSD Checklist-Civilian version
PREP: Prevention and Relationship Enhancement Program
PSQItotal: Pittsburgh Sleep Quality Index
PSS: Perceived Stress Scale-10 item
PTSD: posttraumatic stress disorder
RDAS: Revised Dyadic Adjustment Scale
RSES: Response to Stressful Experiences Scale
SCStotal: Self-Compassion Scale
VA: US Department of Veterans Affairs

Original Paper

Design and Testing of BACRA, a Web-Based Tool for Middle Managers at Health Care Facilities to Lead the Search for Solutions to Patient Safety Incidents

Irene Carrillo[1], BSc, MSc; José Joaquín Mira[1,2], PhD; Maria Asuncion Vicente[3], PhD; Cesar Fernandez[3], PhD; Mercedes Guilabert[1], PhD; Lena Ferrús[4], MHA, PhD; Elena Zavala[5], MD; Carmen Silvestre[6], PhD; Pastora Pérez-Pérez[7], MPH

[1]Health Psychology Department, Miguel Hernández University, Elche, Spain

[2]Alicante-Sant Joan Health District, Consellería de Sanidad Universal y Salud Pública, Alicante, Spain

[3]Systems Engineering and Automation Department, Miguel Hernández University, Elche, Spain

[4]Consorci Sanitari Integral, L'Hospitalet de Llobregat, Barcelona, Spain

[5]Donostia University Hospital, Donostia, Spain

[6]Navarra Health Service - Osasunbidea, Pamplona, Spain

[7]Patient Safety Observatory, Andalusian Agency for Health Care Quality, Seville, Spain

Corresponding Author:
Irene Carrillo, BSc, MSc
Health Psychology Department
Miguel Hernández University
Universidad s/n
Elche, 03202
Spain
Phone: 34 966658984
Fax: 34 966658984
Email: icarrillo@umh.es

Abstract

Background: Lack of time, lack of familiarity with root cause analysis, or suspicion that the reporting may result in negative consequences hinder involvement in the analysis of safety incidents and the search for preventive actions that can improve patient safety.

Objective: The aim was develop a tool that enables hospitals and primary care professionals to immediately analyze the causes of incidents and to propose and implement measures intended to prevent their recurrence.

Methods: The design of the Web-based tool (BACRA) considered research on the barriers for reporting, review of incident analysis tools, and the experience of eight managers from the field of patient safety. BACRA's design was improved in successive versions (BACRA v1.1 and BACRA v1.2) based on feedback from 86 middle managers. BACRA v1.1 was used by 13 frontline professionals to analyze incidents of safety; 59 professionals used BACRA v1.2 and assessed the respective usefulness and ease of use of both versions.

Results: BACRA contains seven tabs that guide the user through the process of analyzing a safety incident and proposing preventive actions for similar future incidents. BACRA does not identify the person completing each analysis since the password introduced to hide said analysis only is linked to the information concerning the incident and not to any personal data. The tool was used by 72 professionals from hospitals and primary care centers. BACRA v1.2 was assessed more favorably than BACRA v1.1, both in terms of its usefulness (z=2.2, *P*=.03) and its ease of use (z=3.0, *P*=.003).

Conclusions: BACRA helps to analyze incidents of safety and to propose preventive actions. BACRA guarantees anonymity of the analysis and reduces the reluctance of professionals to carry out this task. BACRA is useful and easy to use.

(*J Med Internet Res 2016;18(9):e257*)   doi:10.2196/jmir.5942

KEYWORDS

patient safety; risk management; root cause analysis; hospital; primary care; frontline health professionals; middle managers

XSL•FO
RenderX

Introduction

Analyzing the causes of unsafe care helps reduce the number of incidents that may cause harm to patients [1-3]. This is the primary reason that incident reporting systems (IRS) for patient safety exist [4].

Incident Reporting Systems

Incident reporting systems were introduced into the health care sector in the late 1990s. Similar mechanisms can be found in high-risk sectors, such as the nuclear, railway, and aviation industries. They are essentially mechanisms designed to record critical incidents anonymously [3]. Once analyzed, the information collected leads to improved safety. These IRS systems are usually voluntary, but in some countries or industries they are compulsory.

The Australian Incident Monitoring System, the Sentinel Events Reporting Program, and the New York Patient Occurrence Reporting and Tracking System (NYPORTS) in the United States are three of the first programs designed to learn from incidents [5-7]. In Canada, Europe, and Latin America, IRS designed under similar premises are also in operation, but each system tends to follow different regulations, taxonomies, types of incidents reporting, system management procedures, incentives for reporting, and methods and resources for analyzing information collected [8,9].

Incident reporting systems provide a mechanism to identify risks (in this context how and why patients can be harmed) and surface learning opportunities in different organizations, based on their own experience, toward the objective of reducing the frequency of safety incidents [10]. They can and should also be used to share lessons within and across organizations concerned with improving safety. In the end, such mechanisms are designed to enhance patient safety culture in health care organizations [9].

Incident reporting systems represent a central data collection of incidents with the aim to define fields where action is needed most. These systems are well established, but sometimes they do not succeed in developing an action plan to implement corresponding measures conceived to prevent recurring incidents. The success of IRS requires the involvement of frontline health care providers in the action plan (analyzing causes and proposing solutions related to specific incidents).

In order for IRS to be effective, it is critical to overcome the distrust of professionals who fear the possible consequences of reporting incidents [11]. As well, it is essential to make certain that appropriate procedures and resources for quickly analyzing the causes of incidents are already in place [9] and that systems exist to disseminate the conclusions drawn from incident analysis such that similar incidents among health care professionals may be avoided in the future [10-12].

Incident Analysis for Patient Safety

Root cause analysis, critical incident analysis, and incident simulations are the most useful techniques for investigating what happened [13-15]. Analyzing reported incidents requires time, knowledge, and confidence in the confidentiality of the use of the information [10,16]. However, this is not always possible because as the number of reported incidents increases, the ability to analyze their causes is reduced. There are other care obligations that frontline personnel prioritize more than incident analysis [11]. This prevents middle managers and professionals who are nearest the incident from always being able to participate in its analysis and to propose alternatives or changes to prevent recurrence [11,17,18].

Patients' Needs After an Adverse Event

Patients who have suffered an adverse event (AE) should receive information about what and how the incident occurred, and about the measures adopted to prevent recurrence [19,20]. New tools are needed to make certain that patients are informed promptly of countermeasures following incidents.

Why This Study

Senior and middle managers have direct responsibility for incidents that can be analyzed and assessed in the interest of preventing recurrence [10,21]. For various reasons, however, they do not always make this happen [22,23].

This study's objective was to develop a tool that helps middle managers and frontline professionals carry out immediate analysis of the causes of incidents related to patient safety whereby they may propose and implement solutions to prevent recurrence. This tool should provide them with the following: a guarantee that the tool adheres to relevant legal regulations; that it engages middle managers and their teams; that it permits appropriate identification of harmful incidents and near misses, probing their immediate and latent causes; and that it conducts a dynamic and agile analysis of these incidents.

Methods

Study design of a tool to identify preventive actions for the improvement of patient safety based on the information collected and analyzed from the experiences of previous incidents. Figure 1 describes the steps followed to establish the criteria for this new tool's design.

XSL•FO
RenderX

Figure 1. Steps for the design of the incident analysis tool.
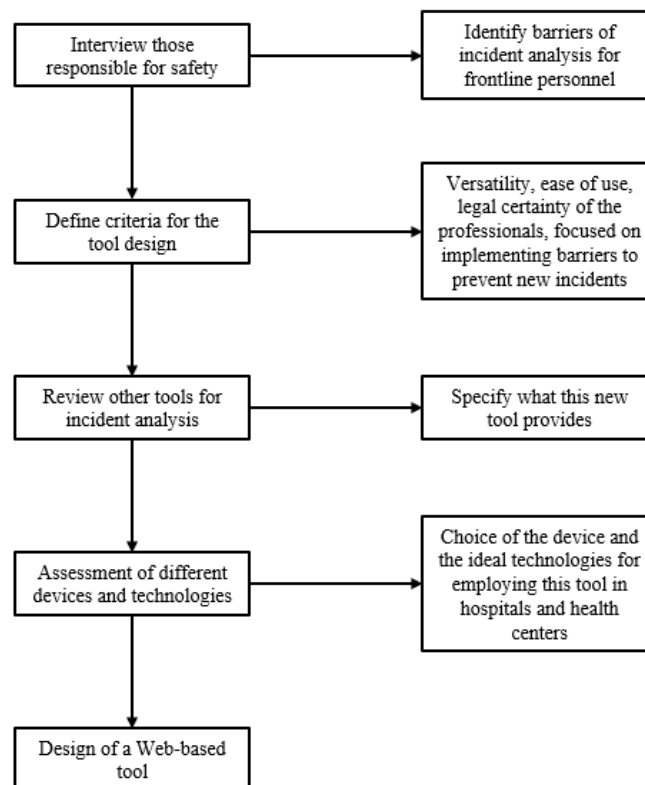
Incident Analysis Tool Requirements

To develop the tool, prior research on the barriers for reporting safety incidents, and incident analysis techniques, were considered first [11,24,25]. This information was expanded by conducting eight semistructured interviews. These interviews were conducted with individuals responsible for safety at hospitals and primary health care centers. They described the difficulties they face in analyzing patient safety incidents and involving professionals and middle managers in the search for solutions to achieve a safer environment for patients.

With this information, a series of criteria were then established with respect to design, navigability, information security and confidentiality, and structure for the analysis of incidents. Based on this, we proposed a feasible projection of what this Web-based tool should deliver in the search for solutions to safety incidents at hospitals and primary care centers.

A review was carried out regarding the tools to conduct an analysis of the causes, consequences, and search for solutions to the incidents, in a broad-based effort to manage risks to patient safety at hospitals and primary care centers. These tools were assessed using criteria established independently by IC, MG, and JJM, with the goal of identifying the characteristics that a new tool should have. From this review, it was determined that the protocol of incident analysis based on the Harvard study [26], the 5 Whys [27], characteristics of root cause analysis, the prioritization of risks employed in failure mode and effect analysis [28], and the matrix employed by the plan-do-check-act (PDCA) tool [29] offered ideal models for this new tool. Due to these characteristics, the tool was renamed BACRA (in

Spanish *Basado en Análisis Causa-RAíz meaning* "based on root cause analysis").

Different alternatives were considered for the development of the incident analysis tool: creating a mobile app for Android and/or iOS tablets and mobile phones (discarded because the necessary devices were not available at most health care centers), developing an executable program (discarded because of the difficulty of installing unofficial apps at health care centers), developing a portable document format (PDF; discarded due to its limitations for generating dynamic content and questions based on previous responses), and a Web form that permitted a sufficient level of flexibility and was accessible from any computer with an Internet connection. The latter was the option chosen.

From this information, a BACRA beta version was developed, a Web tool based on root cause analysis to search for solutions to incidents of patient safety with leadership from middle managers.

BACRA Design

The tool's beta version was presented to 43 professionals (middle managers of nursing and surgical medicine, intensive care units, blood banks, laboratories, radiology, mental health, pediatrics, surgery, orthopedics, gynecology, medicine, and primary care). Their feedback helped to improve data access and privacy, and their assessments and suggestions were kept in mind to improve the tool, its user friendliness, and its final result in the form of a summarizing table. In this redesign (BACRA v1.1), fields were added in the solutions table and the app was also personalized to register certain specific types of incidents relevant for hospitals or primary care centers.

BACRA v1.1 was presented to 43 other middle managers, who provided ideas for improving its design (BACRA v1.2). This permitted the elimination of unnecessary information and the introduction of small changes to the types and causes of harm addressed, and allowed the improvement of help texts.

In January 2016, once the tool improvement proposals were introduced, BACRA v1.2 became available to users [30]. It can be used with the main browsers (Chrome, Safari, Microsoft Explorer, Firefox) and operating systems (Windows, OSX, Linux).

Evaluation of the Web Tool

BACRA v1.1 was used by 13 frontline professionals to analyze distinct types of incidents, whereas its final version (BACRA v1.2) was used by 59 frontline professionals to analyze incidents both with and without harm to patients. Once the analysis was finished, all users had the opportunity to voluntarily assess the Web-based tool's utility and ease of use.

The evaluation results of both versions were compared using the nonparametric Mann-Whitney U test.

Ethics

This study was approved by the Ethics Committee of Clinical Research at the San Juan de Alicante University Hospital, Alicante, Spain.

Results

Bases for BACRA Design

The professionals interviewed pointed out the following main difficulties related to analysis of reported incidents: the lack of time, coupled with the belief that solutions should be proposed by the services responsible for the area of patient safety; the lack of procedures in primary care for addressing this problem; the difficulty in getting middle managers involved in root cause analysis; the delay in communicating analysis incident results; and the legal consequences for professionals. In their opinion, after identifying the type of incident and whether it had caused harm to patients, the tool should help identify causes and consequences in order to gather the basic knowledge necessary to propose solutions intended to prevent recurrence of similar incidents, including a plan of action.

BACRA Structure

BACRA v1.2 is structured in seven tabs: (1) general information about the tool, (2) hide/show, (3) what consequences did the incident have, (4) when and how did it occur, (5) why did it occur and root of the incident, (6) how could it have been prevented and solutions and plan of action, and (7) printout of the report. A helpline for users who needed guidance was made available. In Textbox 1, the content of each tab is specified.

Figure 2 shows a diagram of the app's different tabs. Moving from top to bottom is the recommended order for the introduction of information, but it is possible to move freely between the various tabs in any order. Each tab's contents corresponds to the sections described in Table 1.

The information tab is always accessible from any point in the analysis (found at the far left, labeled "Info"; see Figure 2). Withdrawing the analysis is also possible at any time by opening the "Hide/Show" tab and clicking "restore" (erases all data) (see the ascending path of "protect analysis" on the right). Once the analysis is finished, and the reports have been printed, erasing all data is recommended (ascending "new analysis" path on the right).

Before beginning a new analysis, and in order to ensure confidentiality when computers are shared, each user can create a unique password on the "Hide/Show" page (password access; only known to him/her) that will allow access to the analysis in progress for a specific incident. This password is not linked to the person conducting the analysis, but rather to the incident in question. Therefore, the app does not include personalized access whereby the user can consult his/her incidents in progress; instead, this person must generate a new password for each incident they wish to analyze. This way, complete confidentiality of the person conducting the analysis is ensured, protecting the professional and reducing the distrust toward these types of systems. The type of center, either hospital or primary care, must be indicated on the "What Happened?" page. Depending on the choice of center (hospital or primary care), the screen will contain information specific to the type chosen. In addition, the "Solutions" page is dynamically generated with the data introduced in the preceding steps.

Figure 3 shows the home page, where information on using the app is offered. The recommended route for introducing information is visible, but navigating freely between any tab is possible.

Figure 4 shows the manner in which the initial information about the type and nature of the harm is introduced. A list of possible incidents is shown, organized by categories in order to be located easily. The list is specific to the type of center where the analysis is being made, either hospital or primary care. Any number of desired options can be selected and, furthermore, any type of patient harm that does not appear on the list can be introduced as free text.

Figure 5 corresponds to the "Causes" page. Here, the reasons for the incident and its roots are described. A list of causes is shown, organized by categories. When applicable, additional information may be added, indicating whether the cause is immediate (active error by professionals that is directly related to patients) or latent (system, organization, or device failure).

Table 1. Criteria for BACRA to satisfy and analysis of other existing tools.

Criteria	App 1[a]	App 2[b]	App 3[c]	BACRA
Permits incident analysis by a small group (3-5 persons)	Yes	Yes	Yes	Yes
Permits incident analysis in less than 20 minutes	No	No	No	Yes
Uses international taxonomy with help menus in order to correctly interpret the terms	Yes	Yes	Yes	Yes
Permits analysis of adverse events and near errors at hospitals and primary care	Yes	Yes	Yes	Yes
Ensures the privacy and confidentiality of the information	Yes	Yes	Yes	Yes
Offers full guarantees for the legal certainty of the professionals (no data recorded)	No	No	No	Yes
Permits analyzing immediate and latent causes of incidents	No	No	No	Yes
Involves middle managers in the search for solutions	Yes	Yes	No	Yes
Focuses on the search for solutions to prevent recurrence of the same incident	Yes	Yes	No	Yes
Includes how to implement solutions and how to verify whether the anticipated result is obtained	No	No	No	Yes

[a]TPSC Cloud (The Patient Safety Company Cloud).

[b]Sistema de Gestión de Incidentes de Seguridad—Junta de Andalucía.

[c]SiNASP-Sistema de Notificación y Aprendizaje para la Seguridad del Paciente (Learning and Reporting System for Patient Safety).

Figure 2. BACRA tool flowchart.
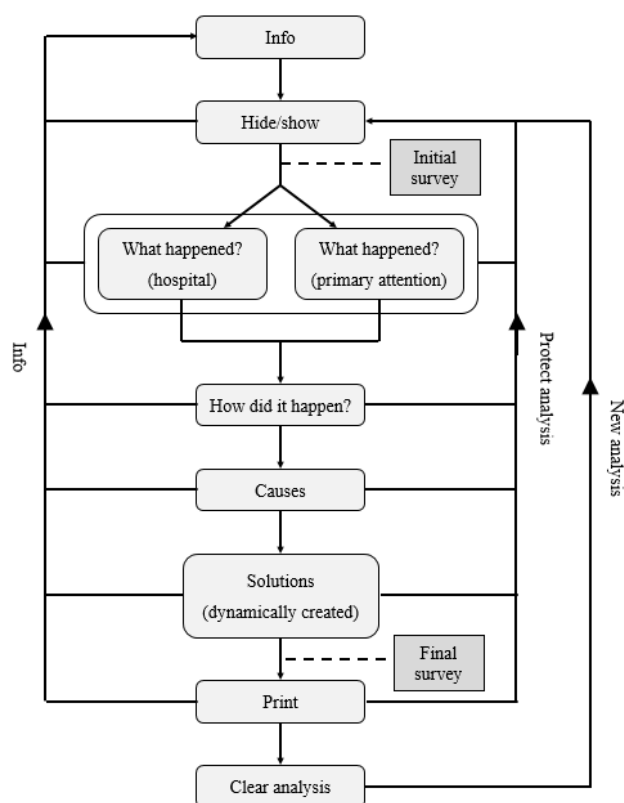
Textbox 1. BACRA content

(1) Info: information

- What BACRA is

- What BACRA offers

- How to use BACRA

(2) Hide/Show: hide the form so that nobody else can access it (password access)

(3) What Happened: what consequences did the incident have?

- Care level selection (hospital or primary care)

- Type of harm

- Has the incident been reported?

- Nature of harm

- Related with nosocomial infection

- Related to procedures

- Related with care

- Related with medication

- Others

- Measures adopted with the patient to remedy the harm and to prevent recurrence of another AE related to the first

- Impact (severity and autonomy of the patient)

(4) How: when and how did it occur?

- Date and time of the occurrence and its detection

- Chronology of the facts

(5) Causes: why did this happen? (root of the incident)

- 5 Whys technique

- Immediate and latent causes

- Use of resources and equipment

- Organization and culture of safety

- Factors attributable to professional action

- Intrinsic risk factors for the patient

(6) Solutions: how could it have been avoided? Solutions and plan of action

Risk priority number (RPN)

(7) Print: print report in PDF

Figure 3. Home page and navigation modes.
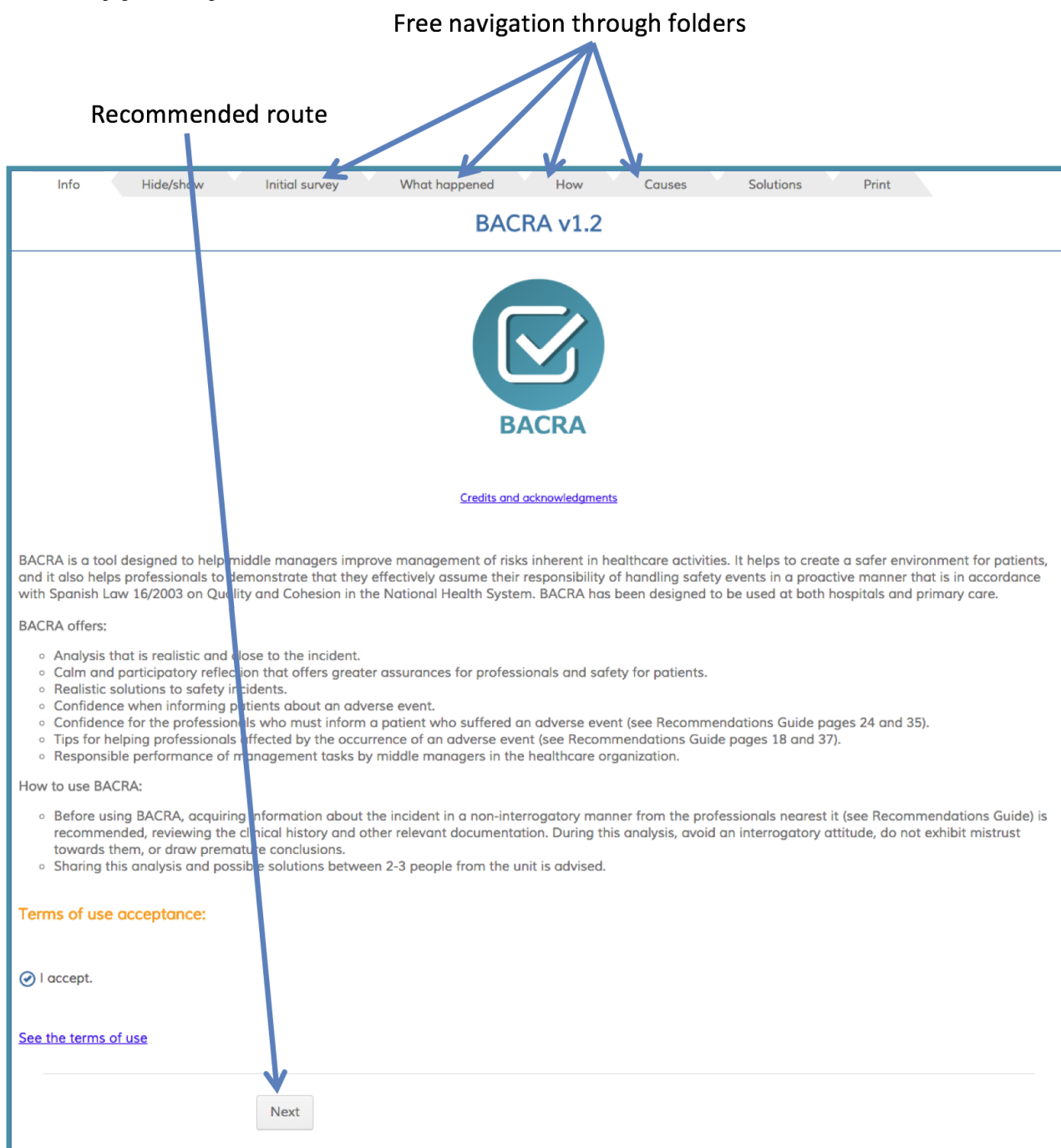
Figure 4. Screenshot of "Consequence" page: type and nature of harm.

Figure 5. Screenshot of "Causes" page: root of the incident.

XSL•FO

RenderX

Figure 6. Screenshot of "Solutions" page: final result of the analysis.

Figure 6 shows the final result of the analysis. A portion of the data is generated dynamically based on the information introduced in previous steps. Each of the incident's causes must be assessed in this tab in terms of their severity, probability of occurrence, and probability of detection. From these values, the app automatically calculates the risk priority number (RPN). For each cause, it asks for solutions to be proposed, which must have an associated date of completion, a professional responsible for implementation, and an outcome measure that permits verification that the proposed solution was successful.

The tool's final screen allows the printing of a final report that contains information that is detailed depending on where the report is destined to go. Specifically, two options exist: a report for the head of the unit, containing information drawn exclusively from the "Solutions" page (the final result of the analysis), or a report for the center's safety committee, containing additional data from the "What Happened?", "How?", and "Causes" pages.

Comparison With Other Tools.

Table 1 shows BACRA's main characteristics compared with other tools designed for reporting incidents for posterior analysis.

Assessment of the Different Versions

BACRA v1.1 was assessed by 12 of 13 professionals who used it (a response rate of 92%) and BACRA v1.2 was assessed by 47 of 59 professionals who used it (a response rate of 80%). Most of these professionals had reported an incident using other reporting systems at their respective health centers (86%, 62/72).

On a scale of zero to five, BACRA v1.2 was rated higher than BACRA v1.1 in both usefulness (v1.1: mean 3.4, SD 1.6; v1.2: mean 4.3, SD 0.9; $z=2.2$, $P=.03$) and ease of use (v1.1: mean 2.8, SD 1.5; v1.2: mean 4.2, SD 1.0; $z=3.0$, $P=.003$).

Discussion

Principal Results

BACRA v1.2 was designed to encourage the implementation of preventive measures that impede the repetition of incidents producing harm in patients, as well as incidents that have not yet caused harm but might do so in future due to repetition. This Web-based tool was devised to provide an appropriate response to the difficulties described in the literature, as well as those described by frontline professionals, that have made it difficult to analyze incidents of safety and to seeking solutions. This new instrument promotes health care provider involvement, seeking alternatives to avoid repetition of safety incidents. The data obtained by evaluating the tool justify the changes introduced in v1.2, making it a tool that is useful and easy to use.

BACRA v1.2 is a supplement to an existing IRS to support the frontline health care professional to analyze safety incidents and propose actions to prevent the recurrence of similar incidents. BACRA has been designed to overcome factors limiting reporting and reaching consensus on preventive actions. This tool guarantees anonymity of the analysis and reduces the reluctance of professionals to carry out this task.

The benefits of reporting incidents have been well described, and there is broad consensus that reporting helps improve the management of risks inherent in health care, it strengthens the safety culture, and ultimately increases patient safety [4,5,9]. However, as health care professionals report greater numbers of incidents, those responsible for their analysis become overwhelmed, and the capacity to prevent new incidents may be reduced. Moreover, many times incidents that do not cause harm, or that cause harm that is hardly noticeable, are not analyzed in sufficient detail to prevent similar incidents in the future [11]. Patient safety systems achieve better results when frontline professionals and middle managers get directly involved in incident analysis [31], and especially when they are capable of implementing preventive measures promptly and efficiently. BACRA was designed with this in mind.

Comparison With Previous Studies

Other studies have pointed out certain barriers that make it difficult to report new incidents, thus making it difficult to transform information into action. Among other reasons, health care professionals attribute their own reticence to fear of punitive action, scant familiarity with incident analysis techniques, difficulties in achieving appropriate feedback, and lack of time to report [32-34].

BACRA guarantees anonymity of the analysis (not only of the reporting) and has been designed to assure that middle managers can become more actively involved in proposing preventive actions to avoid the occurrence of new incidents and thus to save patients from suffering avoidable harm. BACRA is used both to analyze real incidents, which caused harm to patients, and to analyze critical incidents that did not cause harm to patients. This tool provides a framework to identify what has occurred and why, enabling caregivers to determine how to resolve issues and to implant barriers to prevent future failures. The BACRA focus is the implementation of an action plan that defines tasks and responsibilities that follow a clear analytical agenda. The aim of BACRA is to improve patient safety, but it also works to enhance the well-being of frontline professionals by contributing to the creation of a safer clinical context.

BACRA's Web format is in line with the preferences of electronic systems found in other studies for the analysis of incidents [35]. It allows analysis that is quick and close to the incident, thus increasing reliability in the identification of incident causes and enhancing the validity of new proposals to prevent future incidents.

Approaches to, and conditions of, incident analysis for patient safety have been studied extensively. These studies have demonstrated the critical importance of leadership, effective dissemination channels, and the capacity for rapid action as crucial to the execution of incident analysis that results in preventive action and that enables caregivers to draw lessons from prior experience [25]. These critical elements were included from the outset in the BACRA design. As well, the design takes into account the importance of providing legal protections to those professionals who report and analyze incidents based on the observation that many countries do not have apology laws, which could encourage reporting by reducing the likelihood of legal consequences.

This Web-based tool employs specific techniques proposed by other analysis methodologies seeking involvement from middle managers that have been deemed beneficial to patient safety [25]. The effectiveness of these tools for analyzing incidents of safety has been analyzed in other studies [35].

Relevance of This Study

The causes of most so-called near errors are not usually analyzed, and so they can continue to cause AEs (reaching the

patient) in practice. A large number of health care professionals are not familiar with techniques for analyzing safety incidents; their care responsibilities limit the time they have available to carry out such analyses and, in many cases, they are wary of the consequences that could result from being seen involved in the analysis of incidents. BACRA strives to respond to all these limitations by offering a guideline for conducting an analysis focused on reaching a consensus on actions designed to prevent the recurrence of similar incidents.

BACRA should be used as a supplement to an existing IRS, and not as a stand-alone tool. BACRA is a tool designed by and for middle managers and frontline professionals that does not require specific training in patient safety. This new tool proposes a friendly framework to define an action plan and for implementing corresponding countermeasures that involve frontline health care professionals. A combination of top-down and bottom-up approaches helps to engage health care teams as a whole. They know the questions and, in most instances, they have solutions to offer that will implant barriers designed to prevent harmful incidents in future.

Tool Limitations

This tool's main limitation is the need for Internet access, which some centers restrict due to security considerations. Response speed problems have also been detected in computers with limited features. To address these limitations in the future, a version of BACRA could be developed that is capable of functioning locally, without access to the Internet.

BACRA is not an incident reporting system and, therefore, should not be used as such. Its focus is centered on identifying preventive actions to avoid the recurrence of incidents that ultimately do harm patients. Sentinel events could require extensive analysis using the root cause technique.

Study Limitations

This study was conducted with professionals experienced in incident reporting and thus familiar with basic issues of patient safety. Other users could require more time to become familiar with the tool. This study did not consider certain variables that might influence incident analysis, such as safety culture or perception of the efficacy of proposed solutions. The effectiveness of the proposals to avoid the repetition of similar incidents was not analyzed.

Recommendations for Practice and Research

The safety culture at health centers can be determinant when implementing this tool. Learning from one's errors is not easy due to questions that are both attitudinal and practical in nature. In order to exploit BACRA's advantages, and learn from the experience toward the end of improving patient safety, it is important to promote a proactive culture of safety that acknowledges the possibility that professionals may commit errors in the course of providing clinical assistance. Effective use of BACRA also requires a commitment to the exercise of responsible behavior that improves patient safety.

By using BACRA at health centers, those responsible for the area of patient safety, in coordination with the middle managers of different care units, can agree on the destination of the results reports that the tool produces. For example, agreement could be reached, under appropriate conditions of confidentiality, to disseminate the proposals of applicable measures for specific types of incidents, thereby fostering shared learning to avoid future risks to patients.

Future research could examine the degree to which BACRA and similar tools are accepted by professionals, both those who use these tools and those who make changes in their clinical practice based on proposals reached consensually following the use of BACRA. The extent that BACRA contributes to strengthening of the culture of safety at centers could also be analyzed, for example, by determining whether application of this tool results in changes to management of the risks inherent to clinical practice.

Authors' Contributions

JJM, LF, IC, and MG conceived the study. EZ, CS, and PP participated in its design. MAV and CF designed BACRA. IC performed the statistical analysis. MG, IC, LF, EZ, CS, and PP coordinated the qualitative research. JJM collected and prepared data to design and improve BACRA. JJM and IC prepared a first version of the original manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1.   Leape LL. Reporting of adverse events. N Engl J Med 2002 Nov 14;347(20):1633-1638. [doi: 10.1056/NEJMNEJMhpr011493] [Medline: 12432059]

XSL•FO
RenderX

2.   Pham JC, Gianci S, Battles J, Beard P, Clarke JR, Coates H, et al. Establishing a global learning community for incident-reporting systems. Qual Saf Health Care 2010 Oct;19(5):446-451. [doi: 10.1136/qshc.2009.037739] [Medline: 20977995]

3.   Staender S, Davies J, Helmreich B, Sexton B, Kaufmann M. The anaesthesia critical incident reporting system: an experience based database. Int J Med Inform 1997 Nov;47(1-2):87-90. [Medline: 9506400]

4.   Leape LL. Why should we report adverse incidents? J Eval Clin Pract 1999 Feb;5(1):1-4. [Medline: 10468378]

5.   Aranaz AJ. [Notification systems and adverse effect reporting]. Rev Calid Asist 2009 Feb;24(1):1-2. [doi: 10.1016/S1134-282X(09)70068-5] [Medline: 19369135]

6.   Spigelman AD, Swan J. Review of the Australian incident monitoring system. ANZ J Surg 2005 Aug;75(8):657-661. [doi: 10.1111/j.1445-2197.2005.03482.x] [Medline: 16076327]

7.   Beckmann U, West LF, Groombridge GJ, Baldwin I, Hart GK, Clayton DG, et al. The Australian Incident Monitoring Study in Intensive Care: AIMS-ICU. The development and evaluation of an incident reporting system in intensive care. Anaesth Intensive Care 1996 Jun;24(3):314-319. [Medline: 8805885]

8.   Bañeres J, Cavero E, López L, Orrego C, Suñol R. Sistemas de registro y notificación de incidentes y eventos adversos. Madrid: Ministerio de Sanidad y Consumo; 2006. URL: http://www.msssi.gob.es/organizacion/sns/planCalidadSNS/pdf/excelencia/opsc_sp3.pdf [WebCite Cache ID 6kNBgwmSf]

9.   Mira JJ, Cho M, Montserrat D, Rodríguez J, Santacruz J. [Key elements in implementing adverse event notification systems in Latin American hospitals]. Rev Panam Salud Publica 2013 Jan;33(1):1-7 [FREE Full text] [Medline: 23440151]

10.   Pham JC, Girard T, Pronovost PJ. What to do with healthcare incident reporting systems. J Public Health Res 2013 Dec 1;2(3):e27 [FREE Full text] [doi: 10.4081/jphr.2013.e27] [Medline: 25170498]

11.   Evans SM, Berry JG, Smith BJ, Esterman A, Selim P, O'Shaughnessy J, et al. Attitudes and barriers to incident reporting: a collaborative hospital study. Qual Saf Health Care 2006 Feb;15(1):39-43 [FREE Full text] [doi: 10.1136/qshc.2004.012559] [Medline: 16456208]

12.   Lederman R, Dreyfus S, Matchan J, Knott JC, Milton SK. Electronic error-reporting systems: a case study into the impact on nurse reporting of medical errors. Nurs Outlook 2013;61(6):417-426.e5. [doi: 10.1016/j.outlook.2013.04.008] [Medline: 23838568]

13.   Reason J. Beyond the organisational accident: the need for "error wisdom" on the frontline. Qual Saf Health Care 2004 Dec 01;13(suppl_2):ii28-ii33. [doi: 10.1136/qshc.2003.009548] [Medline: 1765802]

14.   Lee A, Mills PD, Neily J, Hemphill RR. Root cause analysis of serious adverse events among older patients in the Veterans Health Administration. Jt Comm J Qual Patient Saf 2014 Jun;40(6):253-262. [Medline: 25016673]

15.   Slakey DP, Simms ER, Rennie KV, Garstka ME, Korndorffer JR. Using simulation to improve root cause analysis of adverse surgical outcomes. Int J Qual Health Care 2014 Apr;26(2):144-150 [FREE Full text] [doi: 10.1093/intqhc/mzu011] [Medline: 24521702]

16.   McGraw C, Drennan VM. Evaluation of the suitability of root cause analysis frameworks for the investigation of community-acquired pressure ulcers: a systematic review and documentary analysis. J Clin Nurs 2015 Feb;24(3-4):536-545. [doi: 10.1111/jocn.12644] [Medline: 24963871]

17.   Sanghera IS, Franklin BD, Dhillon S. The attitudes and beliefs of healthcare professionals on the causes and reporting of medication errors in a UK Intensive care unit. Anaesthesia 2007 Jan;62(1):53-61 [FREE Full text] [doi: 10.1111/j.1365-2044.2006.04858.x] [Medline: 17156227]

18.   Ulanimo VM, O'Leary-Kelley C, Connolly PM. Nurses' perceptions of causes of medication errors and barriers to reporting. J Nurs Care Qual 2007 Jan-Mar;22(1):28-33. [Medline: 17149082]

19.   Wu AW, McCay L, Levinson W, Iedema R, Wallace G, Boyle DJ, et al. Disclosing adverse events to patients: international norms and trends. J Patient Saf 2014 Apr 8:1-7. [doi: 10.1097/PTS.0000000000000107] [Medline: 24717530]

20.   Mira JJ, Lorenzo S, en nombre del Grupo de Investigación en Segundas Víctimas. [Something is wrong in the way we inform patients of an adverse event]. Gac Sanit 2015 Sep-Oct;29(5):370-374 [FREE Full text] [doi: 10.1016/j.gaceta.2015.04.004] [Medline: 26026725]

21.   Van Gerven E, Seys D, Panella M, Sermeus W, Euwema M, Federico F, et al. Involvement of health-care professionals in an adverse event: the role of management in supporting their workforce. Pol Arch Med Wewn 2014;124(6):313-320 [FREE Full text] [Medline: 24781784]

22.   Parand A, Dopson S, Renz A, Vincent C. The role of hospital managers in quality and patient safety: a systematic review. BMJ Open 2014 Sep;4(9):e005055 [FREE Full text] [doi: 10.1136/bmjopen-2014-005055] [Medline: 25192876]

23.   Mira JJ, Lorenzo S, Carrillo I, Ferrús L, Pérez-Pérez P, Iglesias F, Research Group on Second and Third Victims. Interventions in health organisations to reduce the impact of adverse events in second and third victims. BMC Health Serv Res 2015 Aug;15:341 [FREE Full text] [doi: 10.1186/s12913-015-0994-x] [Medline: 26297015]

24.   Pfeiffer Y, Manser T, Wehner T. Conceptualising barriers to incident reporting: a psychological framework. Qual Saf Health Care 2010 Dec;19(6):e60. [doi: 10.1136/qshc.2008.030445] [Medline: 20558472]

25.   Benn J, Koutantji M, Wallace L, Spurgeon P, Rejman M, Healey A, et al. Feedback from incident reporting: information and action to improve patient safety. Qual Saf Health Care 2009 Feb;18(1):11-21. [doi: 10.1136/qshc.2007.024166] [Medline: 19204126]

26.   Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 1991 Feb 7;324(6):370-376. [doi: 10.1056/NEJM199102073240604] [Medline: 1987460]

27.   Williams PM. Techniques for root cause analysis. Proc (Bayl Univ Med Cent) 2001 Apr;14(2):154-157 [FREE Full text] [Medline: 16369607]

28.   DeRosier J, Stalhandske E, Bagian JP, Nudell T. Using health care Failure Mode and Effect Analysis: the VA National Center for Patient Safety's prospective risk analysis system. Jt Comm J Qual Improv 2002 May;28(5):248-267. [Medline: 12053459]

29.   Redick EL. Applying FOCUS-PDCA to solve clinical problems. Dimens Crit Care Nurs 1999 Nov-Dec;18(6):30-34. [Medline: 10640052]

30.   Mira JJ, Carrillo I, Vicente MA, Fernández C, Guilabert M, Ferrús L, et al. BACRA. URL: https://form.jotformeu.com/62374021493351 [accessed 2016-09-09] [WebCite Cache ID 6kOKyT3zD]

31.   Mira JJ, Carrillo I, Fernández C, Vicente MA, Guilabert M. The Safety Agenda Mobile App (SAMA) for managing hospital managers' responsibilities in patient safety: study design. JMIR Mhealth Uhealth 2016 (forthcoming). [doi: 10.2196/mhealth.5796]

32.   Mahajan RP. Critical incident reporting and learning. Br J Anaesth 2010 Jul;105(1):69-75 [FREE Full text] [doi: 10.1093/bja/aeq133] [Medline: 20551028]

33.   Schectman JM, Plews-Ogan ML. Physician perception of hospital safety and barriers to incident reporting. Jt Comm J Qual Patient Saf 2006 Jun;32(6):337-343. [Medline: 16776388]

34.   Temelkovski S, Callaghan K. Opportunities to learn from medical incidents: a review of published reports from the Health and Disability Commissioner. N Z Med J 2010 May 14;123(1314):18-30. [Medline: 20581909]

35.   Elliott P, Martin D, Neville D. Electronic clinical safety reporting system: a benefits evaluation. JMIR Med Inform 2014;2(1):e12 [FREE Full text] [doi: 10.2196/medinform.3316] [Medline: 25600569]

Abbreviations

AE: adverse event
IRS: incident reporting system
NYPORTS: New York Patient Occurrence Reporting and Tracking System
PDCA: plan-do-check-act
PDF: portable document format
RPN: risk priority number

XSL•FO
RenderX

Original Paper

A Mobile App to Stabilize Daily Functional Activity of Breast Cancer Patients in Collaboration With the Physician: A Randomized Controlled Clinical Trial

Marco Egbring[1], MD; Elmira Far[1], MD; Malgorzata Roos[2], PhD; Michael Dietrich[3], MD; Mathis Brauchbar[4], PhD; Gerd A Kullak-Ublick[1], MD; Andreas Trojan[5], MD

[1]Department of Clinical Pharmacology and Toxicology, University Hospital Zurich and University of Zurich, Zurich, Switzerland

[2]EBPI, Department of Biostatistics, University of Zurich, Zurich, Switzerland

[3]Department of Orthopedics and Emergency Department, Waid City Hospital, Zurich, Switzerland

[4]Advocacy AG, Zurich, Switzerland

[5]Breast-Center Zurich, Zurich, Switzerland

Corresponding Author:
Andreas Trojan, MD
Breast-Center Zurich
Hardturmstrasse 135
Zurich, 8005
Switzerland
Phone: 41 43 344 33 33
Fax: 41 43 344 33 44
Email: andreas.trojan@ozh.ch

Abstract

Background: The well-being of breast cancer patients and reporting of adverse events require close monitoring. Mobile apps allow continuous recording of disease- and medication-related symptoms in patients undergoing chemotherapy.

Objective: The aim of the study was to evaluate the effects of a mobile app on patient-reported daily functional activity in a supervised and unsupervised setting.

Methods: We conducted a randomized controlled study of 139 breast cancer patients undergoing chemotherapy. Patient status was self-measured using Eastern Cooperative Oncology Group scoring and Common Terminology Criteria for Adverse Events. Participants were randomly assigned to a control group, an unsupervised group that used a mobile app to record data, or a supervised group that used the app and reviewed data with a physician. Primary outcome variables were change in daily functional activity and symptoms over three outpatient visits.

Results: Functional activity scores declined in all groups from the first to second visit. However, from the second to third visit, only the supervised group improved, whereas the others continued to decline. Overall, the supervised group showed no significant difference from the first (median 90.85, IQR 30.67) to third visit (median 84.76, IQR 18.29, P=.72). Both app-using groups reported more distinct adverse events in the app than in the questionnaire (supervised: n=1033 vs n=656; unsupervised: n=852 vs n=823), although the unsupervised group reported more symptoms overall (n=4808) in the app than the supervised group (n=4463).

Conclusions: The mobile app was associated with stabilized daily functional activity when used under collaborative review. App-using participants could more frequently report adverse events, and those under supervision made fewer and more precise entries than unsupervised participants. Our findings suggest that patient well-being and awareness of chemotherapy adverse effects can be improved by using a mobile app in collaboration with the treating physician.

ClinicalTrial: ClinicalTrials.gov NCT02004496; https://clinicaltrials.gov/ct2/show/NCT02004496 (Archived by WebCite at http://www.webcitation.org/6k68FZHo2)

KEYWORDS

collaboration; breast cancer; mobile app; daily functional activity; breast neoplasms

Introduction

Telemedicine services have historically provided education, hardware, and software to patients and have been shown to improve functional status. If symptoms worsen, automatic algorithms to alert the nurse or physician have been demonstrated to be more effective than patient self-reporting alone [1]. The advent of mobile apps and the ease with which these are developed and used enables the stakeholders to collaborate in an increasingly dynamic and efficient manner.

Until now, mobile apps that allow continuous recording of disease- and medication-related symptoms have only occasionally been implemented in the management of patients undergoing chemotherapy [2]. However, symptom self-reporting on paper can only be explored with cumbersome efforts. In particular, therapeutic regimens in cancer patients require intensive monitoring procedures [3] because outpatient chemotherapy may frequently be accompanied by serious and potentially life-threatening adverse effects [4,5]. Well-informed patients spend significant efforts in documentation and management of symptoms as well as reviewing information related to both during the course of their disease and therapeutic interventions [6]. Consequently, involved physicians increasingly need to focus on collection and interpretation of a variety of information presented during the patient's visit at first and after appropriate measures, if indicated.

Mobile apps have raised high expectations in various settings of daily life; their potential is now increasingly being examined in the health care sector. Supporting self-care with information on the disease or symptoms during routine cancer care can be enhanced by remote devices and patient collaboration through online patient groups and other forums [7]. In addition, recent studies have shown that the reporting of symptoms via email reminder and improved previsit preparation of physicians via printouts of symptoms may increase patients' well-being [8]. Moreover, ancillary disease-related counseling with or without use of electronic devices successfully improved patient satisfaction through better understanding of specific symptoms, which in turn decreases patient anxiety [9]. Thus, available data indicate that mobile apps may well improve patient status if applied in a supervised setting in which the physician is empowered by additional information or services. However, the question remains whether self-quantifying apps, such as step counters, will provide added value simply by empowering the patient in an unsupervised setting.

Our objective was to explore the impact of a mobile and Web-based app in both an unsupervised setting and a setting supervised by the treating physician. We conducted a prospective randomized trial and invited early breast cancer patients to continuously record their symptoms according to the Common Terminology Criteria for Adverse Events (CTCAE) v4.0 via a novel electronic study device (mobile or Web app) and to also indicate their Eastern Cooperative Oncology Group (ECOG) Performance Status. Information for supportive care was also displayed by the app depending on the severity of symptoms upon data entry. The cumulative recorded data were made available to the patient in chart form (similar to a stock chart) at any time during the study. In the unsupervised setting, patients were instructed to use the app at any time except during study visits. In the supervised setting, the treating physician accessed the recorded data and patient-derived charts using the Web app during the study visit.

Methods

Study Design

We conducted a single center, three-arm, randomized, controlled, single-blinded interventional study. The protocol was approved by the Swiss Institutional Review Board (KEK-EK-ZH:2013-0200) and registered on ClinicalTrials.gov (NCT02004496). Patients with early breast cancer, aged 18 years and older, and initiating adjuvant or neoadjuvant chemotherapy at the Breast-Center Zürich were eligible to participate upon written informed consent. In addition, participants had to speak German and own a mobile phone.

Study Groups

Eligible participants were recruited consecutively and without preselection during planning of chemotherapy. Participants received an envelope randomly assigning them to one of three study groups. The allocation sequence was concealed by the use of sealed, sequentially numbered, opaque envelopes. Group A (control) was the control group and received regular physician support. Group B (app) patients were instructed to use the mobile app without physician review. Group C (app and physician) patients used the mobile app and reviewed the reported data with the treating physician at scheduled visits. Participants in each group underwent three regular medical oncology visits scheduled on days 1, 21, and 42 during their chemotherapeutic intervention and independently of the study. The physician was single-blinded with regard to groups A and B. At the end of each visit, patients in all groups completed a questionnaire on paper. The observation period comprised 6 weeks and was initiated on the day of the first infusion therapy.

Questionnaire

The questionnaire recorded performance status and daily functional activities according to the ECOG scale in an attempt to quantify daily levels of activity among breast cancer patients. The first five categories were presented as a visual analog scale; the sixth category, death, was omitted from the questionnaire. The ECOG activity score measure is routinely used to evaluate whether patients are eligible for chemotherapy, whether dose adjustments are required, and as an indicator for the need of palliative care. Additional questions were posed to further characterize the quality of care received by patients and the relationship between the patient and physician (Table 1). In addition, 30 preselected adverse events were listed with selectable severity, onset, and duration. The most frequent symptoms were selected according to clinical preview and experience; furthermore, patients could add additional symptoms as free text. Any medical measures undertaken could also be reported as free text. The questionnaire was issued to all groups at each of the three scheduled visits during the 6 weeks of study.

Table 1. Questions regarding the relationship between patient and physician for the three visits for each patient group.

Questions	Group and visit number, median (IQR)[a]								
	Control			App			App and physician		
	1	2	3	1	2	3	1	2	3
Concentration problems during visit	13 (69)	6[b] (77)	17[b] (64)	10 (51)	2[c] (8)	4[c] (10)	3 (29)	4[b,c] (18)	4[c] (12)
Well-informed about therapy	94 (15)	94 (35)	94 (20)	94 (20)	95 (20)	95 (14)	94 (21)	94 (14)	96 (12)
Well-informed about disease	92 (25)	94 (31)	94 (19)	92 (24)	93 (24)	95 (15)	96 (18)	96 (14)	96 (30)
Less likely to disfavor with care	4 (40)	6[b] (73)	6 (17)	4 (10)	2[b,c] (5)	4 (13)	2 (10)	1[c] (8)	4 (6)
Awareness regarding adverse events	96 (23)	94 (14)	94 (15)	96 (21)	96 (18)	98 (13)	100 (13)	100 (14)	100 (15)
Satisfaction with medical care	96 (10)	96 (14)	95 (17)	99 (12)	96 (10)	98 (11)	100 (12)	99 (11)	100 (7)
Trust in data security	96 (14)	99 (13)	99 (13)	98 (13)	95 (11)	100 (8)	100 (11)	100 (7)	100 (8)
Feeling of being taken seriously	98 (18)	96 (17)	94 (19)	96 (17)	96 (11)	99 (11)	100 (12)	100 (8)	99 (8)

[a] Value of 0 indicates complete disagreement, whereas 100 represents complete agreement with the statement in the question.

[b,c] Annotations mark significant differences between the three groups. In this case three combinations of group pairs are possible. Two groups differ significantly in the answer to the question if they have different superscript characters.

Mobile App

We developed a novel open-source mobile and Web app to record daily functional activity and adverse events. This mobile app was made available in the Apple and Google Android stores free of charge. Patients could report daily functional activity or symptoms with indication of severity in the electronic app device similar to the paper questionnaire. The visual analog scale from the questionnaire was substituted with a horizontal slider. Similar to the questionnaire, the label of symptom severity and category according to CTCAE was displayed below the slider. During the visits, nurses reminded the participants according to their randomization to use the app, but no other reminders were issued. Patients could edit a quick list of their preselected symptoms or select any of the 48 symptoms made available from the CTCAE listing. Only for group C was the treating physician enabled access to review and discuss the electronically reported symptoms during scheduled visits. Although the mobile app was publically available for download and use, only study patients who scanned an issued QR code containing a shared password could upload and store data securely. The development was frozen during the trial.

The variables of daily functional activity and severity of symptoms from the questionnaire and the mobile app were transformed to the same interval (0-100), if applicable. Missing data or withdrawn subjects were not excluded from the analysis to prevent selection bias for successfully compliant patients.

Sample Size

We calculated the sample size for the three groups based on a 10% difference in daily functional activity. The significance level, after Bonferroni correction for the comparison of the three groups, was .016 (P=.05/3). A mean of 80% (SD 13%) was estimated by reviewing medical records of randomly selected patients at the Breast-Center Zürich. To detect differences with a power of 91%, a sample size of 50 patients per group was calculated. Given this sample size, a power of 80%, and a mean frequency of 4.4 (SD 3.3) adverse events, we expected to detect a difference in mean frequency of 2.2 adverse events per group. To minimize bias in the groups, only patients under the care of a single physician, who also participated as the study physician, were recruited.

Statistical Analysis

All statistical tests were calculated using SPSS version 22.0 (IBM Corp, Armonk, NY, USA). Two-sided P values of less than .05 were considered to indicate statistical significance. For cases of more than two independent samples, Bonferroni correction was applied. Descriptive statistics, such as median and interquartile range, were computed. The Kruskal-Wallis test was used to identify differences in outcome variables among the three groups at all time intervals. The Mann-Whitney test further investigated differences between groups B and C. The paired Wilcoxon test analyzed the differences between time intervals within one group. Scatterplots and Spearman correlation (ρ) were used to validate the reported data from the app against the questionnaire.

Results

Baseline Characteristics

Between December 2013 and July 2015, 139 patients were enrolled, 12 of whom did not complete the study for various reasons (Figure 1). Dislike of the constant confrontation with the disease was the reason for withdrawal for five of 12 patients who withdrew. The remaining 127 patients completed all three study visits. Baseline characteristics were equally distributed between the groups (Table 2). The most frequent chemotherapy

XSL·FO
RenderX

regimen in all groups was epirubicin/cyclophosphamide (n=32), followed by paclitaxel/trastuzumab (control: n=8; app: n=4; app and physician: n=7), and paclitaxel/carboplatin (control: n=4; app: n=8; app and physician: n=7). In total, six different chemotherapy regimens were reported using seven distinct chemotherapeutic agents. The median observation interval between visit 1 and visit 2 (IQR 6), and also between visit 2 and visit 3 (IQR 8), was 21 days.

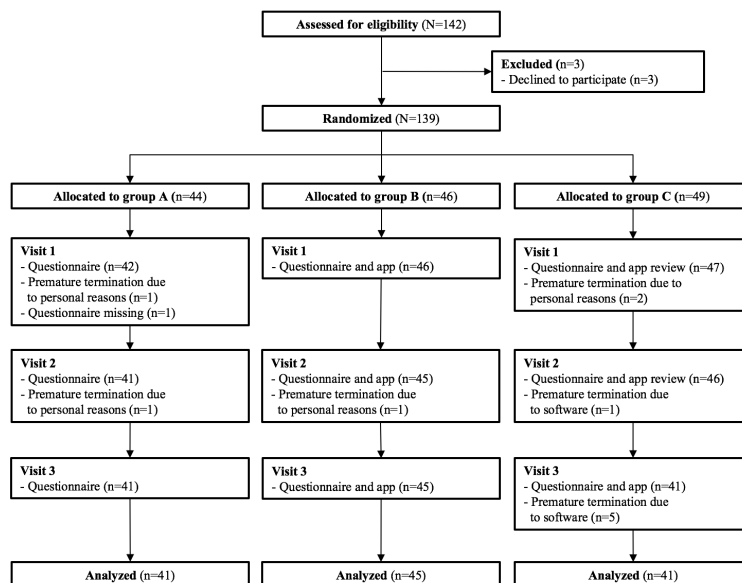**Figure 1.** CONSORT diagram demonstrating the flow of patients.

Table 2. Summary of baseline characteristics of participants.

Characteristics	All patients (N=139)	Control (n=44)	App (n=46)	App and physician (n=49)
Age (years), mean (SD)	53 (13)	56 (15)	50 (10)	53 (12)
Sex (female), n (%)	139 (100)	44 (100)	46 (100)	49 (100)
Chemotherapy, n (%)				
Adjuvant	86 (61.9)	25 (57)	30 (65)	31 (63)
Neoadjuvant	50 (36)	18 (41)	16 (35)	16 (33)
Surgery, n (%)				
Biopsy (sentinel)	50 (36)	17 (39)	16 (35)	17 (35)
Breast conserving	72 (51.8)	23 (52)	27 (59)	22 (45)
Breast ablation	14 (10.1)	3 (7)	3 (7)	8 (16)
Interval (days), median (IQR; range)				
Visits 1-2	21 (6; 6-43)	21 (6; 6-43)	20 (6; 7-27)	21 (2; 7-42)
Visits 2-3	21 (8; 6-42)	21 (9; 7-42)	21 (7; 7-28)	20 (7; 6-28)
Visits 1-3	42 (7; 13-84)	42 (2; 14-84)	41 (9; 26-49)	39 (7; 13-56)

Daily Functional Activity

We collected 381 questionnaires from patients who completed the study. Current daily functional activity was indicated as the ECOG score at the time of the study visit, whereas the worst daily functional activity was defined as the worst rating before visits 1 to 3, respectively.

As shown in Figure 2, both median current and worst daily functional activity declined in all patients during chemotherapy from the first to the second visit. From the second to the third visit, only patients in group C (app and physician) reported improvement in their functional activity, whereas scores in groups A (control) and B (app) continued to decline. From the first to the third visit, group A (current: median 90.24, IQR 19.63 vs median 75.61, IQR 21.95, *P*=.006; worst: median 84.15, IQR 23.93 vs median 71.95, IQR 32.32, *P*=.02) and group B (current: median 90.24, IQR 21.47 vs median 74.39, IQR 21.95, *P*=.02; worst: median 84.76, IQR 36.20 vs median 62.80, IQR 26.83, *P*<.001) showed a significant decline in functional activity in contrast to group C (current: median 90.85, IQR 30.67 vs median 84.76, IQR 18.29, *P*=.72; worst: median 84.15, IQR 39.88 vs median 65.24, IQR 32.32, *P*=.13). However, within the three groups differences in the reported functional activity at the three visits did not reach statistical significance,

irrespective of whether current or worst daily functional activity was analyzed. Post hoc, the sample size for the actual standard deviation of 17% and a power of 91% should have required at least 83 patients per group in order to demonstrate a significant difference between the groups.

Overall, for groups B and C, results from the questionnaire were partially aligned with the data derived from app use. Patients started electronic recording of daily functional activity and symptoms in the app after the first visit, which also was the first day of treatment. The last data entry before each visit corresponded to the current daily functional activity indicated in the questionnaire at the visit. Similar to the results obtained from the questionnaire, the current daily functional activity (last value) declined significantly from the second (median 79.50, IQR 89.00) to the third visit (median 73.00, IQR 85.00) in group B (*P*=.007), but not in group C (median 75.00, IQR 90.00 vs

median 70.00, IQR 84.00, *P*=.90). In addition, the best functional activity score of patients in group B (median 85.50, IQR 94.00 vs median 78.50, IQR 91.00, *P*=.008), but again not in group C (median 80.00, IQR 98.00 vs median 72.00, IQR 91.00, *P*=.34) dropped significantly in the first interval compared with the second interval.

In contrast, the median of the worst daily functional activity recorded in the app showed no significant difference in groups B and C between the first and second interval. The median over the total interval in the app-derived scores was not significantly lower for group B than for group C (median 45.50, IQR 49.00 vs median 45.00, IQR 70.00, *P*=.26) compared to almost identical results derived from the questionnaire for groups B and C (median 62.80, IQR 26.83 vs median 65.24, IQR 32.32, *P*=.07), respectively.

Figure 2. Visual analog scale for current and worst daily functional activity in the questionnaire.
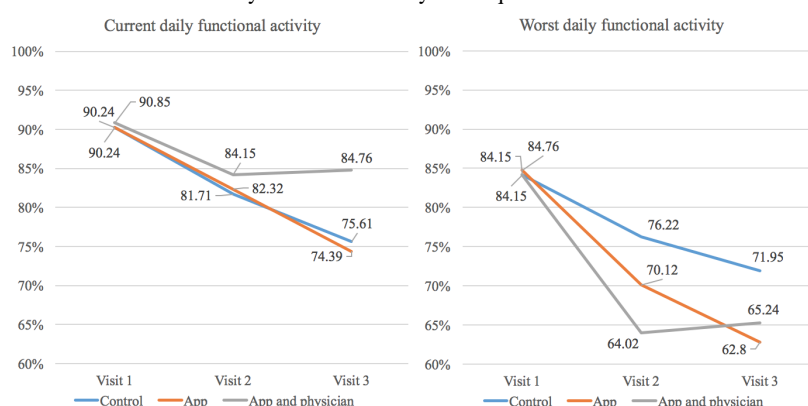

Patient Communication

We further analyzed different aspects of the patient-physician relationship in each group (Table 1). In general, all patients felt very well informed about their disease and treatment, and were highly satisfied with their medical care. During the second visit, only group B reported significantly fewer concentration issues than group A (*P*=.006). In addition, group A was significantly less likely to express dissatisfaction with quality of care at the second visit than group C (*P*=.03). Most importantly, at the last visit, patients in groups B and C reported significantly fewer issues with concentration than patients in group A (*P*=.002).

Patient Empowerment

As the study progressed, patients in all groups were significantly more likely to change their responses to the question about whether they used the Internet to obtain further information about their disease (Figure 3). At the third visit, significantly more patients in groups B (64.29% *P*=.04) and C (70.73%, *P*=.007) confirmed use of the Internet for this purpose compared with group A (41.46%).

Figure 3. Participants who used the Internet to obtain disease-specific information.
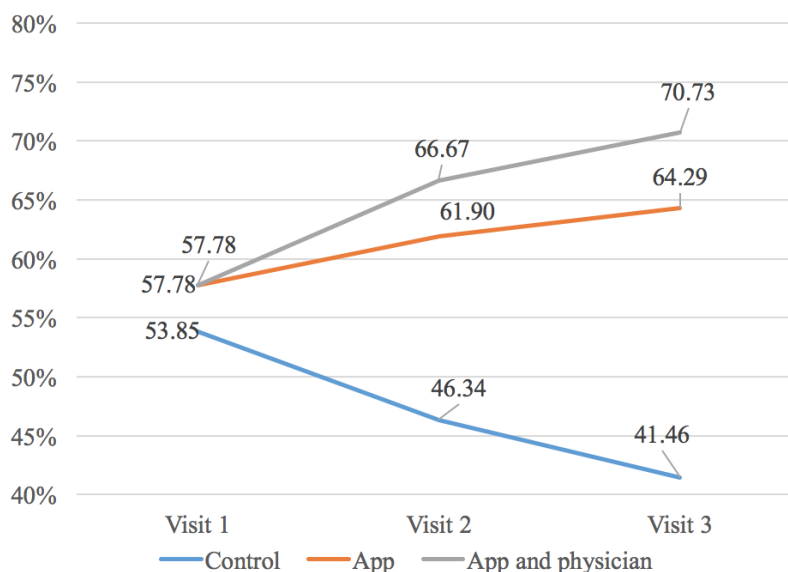
Symptom Reporting

The most frequently reported symptoms in the app were fatigue, hair loss, headache, and hypertension; from the questionnaire, fatigue, dry skin, headache, and sleep disorder were most prevalent. Regarding the number of distinct symptoms reported in the questionnaire and app, only group C showed a significant correlation in the interval from the first to the second visit (ρ=.381, P=.009) and from the second to the third visit (ρ=.362, P=.02), respectively.

Both groups reported more distinct symptoms in the app than in the questionnaire, a trend that was less prominent in group B than in group C. Therefore, the difference in numbers of distinct symptoms reported in the app versus the questionnaire was greater in group C (difference 377=1033−656) than in group B (difference 29=852−823). However, group B reported more symptoms in total (n=4808) in the app than group C (n=4463). Furthermore, group C reported significantly more distinct symptoms (median 13.00, IQR 12.00, P=.04) in the app for the total intervention interval, and more distinct mild symptoms (median 8.00, IQR 11.00, P=.02) in the interval from the first to the second visit compared with group B (distinct total: median 9.00, IQR 13.00; distinct mild: median 5.00, IQR 11.00). As for group A, the amount of distinct symptoms reported in the questionnaire was comparable to group C, whereas group B reported approximately 25% more symptoms.

Discussion

Principal Results

Few data exist that have addressed modalities and effects of electronic symptom reporting in patients undergoing chemotherapy. We demonstrate that a mobile app merits the potential to stabilize the daily functional activity of early breast cancer patients. Supervision and review of patient-reported symptoms in collaboration with their physician encouraged a timely and sincere discussion of symptoms reported in the app. Moreover, despite a rather short intervention period, supervised

patients (group C) experienced a significant benefit with respect to daily functional activity in contrast to the unsupervised patients (group B and controls). These findings are in accordance with data gained from advanced cancer patients with longer follow-up managed by email prompts [8].

Patient-physician collaboration might have influenced the patients in three ways. First, the stabilization of daily functional activity seemed partially linked to a change in behavior, reflected by a more precise recording of symptoms in supervised patients, whereas fewer data entries were recorded than in unsupervised patients. Potentially augmenting existing mechanisms for symptom management along with routine oncology care, supervised app users also seemed to attain the ability to better differentiate and communicate their treatment-related symptoms, which consequentially facilitates appropriate management by the physician.

Furthermore, supervised patients were more likely than unsupervised patients and controls to communicate their dissatisfaction. It seems plausible that an increase in self-confidence may also positively affect daily functional activity. Strengthened self-confidence is likely reflected in decreasing incidence of concentration issues reported during the visits and an increasing use of the Internet to obtain further information about disease and related treatment in both groups using the mobile app. Previous studies also indicate that physicians, as well as the majority of patients, believe that mobile apps facilitate communication [10] and increase the frequency of discussion [11] during consultations.

In addition, patients receiving chemotherapy frequently report cognitive impairments [8]. In our study, both the supervised and unsupervised groups recorded improvements in worst daily functional activity in the app before the visit, although this became less evident from the questionnaires during the visit. The diary character of our mobile app is helpful for recalling disease-related information. This feature may have positive and negative consequences. Patients in group C reported impaired

and probably more accurate scores for worst daily functional activity compared to the controls.

Six patients in group C withdrew from the study because of software reasons. One patient had functional problems with the recording of symptoms in the app. Five patients withdrew their participation from the study to avoid the constant confrontation with their disease. Of note, these patients in group C had the worst reported daily functional activity between the first and second visit. In general, group C was more likely to communicate their dissatisfaction with quality of care than group B and the controls. Maybe also patients in group B would have withdrawn from the study if their communication skills had been comparably strengthened. Furthermore, the daily recording of symptoms plus the supervision of symptoms during the visit in group C might be an additional burden. For some patients, this continuous workup of their symptoms might be too intense.

Limitations

There are several limitations of this study. The three groups did not differ significantly from one another with respect to daily functional activity scores because the sample size turned out to be inadequate for the effect size observed. However, the three groups would have reported significantly different daily functional activity scores in favor of group C if withdrawn patients had not been included in the analysis. Similar findings have been reported for supervised cohorts of patients in previous studies [8,11].

Efforts were made to blind the physician to group A and B randomization. However, the physician may have inferred from the patients' behavior whether they belonged to the control or the unsupervised app group. Unblinding of the physician and the patient to the intervention may have affected the patients' responses to the questions, especially because the patient completed the questionnaire following consultation with the physician. However, patients reported congruent answers in the mobile app outside the clinic.

Different chemotherapeutic regimens may cause different adverse events. We cannot exclude that effects observed on daily functional activity were thereby affected. We maintain, however, that the randomized design of the study and a balanced distribution of treatment regimens renders a systematic bias unlikely.

It should be noted that no alert signals about technical issues and data safety were raised during the entire course of the study.

Conclusion

Mobile apps increasingly contribute to patient education, disease self-management, and remote monitoring of patients [12]. We demonstrate that only a collaborative review of timely reported and naïve, although patient-derived, symptoms has beneficial effects on daily living activity in early breast cancer patients. The use of mobile apps under supervision may enable patients to report adverse events more precisely in the context of increasingly complex cancer therapies and limited resources.

Authors' Contributions

Design: Marco Egbring, Elmira Far, Malgorzata Roos, Mathis Brauchbar, Michael Dietrich, Andreas Trojan, Gerd A. Kullak-Ublick. Coding: Marco Egbring. Medical writing: Elmira Far. Statistics: Elmira Far, Marco Egbring, Malgorzata Roos. Study management: Elmira Far. Manuscript writing: all authors. Final approval of manuscript: all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT EHEALTH Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v18i9e238_app1.pdf]

References

1. Cordova FC, Ciccolella D, Grabianowski C, Gaughan J, Brennan K, Goldstein F, et al. A telemedicine-based intervention reduces the frequency and severity of COPD exacerbation symptoms: a randomized, controlled trial. Telemed J E Health 2015 Aug 10:e1. [doi: 10.1089/tmj.2015.0035] [Medline: 26259074]
2. Breen S, Ritchie D, Schofield P, Hsueh Y, Gough K, Santamaria N, et al. The Patient Remote Intervention and Symptom Management System (PRISMS): a Telehealth-mediated intervention enabling real-time monitoring of chemotherapy side-effects in patients with haematological malignancies: study protocol for a randomised controlled trial. Trials 2015;16:472 [FREE Full text] [doi: 10.1186/s13063-015-0970-0] [Medline: 26481873]
3. Wong S, Bounthavong M, Nguyen CP, Chen T. Outcome assessments and cost avoidance of an oral chemotherapy management clinic. J Natl Compr Canc Netw 2016 Mar;14(3):279-285. [Medline: 26957614]

XSL·FO

RenderX

4. Shapiro CL, Recht A. Side effects of adjuvant treatment of breast cancer. N Engl J Med 2001 Jun 28;344(26):1997-2008. [doi: 10.1056/NEJM200106283442607] [Medline: 11430330]

5. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982 Dec;5(6):649-655. [Medline: 7165009]

6. Klasnja P, Hartzler AC, Unruh KT, Pratt W. Blowing in the Wind: Unanchored Patient Information Work during Cancer Care. CHI Conf Proc 2010:193-202 [FREE Full text] [doi: 10.1145/1753326.1753355] [Medline: 20697453]

7. Gallagher J, O'Donoghue J, Car J. Managing immune diseases in the smartphone era: how have apps impacted disease management and their future? Expert Rev Clin Immunol 2015 Apr;11(4):431-433. [doi: 10.1586/1744666X.2015.1010518] [Medline: 25723710]

8. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]

9. Myers EM, Robinson BL, Geller EJ, Wells E, Matthews CA, Fenderson JL, et al. Randomized trial of a web-based tool for prolapse: impact on patient understanding and provider counseling. Int Urogynecol J 2014 Aug;25(8):1127-1132. [doi: 10.1007/s00192-014-2364-3] [Medline: 24652032]

10. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. JAMA 2002 Dec 18;288(23):3027-3034. [Medline: 12479768]

11. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]

12. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]

Abbreviations

CTCAE: Common Terminology Criteria for Adverse Events
ECOG: Eastern Cooperative Oncology Group

XSL•FO
RenderX

Original Paper

Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry

Sjaan R Gomersall[1], PhD; Norman Ng[1], PhD; Nicola W Burton[1], PhD; Toby G Pavey[1], PhD; Nicholas D Gilson[1], PhD; Wendy J Brown[1], PhD

Centre for Research on Exercise, Physical Activity and Health, School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, Australia

Corresponding Author:
Sjaan R Gomersall, PhD
Centre for Research on Exercise, Physical Activity and Health
School of Human Movement and Nutrition Sciences
The University of Queensland
Building 26B
Blair Drive
Brisbane, 4072
Australia
Phone: 61 733653115
Fax: 61 733656877
Email: s.gomersall1@uq.edu.au

Abstract

Background: Activity trackers are increasingly popular with both consumers and researchers for monitoring activity and for promoting positive behavior change. However, there is a lack of research investigating the performance of these devices in free-living contexts, for which findings are likely to vary from studies conducted in well-controlled laboratory settings.

Objective: The aim was to compare Fitbit One and Jawbone UP estimates of steps, moderate-to-vigorous physical activity (MVPA), and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context.

Methods: Thirty-two participants were recruited using convenience sampling; 29 provided valid data for this study (female: 90%, 26/29; age: mean 39.6, SD 11.0 years). On two occasions for 7 days each, participants wore an ActiGraph GT3X+ accelerometer on their right hip and either a hip-worn Fitbit One (n=14) or wrist-worn Jawbone UP (n=15) activity tracker. Daily estimates of steps and very active minutes were derived from the Fitbit One (n=135 days) and steps, active time, and longest idle time from the Jawbone UP (n=154 days). Daily estimates of steps, MVPA, and longest sedentary bout were derived from the corresponding days of ActiGraph data. Correlation coefficients and Bland-Altman plots with examination of systematic bias were used to assess convergent validity and agreement between the devices and the ActiGraph. Cohen's kappa was used to assess the agreement between each device and the ActiGraph for classification of active versus inactive (≥10,000 steps per day and ≥30 min/day of MVPA) comparable with public health guidelines.

Results: Correlations with ActiGraph estimates of steps and MVPA ranged between .72 and .90 for Fitbit One and .56 and .75 for Jawbone UP. Compared with ActiGraph estimates, both devices overestimated daily steps by 8% (Fitbit One) and 14% (Jawbone UP). However, mean differences were larger for daily MVPA (Fitbit One: underestimated by 46%; Jawbone UP: overestimated by 50%). There was systematic bias across all outcomes for both devices. Correlations with ActiGraph data for longest idle time (Jawbone UP) ranged from .08 to .19. Agreement for classifying days as active or inactive using the ≥10,000 steps/day criterion was substantial (Fitbit One: κ=.68; Jawbone UP: κ=.52) and slight-fair using the criterion of ≥30 min/day of MVPA (Fitbit One: κ=.40; Jawbone UP: κ=.14).

Conclusions: There was moderate-strong agreement between the ActiGraph and both Fitbit One and Jawbone UP for the estimation of daily steps. However, due to modest accuracy and systematic bias, they are better suited for consumer-based self-monitoring (eg, for the public consumer or in behavior change interventions) rather than to evaluate research outcomes. The outcomes that relate to health-enhancing MVPA (eg, "very active minutes" for Fitbit One or "active time" for Jawbone UP) and sedentary behavior ("idle time" for Jawbone UP) should be used with caution by consumers and researchers alike.

XSL•FO

RenderX

Introduction

Regularly participating in physical activity and minimizing time spent in sedentary behavior are associated with a significantly reduced risk of poor health outcomes, including cardiovascular disease, overweight and obesity, and all-cause mortality [1,2]. Despite the health benefits, many individuals are physically inactive and spend large amounts of time sedentary [3,4]; Australians spend 50% to 70% of their waking time being sedentary [5] and almost 60% of Australian adults are classified as insufficiently active (<150 minutes/week of moderate-to-vigorous physical activity) [3]. Strategies for increasing physical activity and reducing time spent sedentary are important to reverse these trends and for preventing poor health outcomes.

Activity trackers are becoming increasingly popular for monitoring physical activity and sedentary behavior, and for promoting positive behavior change [6]. Activity trackers are commonly waist- or wrist-worn devices that include a range of sensors for self-monitoring behavior. These devices typically sync with a Web- or app-based interface, which provides summary data and individual feedback on behaviors. Common types of activity data from these devices include number of steps, time spent in physical activity by intensity, and time spent "idle." The devices also have additional functions that can be used to support behavior change, such as goal setting (eg, 10,000 steps per day), prompts/cues, and social networking and accountability [7]. The uptake of these devices, both in the consumer market and in research, has been rapid [8-10]. It is estimated that more than 50 million smartwatch and health and fitness trackers were sold worldwide in 2015 and it is predicted that this figure will reach more than 80 million in 2016 [11].

Given the commercial availability of next-generation activity trackers and the rapid rate of uptake by consumers and researchers, it is important to understand their accuracy. Previous studies investigating the Fitbit One (worn on the hip) or the wrist-worn Jawbone UP in laboratory settings have demonstrated a high level of accuracy for physical activity outcomes compared with reference methods, with relative differences and correlations ranging from approximately 2% to 20% and ≥.97, respectively [12-15]. However, given their controlled settings, these studies have limited ecological validity. Only two studies have investigated the validity of a Fitbit or Jawbone device in free-living settings [16,17]. Although these studies have also demonstrated acceptable validity, correlations were lower than in controlled settings, ranging from .8 to .9 [16,17]. In addition, no previous studies have investigated the sedentary behavior features of the Jawbone UP (eg, idle time) [18]. Further research is required in free-living conditions, which reflect how these devices are used day to day.

The aim of this study was to compare Fitbit One and Jawbone UP estimates of steps, moderate-to-vigorous physical activity (MVPA), and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context.

Methods

Data were collected as part of a larger, 12-week physical activity intervention study that included three groups that were randomly allocated to wear a Fitbit One, Jawbone UP, or standard pedometer. The aim of this larger study was to compare the efficacy of the three devices to increase physical activity. Outcomes were measured at baseline, mid-, and post-intervention in August, September, and October 2014 using the ActiGraph GT3X+ accelerometer. Data for this substudy were collected at mid- and post-intervention when participants concurrently wore an ActiGraph GT3X+ accelerometer. Demographic and anthropometric data for this study were collected at baseline. Ethical approval for this study was obtained from The University of Queensland Ethics Committee (#2014000766).

Recruitment and Participants

Participants were recruited in July 2014 via convenience sampling at three campuses of a large Australian metropolitan university via an email advertisement to staff that included study information and participant eligibility criteria. The target sample size was 15 participants per group based on sample size calculations for the intervention trial. People indicated interest via return email and were then screened for eligibility via telephone interview. To be eligible, participants had to be healthy, ambulatory, aged between 18 and 65 years, have accumulated less than 150 minutes of MVPA in the past week (assessed using the Active Australia Survey [19,20]), and own or have access to a mobile phone compatible with both the Fitbit One and Jawbone UP. People who met all eligibility criteria were then invited to an individual face-to-face appointment where they provided written informed consent prior to data collection. At the conclusion of the study, participants received an AU $50 gift card gratuity.

Measures

Fitbit One

The Fitbit One (Fitbit Inc, San Francisco, CA, USA) is a small (48.0 × 19.3 × 9.65 mm), light-weight (8 g), consumer-based activity tracker that includes a three-axis accelerometer and an altimeter (Figure 1). Outcomes from this device include steps, floors climbed, distance traveled, calories burned, and active minutes. At the time these data were collected, the active time variable was labeled "very active minutes." This was not well defined by Fitbit Inc; the explanation was limited to time spent in "higher intensity" exercise [21]. For this study, *very active minutes* was defined as time spent in MVPA. The Fitbit One syncs via Bluetooth Low Energy to a mobile device or a computer. Data can be viewed using the Web-based platform

XSL·FO

RenderX

or the Fitbit mobile phone app, and immediate feedback on cumulative data for the current day is provided on the organic light-emitting diode (OLED) display. The device has a battery life of 10 to 14 days and stores 7 days of detailed, minute-by-minute data as well as daily summary data for the preceding 23 days. The device is not waterproof and the manufacturer's instructions indicate that the device can be worn on a belt, bra, or in a pocket.

Figure 1. Fitbit One (A) and Jawbone UP (B).

Jawbone UP

The Jawbone UP (Jawbone, San Francisco, CA, USA) is a light (19-23 g), wrist-worn, consumer-based activity tracker that includes a three-axis accelerometer (Figure 1). Outcomes from this device include steps, distance traveled, active time, calories burned, and longest idle time. No manufacturer definition for "active time" or "longest idle time" could be identified. For this study, *active time* was defined as time spent in MVPA and *longest idle time* was defined as longest sedentary bout. The Jawbone UP syncs via the 3.5 mm headphone jack and is only compatible with mobile devices; therefore, data are viewed using the associated mobile phone app and there is no display on the device for immediate feedback. The wrist-worn activity tracker comes in three sizes (small: 14.0×15.5 cm; medium: 15.5×18.0 cm; large: 18.0×23.0 cm), has a battery life of approximately 10 days, can store up to 9 months of data, and is not waterproof. The manufacturer's instructions indicate that the device can be worn on either the dominant or nondominant wrist, with the wear location reported in the Settings. For this study, participants could choose which wrist they wanted to wear the device on and were encouraged to ensure that the correct location was entered in the Settings.

For both trackers, daily estimates for the outcomes of interest (Fitbit One steps and active minutes; Jawbone UP active time and longest idle time) were extracted from the users' accounts and entered into an Excel spreadsheet by a research assistant. Data were included for days that there was a corresponding valid day of accelerometry data. Participants were instructed to wear the devices during waking hours, removing them for water-based activities or contact sports, but were not required to keep wear logs in order to improve the free-living fidelity of the devices over the 12-week intervention. Participants were able to input activity sessions, such as swimming/contact sports, through the "log workout" function in the Jawbone UP app and the "track exercise" feature in the Fitbit app.

Accelerometry

The comparison instrument for this study was the ActiGraph GT3X+ accelerometer (Pensacola, FL, USA). Participants were asked to wear the accelerometer on their right hip for 7 days at each measurement occasion, except when sleeping, during water-based activities, or engaging in contact sport. Participants were also asked to complete a brief log to record and monitor on/off times, wake and sleep times, and the duration and reason if the monitor was removed for more than 10 minutes. The ActiGraph GT3X+ accelerometer was initialized with a 30 Hz sampling frequency and raw data from .gt3x files were converted to 30-second epoch data files prior to analysis. A *valid day* was defined as a minimum wear time of 10 hours/day, with *non-wear time* defined as 60 minutes or more of consecutive activity counts of zero, with a spike tolerance of 2 minutes and 100 counts/minute [22,23]. For all valid days, daily estimates of steps (steps/day), time spent in MVPA (minutes/day), and longest sedentary bout (minutes/day) were derived from the vertical axis data using ActiLife software version 6 (ActiGraph, Pensacola, FL, USA) using cutpoints of less than 100 counts/minute for sedentary [24] and more than or equal to 2020 counts/minute for moderate-to-vigorous intensity activity [22].

The ActiGraph GT3X+ accelerometer has been shown to have good reliability (intraclass correlation coefficient [ICC]=.97 when tested using a motorized vibration table) [25]. Few studies have been published on the validity of the GT3X+ version of the ActiGraph accelerometer specifically; however, previous versions of the ActiGraph accelerometer (CSA and GT1M) have good waist-worn validity in treadmill walking and running

compared with indirect calorimetry (r=.56, P<.001 and r=.53, P<.05, respectively) in adults [26,27]. A recent study has demonstrated acceptable agreement between steps estimated by the ActiGraph GT3X+ at moderate-high walking speeds in a laboratory setting (ICC .72-.99 compared with direct observation) and in free-living situations (ICC=.90; compared with Yamax Digiwalker) [28].

Demographic Variables and Anthropometry

Written questionnaire items were used to collect information on gender, date of birth, and level of education. Standing height and weight were measured using a stadiometer (217 stadiometer, SECA, Hamburg, Germany) and an electronic scale (Sensa 804, SECA, Hamburg, Germany) according to protocols developed by the International Society for the Advancement of Kinanthropometry [29]. Each variable was measured twice and the mean obtained. If the first and second measures varied by more than 10%, a third was measured and the median of the three values was recorded. The same equipment was used for all participants. Body mass index (BMI) was determined as weight (kg)/height^2 (m).

Statistical Analyses

Descriptive statistics (n, mean, standard deviation, and prevalence) were calculated for demographic and physical measures. Absolute agreement was examined using ICCs and 95% confidence intervals. Correlation was assessed using Pearson correlation coefficient (r) or Spearman rank correlation coefficient (ρ) when data were non-normally distributed with 95% confidence intervals. The strength of correlation coefficients was interpreted based on the following definitions: weak (r=.5), moderate (r=.5-.7), and strong (r $\geq$.7). Bland-Altman plots [30] were used to examine the differences between all outcomes, with mean bias and 95% limits of agreement reported. After visual examination of the plots, linear regression was used to examine whether mean difference and

limits of agreement varied across mean values of Fitbit One or Jawbone UP and ActiGraph outcomes ([Fitbit One or Jawbone UP + ActiGraph outcome]/2) [31]. Cohen's kappa (κ) statistic was used to assess the agreement between devices for classification of active versus inactive based first on achieving 10,000 steps or more per day (default step goal on both devices) and second on achieving 30 minutes/day or more of MVPA, comparable with public health guidelines [32]. For each outcome (steps and MVPA), each day was coded as either 0 for active ($\geq$10,000 steps or $\geq$30 mins of MVPA) or 1 for inactive (<10,000 steps or <30 mins of MVPA). The strength of Cohen's kappa was interpreted based on the following definitions: less than chance agreement (<.0), slight agreement (.01-.20), fair agreement (.21-.40), moderate agreement (.41-.60), substantial agreement (.61-.80), and almost perfect agreement (.81-.99). P values were based on two-sided tests and were considered statistically significant at P<.05. All statistical analyses were conducted in SPSS version 21 (IBM Corp, Armonk, NY, USA). Post hoc power calculations determined that a sample size of N=289 daily comparisons would detect correlations as low as .17 with 80% power and 5% alpha.

Results

Participant Characteristics

A total of 48 participants were recruited for the larger intervention study (n=16 per group). Of the 32 participants allocated to the activity tracker groups, 29 provided valid data for the current analyses, with comparable numbers of participants in the Fitbit One group (n=14) and the Jawbone UP group (n=15). The characteristics of the participants are presented in Table 1. The sample consisted predominantly of middle-aged women (female: 90%, 26/29; age: mean 39.6, SD 11.0 years) who were highly educated (86%, 25/29 completed tertiary education) and with normal-overweight BMI (mean 25.9, SD 5.0 kg/m^2).

Table 1. Participant characteristics.

Characteristics	All (n=29)	Fitbit (n=14)	Jawbone (n=15)
Female, n (%)	26 (90)	12 (86)	14 (93)
Age (years), mean (SD)	39.6 (11.0)	36.1 (12.8)	42.8 (8.1)
Completed tertiary education, n (%)	25 (86)	13 (93)	12 (80)
BMI (kg/m^2), mean (SD)	25.9 (5.0)	26.5 (6.5)	25.4 (3.2)

Comparison Findings

The 29 participants contributed a total of 289 valid days of data for analyses (Fitbit One: n=135 days; Jawbone UP: n=154 days). Descriptive statistics, correlation coefficients, agreement, and Bland-Altman parameters for Fitbit One, Jawbone UP, and ActiGraph are presented in Table 2. According to accelerometry estimates, participants in the Fitbit group accumulated a mean 8497 (SD 2878) steps/day and 36.6 (SD 25.0) minutes/day of MVPA. Mean values for the Jawbone UP group were 7511 (SD 2692) steps/day and 37.4 (SD 22.8) minutes/day of MVPA; the

mean longest sedentary bout for this group was 46.4 (SD 9.8) minutes/day. Overall, correlations for steps and MVPA were strong for both devices, although higher for Fitbit One (r=.85 for steps and ρ=.80 for MVPA) than for Jawbone UP (r=.75 for steps and ρ=.75 for MVPA). The correlation between Jawbone UP longest idle time and ActiGraph longest sedentary bout was poor (ρ=.19). Absolute agreement (ICC) was acceptable for ActiGraph and Fitbit One steps (.90) and MVPA (.72) and Jawbone UP steps (.79). However, agreement was weak between ActiGraph and Jawbone UP estimates of MVPA (.56) and longest idle time (.08).

Table 2. Descriptive statistics, correlations, agreement, and Bland-Altman parameters for Fitbit One and ActiGraph GT3X+.[a]

Statistic	Fitbit One		Jawbone UP		
	Steps	MVPA	Steps	MVPA	Longest sedentary bout
Mean (SD)	9221 (3416)	17.0 (17.6)	8690 (4029)	75.5 (35.5)	87.1 (45.6)
GT3X+, mean (SD)	8497 (2878)	36.6 (25.0)	7511 (2692)	37.4 (22.8)	46.4 (9.8)
r /ρ (95% CI)[b]	.85 (.80, .89)	.80 (.73, .85)	.75 (.67, .81)	.75 (.67, .81)	.19 (.03, .34)
ICC (95% CI)[c]	0.90 (0.86, 0.93)	0.72 (–0.15, 0.90)	0.79 (0.72, 0.84)	0.56 (–0.20, 0.83)	0.08 (–0.12, 0.27)
Mean difference (SD)[d]	0.2*x–916.1 (1820)	–0.4*x–9.2 (–19.2)	0.5*x–2491.3 (699.0)	0.5*x+10.6 (38.1)	1.7*x–72.3 (44.8)
95% Limits of agreement[d]					
Upper	(0.2*x–916.1)+3567.0	0.2*x+0.5	(0.5*x–2491.3)+5290.0	0.8*x+20.2	(1.7*x–72.3)+87.8
Lower	(0.2*x–916.1)–3567.0	–0.8*x–18.8	(0.5*x–2491.3)–5290.0	0.08*x+1.1	(1.7*x–72.3)–87.8

[a]Days analyzed for Fitbit One: n=135; days analyzed for Jawbone UP: n=154.

[b]Correlations for steps were calculated using Pearson correlation coefficient (r). Correlations for MVPA and longest idle time were calculated using Spearman rank correlation coefficient (ρ) due to non-normally distributed data. All correlations were significant at $P<.05$.

[c]All agreements are significant at $P<.001$ except for longest sedentary bout ($P>.99$).

[d]Where the mean difference or limits of agreement were systematically biased, equations are presented, where x=a given value on the x-axis (mean of device and ActiGraph GT3X+ value).

Bland-Altman plots for steps for both devices are presented in Figure 2. For the estimation of steps, analyses revealed systematic bias for mean difference for both devices, with differences increasing with increasing steps/day (see Figure 2). However, 95% limits of agreement were unbiased for both devices and limits were wider for the Jawbone UP than for Fitbit One (5290 and 3567 steps/day, respectively). When absolute values were calculated using the mean of the x-axis values (mean of device and GT3X+ values; Fitbit One: 8859 steps/day; Jawbone UP: 8100 steps/day), both devices overestimated steps (Fitbit One: mean bias 767, 95% limits of agreement –2800 to 4334; Jawbone UP: mean bias 1178, 95% limits of agreement –4112 to 6468).

For MVPA, systematic bias was evident for both the mean difference and the limits of agreement for both the Fitbit One and the Jawbone UP (see Figure 3). The bias was toward larger mean differences and 95% limits of agreement as values on the x-axis increased. When absolute values were calculated using the mean of the x-axis values (mean of device and GT3X+; Fitbit One: 26.6 minutes/day; Jawbone UP: 56.4 minutes/day), the Fitbit One underestimated MVPA by a mean 19.2 minutes/day (95% limits of agreement –39.2 to 5.5), whereas the Jawbone UP overestimated by a mean of 38.1 minutes/day (95% limits of agreement 5.8-65).

The differences between Jawbone UP and ActiGraph estimates of longest sedentary bout were also biased (see Figure 4), with larger differences when bouts were longer. The limits of agreement were unbiased but wide (mean difference ±88 minutes), varying by up to 150% of the mean estimate according to ActiGraph.

XSL•FO
RenderX

Figure 2. Bland-Altman plots for device steps and ActiGraph steps: Fitbit (panel A; n=135), Jawbone UP (panel B; n=154). The solid line represents the mean difference (steps) between the two measures and the dashed lines are the 95% limits of agreement.
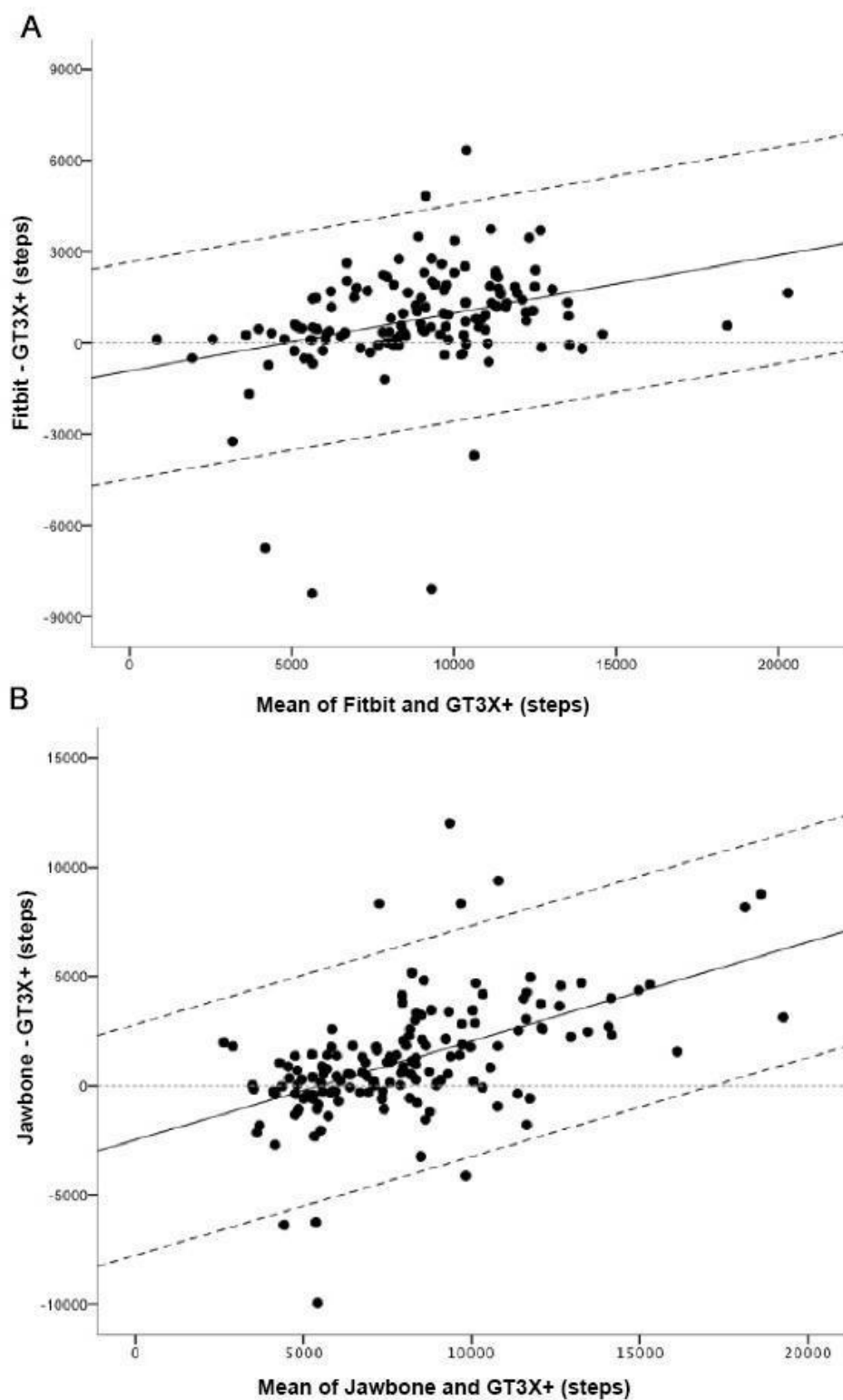
Figure 3. Bland-Altman plots for Fitbit "very active minutes" (panel A; n=135) and Jawbone "active minutes" (panel B; n=154) and ActiGraph MVPA (mins). The solid line represents the mean difference (mins) between the two measures and the dashed lines are the 95% limits of agreement. MVPA: moderate-to-vigorous physical activity.
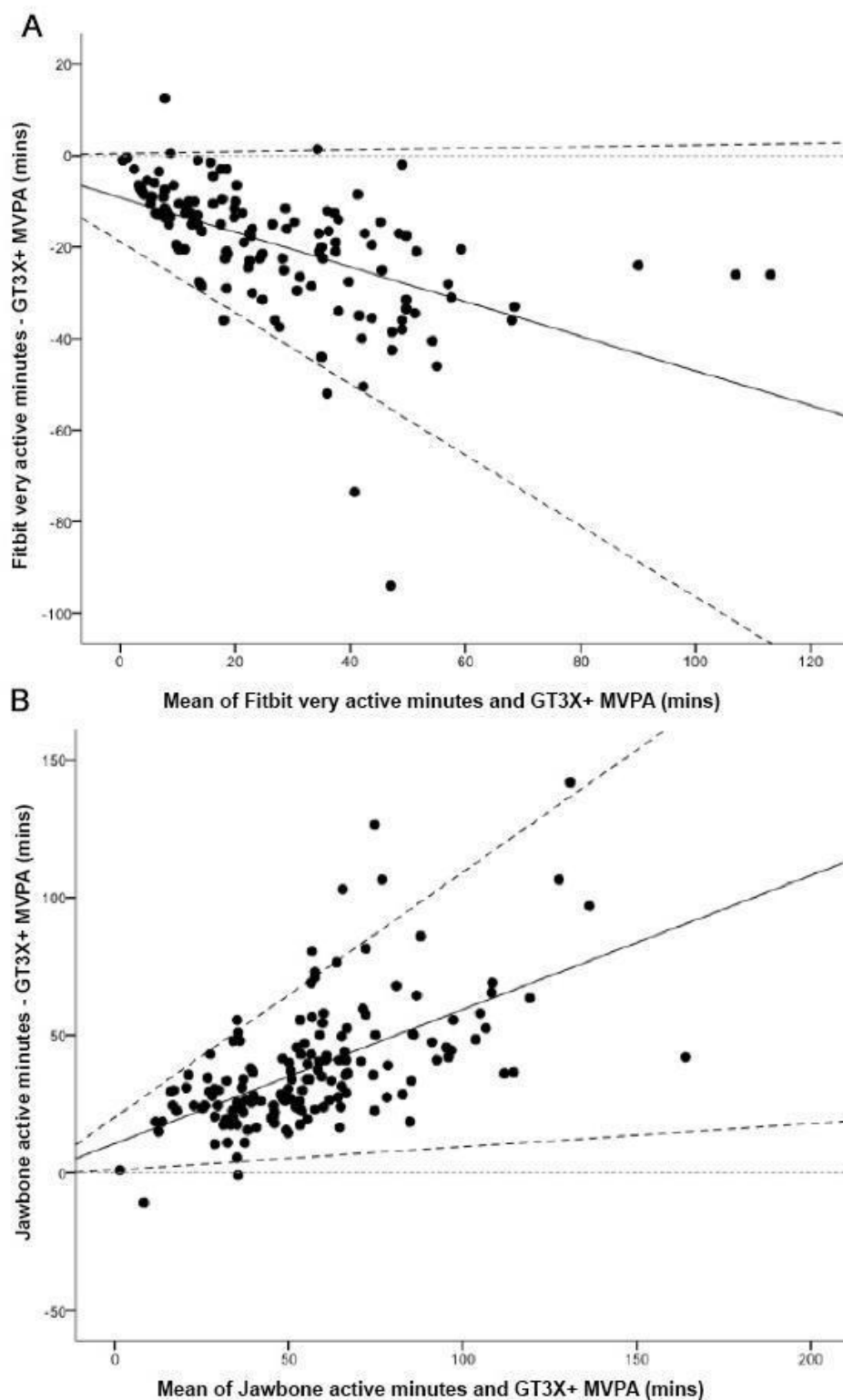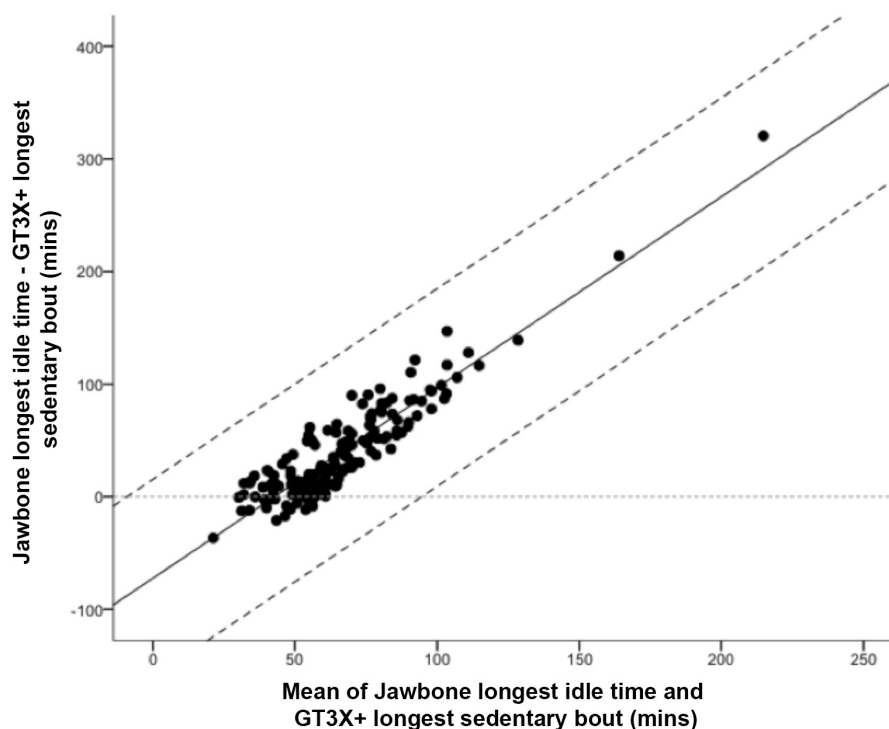
Figure 4. Bland-Altman plot for Jawbone "longest idle time" and ActiGraph "longest sedentary bout." The solid line represents the mean difference (minutes) between the two measures and the dashed lines are the 95% limits of agreement.

Classification of Active Versus Inactive

Using the criterion of at least 10,000 steps per day, agreement between the Fitbit One and ActiGraph for the classification of active versus inactive was substantial (κ=.68, P<.001). The Fitbit One correctly classified 95% (41/43) of days as active and 79% (73/92) of days as inactive. Agreement between the Jawbone UP and ActiGraph was moderate (κ=.52, P<.001). The Jawbone UP correctly classified 90% (25/28) of days as active and 80% (100/126) of days as inactive.

Using the criterion of at least 30 minutes/day of MVPA, agreement between the Fitbit One and ActiGraph was fair (κ=.39, P<.001). The Fitbit One correctly classified 40% (28/70) of days as active and 100% (63/63) of days as inactive (<30 min of MVPA per day). Agreement between the Jawbone UP and ActiGraph was slight (κ=.14, P=.001). The Jawbone UP correctly classified 100% (94/94) of days as active and 12% (7/60) of days as inactive.

Discussion

The aim of this study was to compare Fitbit One and Jawbone UP estimates of steps, MVPA, and sedentary behavior to data from the ActiGraph GT3X+ accelerometer in a free-living context. Both the Fitbit One and Jawbone UP demonstrated acceptable accuracy compared with an ActiGraph GT3X+ accelerometer for the estimation of steps per day; however, there were large over- and underestimates of MVPA. Analyses revealed systematic bias for both devices, with significant linear associations between the mean difference and mean values for steps and MVPA, and for the 95% limits of agreement and mean values for MVPA alone. The validity of the Jawbone UP measure of sedentary behavior ("longest idle time") was poor. Both devices accurately classified more than 80% of the sample days as active or inactive based on the 10,000 steps criterion; however, days were frequently misclassified for meeting public health guidelines of 30 minutes/day of MVPA.

The findings reported in this study suggest that both activity trackers have utility for counting steps in free-living settings, with both devices overestimating daily steps by only 5% to 15% compared with ActiGraph (Fitbit One: 8%; Jawbone UP: 14%). These findings are comparable to those reported in other studies in free-living contexts [16,17], although our correlation for Jawbone UP steps was lower (r=.75 vs r=.97) [17]. This difference in correlation may be due to the larger sample size in our study (n=154 vs n=21 days). The Fitbit One and Jawbone UP have also been previously assessed in laboratory settings, where the correlation with reference measures was considerably stronger (.97-1.00) [13,15,33]. This is likely due to the tightly controlled conditions in a laboratory protocol. No previous studies have reported systematic bias for steps or MVPA and these findings are important as they suggest that the magnitude and direction of the average device error changes with increasing total number of steps/day. However, this appears to have little influence on the classification of participants as active or inactive based on the cutoff of 10,000 steps/day, with excellent agreement for both devices compared with ActiGraph accelerometry.

Both devices were less accurate measuring MVPA than steps, with correlations of .56 to .80 for both devices against ActiGraph data. These findings are comparable to those reported by Ferguson et al [17], who also reported a similar range of correlations for Fitbit One and Jawbone UP estimates of MVPA

compared with ActiGraph accelerometry (.46-.91) in a free-living context. Our findings also demonstrated systematic bias in both devices across mean difference and 95% limits of agreement, indicating that both the difference between devices and the range of error vary across mean values.

It is important to note that despite reasonable correlations for MVPA, compared with ActiGraph, the Fitbit One underestimated MVPA by 46%, misclassifying 60% of days as inactive when they were active, and the Jawbone UP overestimated MVPA by 50%, misclassifying 88% of days as active when they were inactive. The implications of consumer devices over- and underestimating MVPA have significant practical implications. For example, the Jawbone UP overestimated MVPA by a mean 38.1 (SD 22.8) minutes/day and was more likely to classify an inactive day as active. Over a week, this would result in an overestimation of MVPA by a mean of 266.7 (SD 159.6) minutes. Therefore, consumers utilizing these devices will believe that they are engaging in almost twice the recommended dose of physical activity [32], when they are unlikely to be meeting minimum requirements [2].

The large discrepancy in under- and overestimations could be attributable to how MVPA was operationalized. The Fitbit measure of MVPA was the "very active minutes" variable, which is described by the manufacturer as "higher intensity exercise." The Jawbone measure of MVPA was the "active time" variable, for which no manufacturer's definition could be identified. We assumed that these measures could relate to MVPA, given that users are encouraged to accrue time spent in these outcomes for health benefits, in line with the promotion of MVPA as "health-enhancing" physical activity. An alternate interpretation of the Fitbit data could be vigorous activity alone, although the correlation was worse for vigorous alone (ρ=.43 compared with ActiGraph) than for MVPA. After data for this study were collected, Fitbit updated this variable with the label of "active minutes" and a clear definition of "any activity that elicits energy expenditure of ≥3 METs [metabolic equivalents] for a duration of 10 minutes or more." This aligns the Fitbit definition of "active minutes" directly with the US Centers for Disease Control and Prevention recommendations for MVPA [21] and future studies should investigate whether the revised classification improves the accuracy of MVPA estimates. The Jawbone active time data could include activities of intensity less than moderate-vigorous, which could also explain the difference from ActiGraph MVPA data.

The Jawbone UP is one of the few devices available on the market that reports on time spent idle and this is valuable because of the evidence base linking prolonged sedentary behavior to a range of adverse health outcomes [34]. Our findings indicate that the accuracy of these data are poor, with weak correlations between the Jawbone UP variable "longest idle time" per day and ActiGraph-determined "longest sedentary bout" per day. No previous studies have investigated the "idle time" feature of the Jawbone UP and, similar to MVPA, the poor correlations may reflect the relative measures of idle time and sedentary time. Sedentary behavior is now well defined as any waking activity performed in a reclined or seated posture that elicits an energy expenditure of ≤1.5 METs [35]. However,

the manufacturer's definition of "idle time" could not be identified. Our assumption that idle time should relate to sedentary time has good justification because Jawbone UP discourages idle time in the same way that health promotion messages discourage sedentary behavior [32]. Nonetheless, increased transparency from manufacturers regarding exact definitions of their variables and how they are calculated (including both idle time and active time for the Jawbone UP) would significantly improve the ability of researchers to explore the accuracy of these devices. It would also inform decision making about how to use these monitors in research studies.

In a recent systematic review of intervention studies, self-monitoring was reported as a "very promising" tool for reduction of sedentary behavior [36]. Accurate tools for self-monitoring sedentary behavior are critically needed; however, our findings suggest that the Jawbone UP measure of idle time should be used with caution. This is an important finding given that the Jawbone UP measures of idle time and the associated "inactivity alerts" are a point of difference between the many devices that are available.

Limitations of this study include the predominantly female, healthy, middle-aged sample, which limits the generalizability of the findings. The study could not control for wear time of the consumer devices and this may explain in part some of the large absolute differences between the devices and the ActiGraph. Although it is likely that data with very large differences may be attributable to differences in wear time, these data were not excluded from the current analyses because we were not able to verify this objectively. However, the sample had good wear compliance, evidenced by mean daily steps above normative values for this population [3]. In addition, it is possible that comparing devices with two different wear locations (wrist-worn Jawbone UP vs hip-worn ActiGraph) may have influenced the results. However, a recent study has reported that ActiGraph measured physical activity correlates moderately well between wrist and hip sites [37]. Further, this study used the ActiGraph with the cutpoint of 100 counts per minute as the comparison methods for sedentary behavior. Although this approach is commonly used in this field [5,24], future studies should consider using the activPAL (PAL Technologies Ltd, Glasgow, UK) device, a thigh-worn accelerometer/inclinometer that evaluates time spent sedentary based on posture rather than the cutpoint method [38]. Finally, the epoch length of the consumer-based trackers may be different from the ActiGraph. Accelerometry data in this study were processed in 30-second epochs. Additional sensitivity analyses (not shown) using 60-second epochs did not alter the overall findings.

Study strengths include the free-living setting, which improves ecological validity and takes previous laboratory studies into a real-world setting. More specifically, our study was conducted in the context of a physical activity intervention as would typically be used in research activities (eg, for self-monitoring of behavior change). The study also concurrently assessed two of the most popular brands of activity trackers on the market and two popular wear locations, wrist and waist. The reference measure (ActiGraph) was a previously validated device, with a large number of daily observations for comparison. Finally, our thorough evaluation of systematic bias is novel and a

strength of this study. Only one recent study has assessed the possibility of systematic bias for these devices and only for the outcome energy expenditure [39]. It also found potential for systematic bias in the mean difference, with less evident bias for the Fitbit Flex than the Jawbone UP [39].

Consumer-based activity trackers are widely used in the general population and research settings and have considerable potential to facilitate positive health behavior change in individuals, through the self-monitoring of physical activity and sedentary behavior. From both a consumer and researcher perspective, it is critical that these devices measure what they claim to measure.

Our findings suggest that the Fitbit One and Jawbone UP have utility for measuring steps; however, due to modest accuracy and systematic bias, they are better suited as self-monitoring tools (eg, for the public consumer or in behavior change interventions) rather than for evaluation of research outcomes. The outcomes that relate to health-enhancing MVPA (eg, the Fitbit One's "very active minutes" or the Jawbone UP's "active time") and sedentary behavior ("idle time" on the Jawbone UP) should be used with caution for both consumers and researchers alike. Future research should continue to assess the accuracy of these activity and sedentary behavior outcomes as manufacturers refine the measurement capacity of these devices.

Conflicts of Interest

None declared.

References

1. Proper KI, Singh AS, van Michelen MW, Chinapaw MJ. Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. Am J Prev Med 2011 Feb;40(2):174-182. [doi: 10.1016/j.amepre.2010.10.015] [Medline: 21238866]

2. 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]

3. Australian Bureau of Statistics. 4364.0.55.044 - Australian health survey: physical activity, 2011-12 URL: http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/D4495467B7F7EB01CA257BAC0015F593?opendocument [accessed 2016-08-17] [WebCite Cache ID 6jpI7OCnG]

4. Hadgraft NT, Lynch BM, Clark BK, Healy GN, Owen N, Dunstan DW. Excessive sitting at work and at home: Correlates of occupational sitting and TV viewing time in working adults. BMC Public Health 2015;15:899 [FREE Full text] [doi: 10.1186/s12889-015-2243-y] [Medline: 26374514]

5. Healy GN, Wijndaele K, Dunstan DW, Shaw JE, Salmon J, Zimmet PZ, et al. Objectively measured sedentary time, physical activity, and metabolic risk: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Diabetes Care 2008 Feb;31(2):369-371. [doi: 10.2337/dc07-1795] [Medline: 18000181]

6. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]

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

8. Butryn ML, Arigo D, Raggio GA, Colasanti M, Forman EM. Enhancing physical activity promotion in midlife women with technology-based self-monitoring and social connectivity: A pilot study. J Health Psychol 2014 Dec 8:1548-1555. [doi: 10.1177/1359105314558895] [Medline: 25488937]

9. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]

10. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]

11. Statista. 2014. Forecast unit sales of wearables worldwide by category from 2014 to 2016 (in millions) URL: http://www.statista.com/statistics/413225/wearables-worldwide-unit-sales-forecast/ [accessed 2016-05-18] [WebCite Cache ID 6hab0GEy5]

12. Lee J, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014 Sep;46(9):1840-1848. [doi: 10.1249/MSS.0000000000000287] [Medline: 24777201]

13. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport 2014 Sep;17(5):496-500. [doi: 10.1016/j.jsams.2013.10.241] [Medline: 24268570]

XSL•FO

RenderX

14. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]

15. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: an accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140. [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]

16. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]

17. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]

18. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

19. Australian Institute for Health and Welfare. 2003. The Active Australia Survey: a guide and manual for implementation, analysis and reporting URL: http://www.aihw.gov.au/publication-detail/?id=6442467449 [accessed 2016-08-17] [WebCite Cache ID 6jpIWasKd]

20. 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]

21. Fitbit. 2015. What are active minutes? URL: https://help.fitbit.com/articles/en_US/Help_article/What-are-very-active-minutes/ [accessed 2016-01-15] [WebCite Cache ID 6eXzcy4DN]

22. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]

23. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One 2015 Jul;10(7):e0133296 [FREE Full text] [doi: 10.1371/journal.pone.0133296] [Medline: 26192184]

24. Healy GN, Matthews CE, Dunstan DW, Winkler EA, Owen N. Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06. Eur Heart J 2011 Mar;32(5):590-597 [FREE Full text] [doi: 10.1093/eurheartj/ehq451] [Medline: 21224291]

25. Santos-Lozano A, Marín PJ, Torres-Luque G, Ruiz JR, Lucía A, Garatachea N. Technical variability of the GT3X accelerometer. Med Eng Phys 2012 Jul;34(6):787-790. [doi: 10.1016/j.medengphy.2012.02.005] [Medline: 22417978]

26. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S450-S456. [Medline: 10993414]

27. Abel MG, Hannon JC, Sell K, Lillie T, Conlin G, Anderson D. Validation of the Kenz Lifecorder EX and ActiGraph GT1M accelerometers for walking and running in adults. Appl Physiol Nutr Metab 2008 Dec;33(6):1155-1164. [doi: 10.1139/h08-103] [Medline: 19088773]

28. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]

29. Marfell-Jones M, Olds T, Stewart A, Carter L. International Standards for Anthropometric Assessment. Potchefstroom, South Africa: International Society for the Advancement of Kinanthropometry; 2006.

30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]

31. Brown R, Richmond S. An update on the analysis of agreement for orthodontic indices. Eur J Orthod 2005 Jun;27(3):286-291 [FREE Full text] [doi: 10.1093/ejo/cjh078] [Medline: 15947229]

32. Department of Health. Australia's Physical Activity and SedentaryBehaviour Guidelines. Canberra: Australian Government; 2014. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines/$File/FS-Adults-18-64-Years.PDF [WebCite Cache ID 6eXzkun16]

33. Storm FA, Heller BW, Mazzà C. Step detection and activity recognition accuracy of seven physical activity monitors. PLoS One 2015;10(3):e0118723 [FREE Full text] [doi: 10.1371/journal.pone.0118723] [Medline: 25789630]

34. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]

35. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl Physiol Nutr Metab 2012 Jun;37(3):540-542. [doi: 10.1139/h2012-024] [Medline: 22540258]

36. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2015 Sep 16:1-24. [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]

37. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]

38.   Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. Br J Sports Med 2006 Dec;40(12):992-997 [FREE Full text] [doi: 10.1136/bjsm.2006.030262] [Medline: 16980531]

39.   Bai Y, Welk GJ, Nam YH, Lee JA, Lee J, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2015 Jul 6;48(1):151. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]

Abbreviations

BMI: body mass index
ICC: intraclass correlation coefficient
MET: metabolic equivalent
MVPA: moderate-to-vigorous physical activity

A Comparison of Recruitment Methods for an mHealth Intervention Targeting Mothers: Lessons from the Growing Healthy Program

Rachel A Laws[1,2], BSc, MSc, PhD; Eloise-Kate V Litterbach[2,3], BHSc; Elizabeth A Denney-Wilson[2,4], BN, MPH, PhD; Catherine G Russell[2,4], BSc, BHSc, PhD; Sarah Taki[2,4], MPH; Kok-Leong Ong[5], BASc, PhD; Rosalind M Elliott[4], RN, PhD; Sharyn J Lymer[2,6], BPthy, BA, MSc, PhD; Karen J Campbell[1,2], BSc, MPH, PhD

[1]Deakin University, Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Science, Geelong, Australia

[2]Centre for Obesity Management and Prevention Research Excellence in Primary Health Care, Sydney, Australia

[3]Deakin University, School of Exercise and Nutrition Science, Geelong, Australia

[4]Faculty of Health, University of Technology Sydney, Sydney, Australia

[5]La Trobe Analytics Lab, La Trobe University, Melbourne, Australia

[6]Faculty of Pharmacy, University of Sydney, Sydney, Australia

Corresponding Author:
Rachel A Laws, BSc, MSc, PhD
Deakin University
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Science
221 Burwood Highway
Geelong, 3125
Australia
Phone: 61 3 92445574
Fax: 61 92446017
Email: r.laws@deakin.edu.au

Abstract

Background: Mobile health (mHealth) programs hold great promise for increasing the reach of public health interventions. However, mHealth is a relatively new field of research, presenting unique challenges for researchers. A key challenge is understanding the relative effectiveness and cost of various methods of recruitment to mHealth programs.

Objective: The objectives of this study were to (1) compare the effectiveness of various methods of recruitment to an mHealth intervention targeting healthy infant feeding practices, and (2) explore factors influencing practitioner referral to the intervention.

Methods: The Growing healthy study used a quasi-experimental design with an mHealth intervention group and a concurrent nonrandomized comparison group. Eligibility criteria included: expectant parents (>30 weeks of gestation) or parents with an infant <3 months old, ability to read and understand English, own a mobile phone, ≥18 years old, and living in Australia. Recruitment to the mHealth program consisted of: (1) practitioner-led recruitment through Maternal and Child Health nurses, midwives, and nurses in general practice; (2) face-to-face recruitment by researchers; and (3) online recruitment. Participants' baseline surveys provided information regarding how participants heard about the study, and their sociodemographic details. Costs per participant recruited were calculated by taking into account direct advertising costs and researcher time/travel costs. Practitioner feedback relating to the recruitment process was obtained through a follow-up survey and qualitative interviews.

Results: A total of 300 participants were recruited to the mHealth intervention. The cost per participant recruited was lowest for online recruitment (AUD $14) and highest for practice nurse recruitment (AUD $586). Just over half of the intervention group (50.3%, 151/300) were recruited online over a 22-week period compared to practitioner recruitment (29.3%, 88/300 over 46 weeks) and face-to-face recruitment by researchers (7.3%, 22/300 over 18 weeks). No significant differences were observed in participant sociodemographic characteristics between recruitment methods, with the exception that practitioner/face-to-face recruitment resulted in a higher proportion of first-time parents (68% versus 48%, *P*=.002). Less than half of the practitioners surveyed reported referring to the program often or most of the time. Key barriers to practitioner referral included lack of time, difficulty remembering to refer, staff changes, lack of parental engagement, and practitioner difficulty in accessing the app.

Conclusions: Online recruitment using parenting-related Facebook pages was the most cost effective and timely method of recruitment to an mHealth intervention targeting parents of young infants. Consideration needs to be given to addressing practitioner barriers to referral, to further explore if this can be a viable method of recruitment.

KEYWORDS

recruitment; mHealth; parents; social media; obesity prevention; infant feeding; children; infants; practitioners; primary health care

Introduction

Mobile health (mHealth) apps hold great promise as an effective delivery mode for evidence-based public health interventions. Interventions using mHealth are appealing for a number of reasons: first, a large proportion of the population has access to (and use) apps on their mobile phones; current data indicate that 75% of the world population has access to a mobile phone [1]. A 2012 global survey reported that mobile phone ownership in the United States encompassed 94% of the population, with similar levels of ownership in the United Kingdom (97%), Australia (86%), China (89%), and India (89%). Smartphone ownership ranges from 60-70% in these countries with the exception of India, where just one in ten people own a smartphone [2]. Second, interest in (and use of) mHealth apps for the management and promotion of health is widespread. Globally, there are over 97,000 health-related apps and approximately 1000 new apps are published every month [3]. A recent survey in the United States reported that 35% of mobile phone users downloaded apps to track or manage their health [4]. Third, evidence suggests that mobile phones are uniquely positioned to bridge gaps in health disparities and enable access to information across demographic groups [5]. Lastly, apps designed for use on mobile phones also provide the advantages of programming flexibility (ie, they can be designed with multiple functions) and can provide *around the clock* high quality information and personalized support, at low cost to both the user and the health provider [6].

As a novel use of emerging technology, mHealth is a relatively new field of research with a need for high quality studies to evaluate the effectiveness of mHealth interventions. Conducting mHealth studies presents unique challenges for researchers, one of which is understanding how best to recruit and retain participants [7], and whether new or novel recruitment methods are required. Recruiting an adequate sample size in a timely and cost effective manner is a critical issue, as it ensures that studies are adequately powered to measure effects and are conducted in an efficient manner [7,8]. Furthermore, the recruitment of diverse samples, including typically under-represented groups, is important to improve the external validity of mHealth studies [7,8].

To date, mHealth studies have used a range of recruitment approaches, including online (advertising on search engines, websites, online forums, direct emails, and social media) and traditional methods of recruitment (flyers, newspaper ads, billboards, TV and radio ads, and direct mail) [7]. Another method of recruitment that may be appropriate for health-related interventions is referral by health practitioners who feel that particular mHealth programs are trusted and fit well with their practices. While Internet-supported therapeutic interventions are gaining popularity [9], there is a lack of studies reporting on the outcomes of practitioner referral to mHealth programs.

There is a paucity of research reporting on recruitment to mHealth programs [7]. We were only able to identify one other study [10] that reported on the reach and cost of various methods of recruitment to an mHealth program. This study did not compare participant characteristics by recruitment method, which is an important issue given the potential differential reach of online versus more traditional methods of recruitment. More is known about recruitment to web-based interventions, with a number of studies [11-20] reporting that the use of online recruitment strategies such as Facebook, search engine advertisements, and promotion on relevant websites, are effective. However, only a few studies of web-based interventions [11,14,15,21] have compared the reach and costs of using online versus traditional methods of recruitment, with conflicting findings. Furthermore, while one study [15] reported no difference in participant characteristics by recruitment method, others [11,14,21] reported online recruitment to be more effective in recruiting *hard to reach* or more *at risk* groups. More research is needed to understand both the reach and costs of various methods of recruitment to both web-based and mHealth interventions, in order to inform the design of future trials.

We have recently developed an mHealth intervention for parents of young infants (growing healthy) that encourages healthy infant feeding practices, with a focus on socioeconomically disadvantaged parents. The program consists of an app and website [22], providing parents with a *one-stop shop* for evidence-based advice and strategies that are consistent with national guidelines pertaining to infant feeding in the first nine months of life. Participating parents received three push notifications or short message service (SMS) text messages (for those without smartphones) regarding infant feeding and related topics, which were relevant to the age of their infant, for each week of the intervention. Messages were tailored to parents' feeding mode (breast, formula, or mixed feeding) with links to more information on the app or website. Further details about the development of the program have been published elsewhere [23]. A feasibility study of the program has been conducted [23] using a quasi-experimental design, with an mHealth intervention group and a concurrent nonrandomized comparison group. The study used a range of recruitment methods, including traditional approaches (face-to-face recruitment by researchers), practitioner referral, and online recruitment, providing a unique

opportunity to examine the reach and costs of these various recruitment approaches.

The aims of this paper were to (1) compare the recruitment rate, costs, and characteristics of participants recruited using a range of recruitment approaches to an mHealth intervention targeting parents with young infants, and (2) to explore factors influencing practitioner referral to the program. These analyses will provide important new insights into recruitment to mHealth interventions, and further our understanding of utilizing health practitioners for referral to such programs.

Methods

### Study Design and Sample Size

The methods of this study have been published previously [23]. Briefly, the growing healthy study utilized a quasi-experimental design with an mHealth intervention group and a concurrent nonrandomized comparison group. The study aimed to recruit approximately 200 parent/child dyads to the intervention arm and a similar number to the comparison arm, with a focus on recruiting parents from socioeconomically disadvantaged regions. As this was a feasibility study, the sample size was not based on a statistical power calculation; rather, the purpose of the study was to test implementation feasibility, and sample size was therefore tailored to logistical limitations of the time and funds available to support recruitment. The data gathered in this study will provide evidence to guide sample size calculations for a subsequent randomized controlled trial. This paper will focus on recruitment to the intervention arm; details of recruitment to the comparison arm have been published [23].

Recruitment Methods - Growing healthy Program

Eligibility criteria for participation in the study included: expectant parents (>30 weeks of gestation) or parents/primary care giver with an infant <3 months old, ability to read and understand English, own a mobile phone, ≥18 years old, and living in Australia. Participants were excluded if their infant was born prematurely (before 37 weeks) or had a disability with the potential to impact on infant feeding. Recruitment to the growing healthy program commenced in December, 2014 and continued for a 12-month period. The initial method of recruitment was via practitioner referral and entailed face-to-face recruitment only. However, due to the slow rate of recruitment using these approaches over the initial 6-month recruitment period, online advertising was used in the final six months of recruitment, with the aim of boosting enrolments. Further details of these recruitment methods are described below and outlined in Figure 1.

Figure 1. Overview of the recruitment and enrolment process for the Growing healthy program.
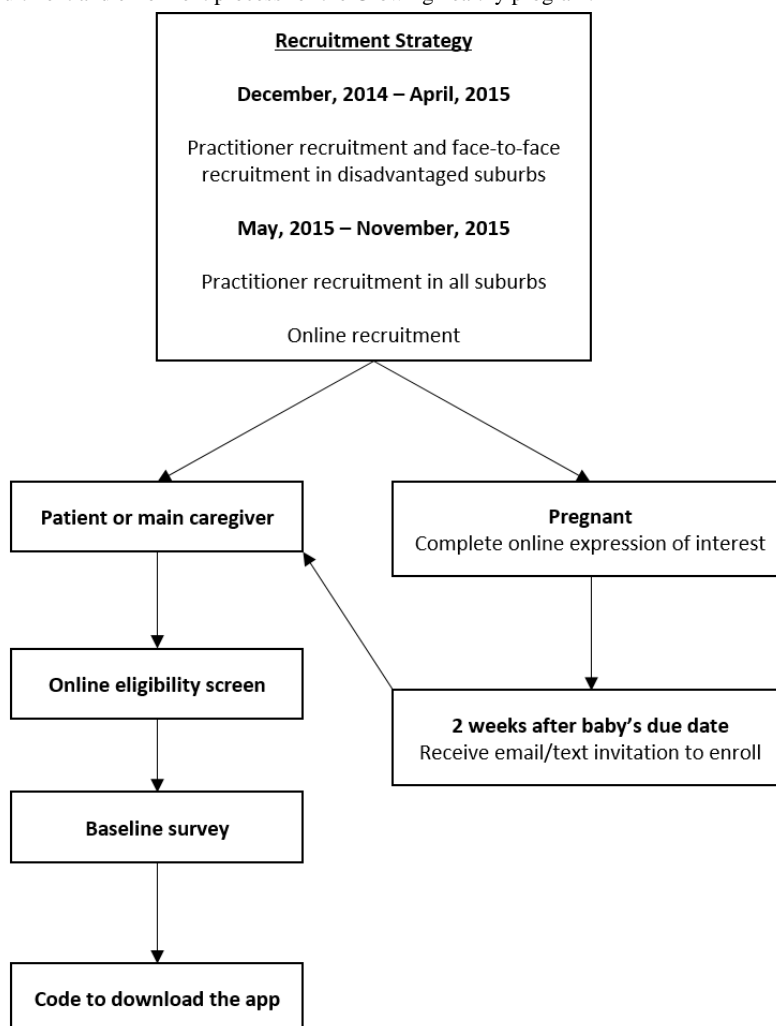
Textbox 1. Practitioner recruitment methods.

- Handing out program brochures at routine appointments

- Displaying posters in waiting rooms in participating clinics/centers/practices

- Asking interested parents to complete an expression of interest form (MCH Services only). Using these forms, parents provided contact details and gave permission for the research team to email information about the study directly to them. This approach was used in response to MCH nurses' concerns that interested parents with young infants may need a reminder to enroll

- Sending a letter of invitation from the practice (general practice only) inviting eligible potential participants (women registered in the practice in final trimester of pregnancy or having an infant less than three months old) to enroll in the program

Practitioner Referral

Practitioners were engaged in recruiting parents to the growing healthy program from three primary health care settings: (1) Maternal and Child Health (MCH) nurses (n=87) in two local government areas in Melbourne, Victoria; (2) midwives (n=10) from outpatient antenatal services at a large Melbourne hospital; and (3) practice nurses (n=8) from four general practices in the Illawarra/Shoalhaven Medicare location in New South Wales. Study sites were selected if they had a high relative level of socioeconomic disadvantage in the surrounding communities (based on the Index of Relative Socio-Economic Advantage and Disadvantage [24]) as well as a relatively high birth rate, and if the sites had been involved in previous studies and were within reasonable proximity to the study researchers. Practitioners attended a face-to-face briefing session with the research team, which included a demonstration of the growing healthy app and recruitment strategies. Practitioners were also offered a code to download the app. Practitioners were requested to promote the program to potential participants using one or more of the methods outlined in Textbox 1.

The ethics approval for practice nurses in general practice was conditional on the use of passive recruitment strategies only (ie, display of posters/brochures in waiting areas of practices and sending letters to potential participants). The use of more active approaches, such as practice nurses providing a brochure to a potential participant, were considered by this committee as potentially coercive due to the existing practitioner-patient relationship. However, ethics committees overseeing the study for MCH nurses and midwives approved both passive and active promotion by practitioners, with the condition that practitioners emphasized that the choice to participate was completely voluntary and would not affect the care provided.

Within the MCH Services in Victoria, only practitioners working in the most disadvantaged communities (defined as having an Index of Relative Socio-Economic Advantage and Disadvantage score of less than a 1000 [24]) were initially invited to recruit parents to the program (December, 2014 - April, 2015). However, due to the slow rate of recruitment over the first five months, all MCH nurses working in the area were subsequently invited to recruit parents to the program for the remaining seven months of recruitment (May, 2015 - November, 2015).

Face-To-Face Recruitment

Within participating MCH Services in Victoria, a research assistant attended first-time parent groups (n=22, two to eight parents per group) in selected low socioeconomic suburbs to inform parents of the growing healthy program. Recruitment involved providing parents with a copy of the program brochure and collecting the names and email addresses of interested parties. These parents were subsequently emailed a web link inviting them to enroll in the program.

Online Recruitment

Online recruitment for the intervention arm commenced in May, 2015 in response to the slow rate of recruitment using practitioner and face-to-face approaches. Online methods involved advertising the program on a range of popular Australian parenting websites and forums, including one advertisement on a parenting website and five Facebook status updates on Facebook pages targeting (and widely followed by) parents of young children. The main factors influencing the choice of website/Facebook pages included: Australian-based groups, groups having a large number of followers, and the advertising costs being within the project budget (costs of advertisements ranged from AUD $65 to $440 each, with costs totaling $832). Additionally, a capped price (AUD $200) official Facebook advertising package was purchased. This initiative ran for five days and could only be seen by those who were using Facebook on a computer (rather than a mobile device) on the side bar of a Facebook newsfeed. Facebook advertising on mobile devices was not possible because the project did not have a public Facebook page established, which is a requirement for Facebook advertising on mobile devices. Figure 2 shows an example of a Facebook advertisement and the text used in online advertising.

In total, online advertisements ran for eight weeks. The snowball effect of social media (ie, people tagging friends and families in the Facebook post) continued to promote the app, resulting in recruitment of a substantial number of parents for a number of weeks following the completion of active advertising. A record of the online recruitment processes was kept, including a list of the date and site of each of the online advertisements and the associated costs.

Figure 2. Facebook advertisement.

Growing healthy program
growinghealthy.org.au
Got a baby less than 3 months old? Join for a
FREE APP & GIFT VOUCHER! Places are
limited.

Text provided for online advertising

Growing healthy
Deakin University is currently running an online study for parents with babies less than 3
months old; Growing healthy. The program provides *access to a FREE app and/or a website
containing expert information and support for parents and caregivers on feeding their
baby in the first 9 months.* Each week participants will receive 3 text messages on these
topics relevant to the age of your baby with links to more information. Participants will be
able to refer to the app/website when they need to and connect with other parents on the
Growing healthy Facebook page.
You can use Growing healthy if you:
   • are a mum/main caregiver of a baby aged under 3 months
   • own any type of mobile phone
   • can speak and read English
   • are aged 18 years or older
Don't have a smartphone? No problem! As long as you have a mobile phone, you can still
join Growing healthy.
How do I join?
Go to www.growinghealthy.org.au and click **'join now'.** If you would prefer to complete
initial survey over the phone, please call 1800 915 595.

Participant Enrolment and Data Collection

Participants enrolled via the program website [22], which involved completing an eligibility screening form, providing consent, and completing a baseline survey that included questions regarding sociodemographic information and how participants heard about the study [23]. To compensate participants for the time involved in completing this survey, participants received an AUD $20 gift voucher. Participants received a code to download the app (at no cost) from the App Store (iPhone users) or Google Play (Android users), or a login for the website (for those without a smartphone capable of supporting the app). Pregnant women that were interested in participating in the study registered their interest on the study website. These mothers immediately received an SMS text message/email inviting them to enroll in the study upon the birth of their baby; a reminder SMS text message/email was sent two weeks after their baby's due date.

Recruitment Costs

Several pieces of information were collected to calculate the costs for each recruitment method, as outlined in textbox 2.

Researcher time consisted mainly of research assistants (AUD $42 per hour), but also involved an administrative assistant (AUD $22 per hour) and a PhD student (AUD $14 per hour). Two research fellows (AUD $59 per hour) led the recruitment of practitioners and were involved in the briefing sessions, along with a senior researcher (AUD $70 per hour). All costs were inclusive of *on-costs* such as superannuation and leave, and based on salary scales at Deakin University and University of Technology Sydney. The total costs were calculated for each recruitment method and divided by the number of participants recruited using each method, to calculate a cost per enrolled participant for each type of recruitment. Sensitivity analysis was conducted regarding the costs of researcher time.

Textbox 2. Costs associated with each recruitment method.

Practitioner referral:

• costs of recruitment materials provided to practitioners, researcher time for attending the briefing session, emails to parents expressing interest, and sending of invitation letters from practices. The costs of the practitioners' time spent recruiting were not included because this was provided *in kind* as part of practitioners' routine practice

Face-to-face recruitment:

• Researcher time for travel and attending first-time parent groups, sending of follow-up emails, and travel costs

Online recruitment:

• Researcher time spent searching for, and placing advertisements in, relevant parenting websites and forums, and direct advertising costs

---

Practitioner Survey and Interviews

Nine months after recruitment commenced, all participating practitioners were invited to complete a short five-to-ten minute online survey. The survey asked about how frequently the practitioners referred to the program, the methods used in promoting the program, perspectives on the barriers to referring participants, views about the program content and credibility, and perceptions of the sustainability of continued referral of parents to the program. Practitioners who were invited to complete the survey, but had not yet done so, were sent a reminder email on three separate occasions. Practitioners participating in the survey, along with their service managers, were also invited to participate in individual semi-structured telephone interviews to further explore issues pertaining to referral to the program.

Data Analyses

Differences in participants' characteristics (obtained from baseline survey) between recruitment methods were tested using Pearson's Chi-square test statistics for categorical data, and independent t-tests for normally distributed continuous variables using IBM SPSS Statistic version 22 [25]. Recruitment rates were calculated for each recruitment method as the number of eligible participants divided by recruitment time, and an exact 95% confidence interval was calculated based on the relationship between Poisson and chi-square distributions [26].

Practitioner survey data were analyzed descriptively using Microsoft Excel. Practitioner interviews were audio recorded (with permission) and transcribed verbatim. Transcripts were analyzed thematically by one author (RL), using NVivo 10 [27] to organize codes. Codes were cross checked by another author (EL) for consistency in coding approach and interpretation. Minor differences were noted in coding, each of which was resolved through discussion.

Ethics and Study Approvals

This study was approved by Deakin Human Research Ethics Committee (HREC) (2014_093), University of Technology Sydney HREC (ETH15-0110 for New South Wales participants) and the Victorian Department of Education and Training.

Results

Enrolment Status

A total of 585 individuals commenced the baseline survey; 171 (29.2%) failed to complete the survey and 32 participants completed it twice (duplicate surveys). A further 82 (82/585, 14.0%) individuals were considered ineligible (mainly because their baby was older than three months), resulting in 300 enrolled participants (51.3% of those who commenced the baseline survey).

Rates and Costs of Various Recruitment Strategies

Just over 50% of participants (151/300) were recruited online. This approach proved to be the quickest and cheapest method of recruitment, at an average cost of AUD $14 per participant recruited over a 22-week period (Table 1). Of the online sources of recruitment, advertising on the Facebook pages of popular parenting websites recruited the most participants at the lowest cost (Table 2). An official Facebook advertisement was less successful, resulting in 150 clicks and five participants recruited, despite reaching over 16,000 women of a child bearing age.

Overall, 29.3% of participants (88/300) were recruited by practitioners during a time span of just under one year, with most participants being recruited by MCH nurses (70/88) and few being recruited by practice nurses (16/88) or midwives (2/88). The costs for each participant recruited was AUD $586 for practice nurses, AUD $268 for midwives, and AUD $77 for MCH nurses (Table 1). In terms of face-to-face recruitment, a total of 22 first-time parent groups were visited by the research team over an 18-week period, resulting in 51 expressions of interest and 23 enrolments, at a cost of AUD $100 per participant. Interestingly, 12.7% (38/300) of participants were recruited through recommendations by family and friends, which could have flowed from online or practitioner referrals.

Comparison of Participant Characteristics by Recruitment Approach

There were no significant differences in participant characteristics by recruitment method, with the exception that participants referred to the program by their practitioner, or recruited by researchers face-to-face, were more likely to be first-time parents/primary care givers compared to those recruited online (Table 3). Approximately 52% (130/251) of participants had no university-level education and approximately one-fifth (53/251) of participants had a high school education or less. There were no sociodemographic differences between participants recruited by practitioners in the first wave of recruitment (disadvantaged suburbs only) compared to those recruited by practitioners in the second wave (all suburbs in the selected areas, data not shown).

Table 1. Rates and costs of recruitment strategies.

Recruitment method	Number (%) of participants recruited	Length of recruitment period	Total cost (AUD $)	Cost per participant (AUD $)	Mean number (Confidence Interval) participants recruited per week
Online recruitment	151 (50.3)	22 weeks	$2082	$13.79	6.9 (5.8-8.0)
Practitioner recruitment					
MCH[a] nurse	70 (23.3)	45 weeks	$5371	$76.73	1.6 (1.2-2.0)
Practice nurse	16 (5.3)	46 weeks	$9371	$585.69	0.3 (0.2-0.6)
Midwife	2 (0.7)	10 weeks	$536	$268.00	0.2 (0-0.7)
Face-to-face recruitment	23 (7.7)	22 weeks	$2310	$100.43	1.0 (0.7-1.6)
Word of mouth	38 (12.7)	42 weeks	n/a	n/a	0.9 (0.7-1.4)

[a] MCH: Maternal and Child Health.

Table 2. Online sources of recruitment (n=151) for the growing healthy program.

Source	Number (%)	Total cost (AUD $)	Direct advertising cost per participant (AUD $)
Parenting website Facebook pages	102 (67.5)	$650	$6.37
Facebook advertising	5 (3.3)	$200	$40
Online mothers group Facebook pages	40 (26.5)	$182	$4.55
Facebook (unspecified)	3 (2.0)	n/a	n/a
Internet search	1 (0.7)	n/a	n/a

Table 3. Comparison of participant characteristics by recruitment strategy for the growing healthy (mHealth) intervention (n=262).

Participant characteristics	Recruitment method		
	Online (n=151)	Practitioner or face to face (n=111)	*P*value
Participant age in years, mean (SD), range	30.05 (4.85), 18-46	31.10 (4.50),18-41	.08
Country of birth, n (%)			
Australia	133 (88.1)	89 (80.2)	.12
New Zealand	3 (2.0)	3 (2.7)	
United Kingdom	5 (3.3)	2 (1.8)	
Other	10 (6.6)	17 (15.3)	
Aboriginal or Torres Strait Islander, n (%)			
Yes	5 (3.3)	1 (0.9)	.20
No	146 (96.7)	110 (99.1)	
Marital status, n (%)			
Married	111 (73.5)	83 (74.8)	.70
Defacto relationship	32 (21.2)	25 (22.5)	
Separated	1 (0.7)	0 (0)	
Divorced	0 (0)	0 (0)	
Never married	7 (4.6)	3 (2.7)	
Widowed	0 (0)	0 (0)	
Average household gross weekly income (AUD $), n=223, n (%)			
Below average (<$1000/week)	16 (11.8)	15 (17.2)	.08
Average ($1000-<$1500/week)	51 (37.5)	21 (24.1)	
Above average ($1500-<2000/week)	30 (22.1)	29 (33.3)	
Higher income (>$2000/week)	39 (28.7)	22 (25.3)	
Highest level of education n=251, n (%)			
High school education or less	30 (21.0)	23 (21.3)	.96
Trade/certificate/diploma	43 (30.1)	34 (31.5)	
Degree or higher degree	70 (49.0)	51 (47.2)	
Employment status, n (%)			
Keeping house and/or raising children full time	130 (86.1)	90 (81.1)	0.23
Working full time	8 (5.3)	12 (10.8)	
Working part time/casual	8 (5.3)	5 (4.5)	
Studying (full or part time)	4 (2.6)	1 (0.9)	
Unemployed/laid off	1 (0.7)	3 (2.7)	
Health care card, n (%)			
Yes	29 (19.2)	13 (11.7)	.10
No	122 (80.2)	98 (88.3)	
Self-rated health status, n (%)			
Excellent	14 (9.3)	11 (9.9)	.19
Very good	58 (38.4)	52 (46.8)	
Good	67 (44.4)	35 (31.5)	
Fair	12 (7.9)	13 (11.7)	
Poor	0 (0)	0 (0)	

XSL•FO
RenderX

Participant characteristics	Recruitment method		
	Online (n=151)	Practitioner or face to face (n=111)	P value
Smoking status, n (%)			
Never smoked	95 (62.9)	70 (63.1)	.69
Past smoker	48 (31.8)	32 (28.8)	
Smoke occasionally	4 (2.6)	3 (2.7)	
Smoke regularly	4 (2.6)	6 (5.4)	
Baseline feeding status, n (%)			
Breastfeeding	103 (68.2)	65 (58.6)	.19
Formula feeding	25 (16.6)	20 (18.0)	
Mixed feeding	23 (15.2)	26 (23.4)	
First time parent, n (%)	73 (48.3)	75 (67.6)	.002*

*$P<.05$.

Feedback from Practitioners - Survey and Interview Findings

A total of 37 of 87 (43%) practitioners completed the online survey to provide feedback on the intervention and the referral process. Four qualitative interviews were subsequently conducted with two MCH nurses (who championed the program in the areas that they worked) and two MCH Service managers in the participating areas. While most practitioners surveyed (76%, 66/87) *agreed* or *strongly agreed* that the growing healthy program was a credible source of information and support for parents on infant feeding, and that there was a need for such programs, less than half of the survey respondents actually referred parents to the program *often* or *most of the time* (Table 4). The primary method of promoting the program was handing out brochures to parents (76%, 66/87) and displaying posters in waiting rooms (60%, 52/87). Only two of the 37 surveyed practitioners showed parents the app on their own device, and only six showed parents the website. Interestingly, only six of the 37 practitioners surveyed had downloaded the app. Nearly half (43%, 37/87) reported not downloading the app or viewing the website (Table 4). A number of practitioners reported problems downloading the app, and (as discussed in the qualitative interviews) this may have affected nurses' confidence in the program:

> *We had real problems with the app. So for the nurses to feel confident about recruiting people when the app wasn't working and getting on to the site, was quite difficult…* [Practitioner 4]

The survey indicated that a lack of time was the main reason for not referring patients to the app, and this was also discussed in the qualitative interviews:

> *We often get asked to do so many things and we also have quite a number of tasks that we need to achieve under each of the key age and stage visits anyway.* [Practitioner 2]

Other barriers included practitioners finding it difficult to remember to refer to the app (43%, 37/87), and the study enrolment process itself, such as the number of steps involved in accessing the app (28%, 24/87; Table 4). Qualitative interviews also highlighted that staff changes, including the use of casual relief staff, was a barrier at one site. During the interviews practitioners also indicated that parents were not always responsive to a referral due to the demands of having a newborn baby, and the lack of uptake or uncertainty about uptake was disheartening, as discussed by one nurse:

> *Once the baby is here, there's sleep deprivation, they're busy and they're coming to Maternal and Child Health and getting lots of different information. So maybe there's a bit of information overload…I just remember feeling a bit dejected that people weren't using it.* [Practitioner 3]

During the interviews, practitioners made a number of useful suggestions for improving program promotion in the future. These suggestions included engaging parents early (such as targeting parents antenatally when they had more time to engage and download the app), offering the app to parents with babies and toddlers, as well as the use of fridge magnets (instead of paper brochures) to act as ongoing visible reminders for parents to enroll in the program. One interviewee highlighted that vulnerable parents frequently change mobile phones and phone numbers, and that this may be a barrier to parent participation in mHealth programs. Suggestions made to improve practitioner promotion of the program included the integration of promotional materials into the standard *information pack* provided to parents by nurses at each visit, whole of service recruitment (rather than recruitment in selected suburbs), and to provide regular reminders to nurses to promote the program. The use of SMS text reminders or a nurse version of the app that provided nurse-specific push notifications was suggested. Managers also discussed the importance of broader endorsement of the program by the MCH Service funding body, for promotion of the program to become standard practice, as well as the engagement of other local health and social services in promoting the app.

Table 4. Practitioner feedback on program promotion (survey findings).

Promotion of Program	Number (%)
Frequency of promotion or referral (n=37)	
Most of the time (more than 75% of eligible parents)	4 (11)
Often (51-75% eligible parents)	12 (32)
Sometimes (26-50% eligible parents)	11 (30)
Rarely (1-25% of eligible parents)	8 (22)
Never	2 (5)
Main reason never or rarely promoted the program (n=9)	
Sometimes no program materials	1 (11)
Lack of time	4 (44)
Do not see parents personally	2 (22)
Did not like brochure with bottle feeding baby	1 (11)
No reason	1 (11)
Not interested	1 (11)
Most clients have poor English skills	1 (11)
Views about the referral process (agree or strongly agree)	
I had adequate information about the program in order to refer parents (n=30)	27 (90)
I found it difficult to remember to refer parents to the program (n=30)	13 (43)
I found the referral process worked well (n=30)	16 (53)
Some parents found it difficult to enroll in the program (n=29)	8 (28)
Method of promotion (n=37)	
Gave brochures to parents	28 (76)
Posters in waiting room or clinic areas	22 (60)
Asked interested parents to complete expression of interest	12 (32)
Showed parents the website	6 (16)
Showed parents the app	2 (5)
Discussed/encouraged in first-time parent groups	3 (8)
Brochure in waiting room	1 (3)
Supported staff to show parents the website	1 (3)
Did not promote the program	2 (5)
Encountered problems referring parents to program (n=30)	5 (17)
Practitioner access to program (n=35)	
Downloaded the app	2 (6)
Viewed the website	13 (37)
Downloaded app and viewed website	4 (11)
Neither downloaded app or viewed website	16 (46)

Discussion

To our knowledge, this is the first study to report on recruitment outcomes for an mHealth intervention targeting parents of young infants, and is unique in comparing outcomes of online versus practitioner-led recruitment. We found online recruitment using parenting-related Facebook pages to be a more effective method of recruitment compared to practitioner-led referral that took more than twice as long, and contributed only 29.3% (88/300) to the total sample, at a substantially higher cost than online recruitment. Face-to-face recruitment by researchers recruited less than 10% (22/300) of the sample, was time consuming, and resulted in higher recruitment costs than online recruitment.

Our findings suggest that consideration should be given to online recruitment strategies in future mHealth interventions

(particularly those targeting parents). In particular, strategies using social media such as parenting-related Facebook pages appear particularly successful. There is growing evidence that parents are high users of social media platforms for parenting information and support [28-30], and use online resources as their primary source of lifestyle information [31]. The limited success of the official Facebook advertisement may reflect the fact that the advertisement did not appear on mobile devices, and was only advertised for a one-week period. With recent statistics [32] indicating that more than 50% of people only use Facebook on their mobile phones, this may have been a major limiting factor. This study did not use other forms of social media, such as Twitter or other online advertising (ie, search engine advertisements), so the effectiveness of these approaches in recruiting parents to mHealth interventions remains untested.

A key challenge for mHealth researchers is keeping abreast of the latest online recruitment options. A recent technological development called ResearchKit [33] may help with this problem. ResearchKit is open-sourced and agnostic to the type of mobile operating system being used. As a result, mHealth projects leveraging ResearchKit can reach out to millions of mobile phone users in a short period of time, because ResearchKit natively brokers the recruitment target using the demographic information that individual mobile operating systems already have from their users. This system also reduces costs in a number of ways. First, the cost to develop the technology to support the mHealth recruitment process is reduced. Second, the overall cost of other aspects of recruitment is also reduced, as ResearchKit manages informed consent and facilitates the capture of digital health data. In summary, the open-sourced and platform-agnostic nature of ResearchKit eliminates the need for individual mHealth research projects to *reinvent the wheel* on common technological components of recruitment, and natively allows access to a larger population in which recruitment is carried out.

There were a number of factors that may have reduced the effectiveness of practitioner-led recruitment in this study. There was a two-month delay between the practitioner briefing session and commencement of recruitment due to a technical issue with the data collection system. It is quite likely that recruitment would have been more successful if briefing coincided with the start of recruitment. There were also delays in program brochures being distributed to the nurse teams by the service managers, due to more pressing staff issues. There was only limited email contact between the research team and practitioners during the recruitment period due to staff time pressures, providing few opportunities to remind practitioners about the study. These circumstances reflect the reality of working with busy primary health care practitioners who were asked to add program recruitment to their long list of issues to cover in appointments with parents. Furthermore, parents recruited by practitioners were required to remember to go to the study website to enroll, whereas online participants were able to do this at the *click of a button* directly from the online advertisement. The findings from our survey with practitioners also highlighted that very few practitioners actually showed parents the website or app to help promote the program, in contrast to some of our online advertising in which a picture of

the app homepage was visible. Over 40% of practitioners surveyed had not downloaded the app or viewed the website, which may reflect a lack of time, interest, technical problems experienced, and possibly the age of the nurses (with nearly 70% over 50 years of age [34]). These issues reflect the findings of recent research that reported a lack of familiarity with mHealth technologies, fear of loss-of-information control, concern about practitioner-patient relationships, and medicolegal risks to be amongst the barriers limiting antenatal health professionals from engaging with mHealth programs [35,36].

Recruitment rates by practice nurses in this study were very low, and as a result costs per participant recruited were high, partially reflecting the large amount of researcher time involved in recruiting and engaging practices. At the time of the study, primary health care organizations were undergoing a restructuring, making it difficult to engage general practices. As a result, only four practices and eight practice nurses were recruited. Furthermore, ethical requirements prevented nurses from directly promoting the program to parents; instead promotion of the program occurred via direct mail to potentially eligible parents, and displays of the program brochure and poster in clinic waiting areas. In contrast, over 80 MCH nurses were involved in the study, and were able to directly promote the program, resulting in higher yields and reduced costs per participant. The organization of MCH Services also allowed for more time-efficient briefing and engagement with practitioners. There was potentially higher buy-in amongst MCH nurses, given the close alignment of the program with their service goals compared to practice nurses who have a generalist role.

Despite our finding that online recruitment was less costly than practitioner-led recruitment, primary health care practitioners are still likely to have an important role to play in promoting and reinforcing mHealth interventions to parents, because parents access these services so frequently. Data from Victoria, Australia suggest that on average parents make 11 visits to general practitioners and 14 visits to MCH nurses in the first year of their child's life, and most of these visits are unrelated to illness [37]. The findings from this study suggest that promotion of mHealth interventions need to be integrated into standard procedures to become a streamlined part of routine practice. Practitioners also need regular prompts and reminders to engage with parents using such interventions. Consideration should be given to mobile forms of reminders, such as SMS texts or a practitioner version of the app with tailored reminders. Broader endorsement of mHealth programs by practitioner governing and funding bodies is also required for promotion of such programs to become standard practice. The cost of practitioner-led recruitment could also be decreased by broader promotion of the program to practitioner governing bodies and associations, and online promotional videos rather than face-to-face briefing sessions.

Interestingly, despite practitioner recruitment initially occurring in more disadvantaged areas, there were no significant differences in the measured sociodemographic characteristics between participants recruited online versus those recruited by their practitioner or face-to-face by researchers. There was, however, a higher proportion of first-time parents recruited

using practitioner-led and face-to-face recruitment approaches. This finding likely reflects face-to-face recruitment at first-time parent groups and practitioners having greater contact with first-time parents and/or more promotion of the program to these parents. Interestingly, more than half of those who enrolled online were second time parents, suggesting a high demand for the program beyond first-time parents. A number of early obesity prevention interventions [38-41] have targeted only first-time parents, with the rationale that the intervention effect may be larger in parents with no previous experience with infant feeding, although this remains untested and will be explored in the outcomes of this study.

The findings of this study are largely in line with the TXT2BFiT mHealth trial [10] that targeted weight gain prevention in adults 18-35 years of age. This study also reported a slow rate of recruitment via health practitioners (in this case general practitioners). The TXT2BFiT trial also found that paid official Facebook advertisements resulted in low uptake and high cost compared to other online advertising, which concurs with our findings. Our findings are also in line with some previous research of web-based interventions (mainly in the smoking cessation field) that have demonstrated that online advertising produces a greater yield of participants [11,14,15,21] than traditional methods of recruitment, and at a lower cost [11,15,21]. There are conflicting findings regarding the impact of recruitment strategies on participants' baseline characteristics. In findings similar to ours, a previous study examining smoking cessation [15] also found no significant differences for baseline characteristics between recruitment type (traditional versus online recruitment). However, other studies have reported online recruitment to be more effective in recruiting *hard to reach* or more *at risk* groups than traditional recruitment approaches [11,14,21].

This study has a number of strengths and weaknesses. The findings of the study are relevant to the recruitment of parents of young infants to mHealth interventions, and cannot be generalized outside of this context. Different research staff were involved in various recruitment strategies (eg, online recruitment only involved a research assistant, while practitioner recruitment involved more senior researchers to promote practitioner buy-in). While this factor may influence the interpretation of costings, a sensitivity analysis revealed no difference in costing outcomes when only research assistants were used in the costing calculations. The study relied on self-reports of how participants heard about the study, and this may be subject to recall bias. However, because recruitment via practitioners occurred in specific geographical areas and commenced prior to any online advertising, we were able to cross-reference this data against

participant addresses and time of recruitment for clarification in some cases. We were unable to determine if word of mouth recruitment resulted from practitioner or online promotion, and thus participants who said they heard about the study from family or friends were excluded from the costing and comparison of participant characteristics. We did observe a rise in word of mouth recruitment that corresponded with online advertising, suggesting that online recruitment may provide an easy method of referring others (ie, tagging friends/family in Facebook posts).

Another limitation of the study was the low response rate to the practitioner survey, and that only four practitioners agreed to be interviewed. This limitation may have resulted in response bias, with those harboring the strongest views about the app (either positive or negative) being more likely to participate. However, the integration of the survey and qualitative findings is a strength, and together provide additional insights into factors influencing practitioner recruitment. The four practitioners that were interviewed can be considered *key informants*; two of whom were MCH Service managers who were aware of the broader views of MCH nurses in their respective services. The other two participants were practitioner *champions* for the project in their respective services, and again were aware of some of the broader practitioner views about the program via informal discussions with their colleagues. We have yet to test whether there is any difference in retention or outcomes between the various recruitment strategies, and this issue will be a focus for future analyses.

Conclusion

This study provides new insights into the relative effectiveness of various recruitment strategies for a parenting-related mHealth intervention. Our findings suggest that mHealth interventions targeting parents should consider online methods of recruitment through Facebook pages linked to popular parenting websites, and that this tactic is likely to result in researchers meeting their required sample sizes in a shorter timeframe and at a lower cost. Participant characteristics were similar for practitioner-led and online recruitment, with the exception that participants recruited by practitioners were more likely to be first-time parents. While practitioner-led recruitment took longer and recruited fewer participants at a higher cost, this approach should not be dismissed because of the high reach of primary health care practitioners (particularly those working closely with families) and the potential to reinforce program content. Addressing practitioner barriers to referral, including improving access to the mHealth intervention, regular reminders, and integration into routine practice, is likely to be important in enhancing practitioner recruitment to mHealth programs.

Conflicts of Interest

None declared.

References

1. The World Bank. International Bank for Reconstruction and Development/The World Bank. Washington DC: World Bank; 2012. Information and Communications for Development 2012: Maximizing Mobile URL: http://siteresources.worldbank.org/EXTINFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/IC4D-2012-Report.pdf [accessed 2016-08-25] [WebCite Cache ID 6k1zb2O3I]

2. Nielsen Holdings. The Mobile Consumer: A global snapshot February 2013.: Nielsen Holdings; 2013. URL: http://www.nielsen.com/us/en/search.html?q=The+Mobile+Consumer%3A+A+global+snapshot&q1=Report&sp_cs=UTF-8&view=xml&x1=contenttypetag&i=1& [accessed 2015-11-30] [WebCite Cache ID 6k1zh4Mem]

3. Research2Guidance. Mobile Health Market Report 2013-2017. The commercialization of mHealth applications (Vol 3.).: Research2Guidance; 20. URL: http://research2guidance.com/product/mobile-health-market-report-2013-2017/ [accessed 2016-08-16] [WebCite Cache ID 6jogcfqGI]

4. Duggan M. Pew Research.: Pew Research; 2013. Cell Phone Activities 2013 URL: http://www.pewinternet.org/2013/09/19/cell-phone-activities-2013/ [accessed 2016-08-16] [WebCite Cache ID 6jogzevg2]

5. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]

6. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]

7. Lane T, 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]

8. Middelweerd A, Mollee J, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]

9. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]

10. Partridge S, Balestracci K, Wong A, Hebden L, McGeechan K, Denney-Wilson E, et al. Effective strategies to recruit young adults into the TXT2BFiT mHealth randomized controlled trial for weight gain prevention. JMIR Res Protoc 2015;4(2):e66 [FREE Full text] [doi: 10.2196/resprot.4268] [Medline: 26048581]

11. Alley S, Jennings C, Plotnikoff R, Vandelanotte C. An evaluation of web- and print-based methods to attract people to a physical activity intervention. JMIR Res Protoc 2016;5(2):e94 [FREE Full text] [doi: 10.2196/resprot.4826] [Medline: 27235075]

12. Bailey J, Pavlou M, Copas A, McCarthy O, Carswell K, Rait G, et al. The Sexunzipped trial: optimizing the design of online randomized controlled trials. J Med Internet Res 2013;15(12):e278 [FREE Full text] [doi: 10.2196/jmir.2668] [Medline: 24334216]

13. Bold K, Hanrahan T, O'Malley S, Fucito LM. Exploring the utility of web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]

14. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]

15. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]

16. Jones RB, Goldsmith L, Hewson P, Williams CJ. Recruitment to online therapies for depression: pilot cluster randomized controlled trial. J Med Internet Res 2013;15(3):e45 [FREE Full text] [doi: 10.2196/jmir.2367] [Medline: 23462072]

17. Lohse B, Wamboldt P. Purposive facebook recruitment endows cost-effective nutrition education program evaluation. JMIR Res Protoc 2013;2(2):e27 [FREE Full text] [doi: 10.2196/resprot.2713] [Medline: 23948573]

18. Morgan A, Jorm A, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]

19. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]

XSL•FO
RenderX

20. Villanti A, Jacobs M, Zawistowski G, Brookover J, Stanton C, Graham AL. Impact of baseline assessment modality on enrollment and retention in a Facebook smoking cessation study. J Med Internet Res 2015;17(7):e179 [FREE Full text] [doi: 10.2196/jmir.4341] [Medline: 26183789]

21. Graham A, Milner P, Saul J, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]

22. Growing healthy. URL: http://www.growinghealthy.org.au [accessed 2016-02-24] [WebCite Cache ID 6fWc7XSDj]

23. Denney-Wilson E, Laws R, Russell CG, Ong K, Taki S, Elliot R, et al. Preventing obesity in infants: the Growing healthy feasibility trial protocol. BMJ Open 2015 Nov;5(11):e009258 [FREE Full text] [doi: 10.1136/bmjopen-2015-009258] [Medline: 26621519]

24. Australian Bureau of Statistics. Socio-Economic Indexes for Areas. 2016. URL: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa [accessed 2016-02-23] [WebCite Cache ID 6fWcmmjaj]

25. IBM Corporation. SPSS Statistic Version 22, Released 2013. Armonk, NY: IBM Corporation; 2013. URL: http://www.ibm.com/au-en/ [accessed 2016-08-16] [WebCite Cache ID 6jojIGSiC]

26. Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). Am J Epidemiol 1990 Feb;131(2):373-375. [Medline: 2296988]

27. QSR International. QSR International Pty Ltd.: QSR International Pty Ltd; 2012. NVivo qualitative data analysis software; version 10 URL: http://www.qsrinternational.com/nvivo-product [accessed 2016-08-26] [WebCite Cache ID 6k3vErOeB]

28. Doub A, Small M, Birch LL. A call for research exploring social media influences on mothers' child feeding practices and childhood obesity risk. Appetite 2016 Apr 1;99:298-305. [doi: 10.1016/j.appet.2016.01.003] [Medline: 26767614]

29. Duggan M, Lenhart A, Lamp C, Ellison N. Pew Research Centre. 2015. Parents and Social Media URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2016-02-23] [WebCite Cache ID 6fWdD5N0K]

30. Dworkin J, Connell J, Doty J. A literature review of parents' online behavior. Cyberpsychology 2013;7(2):1. [doi: 10.5817/CP2013-2-2]

31. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [Medline: 23899373]

32. DMR. DMR. 2016. By the numbers: 70 Amazing Facebook Mobile Stats URL: http://expandedramblings.com/index.php/facebook-mobile-app-statistics/ [accessed 2016-08-16] [WebCite Cache ID 6jokHlI0P]

33. Apple Inc.. Research Kit.: Apple Inc; 2016. URL: http://www.researchkit.org [accessed 2016-02-23] [WebCite Cache ID 6fWcNwfls]

34. Laws R, Campbell K, van der Pligt P, Ball K, Lynch J, Russell G, et al. Obesity prevention in early life: an opportunity to better support the role of maternal and child health nurses in Australia. BMC Nurs 2015;14:26 [FREE Full text] [doi: 10.1186/s12912-015-0077-7] [Medline: 25972765]

35. Willcox J, van der Pligt P, Ball K, Wilkinson S, Lappas M, McCarthy E, et al. Views of women and health professionals on mHealth lifestyle interventions in pregnancy: a qualitative investigation. JMIR Mhealth Uhealth 2015;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]

36. Soltani H, Furness P, Arden M, McSeveny K, Garland C, Sustar H, et al. Women's and midwives' perspectives on the design of a text messaging support for maternal obesity services: an exploratory study. J Obes 2012;2012:835464. [doi: 10.1155/2012/835464] [Medline: 22900153]

37. Goldfeld SR, Wright M, Oberklaid F. Parents, infants and health care: utilization of health services in the first 12 months of life. J Paediatr Child Health 2003;39(4):249-253. [Medline: 12755928]

38. Campbell K, Lioret S, McNaughton S, Crawford D, Salmon J, Ball K, et al. A parent-focused intervention to reduce infant obesity risk behaviors: a randomized trial. Pediatrics 2013 Apr;131(4):652-660 [FREE Full text] [doi: 10.1542/peds.2012-2576] [Medline: 23460688]

39. Daniels L, Magarey A, Battistutta D, Nicholson J, Farrell A, Davidson G, et al. The NOURISH randomised control trial: positive feeding practices and food preferences in early childhood - a primary prevention program for childhood obesity. BMC Public Health 2009;9:387 [FREE Full text] [doi: 10.1186/1471-2458-9-387] [Medline: 19825193]

40. Taylor B, Heath A, Galland B, Gray A, Lawrence J, Sayers R, et al. Prevention of Overweight in Infancy (POI.nz) study: a randomised controlled trial of sleep, food and activity interventions for preventing overweight from birth. BMC Public Health 2011;11:942 [FREE Full text] [doi: 10.1186/1471-2458-11-942] [Medline: 22182309]

41. Wen LM, Baur LA, Simpson JM, Rissel C, Wardle K, Flood VM. Effectiveness of home based early intervention on children's BMI at age 2: randomised controlled trial. BMJ 2012;344:e3732 [FREE Full text] [Medline: 22735103]

Abbreviations

HREC: Human Research Ethics Committee
MCH: Maternal and Child Health
mHealth: mobile health
SMS: short message service

XSL·FO
RenderX

XSL•FO
RenderX

Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study

Ryan R Kroll[1], MD; J Gordon Boyd[1,2], MD, PhD, FRCPC; David M Maslove[1,2], MD, MS, FRCPC

[1]Department of Medicine, Queen's University, Kingston, ON, Canada

[2]Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada

Corresponding Author:
David M Maslove, MD, MS, FRCPC
Department of Critical Care Medicine
Queen's University
Kingston General Hospital
76 Stuart St.
Kingston, ON, K7L 2V7
Canada
Phone: 1 613 549 6666 ext 7471
Fax: 1 613 548 1325
Email: david.maslove@queensu.ca

Abstract

Background: As the sensing capabilities of wearable devices improve, there is increasing interest in their application in medical settings. Capabilities such as heart rate monitoring may be useful in hospitalized patients as a means of enhancing routine monitoring or as part of an early warning system to detect clinical deterioration.

Objective: To evaluate the accuracy of heart rate monitoring by a personal fitness tracker (PFT) among hospital inpatients.

Methods: We conducted a prospective observational study of 50 stable patients in the intensive care unit who each completed 24 hours of heart rate monitoring using a wrist-worn PFT. Accuracy of heart rate recordings was compared with gold standard measurements derived from continuous electrocardiographic (cECG) monitoring. The accuracy of heart rates measured by pulse oximetry (Spo_2.R) was also measured as a positive control.

Results: On a per-patient basis, PFT-derived heart rate values were slightly lower than those derived from cECG monitoring (average bias of −1.14 beats per minute [bpm], with limits of agreement of 24 bpm). By comparison, Spo_2.R recordings produced more accurate values (average bias of +0.15 bpm, limits of agreement of 13 bpm, $P<.001$ as compared with PFT). Personal fitness tracker device performance was significantly better in patients in sinus rhythm than in those who were not (average bias −0.99 bpm vs −5.02 bpm, $P=.02$).

Conclusions: Personal fitness tracker–derived heart rates were slightly lower than those derived from cECG monitoring in real-world testing and not as accurate as Spo_2.R-derived heart rates. Performance was worse among patients who were not in sinus rhythm. Further clinical evaluation is indicated to see if PFTs can augment early warning systems in hospitals.

Trial Registration: ClinicalTrials.gov NCT02527408; https://clinicaltrials.gov/ct2/show/NCT02527408 (Archived by WebCite at http://www.webcitation.org/6kOFez3on)

Introduction

Over the last 5 years, consumer interest in self-monitoring and personal health tracking has grown considerably [1-4]. What began as a small movement among self-described "Quantified Self" enthusiasts has grown into an industry that is worth an estimated US $9 billion worldwide and is projected to grow to US $30 billion by 2018 [5]. This growth is largely driven by consumer interest in recording and reviewing high-frequency data about activity levels and general health in order to modify personal habits and promote healthy lifestyles. Data are

generated by so-called wearables, small electronic devices that contain sensors and computing capabilities, which can be worn on a part of the body or integrated into clothing [5].

There has been growing enthusiasm for the potential use of wearable devices to improve health care delivery [4,6]. A number of different wearable sensors have been developed, which generate data that could potentially be useful in health care [7-9]. For instance, accelerometers have been incorporated into wearable devices to track physical activity such as walking, running, or climbing stairs and have also been used to evaluate sleep quality [10]. Wearable devices have seen only limited deployment in patient care settings, but their presence in clinics and hospitals is expected to grow significantly in the coming years [5].

Current clinical uses for wearable devices are mostly limited to outpatient and ambulatory settings, with a focus on the management of chronic diseases [11-13]. Applications include long-term ambulatory electrocardiogram (ECG) monitoring, optimizing pulmonary rehabilitation in patients with chronic obstructive pulmonary disease, and monitoring motor function in stroke patients as well as patients with Parkinson disease [11].

There is ample opportunity to leverage the sensing capabilities of wearable devices in the inpatient setting as well. Many newer wearable devices use photoplethysmography (PPG) to record heart rate by measuring differential reflection of light from the skin, based on the pulsatility of superficial blood vessels [14]. Heart rate sensing devices may be useful in extending the reach of vital signs monitoring in hospitals, which is typically limited by constraints on human resources. These signs, including heart rate, are monitored only a few times each day in ward settings. More frequent monitoring of heart rate stands to improve timely identification of deteriorating health of patients, increasing the chances that costly admission to the intensive care unit (ICU) can be avoided [15-19]. Heart rate surveillance also has the potential to identify patients with poorly controlled pain, to recognize incident arrhythmias, to detect sympathomimetic states such as alcohol withdrawal, and to generate more granular datasets for clinical research. A low-cost system capable of hospital-wide heart rate monitoring would therefore be valued in a time when health care expenditures are under increasing scrutiny [20].

The ability of wearable PPG sensors to reliably measure heart rate in the outpatient population has been demonstrated in at least one study [12]; however, their accuracy in hospital inpatients has not been firmly established. Equally uncertain are the accuracy and reliability of heart rate data derived from less costly wearables, such as commercially available personal fitness trackers (PFTs).

In order to address these and other questions, we examined the accuracy of heart rate measurements derived from PFTs. We focused on patients in the ICU as this cohort is closely monitored using continuous ECG (cECG) monitoring, which provides a gold standard measurement of heart rate. Because the degree of agreement that would be sufficient for clinical applications is not well defined, we also examined the agreement between cECG-derived heart rate measurements and a more widely accepted method of heart rate measurement, namely, pulse oximetry (SpO_2) monitoring.

Methods

Study Setting and Patients

We used the Fitbit Charge HR (Fitbit, San Francisco, CA) PFT to monitor heart rate in 50 patients admitted to the ICU at Kingston General Hospital (KGH), a tertiary academic medical center in Ontario. The 33-bed ICU at KGH is a mixed medical, surgical, trauma, and neurosciences unit. The PFT device studied is a wrist-worn device resembling a watch, which uses PPG to detect periodic changes in blood flow beneath the sensor, thereby deriving heart rate measurements. Heart rate values are recorded every 5 minutes. The Fitbit Charge HR is a commercially available PFT and is not currently regulated by the US Food and Drug Administration.

In order to study a cohort of patients that would best resemble hospital ward patients, we included only stable patients who were not receiving mechanical ventilation, continuous analgesia, or sedation. To reduce the risk of transmitting nosocomial infections, we excluded patients under contact precautions for methicillin-resistant *Staphylococcus aureus* and *Clostridium difficile* infections. We further excluded patients with the potential for vascular compromise of the arm on which the device was to be placed, including those with deep venous thrombosis of the upper extremity, peripherally inserted central catheters, radial arterial lines, dialysis fistulas, and severe upper extremity trauma or fracture. Patients were monitored only once for a total duration of 24 hours.

Data Capture

The study used 6 separate PFTs (3 size large, 3 size extra-large), each of which was assigned a unique email address and log-in credentials for the Fitbit website. An automated R script was used to download and process PFT data from the Fitbit website. Heart rate data are recorded by the PFT every 5 minutes. To provide a gold standard measurement of heart rate, we recovered data from the ICU bedside monitors using specialized software (BedMasterEX, Excel Medical, Jupiter, FL). Data included heart rate values, as well as heart rate data derived from continuous SpO_2 monitoring (SpO_2.R), both recorded every minute. These data were acquired as XML files and processed using an automated Python script to derive minute-level heart rate data. We synchronized bedside monitor data and PFT data using a correction factor that accounted for the difference between each device's internal clock.

Statistical Analysis

We analyzed heart rate data in aggregate across all patients, as well as on a per-patient basis. We determined the difference between cECG and PFT readings measured simultaneously, the median of differences over a 24-hour period, the interquartile range (IQR) of differences, and the Pearson correlation coefficient between cECG and PFT measurements. We used a Wilcoxon signed rank test to determine if the distribution of cECG-derived heart rates differed from that of the PFT-derived heart rates. Finally, we used Bland-Altman analysis to measure

the agreement between the PFT and cECG methods of heart rate monitoring, as well as the bias of the PFT relative to cECG.

We calculated all the above-mentioned metrics for cECG-Spo$_2$.R pairs in order to compare the accuracy of PFT measurements with that of a well-established and widely used alternative for heart rate measurement. On the basis of the mechanism of sensing used by the PFTs, we hypothesized that accuracy would differ in patients not in sinus rhythm and conducted a subgroup analysis to test this effect. Rhythm status was based on examination of cECG recording both at the time of device application and at the time of device removal, with patients designated as being in sinus rhythm only if this was present at both time points. To assess the potential for degradation in PFT performance over time, we compared the accuracy in the first 20 patients with that of the last 20 patients.

The study was approved by the Health Sciences Research Ethics Board of Queen's University (DMED-1818-15) and is registered with ClinicalTrials.gov (NCT02527408). Patients or their substitute decision makers provided informed consent. All study data were deidentified. The study did not receive funding from the device manufacturer or from any other source. All statistical analyses were done using R (v 3.2.2).

Results

Patients

Between August 2015 and January 2016, we enrolled a convenience sample of 50 patients meeting our enrollment criteria. Patients were admitted with a variety of medical and surgical conditions and were of low clinical acuity at the time of monitoring (Table 1).

Table 1. Characteristics of patients included in the study (N=50)

Characteristics	Values
Mean heart rate, beats per minute	88.3
Mean age, years	64
Sex, n (%)	
Male	26 (52)
Female	24 (48)
Admission diagnosis, n (%)	
Respiratory	12 (24)
Sepsis	7 (14)
Surgical	7 (14)
Neurologic	11 (22)
Trauma	3 (6)
Cardiovascular	6 (12)
Medical	4 (8)
Sinus rhythm, n (%)	
At start of monitoring	43 (86)
At end of monitoring	42 (84)
Personal fitness tracker size used, n (%)	
Large	23 (46)
Extra large	27 (54)

Data Acquisition

The PFT device was removed prematurely in 2 cases; in one case a patient was discharged from the ICU early, and in another case a patient developed a diffuse drug rash. In 4 cases cardiorespiratory monitoring was discontinued early, resulting in incomplete comparison data. Personal fitness tracker devices in these cases continued to collect data for the full 24-hour period. Excluding the 2 patients whose devices were removed early, PFTs showed a high degree of data capture (mean 98% of eligible data).

Heart Rate Accuracy

We analyzed a total of 12,358 cECG-PFT heart rate pairs and 56,385 cECG-Spo$_2$.R heart rate pairs. Most of the 24-hour heart rate recordings conformed to a skewed or bimodal distribution (Multimedia Appendix 1). In the pooled analysis (Figure 1 and Multimedia Appendix 2), the median difference between PFT-derived heart rates and cECG-derived heart rates was 1 beat per minute (bpm), with 73% of readings within 5 bpm of the cECG value. The correlation with cECG heart rate values was .74, and the distribution of PFT-derived heart rate values was significantly different from that of the cECG values ($P<.001$). By comparison, Spo$_2$.R-derived heart rates more

closely approximated cECG, with a median difference of 0 bpm, correlation coefficient of .91, and 89% of readings within 5 bpm of the cECG value. The Spo_2.R and cECG heart rate distributions were similar (P=.18). Visual inspection of the Bland-Altman plots revealed a tendency for the PFT to underestimate heart rate values in the range of approximately 75 to 120 bpm (Figure 1, part C). There was greater bias with the PFT method compared with the Spo_2.R method (–4.7 bpm, 95% CI –4.91 to –4.44 bpm, vs –0.2 bpm, 95% CI –0.30 to –0.16 bpm). The limits of agreement were wider with the PFT method compared with the Spo_2.R method: –31 (95% CI –31.22 to –30.40 bpm) to 21 bpm (95% CI 21.06-21.87 bpm) versus –17 (95% CI –17.01 to 16.77 bpm) to 16 bpm (95% CI 16.31-16.55 bpm; see Multimedia Appendix 3).

Per-Patient Analysis

Scatterplots for individual patients are provided in Multimedia Appendix 4. Summary statistics are presented in Table 2 and Figure 2. Although the median heart rate difference was 0 for both the PFT device and Spo_2.R readings, when compared with cECG, there was a statistically significant difference between these 2 groups (P=.003). On average, PFT recordings yielded a higher IQR, lower Pearson correlation coefficient, larger bias, and wider limits of agreement than Spo_2.R recordings (P<.001 for all comparisons).

Figure 1. Results of the pooled analysis comparing continuous electrocardiogram (cECG)-derived heart rates and personal fitness tracker (PFT)-derived heart rates (in red), as well as cECG-derived heart rates and pulse oximetry heart rates (SpO_2.R, in blue). A and B, Scatterplots showing simultaneous heart rate measurements from cECG (x-axis) compared with alternative methods (y-axis). C and D, Bland-Altman plots for heart rate measured by PFT and SpO_2.R compared with cECG. Mean heart rate is shown on the x-axis, with the difference between heart rates shown on the y-axis. The solid horizontal line represents a difference between measurements of 0, while the dashed lines represent the observed mean difference (bias) and limits of agreement. HR: heart rate; bpm: beats per minute.
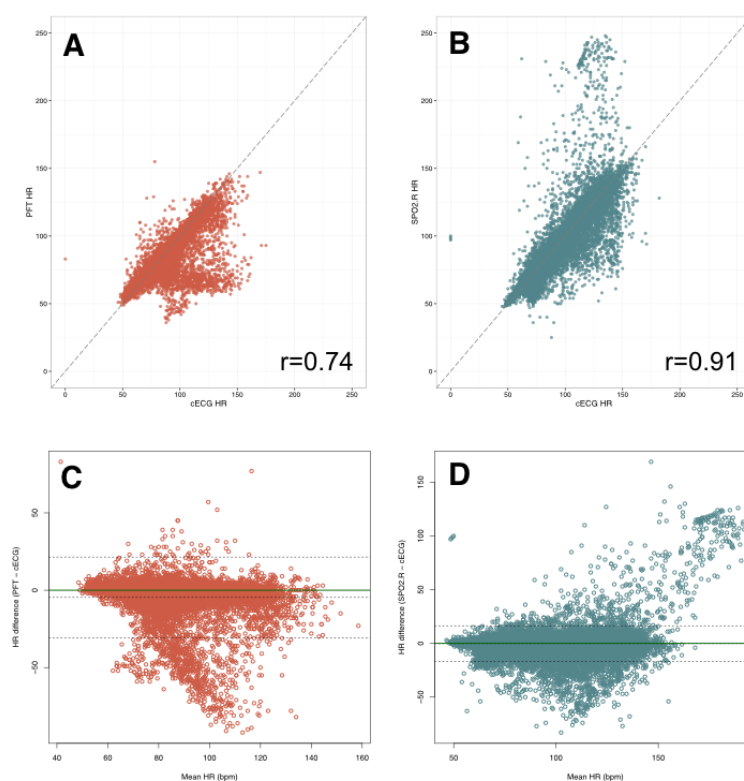
Table 2. Results of the per-patient analysis of heart rate recording accuracy.

Variable	PFT[a]	Spo$_2$.R[b]	P value
Pairs	281	1328	<.001
Zeros (median)	0	0	<.001
Median difference (bpm[c])	0	0	.003
Interquartile range (bpm)	4	1	<.001
Correlation coefficient	.57	.89	<.001
Wilcoxon P value	1.52E–06	5.06E–11	.57[d]
Bias (bpm)	−1.14	0.15	<.001
Limits of agreement (bpm)	23.88	13.00	<.001

[a]PFT: personal fitness tracker.

[b]Spo$_2$.R: pulse oximetry heart rate.

[c]bpm: beats per minute.

[d]comparing the number of recordings with Wilcoxon P value < .05.

Our subgroup analyses compared 8 patients who were not in sinus rhythm with 40 patients in sinus rhythm. Median heart rate difference, IQR, Pearson correlation coefficient, bias, and limits of agreement were all significantly worse in patients with rhythms other than sinus (P<.05 for all comparisons, Table 3). An example of poor PFT performance is shown in Figure 3. Of the 5 recordings showing the worst PFT performance, 4 were from patients not in sinus rhythm (Figure 4). There was no difference in the correlation between PFT heart rates and cECG-derived heart rates between the first 20 patients enrolled and the last 20 patients (mean Pearson correlation coefficient .51 vs .46, P=.61). Individual PFT devices were used between 5 and 13 times (mean 9 times).

Table 3. Heart rate measurement accuracy in patients in sinus rhythm compared with those not in sinus rhythm.

Measurement	Sinus rhythm (n=40)	Nonsinus rhythm (n=8)	P value
Median difference (bpm[a])	0	3.5	.04
Interquartile range (bpm)	4	8.6	.01
Correlation coefficient	.58	.23	<.001
Bias (bpm)	−0.99	−5.02	.02
Limits of agreement (bpm)	22.9	46.4	.049

[a]bpm: beats per minute.

Figure 2. Summary of the per-patient analysis of the differences between personal fitness tracker (PFT, shown in red) and pulse oximetry heart rate values (SpO$_2$.R, shown in blue) as compared with continuous electrocardiogram (cECG)–derived heart rate. A, Number of heart rate pairs analyzed. B, Number of zero measurements recorded. C, Median difference between the device-derived heart rate and cECG-derived heart rate. D, Interquartile range (IQR) of the differences. E, Pearson correlation coefficient (PFT vs cECG and SpO$_2$.R vs cECG). F, P values indicating the likelihood that the distribution of heart rate values derived from the other devices differed from that derived from cECG (Wilcoxon signed rank test). G, Mean difference between device-derived heart rate and cECG-derived heart rate from Bland-Altman analysis. H, Limits of agreement between device-derived heart rate and cECG-derived heart rate from Bland-Altman analysis. For each boxplot, individual patients are represented by an individual point. All comparisons showed statistically significant differences by Wilcoxon rank sum test ($P<.01$), with the exception of the comparison of P values ($P=.57$).
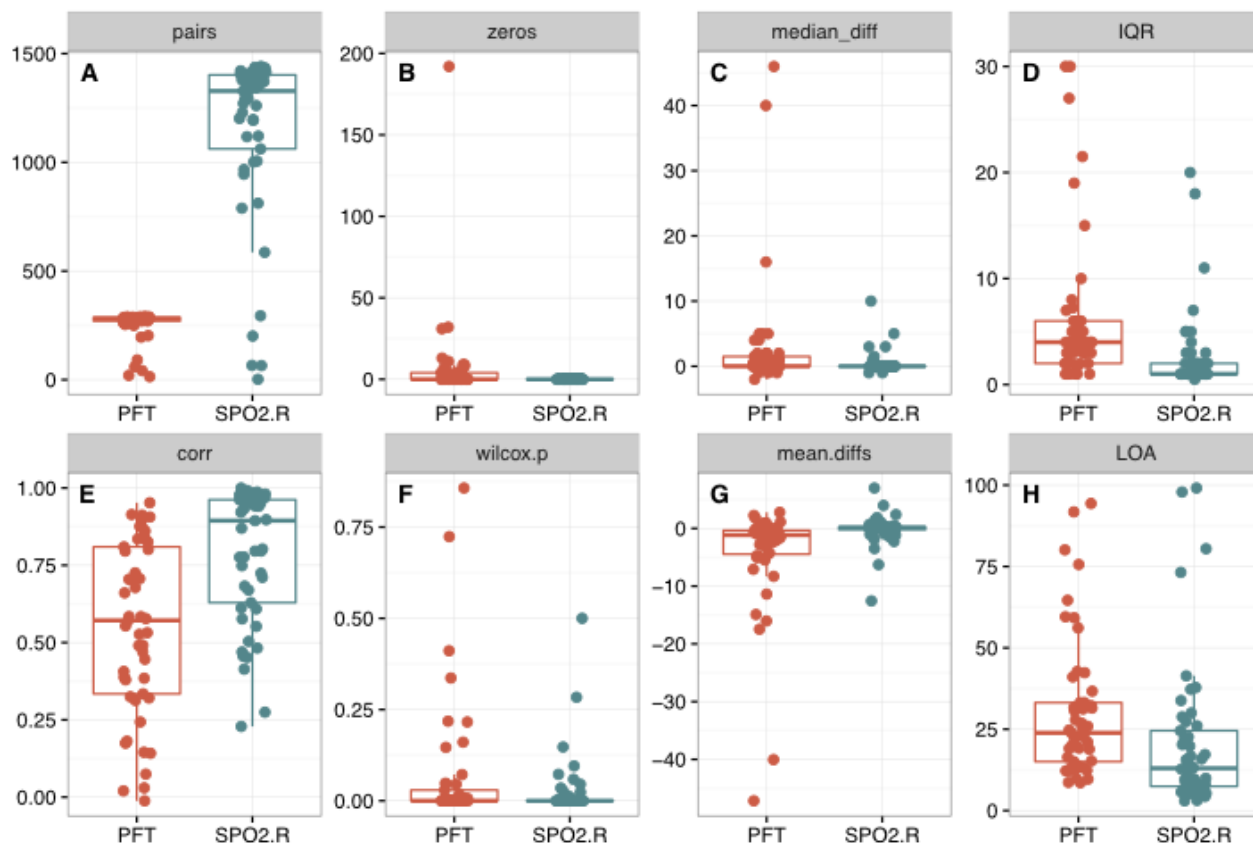

Figure 3. Scatterplots showing simultaneous heart rate (HR) measurements derived from continuous electrocardiogram (cECG; x-axis) compared with heart rate from personal fitness tracker (PFT, left) and pulse oximetry (SpO$_2$.R, right) in a patient with atrial fibrillation. In the absence of normal sinus rhythm, the PFT consistently underestimated heart rate.
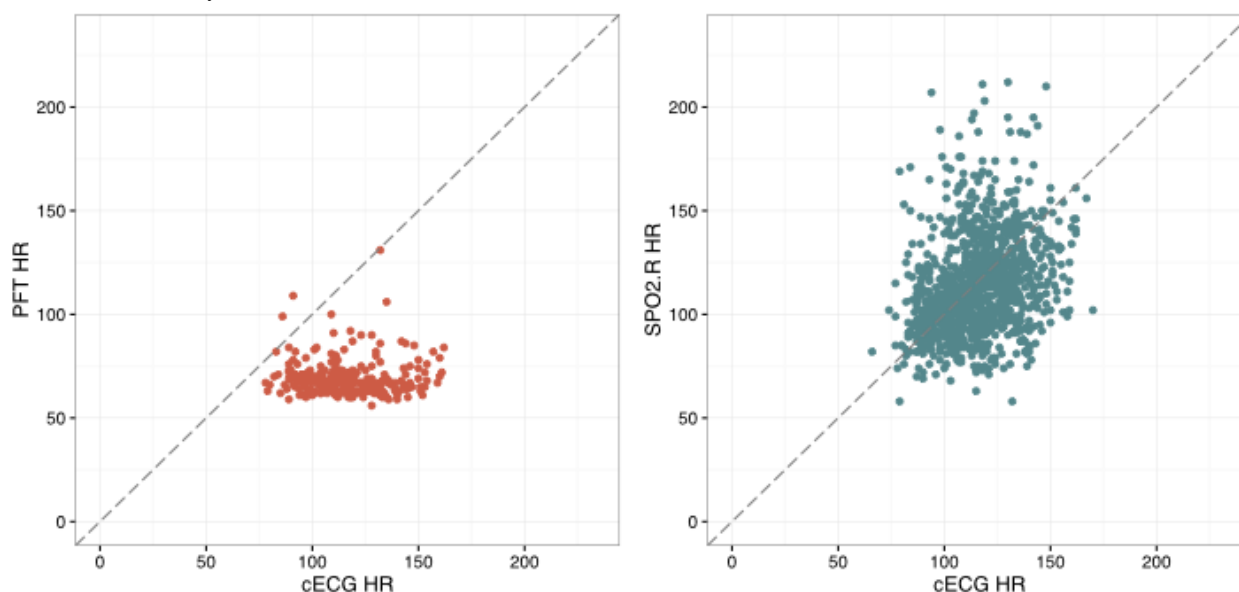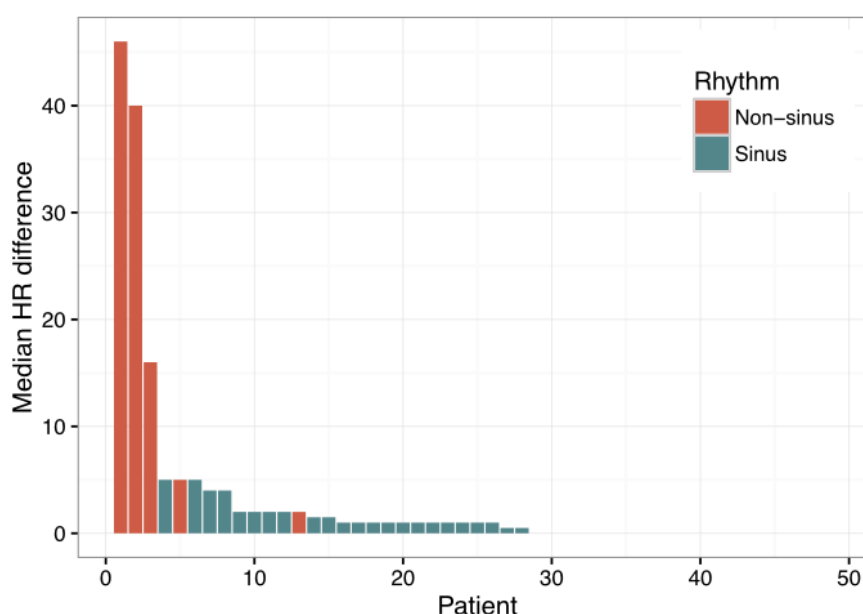
Figure 4. Patient recordings arranged according to median difference between personal fitness tracker–derived heart rate and heart rate from continuous electrocardiogram. The least accurate recordings were from patients who were not in sinus rhythm. HR: heart rate.

Discussion

Wearables in Health Care

Although the use of wearables in health care has garnered considerable attention in recent years, few objective studies exist resulting in a substantial dearth of clinical evidence regarding their use. A recent PubMed search of the term "wearable technology" revealed nearly 1000 articles published in the last 5 years, only 3% of which were clinical trials [21]. None of these included acutely ill patients. Despite the absence of evidence regarding their accuracy, data from PFTs have been used in acute care settings, including in one recently published case of a patient presenting to an emergency department who received electrical cardioversion for stable atrial fibrillation [22]. Data derived under real-world conditions from clinical settings are needed to better define the role of wearable devices in general, and commercially available fitness trackers in particular, in the delivery of acute care medicine. We conducted an observational study of heart rate monitoring accuracy of a commercially available PFT in order to provide objective evidence regarding the accuracy of its heart rate monitoring capabilities among hospitalized patients.

Principal Findings

We found that, overall, Fitbit PFT-derived heart rate measurements were less accurate and consistent than heart rate values recorded by continuous pulse oximetry. There was, however, considerable between-patient heterogeneity, with PFT heart rate values proving highly accurate in some cases and less so in others. With heart rate values analyzed on a per-patient basis, the differences between the PFT and pulse oximetry methods were less pronounced. The accuracy of PFT-based heart rate monitoring was poor among patients not in sinus rhythm.

Our results show that, on average, the PFT devices tested tended to underestimate heart rate values slightly, particularly with

heart rate values in the range of 75 to 120 bpm. The clinical implications of this degree of bias are uncertain and likely depend on the intended purpose of the monitoring. A difference of the magnitude observed might be acceptable for detecting acute clinical deterioration, which is often accompanied by marked changes in heart rate, but may not be adequate for identifying more subtle physiological derangements.

Wrist-worn heart rate sensing devices have the potential to enhance inpatient safety by identifying episodes of clinical deterioration faster than current nurse-driven vital signs monitoring practices allow. With only a small minority of hospitalized patients receiving cECG monitoring in intensive care settings, most have heart rate measurements taken only 2 to 3 times in a 24-hour period. Early warning systems (EWSs) have been shown to accurately predict cardiac arrest and hospital mortality, with some studies suggesting a reduction in these events following EWS implementation [15]. Heart rate is a common variable factored into most EWS algorithms [15,17]. Derangements in heart rate in general, and tachycardia in particular, have consistently been shown to predict impending clinical deterioration [15-19]. Early warning system variants can be complicated and difficult to use on a practical level [15,23-25]. Commercially available PFTs suggest a potential solution to address shortfalls by supplementing the monitoring of ward patients with frequent heart rate measurements generated automatically.

There are a number of potential advantages to augmenting hospital monitoring practices using wrist-worn PPG-based heart rate sensors such as the one we studied. A cost advantage may be achievable given that the device we tested retails for approximately US $170 and that we were able to reuse devices on average 9 times without seeing a decrement in performance. By comparison, conventional heart rate monitoring on inpatient wards (ie, telemetry) has been estimated to incur expenses of just under US $40 per patient per day in direct costs and as much as an additional US $170 in opportunity costs [26].

Although PFTs do not measure any additional vital signs, they do record movement data that can be used to monitor physical rehabilitation [11]. Personal fitness trackers could therefore provide benefit through the continuum of an illness episode, by providing enhanced heart rate monitoring during the acute phases, accurate tracking of mobility during convalescence, and ongoing feedback to both patient and clinician following discharge.

The use of wrist-worn devices for heart rate monitoring in hospitals also has potential disadvantages. Consumer-grade PFTs do not provide information regarding respiratory rate or blood pressure, both of which have been shown to add value in EWS [17]. Wrist-worn PPG devices might also be susceptible to errors in heart rate measurement owing to the phenomenon of the pulse deficit, in which beat-to-beat variability in stroke volume alters the amplitude of the pulse. This can be seen in atrial fibrillation, as well as other physiological conditions of acute illness such as cardiac tamponade, status asthmaticus, and various shock states, and may explain the significant decrement in heart rate sensing accuracy seen in our subgroup of patients who were not in sinus rhythm. Heart rate reporting might therefore be less accurate in the patients for whom the recognition of clinical deterioration is most needed, namely, those developing hemodynamic instability. Whether a degradation of signal quality could be used to identify physiological decompensation remains unknown.

Our study has a number of strengths. We examined the use of PFTs in a sizeable cohort of hospitalized patients under real-world conditions. Devices were adjusted only once at the time of application and were not reassessed for the duration of the 24-hour recording period by either study personnel or clinical staff. We used high-frequency data captured from continuous bedside monitoring to provide an accurate gold standard assessment of heart rate and analyzed PFT performance on both a pooled and per-patient level.

Limitations

One of the potential limitations of our study arises from the fact that the PFT-derived and cECG-derived heart rate values were obtained from different devices, with separate internal clocks. Although correction factors were used to synchronize the time stamps from the 2 heart rate sources, it is possible that in some cases the heart rate values that were treated as simultaneous were in fact separated by a short time interval. As the PFT device only recorded heart rate measurements every 5 minutes—an interval longer than the maximal device time discrepancy observed—the impact of any potential asynchrony was likely minimal.

Our study was conducted in the ICU, where cECG monitoring provides a gold standard comparator for heart rate. The extent to which our results can be generalized to hospitalized patients on the wards is therefore not certain; however, all patients enrolled were stable and were receiving ward-level care at the time of monitoring. Finally, our subgroup analysis included a relatively small number of patients not in sinus rhythm, thereby limiting the statistical power of the results.

Our study used one particular type of PFT, namely, the Fitbit Charge HR. Although many consumer-grade PFTs have similar intended functionality and use similar heart rate sensing technology, our results cannot necessarily be generalized to other wearable devices. Given that the different performance characteristics of various PFTs are not known, a study in which a mix of devices is used would be vulnerable to unwarranted mixing of effects or would require an increase in sample size proportional to the number of different devices tested.

Comparison With Prior Work

Our study is the first to report on the accuracy of heart rate recordings from wearable devices among hospital inpatients. Previous work has focused on the technical and engineering aspects of PPG-based wearable heart rate sensors, as well as discussion of their potential uses in health care settings [7,8,11]. Studies regarding the accuracy of wearables have largely focused on activity tracking and have been done using healthy volunteers [27,28]. Our study differs from previous clinical evaluations of wearable devices [29,30] in its focus on heart rate monitoring rather than activity tracking, as well as its inclusion of inpatients rather than ambulatory patients.

Conclusions

The health care sector is expected to drive a large proportion of sales of wearable devices in the coming years [5]. Optimal deployment and value from these devices will require clinical trials conducted under real-world conditions, to test the feasibility, accuracy, and costs associated with their use in health care settings. Our study suggests a potential role for PFTs in monitoring heart rate among inpatients; however, recording accuracy was not as high as with pulse oximetry and lagged substantially among patients not in sinus rhythm. Our findings suggest that future work should focus on identifying which patients are most suitable for PFT-derived heart rate monitoring, as well as software development to optimize recording accuracy in a wide range of illness states, including those associated with a pulse deficit. Although our results suggest that PFT-based heart rate monitoring may be highly accurate in some cases, prospective clinical trials are needed to evaluate their capacity to improve clinical outcomes as part of a larger strategy of enhanced hospital-based monitoring.

Authors' Contributions

RRK collected data for the study, interpreted the results, and helped draft the manuscript. DMM and JGB developed the study concept, developed the trial design, and interpreted the results. DMM collected data for the study, performed the data analysis, and helped draft the manuscript. All the authors contributed to revisions of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Frequency distributions for continuous electrocardiogram–derived heart rate (black), personal fitness tracker heart rate (red), and pulse oximetry heart rate (green) for each patient.

[PDF File (Adobe PDF File), 1MB - jmir_v18i9e253_app1.pdf]

Multimedia Appendix 2

Table showing the results of the pooled analysis of heart rate comparisons.

[PDF File (Adobe PDF File), 12KB - jmir_v18i9e253_app2.pdf]

Multimedia Appendix 3

Histogram showing the distribution of the obtained heart rate differences (x-axis) of both personal fitness tracker and pulse oximetry (SpO_2) when compared with continuous electrocardiogram, in beats per minute.

[PNG File, 48KB - jmir_v18i9e253_app3.png]

Multimedia Appendix 4

Scatterplots (continuous electrocardiogram–derived heart rate vs device-derived heart rate) for each patient.

[PDF File (Adobe PDF File), 921KB - jmir_v18i9e253_app4.pdf]

References

1. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48. [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
2. Fawcett T. Mining the Quantified Self: Personal Knowledge Discovery as a Challenge for Data Science. Big Data 2015 Dec;3(4):249-266. [doi: 10.1089/big.2015.0049] [Medline: 27441406]
3. Kostkova P. Grand challenges in digital health. Front Public Health 2015;3:134 [FREE Full text] [doi: 10.3389/fpubh.2015.00134] [Medline: 26000272]
4. Kostkova P, Brewer H, de LS, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
5. Salah H, MacIntosh E, Rajakulendran N. MaRS Market Insights. 2014 Mar 26. Wearable Tech: Leveraging Canadian Innovation to Improve Health URL: https://www.marsdd.com/mars-library/wearable-tech-leveraging-canadian-innovation-to-improve-health/ [accessed 2016-09-09] [WebCite Cache ID 6kPHdLyp0]
6. Savage N. Mobile data: Made to measure. Nature 2015 Nov 5;527(7576):S12-S13. [doi: 10.1038/527S12a] [Medline: 26536217]
7. Pantelopoulos A, Bourbakis N. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Trans. Syst., Man, Cybern. C 2010 Jan;40(1):1-12. [doi: 10.1109/TSMCC.2009.2032660]
8. Asada HH, Shaltis P, Reisner A, Rhee S, Hutchinson RC. Mobile monitoring with wearable photoplethysmographic biosensors. IEEE Eng Med Biol Mag 2003;22(3):28-40. [Medline: 12845817]
9. Di RM, Rizzo F, Parati G, Brambilla G, Ferratini M, Castiglioni P. MagIC System: a New Textile-Based Wearable Device for Biological Signal Monitoring. Applicability in Daily Life and Clinical Setting. Conf Proc IEEE Eng Med Biol Soc 2005;7:7167-7169. [doi: 10.1109/IEMBS.2005.1616161] [Medline: 17281930]
10. van Langenberg DR, Papandony MC, Gibson PR. Sleep and physical activity measured by accelerometry in Crohn's disease. Aliment Pharmacol Ther 2015 May;41(10):991-1004. [doi: 10.1111/apt.13160] [Medline: 25783784]
11. Bonato P. Clinical applications of wearable technology. Conf Proc IEEE Eng Med Biol Soc 2009;2009:6580-6583. [doi: 10.1109/IEMBS.2009.5333997] [Medline: 19964699]

12.   Chuang C, Ye J, Lin W, Lee K, Tai Y. Photoplethysmography variability as an alternative approach to obtain heart rate variability information in chronic pain patient. J Clin Monit Comput 2015 Dec;29(6):801-806. [doi: 10.1007/s10877-015-9669-8] [Medline: 25708672]

13.   Petersen C. Patient-generated health data: a pathway to enhanced long-term cancer survivorship. J Am Med Inform Assoc 2015 Dec 29:456-461. [doi: 10.1093/jamia/ocv184] [Medline: 26714765]

14.   Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable Photoplethysmographic Sensors—Past and Present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]

15.   Smith ME, Chiovaro JC, O'Neil M, Kansagara D, Quiñones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]

16.   Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs*. Crit Care Med 2012 Jul;40(7):2102-2108 [FREE Full text] [doi: 10.1097/CCM.0b013e318250aa5a] [Medline: 22584764]

17.   Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med 2016 Feb;44(2):368-374. [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]

18.   Churpek MM, Yuen TC, Winslow C, Hall J, Edelson DP. Differences in vital signs between elderly and nonelderly patients prior to ward cardiac arrest. Crit Care Med 2015 Apr;43(4):816-822 [FREE Full text] [doi: 10.1097/CCM.0000000000000818] [Medline: 25559439]

19.   Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848 [FREE Full text] [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]

20.   Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]

21.   NCBI. US National Library of Medicine 2016 URL: http://www.ncbi.nlm.nih.gov/pubmed [WebCite Cache ID 6heFmA5ga]

22.   Rudner J, McDougall C, Sailam V, Smith M, Sacchetti A. Interrogation of Patient Smartphone Activity Tracker to Assist Arrhythmia Management. Ann Emerg Med 2016 Mar 29;68(3):292-294. [doi: 10.1016/j.annemergmed.2016.02.039] [Medline: 27045694]

23.   Mitchell IA, McKay H, Van LC, Berry R, McCutcheon C, Avard B, et al. A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients. Resuscitation 2010 Jun;81(6):658-666. [doi: 10.1016/j.resuscitation.2010.03.001] [Medline: 20378235]

24.   Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526 [FREE Full text] [Medline: 11588210]

25.   Moon A, Cosgrove JF, Lea D, Fairs A, Cressey DM. An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR. Resuscitation 2011 Feb;82(2):150-154. [doi: 10.1016/j.resuscitation.2010.09.480] [Medline: 21056524]

26.   Ward MJ, Eckman MH, Schauer DP, Raja AS, Collins S. Cost-effectiveness of telemetry for hospitalized patients with low-risk chest pain. Acad Emerg Med 2011 Mar;18(3):279-286 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01008.x] [Medline: 21401791]

27.   El-Amrawy F, Nounou MI. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]

28.   Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]

29.   Appelboom G, Taylor BE, Bruce E, Bassile CC, Malakidis C, Yang A, et al. Mobile Phone-Connected Wearable Motion Sensors to Assess Postoperative Mobilization. JMIR Mhealth Uhealth 2015;3(3):e78 [FREE Full text] [doi: 10.2196/mhealth.3785] [Medline: 26220691]

30.   Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]

Abbreviations

bpm: beats per minute
ECG: electrocardiogram
EWS: early warning system
cECG: continuous electrocardiogram
ICU: intensive care unit

IQR: interquartile range
KGH: Kingston General Hospital
PFT: personal fitness tracker
PPG: photoplethysmography
SpO$_2$: pulse oximetry
SpO$_2$.R: pulse oximetry heart rate

XSL•FO
RenderX

<u>Original Paper</u>

Adolescent Female Text Messaging Preferences to Prevent Pregnancy After an Emergency Department Visit: A Qualitative Analysis

Lauren Stephanie Chernick[1], MSc, MD; Rebecca Schnall[2], RN, MPH, PhD; Melissa S Stockwell[3,4], MPH, MD; Paula M Castaño[5], MPH, MD; Tracy Higgins[2], RN, MA; Carolyn Westhoff[4,5], MS, MD; John Santelli[3,4], MPH, MD; Peter S Dayan[1], MSc, MD

[1]Division of Pediatric Emergency Medicine, Department of Pediatrics, Columbia University Medical Center, New York, NY, United States

[2]Department of Nursing, Columbia University Medical Center, New York, NY, United States

[3]Division of Child and Adolescent Health, Department of Pediatrics, Columbia University Medical Center, New York, NY, United States

[4]Department of Population and Family Health, Mailman School of Public Health, New York, NY, United States

[5]Department of Obstetrics-Gynecology, Columbia University Medical Center, New York, NY, United States

Corresponding Author:
Lauren Stephanie Chernick, MSc, MD
Division of Pediatric Emergency Medicine
Department of Pediatrics
Columbia University Medical Center
3959 Broadway
CHN 1-116
New York, NY, 10032
United States
Phone: 1 212 305 9825
Fax: 1 212 305 0682
Email: lc2243@columbia.edu

Abstract

Background: Over 15 million adolescents use the emergency department (ED) each year in the United States. Adolescent females who use the ED for medical care have been found to be at high risk for unintended pregnancy. Given that adolescents represent the largest users of text messaging and are receptive to receiving text messages related to their sexual health, the ED visit represents an opportunity for intervention.

Objective: The aim of this qualitative study was to explore interest in and preferences for the content, frequency, and timing of an ED-based text message intervention to prevent pregnancy for adolescent females.

Methods: We conducted semistructured, open-ended interviews in one urban ED in the United States with adolescent females aged 14-19 years. Eligible subjects were adolescents who were sexually active in the past 3 months, presented to the ED for a reproductive health complaint, owned a mobile phone, and did not use effective contraception. Using an interview guide, enrollment continued until saturation of key themes. The investigators designed sample text messages using the Health Beliefs Model and participants viewed these on a mobile phone. The team recorded, transcribed, and coded interviews based on thematic analysis using the qualitative analysis software NVivo and Excel.

Results: Participants (n=14) were predominantly Hispanic (13/14; 93%), insured (13/14; 93%), ED users in the past year (12/14; 86%), and frequent text users (10/14; 71% had sent or received >30 texts per day). All were interested in receiving text messages from the ED about pregnancy prevention, favoring messages that were "brief," "professional," and "nonaccusatory." Respondents favored texts with links to websites, repeated information regarding places to receive "confidential" care, and focused information on contraception options and misconceptions. Preferences for text message frequency varied from daily to monthly, with random hours of delivery to maintain "surprise." No participant feared that text messages would violate her privacy.

Conclusions: Adolescent female patients at high pregnancy risk are interested in ED-based pregnancy prevention provided by texting. Understanding preferences for the content, frequency, and timing of messages can guide in designing future interventions in the ED.

XSL•FO
RenderX

Introduction

Approximately 15 million adolescents use an emergency department (ED) each year in the United States [1]. Adolescent females in the ED have a risk of pregnancy that is up to 3 times higher than the general population; this risk is associated with reduced contraception use and lack of a primary care doctor [2,3]. Despite female adolescents expressing an interest in learning about contraception while in the ED, current methods of referral to reproductive health preventive care services from the ED show limited success [4,5].

Mobile technology has the potential to play a role in the reproductive health among adolescents [6]. Text messaging is a fast, convenient, low-cost, and scalable way of sending information [7]. In the United States, nearly 3 quarters of adolescents have access to a mobile phone, with teen females sending and receiving an average of 4050 texts per month [7,8].

In the outpatient setting, the use of text messaging to improve adolescent reproductive health shows promise; however, data exploring pregnancy prevention interventions using text messages from the ED are limited [9]. Data are needed to understand adolescents' preferences in order to develop an engaging and acceptable text messaging intervention [10]. Our objective was to study interest in and preferences for the content, structure, and timing of an ED-based text messaging pregnancy prevention intervention for adolescent females at high risk of pregnancy.

Methods

We conducted semistructured interviews from June to October 2013 at an urban tertiary care pediatric ED. The Institutional Review Board approved the study with written informed consent from participants and a waiver of parental permission.

Study Participants

We enrolled a convenience sample of females aged 14-19 years who presented to the ED. Eligibility required (1) being sexually active with a male partner in the past 3 months, (2) having a reproductive health complaint, and (3) being at high risk for pregnancy that is defined as nonuse of contraception at the last intercourse and currently not using any hormonal contraception or an intrauterine device. We excluded patients if pregnant, trying to become pregnant, too sick per the attending physician, cognitively impaired, in foster care or a ward of the state, or if they did not own a mobile phone. We enrolled only English-speaking patients as prior studies demonstrated that the adolescent Hispanic population in our ED is bilingual [2].

Study Procedures

The research team identified potentially eligible patients using an electronic tracking board. If a patient met the inclusion criteria, the research team privately explained the study to the patient and obtained written consent. After obtaining consent, participants completed a paper-based questionnaire in the ED regarding demographics, access to care, sexual behaviors, and pregnancy intentions. One of the 2 trained interviewers (LC or RS) conducted and audio-recorded interviews in a private ED room. All interviews were then transcribed by a professional service.

Interview Guide

The study team iteratively wrote the interview guide. Questions explored participants' interest in receiving pregnancy prevention texts from the ED. Probes included preference for texting versus other forms of mobile technology (email, Internet, and phone calls) and concerns of confidentiality. Questions then focused on the preferred content, timing, and frequency of text messages. Participants viewed a mobile phone with a sample of 10-15 messages and provided their preferences regarding the content and structure as well as preferences for frequency and timing.

Text Messages

Based on previous literature, we created text messages with both an educational component and action component [11]. The educational component, or "hooks", was a motivational or informative statement such as a pregnancy prevention message or information on contraceptives. We used the Health Belief Model to design the message "hook" [12]. Each message addressed one of the following: perceived severity or susceptibility of being or becoming pregnant, perceived barriers or benefit of starting contraceptives, and self-efficacy statements. The "prompt" was an action the adolescent could take at that moment, such as a link to a family planning clinic website or website with contraception advice. "Prompts" were modeled as "Cues to action," also part of the Health Belief Model. We iteratively developed 60 texts that included a "hook" and "prompt." Examples of texts included the following:

ERdoc: All services are PRIVATE at the Family Planning Clinic. You don't need a parent to receive care. (link to the local family planning clinic website)

ERdoc: Is NOW the right time to become pregnant? Talk to your partner. Or have him watch this. (link to Besider.org)

ERdoc: 1 Depo shot = 3 months of not having to worry about getting pregnant. Talk to your doctor. Learn now. (link to stayteen.org)

Data Analysis

We sought IRB approval to conduct 20 interviews; however, interviews concluded sooner when saturation was reached such that no new themes emerged. Two investigators (LC and TH) coded the transcripts using Excel (Microsoft Corporation) and the qualitative software NVivo version 10 (QSR International), each independently generating a set of codes. A codebook was then developed and used to code the transcripts using a thematic

XSL•FO

RenderX

analysis. Study team members discussed discrepancies in coding until consensus was achieved.

Results

We conducted 14 interviews, and no new themes emerged after the twelfth interview. Participants were predominantly older adolescents, Hispanic, insured, and frequent ED users (Table 1).

All but one participant had an unlimited texting data plan, 79% (11/14) used passwords to protect their phones, and 71% (10/14) sent or received 30 or more texts per day. A summary of our themes, definitions, and exemplary quotes can be found in Table 2.

Receptivity to Receiving Texts

Interested in Receiving Information via Texting Rather Than by Phone or Email

Texting was perceived as pervasive, easy, and informative. Participants noted often that "tech-heavy" teenagers do not listen to voicemails but "check the text messages incessantly."

Considered Texts to Be Safe, Because They Are Confidential and Will Not Cause Embarrassment or Harm

Messages were perceived to be private, "something between me and the doctor." Among participants who admitted that a parent or boyfriend might see their phone, it was felt that this person would be supportive, or it would be easy to deny knowing the sender.

Texts Sent by a Physician Are Trustworthy and Motivating

Participants showed interest in learning about "something new" from physicians. Messages from a trustworthy source seemed to drive ambition to go to a clinic or try a contraceptive.

Preferred Content of Text Messages

Chose Messages That Contain Sexual Health Facts Not Previously Known to Them

Participants favored educational messages about different types of contraceptives and supported messages that dispelled misconceptions, such as the message, "Many contraception methods cause no weight gain." The majority interviewed also stressed the importance of writing not only about pregnancy prevention but also sexually transmitted diseases, as one female affirmed that, even with the Depo shot, "you can still catch an infection."

Emphasized the Importance of a Respectful Staff at Free and Confidential Clinics

There was a request for information about confidential clinics, their telephone numbers, and repeated directions to clinics, for those who are "too shy to call." Clinics with walk-in policies were preferred, where you could simply "show up."

Favored Messages Commenting on the Belief in One's Capacity to Achieve a Behavior

Females responded favorably to messages about self-efficacy and "leading them in the right route." They preferred reminders "to remember it's your own life and you do what you want to do and you can say no (to sex without condoms) and that's okay."

Interested in Links to Videos, Quizzes, Websites, or Events

Most females said they would click on a link, especially when "sitting at home on your computer... (with) time your hands." There was interest in sharing links with friends or boyfriends.

Table 1. Characteristics of female adolescent participants in the emergency department (ED) (n=14).

Characteristics	n (%)
Age (18-19 years)	10 (71%)
Hispanic ethnicity	13 (93%)
Insured	13 (93%)
In college	8 (57%)
Prior pregnancy	3 (21%)
Used any effective contraception in the past	11 (79%)
Currently in a sexual relationship	12 (86%)
Saw a primary care provider in the past year	14 (100%)
Received medical care in ED in the past year	13 (93%)

XSL•FO
RenderX

Table 2. Summary of codebook organizing themes, definitions, and exemplary quotes.

Theme	Definition	Quote
Receptivity to texting from the ED		
Preference	Interested in receiving information via text messages rather than by phone or email	It will be easier than email or calling on the phone because I might not get the phone or signal...I am always texting. [Participant #3]
Comfort	Considered text messages to be safe, because they are confidential and will Not cause embarrassment or harm	If my mom passes by while my phone is charging...and she sees that (text)...'Oh Mom, it's just the clinic. They have a new texting thing. Yeah Mom, come on. We're not living in the 90s anymore'...She'd actually be very happy with that. [Participant #7]
MD-patient relationship	Text messages sent by a physician are trustworthy and motivating	I think she (older sister) would be comfortable with it because I am actually talking to a doctor, not like a regular person. [Participant #2]
Preferred content of text messages		
Information	Chose messages that contain sexual health facts not previously known to them	It'll be good to write (about) STDs and stuff, because it's not only about pregnancy. [Participant #1]
Services	Emphasized the importance of a respectful staff at free and confidential clinics	You have to make sure you tell them that their information is confidential. And that it's free...because most of my friends don't have insurance. [Participant #11]
Self-efficacy	Favored messages commenting on the belief in one's capacity to achieve a behavior	You are responsible for your own body. [Participant #10]
Links	Interested in links to videos, quizzes, websites, or events	If you want more information, you can just click on the link. [Participant #13]
Preferred text message structure		
Word choice	Preferred simple, nonaccusatory words	If you choose to have unprotected sex, you are choosing to get pregnant...I think that one might be a little bit too accusatory...I don't like being told what to do. [Participant #9]
Order	Messages should have a certain order and contain a header	Because sometimes we get those...random messages from like my phone company…And I just delete them. [Participant #6]
Customization	Commented on personalized text messages	Those girls...think they're going to be with that boyfriend for the rest of their lives. And the boyfriend tells them "Oh, we don't need to use condoms because it's only me"...They don't want to get pregnant because they have hopes and dreams...But then there is the totally different girl who's sleeping with anyone. And she doesn't want to get pregnant and she doesn't use contraception either. So maybe those are two separate girls that you can have two separate sorts of text messages for. [Participant #14]
Preferred schedule of text message delivery		
Random versus selected times	Favored random timing of messages over selected time	I think definitely send them randomly for the element of surprise...if people are expecting it and know it's coming, they going to like, 'Oh, I know what this is. I don't need it, delete it.' [Participant #9]
Day	Reached no consensus on which days text messages should be sent	Maybe on the weekends, because usually the weekends are when mistakes happen. [Participant #11]
Time	Selected texts to be sent in the afternoon and early evening	Definitely not after 8 o'clock because that's just creepy...after 8 o'clock is when the condom commercials come on, and sex...stuff like that. [Participant #7]
Frequency	Ranged in opinions of message delivery from twice a day to once a month	Don't nag them like every day. [Participant #10]

XSL•FO

RenderX

Preferred Text Message Structure

Preferred Simple, Nonaccusatory Words

Repeatedly, participants found texts with the word "you" to be offensive and condemning, such as, "Every minute another teenage girl becomes pregnant. Do you want to be one of them?" They preferred messages that provided information rather than made them feel as though they were being told what to do. In addition, they selected messages that were "direct and to the point," much like facts on a "Snapple" bottle top. However, they also wanted "supporting detail" such as a link with additional trustworthy information. Importantly, they preferred the word "birth control" rather than "contraception" and rejected abbreviations, especially when English was not their first language.

Messages Should Have a Certain Order and Contain a Header

Females approved messages with a header statement, like "ERDoc." This provided them with knowledge regarding the sender—"a reliable professional source." While they preferred repetitive information, they also voiced that each message could continue where the last message left off, as if all the messages were part of a larger dialogue.

Commented on Personalized Messages

Finally, participants favored interactive, personalized messages. They liked the idea of being able to ask questions through texting and receiving answers such as saying they do not want birth control pills and receiving information on another option or requesting messages less or more often.

Preferred Schedule of Text Message Delivery

Favored Random Timing of Messages Over Selected Time

Participants explained that the element of surprise would stimulate interest. However, one possible benefit of receiving the messages the same time every day was knowing that the texts were coming from a doctor, thus creating trust.

Reached No Consensus on Which Days Texts Should Be Sent

Some participants thought message delivery on a weekend was impactful because "that is when mistakes happen." On the contrary, weekends were the time females were busy and thus might not pay attention to their phones.

Selected Texts to Be Sent in the Afternoon and Early Evening

Participants preferred receiving messages neither too early in the morning nor too late at night.

Ranged in Opinions of Message Delivery From Twice a Day to Once a Month

Females varied widely in their judgment of how often messages should be sent. Some perceived daily messages as a "nag," whereas others viewed repetition beneficial in case they missed a prior text. In addition, messages sent weekly were similar to "sale emails," which they might disregard as spam.

Discussion

Principal Findings

In this study, adolescent females in the ED were uniformly receptive to using text messaging to promote pregnancy prevention. Text messages were considered confidential and truthful when coming from a medical source. Adolescents preferred messages that supported their ability to achieve a positive behavior, provided repeated information about locations of confidential services, were both personalized and interactive, and provided accompanying links on sexual health information previously unknown to them. Preferred word choice was nonaccusatory and direct. Although there was no consensus on the timing and frequency of messaging, there was preference for the random delivery of messages in the afternoon and evening.

The use of texting interventions has shown promise as a means to deliver health information and change health behaviors among ED patients, including reducing binge drinking, increasing postED follow-up, and increasing medication adherence [13-15]. One study noted improved ability to contact adolescents regarding sexually transmitted diseases [16]. In a study focused on violence prevention, Ranney and colleagues found that adolescents supported ED-based preventive health texting interventions, with an emphasis on personalized and tailored messages [10].

Prior research of adolescents in the outpatient (nonED) setting has found similar text message preferences regarding content, structure, tone, and confidentiality [6,17]. Adolescents in our study and several prior studies favored messages that focused on future opportunities and promoted self-efficacy [18]. We found our adolescents favored messages that focused on what they could do to prevent pregnancy as well as ones that addressed perceived barriers to obtaining contraception. These preferences match the behavioral health literature in which messages focusing on engaging in a particular behavior (gain-frame) appear more effective in changing a preventive behavior than those highlighting the consequences of failing to engage in a behavior (loss-frame) [19]. Finally, some of the variability in responses we noted and those noted in earlier literature suggests the need for customized message delivery and content for certain subgroups of patients; this may increase receptivity and self-referential processing of information [10,20].

Limitations

Limitations to our study include enrollment of a convenience sample of predominantly Hispanic patients. This study explored texting from the ED; we did not specifically compare acceptability of texting from the ED versus other health care sites. In addition, we did not ask in-depth questions to determine the specific processes necessary to create and implement interactive, tailored messages. Furthermore, while this study included females aged 14-19 years, all participants were 16 years and older; females aged 15 and younger may have different opinions. Finally, we interviewed only females. Adolescent males may have unique views regarding ED-based pregnancy prevention interventions and the use of text messaging for the ED; this is an area for future study.

Conclusions

Adolescent female patients at high pregnancy risk are interested in ED-based pregnancy prevention provided by text messages. We identified key preferences for the content, frequency, and timing of texts, which can guide the design of future interventions in the ED.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Ziv A, Boulet JR, Slap GB. Emergency department utilization by adolescents in the United States. Pediatrics 1998 Jun;101(6):987-994. [Medline: 9606224]
2. Chernick L, Kharbanda EO, Santelli J, Dayan P. Identifying adolescent females at high risk of pregnancy in a pediatric emergency department. J Adolesc Health 2012 Aug;51(2):171-178. [doi: 10.1016/j.jadohealth.2011.11.023] [Medline: 22824448]
3. Todd CS, Mountvarner G, Lichenstein R. Unintended pregnancy risk in an emergency department population. Contraception 2005 Jan;71(1):35-39. [doi: 10.1016/j.contraception.2004.08.005] [Medline: 15639070]
4. Chernick LS, Westhoff C, Ray M, Garcia M, Garth J, Santelli J, et al. Enhancing referral of sexually active adolescent females from the emergency department to family planning. J Womens Health (Larchmt) 2015 Apr;24(4):324-328. [doi: 10.1089/jwh.2014.4994] [Medline: 25860108]
5. Miller MK, Hornberger L, Sherman AK, Dowd MD. Acceptability of sexual health discussion and testing in the pediatric acute care setting. Pediatr Emerg Care 2013 May;29(5):592-597. [doi: 10.1097/PEC.0b013e31828e646f] [Medline: 23611917]
6. Perry RC, Kayekjian KC, Braun RA, Cantu M, Sheoran B, Chung PJ. Adolescents' perspectives on the use of a text messaging service for preventive sexual health promotion. J Adolesc Health 2012 Sep;51(3):220-225. [doi: 10.1016/j.jadohealth.2011.11.012] [Medline: 22921131]
7. Nielsen. 2010. US Teen Mobile Report: Calling Yesterday, Texting Today, Using Apps Tomorrow URL: http://blog.nielsen.com/nielsenwire/online_mobile/u-s-teen-mobile-report-calling-yesterday-texting-today-using-apps-tomorrow/ [WebCite Cache ID 6io1bs2yj]
8. Lenhart A. Pew Research Center. Teens, Social Media and Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2016-06-07] [WebCite Cache ID 1467833666333550]
9. Suffoletto B, Akers A, McGinnis KA, Calabria J, Wiesenfeld HC, Clark DB. A sex risk reduction text-message program for young adult females discharged from the emergency department. J Adolesc Health 2013 Sep;53(3):387-393. [doi: 10.1016/j.jadohealth.2013.04.006] [Medline: 23707402]
10. Ranney ML, Choo EK, Cunningham RM, Spirito A, Thorsen M, Mello MJ, et al. Acceptability, language, and structure of text message-based behavioral interventions for high-risk adolescent females: a qualitative study. J Adolesc Health 2014 Jul;55(1):33-40 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.017] [Medline: 24559973]
11. Armstrong AW, Watson AJ, Makredes M, Frangos JE, Kimball AB, Kvedar JC. Text-message reminders to improve sunscreen use: a randomized, controlled trial using electronic monitoring. Arch Dermatol 2009 Nov;145(11):1230-1236. [doi: 10.1001/archdermatol.2009.269] [Medline: 19917951]
12. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]
13. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-72.e4 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
14. 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):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
15. Arora S, Burner E, Terp S, Nok Lam C, 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]

16.   Huppert JS, Reed JL, Munafo JK, Ekstrand R, Gillespie G, Holland C, et al. Improving notification of sexually transmitted infections: a quality improvement project and planned experiment. Pediatrics 2012 Aug;130(2):e415-e422 [FREE Full text] [doi: 10.1542/peds.2011-3326] [Medline: 22753557]
17.   Selkie EM, Benson M, Moreno M. Adolescents' Views Regarding Uses of Social Networking Websites and Text Messaging for Adolescent Sexual Health Education. Am J Health Educ 2011 Dec;42(4):205-212 [FREE Full text] [Medline: 22229150]
18.   Gold J, Lim MS, Hellard ME, Hocking JS, Keogh L. What's in a message? Delivering sexual health promotion to young people in Australia via text messaging. BMC Public Health 2010;10:792 [FREE Full text] [doi: 10.1186/1471-2458-10-792] [Medline: 21190584]
19.   Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
20.   Smith KL, Kerr DA, Fenner AA, Straker LM. Adolescents just do not know what they want: a qualitative study to describe obese adolescents' experiences of text messaging to support behavior change maintenance post intervention. J Med Internet Res 2014;16(4):e103 [FREE Full text] [doi: 10.2196/jmir.3113] [Medline: 24713407]

<u>Original Paper</u>

Can Facebook Be Used for Research? Experiences Using Facebook to Recruit Pregnant Women for a Randomized Controlled Trial

Laura M Adam[1], BSC, RD; Donna P Manca[2], MD, MCISc, FCFP; Rhonda C Bell[1], PhD

[1]Department of Agricultural, Food & Nutritional Science, Faculty of Agricultural, Life & Environmental Sciences, University of Alberta, Edmonton, AB, Canada

[2]Department of Family Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Rhonda C Bell, PhD
Department of Agricultural, Food & Nutritional Science
Faculty of Agricultural, Life & Environmental Sciences
University of Alberta
4126B Li Ka Shing Centre for Health Innovation
Edmonton, AB, T6G 2H3
Canada
Phone: 1 780 492 7742
Fax: 1 780 492 2011
Email: rhonda.bell@ualberta.ca

Abstract

Background: Recruitment is often a difficult and costly part of any human research study. Social media and other emerging means of mass communication hold promise as means to complement traditional strategies used for recruiting participants because they can reach a large number of people in a short amount of time. With the ability to target a specified audience, paid Facebook advertisements have potential to reach future research participants of a specific demographic. This paper describes the experiences of a randomized controlled trial in Edmonton, Alberta, attempting to recruit healthy pregnant women between 8 and 20 weeks' gestation for participation in a prenatal study. Various traditional recruitment approaches, in addition to paid Facebook advertisements were trialed.

Objective: To evaluate the effectiveness of paid advertisements on Facebook as a platform for recruiting pregnant women to a randomized controlled trial in comparison with traditional recruitment approaches.

Methods: Recruitment using traditional approaches occurred for 7 months, whereas Facebook advertisements ran for a total of 26 days. Interested women were prompted to contact the study staff for a screening call to determine study eligibility. Costs associated with each recruitment approach were recorded and used to calculate the cost to recruit eligible participants. Performance of Facebook advertisements was monitored using Facebook Ads Manager.

Results: Of the 115 women included, 39.1% (n=45) of the women who contacted study staff heard about the study through Facebook, whereas 60.9% (n=70) of them heard about it through traditional recruitment approaches. During the 215 days (~7 months) that the traditional approaches were used, the average rate of interest was 0.3 (0.2) women/day, whereas the 26 days of Facebook advertisements resulted in an average rate of interest of 2.8 (1.7) women/day. Facebook advertisements cost Can $506.91 with a cost per eligible participant of Cad $20.28. In comparison, the traditional approaches cost Cad $1087, with approximately Cad $24.15 per eligible participant. Demographic characteristics of women were similar between the 2 recruitment methods except that women recruited using Facebook were significantly earlier in their pregnancy than those recruited using traditional approaches (*P*<.03).

Conclusions: Paid Facebook advertisements hold promise as a platform for reaching pregnant women. The relative ease of placing an advertisement, the comparable cost per participant recruited, and the dramatically improved recruitment rates in comparison with traditional approaches highlight the importance of combining novel and traditional recruitment approaches to recruit women for pregnancy-related studies.

Trial Registration: ClinicalTrials.gov NCT02711644; https://clinicaltrials.gov/ct2/show/NCT02711644 (Archived by WebCite at http://www.webcitation.org/6kKpagpMk)

XSL•FO
RenderX

Introduction

The recruitment portion of human research studies is often expensive and resource intensive. Finding interested and eligible participants can be challenging and can take longer than expected. Extending the recruitment period can negatively affect time-sensitive study funding, delay data collection and analyses, and ultimately delay the release of evidence necessary to change practice. For prenatal studies, an additional challenge is to find women at an appropriate gestational age.

Traditional recruitment approaches for prenatal studies include printed posters and brochures, radio, television, and newspaper advertisements, word of mouth, and approaching pregnant women in clinics [1]. Placing materials and/or study staff within family physician offices, public health centers, and community centers requires the investigators to establish working relationships with other individuals or agencies, which also takes time. Several investigators have highlighted the potential of supplementing traditional recruitment approaches with social media–based approaches to improve effectiveness of recruitment for clinical studies involving pregnant women [2,3]. Social media, including Facebook, is more than a way to connect with friends and has evolved into a platform for sharing information [3] as well as an effective location for crowdfunding for various causes [4]. With an average of 1.09 billion users daily [5], Facebook is the social media site that individuals engage in most often [6]. The greatest proportion of users are women between the ages of 18 and 49 years [6], which highlights its potential to recruit participants for prenatal studies. Women are likely to engage on Facebook by "sharing statuses" and "liking" posts, which may allow information to be quickly passed along to Facebook friends [4,7].

Edmonton, Alberta, has been noted previously to be a difficult center in which to access and recruit pregnant women for research [1], and Facebook advertisements have proved useful for other studies recruiting participants from "hard to reach" populations. For example, parents of 13- to 17-year-olds [8], immigrants with language barriers [9], individuals at high risk for human immunodeficiency virus infection [10], adolescents [11,12], and young adult veterans [13] have all been successfully recruited using this approach. Advertising to potential participants via Facebook is novel within maternal health research. One study found social media effective for recruiting women before conception [3], while 3 other studies used paid Facebook advertisements to successfully recruit participants for preconception [14] or prenatal studies [15,16]. Each of these pregnancy-related studies involved Internet-based or telephone interventions; thus, there is little known about the potential for Facebook to be used to recruit pregnant women to a randomized controlled trial (RCT) involving face-to-face clinic visits.

The overall objective of the RCT was to examine the efficacy of supportive prenatal counseling versus standard prenatal care in promoting appropriate weight gain and dietary intake among pregnant women. The recruitment goal was 70 healthy pregnant women between 8 and 20 weeks' gestation, living in the greater Edmonton area. The purpose of this paper was to evaluate the effectiveness of paid advertisements on Facebook as a platform for recruiting pregnant women to this RCT in comparison with traditional recruitment approaches.

Methods

Recruitment

Recruitment for the RCT using traditional approaches, including printed posters and brochures, word of mouth, newspaper advertisements, local television news health report, booths at mommy and baby fairs, and advertisements in physicians' offices, started in July 2015. A Facebook account was created to distribute paid advertisements that ran intermittently from October 6 to December 1, 2015, for 26 nonconsecutive days. Advertisements were targeted to Facebook users based on the following criteria: female, 23-40 years of age, living in the Edmonton + 25-mile radius geographic area, and "Interests" related to pregnancy. The "Interests" used were Childbirth, Infant, Maternity clothing, Mommy connections, Parent, Prenatal, Prenatal care, Prenatal development, Motherhood, Today's Parent, Ultrasound, Prenatal nutrition, Family, and Parenting (Figure 1). Advertisements were managed with a set lifetime budget and Facebook performed the automatic bidding. Advertisements were optimized to get the most number of clicks to the study website at the lowest cost, with fees charged per impression. All Facebook advertisements used the same wording with the headline "Be healthy for baby and you!" and the description text: "Are you less than 20 weeks pregnant? Join a prenatal research study at the University of Alberta and have free access to a Registered Dietitian!"; multiple photographs were tested (Figure 2). Interested women were prompted to click on the "Learn More" button, which took them to the study website [17] where details about the study were provided, including the study contact information. Potential participants were encouraged to contact study staff, after which an appointment was set for screening to determine participant eligibility. Informed consent was provided by the participant in writing before the start of data collection at the baseline visit.

Performance of Facebook advertisements was monitored regularly and optimized in real time using Facebook Ads Manager. Following the trial of multiple paid Facebook advertisement campaigns, the advertisement performance was assessed by examining number of clicks and the amount of engagement. Engagement can be defined as interactions with the study advertisement such as likes, comments, or shares. Funds were reallocated to the Facebook advertisements with the highest number of clicks and highest rate of engagement. Advertisement performance was most affected by the location of the advertisement and the "Interests" specified. Advertisements located on the mobile newsfeed performed

better than those on desktop newsfeed; therefore, desktop advertisements were turned off after 9 days. The "Interests" *Prenatal nutrition*, *Family*, and *Parenting* were removed after 9 days owing to poor performance determined through analytics produced from a free trial of AdEspresso, a platform that provides additional statistics to optimize paid Facebook advertisements [18].

Costs associated with each recruitment approach were recorded and used to calculate the cost to the study to recruit eligible participants. The time required by staff to support each of these approaches was not captured and therefore not included in the cost calculations. The local television news report (ie, news reporter, film crew, television airing of segment) was donated in kind, and therefore the costs associated with it were also not included in determination of costs.

Statistical Analysis

Advertising performance statistics were collected from information provided by Facebook Ads Manager. Demographic information was self-reported by participants who were eligible and had consented to study participation. These data were collected through the Research Electronic Data Capture (REDCap) Consortium member site housed at the University of Alberta [19]. We used *t* tests or Pearson chi-square tests, as appropriate, to investigate differences between groups recruited on Facebook and through traditional means. A *P* value of .05 was considered statistically significant, and all data were analyzed using Stata 14.1 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX, USA: StataCorp LP).

Ethics approval for the RCT was obtained from the Health Research Ethics Board–Health Panel at the University of Alberta (study ID number: Pro00054360). This study is registered at ClinicalTrials.gov (ID: NCT02711644).

Figure 1. A screenshot that shows the set-up of a paid Facebook advertisement.

Figure 2. A screenshot that shows the paid Facebook advertisements on a mobile phone.

Results

Recruitment

A total of 11 women (11/126, 8.7%) did not indicate how they heard about the RCT and were excluded from further analyses. Of the 115 women included in analysis, 39.1% (n=45) of the women who contacted study staff heard about the study through Facebook, whereas 60.9% (n=70) of them heard about it through traditional approaches (Figure 3). Paid Facebook advertisements were received by 44,439 people on their Facebook newsfeed and 1001 Facebook users advanced to the study website, resulting in a click-through rate of 2.3% to the study website. Because of the nature of traditional approaches, the reach of recruitment efforts is not known.

During the 215 days (~7 months) that the traditional approaches were used, the average rate of interest was 0.3 (0.2) women/day. The local television news report resulted in 18 women contacting the study and was the most successful component of the traditional approach. When the local television news report is excluded, the remaining traditional approaches had an average interest rate of 0.2 (0.1) women/day. In comparison, the 26 days of Facebook advertisements resulted in an average interest rate of 2.8 (1.7) women/day. The traditional approaches alone (ie, before launching Facebook advertisements) had an overall interest rate of 8.7 women/month. Adding Facebook advertisements to traditional recruitment approaches increased the overall interest rate to 29.7 women/month. Of note, the Facebook advertisements ran for 26 nonconsecutive days to avoid advertisement fatigue. The maximum number of consecutive days the advertisement was displayed was 8. Advertisements were stopped on day 8 as the performance statistics dropped on days 6-8.

The 45 women who contacted study staff in response to the Facebook advertisement resulted in 40 women who were screened and 25 who were eligible and agreed to participate in the study (55.6% of interested). This resulted in a recruitment rate of 0.96 eligible participants/day. Of the 70 women who expressed interest through traditional approaches, 64 were screened and 45 were eligible and agreed to participate (64.2% of interested), resulting in a recruitment rate of 0.21 eligible participants/day. The calculated (hypothetical) amount of time needed to recruit 70 women using only traditional approaches is 334 days and could be shortened to 73 days using Facebook advertisements.

Facebook advertisements cost Cad $506.91, with a cost of Cad $0.28 per click. Other forms of engagement with the Facebook advertisements included 55 likes, 24 comments, and 28 shares. Most of the comments consisted of the names of Facebook friends who would be notified of the tag. A few comments required response from the study team as they asked questions regarding the eligibility criteria. Facebook advertisements had a cost per eligible participant of Cad $20.28. In comparison, the traditional approaches cost Cad $1087 for all methods combined (ie, printing, media advertisements, mileage to deliver brochures to different venues, participation at mommy and baby fairs). Traditional approaches cost approximately Cad $24.15 per eligible participant.

Participant Characteristics Relative to Their Method of Recruitment

Women recruited to the randomized controlled trial using traditional approaches and paid Facebook advertisements (n=115). The traditional approaches include methods listed in addition to other methods that consist of online classified advertisements, doctor referrals, blog posts, email newsletters, newspaper advertisements, mommy fairs, and online mom connection groups.

Figure 3. Proportion of interested women for the randomized control trial using Traditional approaches and paid Facebook Advertisements (n=115). The Traditional Approaches include methods listed in addition to 'Other' methods that consist of: Online classified advertisements, Doctor referrals, blog posts, email newsletters, newspaper advertisements, mommy markets and online mom connection groups.
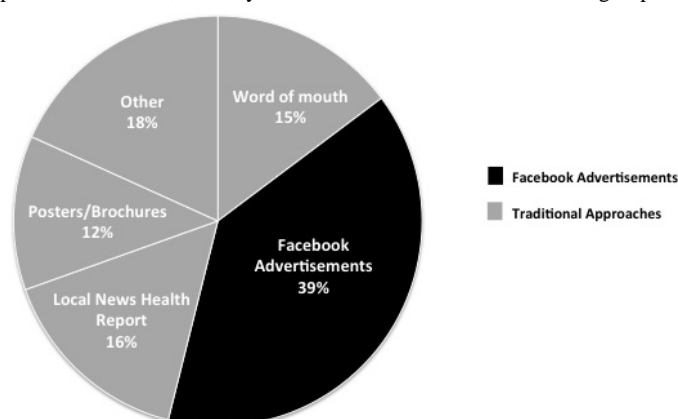
Table 1. Characteristics of women recruited using Facebook advertisements versus traditional approaches.

Characteristic	Recruitment method		P value
	Facebook advertisements (n=25)	Traditional approaches (n=45)	
Age, years, mean (SD)	35 (4.6)	34 (4.5)	.42
Gestational age, weeks, mean (SD)	12.6 (3.7)	14.7 (3.8)	.03
Prepregnancy BMI^a, kg/m^2, mean (SD)	26.4 (5.9)	24.8 (4.9)	.25
Birthplace, n (%)			
Born in Canada	22 (88)	36 (80)	.40
Not born in Canada	3 (12)	9 (20)	
Marital status, n (%)			
Single	1 (4)	2 (4)	.92
Married or common-law	24 (96)	43 (96)	
Education, n (%)			
Less than bachelor's degree	6 (24)	17 (38)	.24
Bachelor's degree or higher	19 (76)	28 (62)	
Household income (Cad $), n (%)			
< $70,000	2 (8)	8 (18)	.26
> $70,000	23 (92)	37 (82)	
Employment hours, n (%)			
Part time	6 (24)	11 (24)	.97
Full time	19 (76)	34 (76)	
Parity, n (%)			
One	7 (28)	18 (40)	.17
Two or more	3 (12)	2 (4)	

^{a}BMI: body mass index.

Discussion

Principal Findings

This study highlights the improved time efficiency achieved by coupling Facebook advertisements with traditional approaches for recruitment of pregnant women to a research study. After a 3-month period of traditional recruitment, we decided to test Facebook recruiting. Along with traditional recruiting, we used Facebook for 26 nonconsecutive days. Had we relied solely on traditional methods of finding interested women, it would have taken close to 1 year to reach our recruitment target, and this was shortened to less than 6 months by adding Facebook advertisements as a recruitment method. This improvement in recruitment rates of pregnant women is similar to that reported by a study recruiting women before conception. Shere et al found that social media–based recruitment resulted in a 12-fold higher rate of recruits per month [3]. It was suggested that the improved recruitment rates could reflect the fact that these types of social media are "active" because the platforms find women based on their previous Web searches related to the research topic, whereas traditional recruitment approaches may be considered "passive" because women likely come across the research opportunity by chance [3]. Our study adds to the

evidence indicating that social media holds promise for informing this population about research and recruiting them to participate. Mass media as a whole shows value in recruiting for prenatal research studies, as more than half (54.8%; 63/115) of the women who expressed interest in the study became aware through Facebook advertisements or the local news station health report. Ultimately, social media and other emerging means of mass communication hold promise as means to complement traditional strategies used for recruiting participants because they can reach a large number of people in a short amount of time.

On the surface, the absolute costs of the 2 recruitment methods used in this study were comparable (Cad $24.15/eligible participant for traditional approaches and $20.28/eligible participant for Facebook advertisements). Neither of these costs included those incurred by study staff. Staff or trainee time, mileage, and other related costs can be difficult to measure. The time required to generate and post Facebook advertisements is typically less than that needed to pick up and distribute posters and brochures to multiple sites throughout a large city. One of our traditional approaches, the television health report, was donated "in kind" and involved a local news anchor along with a 1-person film crew. This would have been very costly had the

study not been seen as a valuable news item. The free television health report is considered a traditional approach to recruitment. However, if its contribution to the traditional recruitment rate is removed, the cost per eligible participant for traditional approaches increases to Cad $33.97. Future studies should track resource investments, such as staff time, more thoroughly to better understand the relative savings or costs of social media compared with traditional recruitment approaches.

To our knowledge, no additional studies have been published within the last 2 years examining paid Facebook advertisements as a recruitment tool for prenatal studies. Changes in reach, access, and usage of Facebook over the 5 years that these studies and our study span make it difficult to fairly compare results between studies. Our Facebook click-through ratio was better than the 0.08% reported in the study by Arcia [15] but less than the 5.3% on the most successful newsfeed advertisement reported by Harris et al [14]. Arcia was recruiting nulliparous women at less than 20 weeks' gestation for a Web-based survey [15], whereas Harris et al aimed to recruit young women for completion of a questionnaire on contraception methods [14]. However, because both of these studies recruited between 2011 and 2013, the 2- to 5-year difference makes it difficult to compare them with our study in a fair way. It is possible that changes in the number or characteristics of people who regularly use Facebook along with changes in the features that Facebook provides to paid advertisers could contribute to the observed differences in click-through ratios. Social media certainly has excellent potential to aid recruitment efforts, although the magnitude of change over time remains difficult to quantify.

Our cost per advertisement click was slightly less than those reported by others. In several other studies, cost/click ranged from Cad $0.39/click [20] to Cad $0.45/click [12] and Cad $0.63/click [15]. Differences in the cost per click may vary for different demographic groups. Unique aspects of the target population need to be considered when formulating Facebook advertisements and when choosing topics of interest that help to identify and target these advertisements. Cost per click may continue to vary as Facebook and other social media platforms are modified, potential participants gain experience with using these platforms, or the platforms gain or lose active users. Relative to Facebook and other platforms that promote information sharing, one advertisement may reach more people than originally targeted [21] through features such as a "like, share, or comment." This can positively result in a more extensive social network of people being notified of a study than what was originally selected [22].

This study suggests that Facebook is a promising platform for reaching and recruiting pregnant women to a research study. Facebook advertisements can be more targeted than many of the traditional approaches because it is possible to define specific demographic and geographic characteristics, and information can be directed to those who search terms on the Web that align with the "Interests" specified by the researcher. Specifying "Interests" related to pregnancy likely enhanced the effectiveness of our advertisements because pregnancy is often a life event that is shared on social media platforms, and Facebook can display the advertisement to users based on their recent activity on the Web. Our RCT is the first prenatal study

to describe the "Interests" used in their advertisements. This could be advantageous by reducing the time needed to identify interested participants for a research study. Because most women use the Web to seek health-related information related to pregnancy, Facebook also holds potential as a useful way to distribute public health messages by using "Interests" to target this population [21,23]. An important consideration for Facebook advertisements is striking the balance between advertisement specificity and reach [8]. Targeting an audience with certain characteristics captured by Facebook could allow an advertisement to be cost-effective to reach the target population, but there is a possibility that some individuals will be missed. With approximately 27% of pregnancies being unplanned in Canada, it is probable that Facebook will miss these potential recruits [24]. Individuals who do not use social media or do not have Web-based activity related to pregnancy would also fail to alert the algorithms and may be missed. Therefore, utilizing the combination of traditional recruitment approaches and social media–based approaches is ideal to avoid this selection bias [25].

Previous studies have noted a concern that recruitment through Web-based methods may result in a nonrepresentative population [26]. We found that both approaches recruited women with similar demographic characteristics. Although not significant, a trend may exist of higher education and income within the Facebook group. Our study may not have been powered to detect difference in these populations. Two Australian studies also recruited a fairly nationally representative sample of females within the ages of 18-25 years using Facebook advertisements [11,14]. In our study, the one difference between women recruited through Facebook and women recruited through traditional approaches was that those recruited through Facebook were at an earlier gestational age. This was a promising finding because it is often a challenge to recruit women early in pregnancy [27,28]. Multiple investigators have noted the need for this in prenatal research [25,29-31], and our findings suggest that Facebook advertisements could prove helpful in this regard. Similarly, Richardson et al [29] found that their Internet-based advertisements recruited women significantly earlier in pregnancy compared with other methods.

Limitations

There are several limitations to this study. Our research team had little experience with Facebook advertisements before starting the study. The advertisements might have been more impactful or cost-effective if those formulating them had more experience developing, testing, and monitoring social media platforms and related statistics. Another limitation was that we could not determine whether women saw the advertisement on their personal Facebook page or if it came to them through a friend who saw it on Facebook. In our study, "word of mouth" was labelled as a traditional approach; however, the original study awareness may have originated from Facebook advertisements. This would underestimate the effectiveness of Facebook advertisements as a recruitment tool.

Conclusions

With ever-changing technology, researchers must stay current and utilize innovative approaches to find interested study

participants. The ease of placing an advertisement on Facebook, the comparable cost per participant recruited, and the dramatically improved recruitment rates when Facebook advertisements were added to traditional approaches highlight the importance of combining novel and traditional recruitment techniques to efficiently recruit women to pregnancy-related research studies, even in geographic areas where recruitment is difficult [1]. Future research should identify the best ways to target pregnant women using Facebook advertisements and other forms of social media to capture a broader range of the resources needed and costs associated with different approaches to recruiting these women.

Conflicts of Interest

None declared.

References

1. Manca DP, O'Beirne M, Lightbody T, Johnston DW, Dymianiw D, Nastalska K, APrON study team. The most effective strategy for recruiting a pregnancy cohort: a tale of two cities. BMC Pregnancy Childbirth 2013;13:75 [FREE Full text] [doi: 10.1186/1471-2393-13-75] [Medline: 23521869]
2. Frew PM, Saint-Victor DS, Isaacs MB, Kim S, Swamy GK, Sheffield JS, et al. Recruitment and retention of pregnant women into clinical research trials: an overview of challenges, facilitators, and best practices. Clin Infect Dis 2014 Dec 15;59 Suppl 7:S400-S407 [FREE Full text] [doi: 10.1093/cid/ciu726] [Medline: 25425718]
3. Shere M, Zhao XY, Koren G. The role of social media in recruiting for clinical trials in pregnancy. PLoS One 2014;9(3):e92744 [FREE Full text] [doi: 10.1371/journal.pone.0092744] [Medline: 24671210]
4. Choi J, Lee J. Online Social Networks for Crowdsourced Multimedia-Involved Behavioral Testing: An Empirical Study. Front Psychol 2015;6:1991 [FREE Full text] [doi: 10.3389/fpsyg.2015.01991] [Medline: 26793137]
5. Facebook. 2015. Company Info: Stats URL: http://newsroom.fb.com/company-info/ [accessed 2016-07-25] [WebCite Cache ID 6jGqfWeSD]
6. Duggan M. Pew Research Center. 2015. Mobile Messaging and Social Media 2015 URL: http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/ [WebCite Cache ID 6jGqn5OWm]
7. Hampton K, Sessions GL, Marlow C, Rainie L. Pew Research Center.: Pew Research Center's Internet & American Life Project: Washington DC; 2012. Why Facebook users get more than they give URL: http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Facebook%20users_2.3.12.pdf [accessed 2016-07-25] [WebCite Cache ID 6jGqu124r]
8. 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]
9. Carlini BH, Safioti L, Rue TC, Miles L. Using Internet to recruit immigrants with language and culture barriers for tobacco and alcohol use screening: a study among Brazilians. J Immigr Minor Health 2015 Apr;17(2):553-560. [doi: 10.1007/s10903-013-9934-1] [Medline: 24563138]
10. Jones R, Lacroix LJ, Nolte K. "Is Your Man Stepping Out?" An Online Pilot Study to Evaluate Acceptability of a Guide-Enhanced HIV Prevention Soap Opera Video Series and Feasibility of Recruitment by Facebook Advertising. J Assoc Nurses AIDS Care 2015;26(4):368-386 [FREE Full text] [doi: 10.1016/j.jana.2015.01.004] [Medline: 26066692]
11. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
12. 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]
13. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]

14. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]

15. Arcia A. Facebook Advertisements for Inexpensive Participant Recruitment Among Women in Early Pregnancy. Health Educ Behav 2014 Jun;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]

16. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and Conduct of an Internet-Based Preconception Cohort Study in North America: Pregnancy Study Online. Paediatr Perinat Epidemiol 2015 Jul;29(4):360-371 [FREE Full text] [doi: 10.1111/ppe.12201] [Medline: 26111445]

17. ENRICH.: University of Alberta; 2016. URL: http://enrich.ales.ualberta.ca/ [accessed 2016-07-25] [WebCite Cache ID 6jGqSocFs]

18. AdsEspresso. 2016. URL: https://adespresso.com [accessed 2016-07-25] [WebCite Cache ID 6jGqyPqxI]

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

20. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]

21. Betsch C. Social media targeting of health messages. A promising approach for research and practice. Hum Vaccin Immunother 2014;10(9):2636-2637. [doi: 10.4161/hv.32234] [Medline: 25483481]

22. Pedersen ER, Kurz J. Using Facebook for Health-related Research Study Recruitment and Program Delivery. Curr Opin Psychol 2016 May;9:38-43. [doi: 10.1016/j.copsyc.2015.09.011] [Medline: 26726313]

23. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]

24. Black A, Yang Q, Wu WS, Lalonde AB, Guilbert E, Fisher W. Contraceptive use among Canadian women of reproductive age: results of a national survey. J Obstet Gynaecol Can 2009 Jul;31(7):627-640. [Medline: 19761636]

25. Maghera A, Kahlke P, Lau A, Zeng Y, Hoskins C, Corbett T, et al. You are how you recruit: a cohort and randomized controlled trial of recruitment strategies. BMC Med Res Methodol 2014;14:111 [FREE Full text] [doi: 10.1186/1471-2288-14-111] [Medline: 25260762]

26. Maloni JA, Przeworski A, Damato EG. Web recruitment and internet use and preferences reported by women with postpartum depression after pregnancy complications. Arch Psychiatr Nurs 2013 Apr;27(2):90-95. [doi: 10.1016/j.apnu.2012.12.001] [Medline: 23540519]

27. Webster GM, Teschke K, Janssen PA. Recruitment of healthy first-trimester pregnant women: lessons from the Chemicals, Health & Pregnancy study (CHirP). Matern Child Health J 2012 Feb;16(2):430-438. [doi: 10.1007/s10995-010-0739-8] [Medline: 21210200]

28. Coleman-Phox K, Laraia BA, Adler N, Vieten C, Thomas M, Epel E. Recruitment and retention of pregnant women for a behavioral intervention: lessons from the maternal adiposity, metabolism, and stress (MAMAS) study. Prev Chronic Dis 2013;10:1-7 [FREE Full text] [doi: 10.5888/pcd10.120096] [Medline: 23469765]

29. Richardson JL, Stephens S, Thomas SH, Jamry-Dziurla A, de Jong-van den Berg L, Zetstra-van der Woude P, et al. An International Study of the Ability and Cost-Effectiveness of Advertising Methods to Facilitate Study Participant Self-Enrolment Into a Pilot Pharmacovigilance Study During Early Pregnancy. JMIR Public Health Surveill 2016;2(1):e13 [FREE Full text] [doi: 10.2196/publichealth.5366] [Medline: 27227148]

30. Jackson RA, Stotland NE, Caughey AB, Gerbert B. Improving diet and exercise in pregnancy with Video Doctor counseling: a randomized trial. Patient Educ Couns 2011 May;83(2):203-209. [doi: 10.1016/j.pec.2010.05.019] [Medline: 21459255]

31. Mottola MF, Giroux I, Gratton R, Hammond J, Hanley A, Harris S, et al. Nutrition and exercise prevent excess weight gain in overweight pregnant women. Med Sci Sports Exerc 2010 Feb;42(2):265-272 [FREE Full text] [doi: 10.1249/MSS.0b013e3181b5419a] [Medline: 20083959]

Abbreviations

RCT: randomized controlled trial

Original Paper

Professional Use of Social Media by Pharmacists: A Qualitative Study

Arcelio Benetoli[1,2], MSc; Timothy Frank Chen[1], BPharm, DHP, PhD; Marion Schaefer[3], PhD; Betty B Chaar[1], BPharm, PhD; Parisa Aslani[1], BPharm (Hons), MSc, PhD

[1]Faculty of Pharmacy, The University of Sydney, Sydney, Australia

[2]Department of Pharmaceutical Sciences, State University of Ponta Grossa, Parana, Brazil

[3]Charité University Medicine Berlin, Berlin, Germany

Corresponding Author:
Parisa Aslani, BPharm (Hons), MSc, PhD
Faculty of Pharmacy, The University of Sydney
Pharmacy and Bank Building A15, Science Road, Broadway
Sydney, 2006
Australia
Phone: 61 290366541
Fax: 61 293514191
Email: parisa.aslani@sydney.edu.au

Abstract

Background: Social media is frequently used by consumers and health care professionals; however, our knowledge about its use in a professional capacity by pharmacists is limited.

Objective: Our aim was to investigate the professional use of social media by pharmacists.

Methods: In-depth semistructured interviews were conducted with practicing pharmacists (N=31) from nine countries. Interviews were recorded, transcribed verbatim, and thematically analyzed.

Results: Wikipedia, YouTube, and Facebook were the main social media platforms used. Professional use of social media included networking with peers, discussion of health and professional topics, accessing and sharing health and professional information, job searching, and professional promotion. Wikipedia was the participants' first choice when seeking information about unfamiliar topics, or topics that were difficult to search for. Very few pharmacy-related contributions to Wikipedia were reported. YouTube, a video-sharing platform, was used for self-education. University lectures, "how-to" footage, and professionally made videos were commonly watched. No professional contribution was made to YouTube. Facebook, a general social networking site, was used for professional networking, promotion of achievements, and job advertisements. It also afforded engagement in professional discussions and information sharing among peers.

Conclusions: Participants used social media in a professional capacity, specifically for accessing and sharing health and professional information among peers. Pharmacists, as medicines experts, should take a leading role in contributing to health information dissemination in these user-friendly virtual environments, to reach not only other health care professionals but also health consumers.

Introduction

Social media (SM) encompasses a wide range of websites whose content is created by users [1]. While the Internet has increased public access to all kinds of information, its evolution to Web 2.0 has provided a more participatory environment where users not only access, but also create, edit, and share content online in several formats (eg, text, picture, audio, and video). SM has transformed communication, including health care communication [2]. Consequently, SM is increasingly shaping the way health care professionals work and provide their services. The Internet can be considered as an important tool for pharmacists and other health care professionals for their daily professional activities, particularly as a medium to access

XSL•FO
RenderX

health information [3]. It is believed that SM has further impacted the way health care professionals deliver their services by increasing their knowledge, efficiency, communication with patients, marketing, and communication with colleagues [4]. However, the influence of social media use by pharmacists in a professional capacity has not been fully investigated, with limited literature on the use of social media by pharmacists [5].

Studies conducted to date have focused on specific SM platforms such as Wikipedia [6], blogs [7,8], social networking sites (SNS) [9,10], and Twitter [11], with only one study addressing SM more broadly, (ie, including several different types of platforms) [12]. One study has shown that in 2009, 35% of US pharmacists were using Wikipedia for general purposes and 10% for drug information [6], while another study found that 72% of pharmacists surveyed when attending a regional pharmacy conference in the United States used Wikipedia [12]. Neither study, however, investigated how Wikipedia supported pharmacists' work or their perceptions of its usefulness. In terms of blogging, 44 [8] and 136 [7] pharmacists' blogs were identified and analyzed in two separate studies published in late 2010. These blogs provided information about news and current events in health care, pharmacological discussions, and pharmacists' personal views on their professional and private lives. To date, pharmacists' professional use of SNS appears to be incipient, with personal and social intentions being the main motivations for use [9,10,12]. For instance, 90% of pharmacy preceptors in a US survey reported using Facebook primarily for personal and social reasons [9], and 90% of pharmacists' tweets were predominantly or exclusively for personal purposes [11]. Pharmacists' Facebook use ranged from 50% to 67% [9,10,12]. Its use was more common among younger practitioners (<29 years old), decreasing as age increased [10]. At the time of the study, the Twitter user rate among pharmacists was very low [9,11], with the most recent study revealing that less than 1% of US pharmacists were active on Twitter [11].

Although some research about the use of SM platforms by pharmacists has been conducted, these studies have either used a quantitative survey [6,9,10,12,13] or a direct observation of SM platform approach [7,8,11] with little in-depth exploration of SM use. Therefore, the current study aimed to build on previous research by using a qualitative approach to investigate how pharmacists perceive and use SM professionally.

Methods

Qualitative Methodology

This was an exploratory qualitative study that employed in-depth, semistructured interviews to explore pharmacists' use, opinions, and perceptions of SM. A qualitative approach was considered most suitable given that our aim was to elicit rich and detailed aspects of pharmacists' professional use of SM and explore their perceptions and opinions about it. Another advantage is that qualitative research allows the investigators to explore the viewpoints of individuals in detail, with a small number of participants [14]. Ethics approval was obtained from the University of Sydney Ethics Committee (Project No. 2013/635, approval date: 14/08/2013) prior to commencement of the study.

Participants and Recruitment

This study sought a global perspective from pharmacists' use of SM, so a purposive sampling was used to recruit pharmacists from a range of countries. Initially pharmacists known to the research team were invited, and thereafter participants were recruited using a snowballing technique. This strategy allowed the research team to take advantage of its international professional network (the research team comprised individuals from different countries: Australia, Brazil, and Germany) and identify active SM users within the pharmacy profession with a range of professional roles. Pharmacists from different professional settings and countries took part in this study. All participants received a participant information statement and verbal explanation about the project's aim, the research team and the interviewer, and their participation prior to the interviews. Consent was acquired by having participant sign a standardized consent form.

Data Collection

Semistructured interviews were conducted between November 2013 and December 2014. An interview guide (Multimedia Appendix 1) based on the study aim and objectives was developed and piloted with 5 participants to assess face and content validity. The piloted data were not included in the analysis. The interview guide consisted of topics related to participant understanding of SM, SM usage, and perception of SM use for health care and in the pharmacy profession. All questions were open-ended to allow relevant topics and themes to emerge without constraint. Whenever needed, follow-up questions were asked in order to clarify or expand participants' comments. The interviews were conducted either face-to-face, by telephone, or Skype (voice call or video call), and audio-recorded with the consent of participants. Notes were taken to assist the formulation of prompt questions and support data familiarization for data analysis. The interviews were conducted in dedicated premises at the Faculty of Pharmacy, University of Sydney; they lasted from 30 to 130 minutes and were all conducted by AB, a male pharmacist and PhD candidate trained in qualitative research. Interviews were conducted until saturation was reached.

Data Analysis

All interview recordings were de-identified and transcribed verbatim. One participant asked to see their transcript, but no revision or correction was made. A thematic analysis with an inductive approach was employed to identify themes within the interviews' transcripts. Thematic analysis is not aligned with a particular epistemological, philosophical, or theoretical approach and is a flexible tool to generate themes in qualitative analysis [15]. Besides being a robust and sophisticated research tool, thematic analysis focuses and presents the data in a way that is readily accessible to those who are not part of academic communities [16], which benefits the research findings' dissemination among pharmacy practitioners.

Repeated reading of transcriptions allowed familiarity with the data and knowledge of each interview's content depth and breadth. After immersion within the data, transcriptions were coded line-by-line and collated within each code with the help

XSL·FO

RenderX

of the software NVivo 10 (QSR International). The limited literature available on the topic [5] associated with the adoption of an inductive approach [17] ensured a data-driven analysis. Therefore, the coding process was open, not restricted by theoretical assumptions from the research team. The coding process was dynamic, iterative, and evolved throughout the analysis. Codes with a repeated pattern across the data (ie, codes with similar or nearly similar meanings) were grouped into subthemes and later assembled into overarching themes. Themes were carefully named according to their overall content. The first three interviews were separately coded by 2 researchers,

and the coding labels and identified themes and subthemes were thoroughly discussed and agreed upon. Throughout the analysis, the coding process, including its grouping into themes, was discussed with a senior member of the research team (PA).

Results

In total, 31 pharmacists from nine countries were interviewed (Table 1). Only one participant approached via email did not take part. Most participants were practicing in community pharmacy and academia (in the field of pharmacy practice).

Table 1. Participant demographics.

Demographics	n
Gender	
Female	16
Male	15
Years of practice	
Range	2-37
Median (SD)	13.3 (9.83)
Work setting	
Community	14
Academia	11
Community & academia	2
Pharmacy/health organization/industry	4
Country	
Australia	17
New Zealand	5
United States	2
Brazil	2
Germany	1
Nigeria	1
Thailand	1
Philippines	1
United Kingdom	1
Type of interview	
Face-to-face	10
Telephone	8
Skype	13

The first aspect explored in the interviews was pharmacists' understanding of SM, and subsequently the interview focus shifted to actual use of these platforms with a focus on professional aspects.

Pharmacists' Knowledge and Understanding of Social Media

Most participants considered SM to be an online (Internet) community venue where information is accessed and shared with the aim of fostering communication. Figure 1 illustrates

the concepts that participants associated with SM and portrays a broad description of their understanding. The word cloud was created using the key words (nouns and verbs) participants used when describing SM.

SM was considered as a new way to communicate and keep in touch with a large group of people publicly. Participants believed that this interactivity created a connection among users and allowed people to keep a virtual network of contacts, including a professional network.

XSL•FO
RenderX

As a source of information, SM was perceived to have the advantage that the users had a high level of control, allowing them to choose the content to read, listen, and watch. Some participants expressed that the information shared on SM platforms was created by the users themselves. However, only a few participants could explain the user-generated content aspect of SM: "It's most like a forum for user-generated content. It's the opposite of maybe company-generated websites; you've got the people themselves generating content and sharing it, communicating about it" [Participant 21].

SM interactivity was a feature highlighted by several participants: "You interact with those who posted the information…you can question. It's a quick way to get in touch, with questions and answer. You hadn't such a quick and effective interaction before Facebook" [P24].

Participants were also asked to provide SM examples. This permitted further evaluation of participants' understanding of SM, through a comparison of the examples provided and the definition of SM given. Facebook was cited by all participants as an example and most times was the first example that "sprang

to mind." Popular SM platforms without SNS features (eg, public profile and list of contacts) like YouTube and Wikipedia were much less frequently cited. Twitter, although a microblog, possesses some SNS features and was commonly mentioned. Other common SM examples were LinkedIn, Instagram, Google Plus, and blogs. Virtual worlds, like Second Life, were rarely mentioned.

Would you consider Research Gate (a professional social networking site for academics, researchers, and scientists) social media?" [P10] and "I have a doubt. Is Moodle (a free software whose purpose is to serve as a learning platform via creation of personalized learning environments [18]) considered a social media platform?" [P23]

After obtaining participants' definitions of SM, including examples, the SM operational definition (ie, a group of interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content employing mobile and Web-based technologies [19]) was provided to ensure that everyone had the same understanding before continuing with the interviews.

Figure 1. Word cloud of participants' understanding of social media.

Pharmacists' Use of Social Media Platforms

This theme details how pharmacists used or perceived SM use by their peers in a professional capacity. All participants reported using SM in a professional capacity, though to varying degrees. The most common platforms were Facebook, Wikipedia, and YouTube. Pharmacists used SM to network with fellow pharmacists, to access and share professional news and health- and medicine-related information, to discuss relevant topics, to exert political influence within the pharmacy profession, to promote their careers, and to find or assist peers to get employment. SM was also used to promote pharmacy stores and in the education of pharmacy students.

Use of Social Networking Sites

The use of SNS ranged from personal to professional. Facebook was the most used SNS platform. Although the majority of participants initially stated that their use of SM, particularly SNS, had a personal and social aim, many professional activities were identified during the interviews. In fact, after articulating

their use of SM throughout the interview, some participants realized that their initial statement about SM use was not completely accurate. Some online activities were later categorized as professional ones. Interestingly, some participants were surprised to realize they had an extensive professional use of social media, even though not directly related to patient care.

Most participants reported using only one SNS regularly, Facebook, despite having accounts or profiles on other SNS (eg, LinkedIn, Twitter). Facebook was preferred due to its popularity (many users), simplicity (easy to use), and versatility (great range of features available). Facebook was regarded as a convenient venue, a "one stop shop" [P13], allowing the convergence or integration of different SM platforms.

Some accessed Facebook multiple times a day, using mobile devices (eg, smartphones, tablets). Differently from academia and research, SNS use was commonly restricted during work time in community pharmacies. A few participants reported that the dispensing computers had the Facebook webpage blocked,

or even the use of their personal mobile phones was totally restricted: "In my pharmacy, we have a computer. We have guidelines for the usage of computer and Facebook is blocked in the pharmacy that I work at" [P17] and "So even your mobile phone shouldn't be on you. Your mobile phone is locked in the staffroom and you don't have access to your mobile phone while you're working" [P2].

Personal motivations, namely communication and interaction with family and friends, were the main driving force for participants to join Facebook. Facebook was regarded as an inexpensive, quick, and easy way to connect with friends and family, including those overseas. As used primarily for personal reasons, participants were much more engaged with Facebook than with other dedicated professional platforms. However, eventually they started using Facebook professionally. Networking with colleagues on SNS was perceived to be an important professional interaction allowing some participants to break the isolation they felt, especially those working in regional or rural areas.

Social Media for Learning and Information Sharing

SM platforms were used as professional learning channels for pharmacists. Facebook served as a convenient vehicle to easily access professional and health-related information from pharmacy organizations and peers. Following pharmacy organizations' Facebook pages to receive professional updates and information on their news feed was commonly reported by participants: "I find it's the quickest way of knowing what's going on rather than me going to their [organizations'] website or finding out by email and taking the time to read the email. I think it is much quicker" [P10].

Participants were also active in sharing information. A variety of sources such as health authorities, health providers, pharmacy organizations, and traditional media (eg, newspaper articles and television reports) were used. Pharmacists described being engaged in posting information, often providing the primary source of the information: "It [the news] can be from a TV channel website, newspaper, magazine… but I always provide the link to the source of information" [P24] and "I believe we are able to understand the primary source of information, so I find information from those sources and I deliver that to my [group] members" [P31].

With regard to Twitter usage, participants either did not use it, or they heavily used it for professional purposes. Those who were tweeting did so in order to receive, provide, and share professional news, raise awareness about pharmacy issues and pharmacy services development, and/or promote their organization or institution. Twitter was also used as a search engine for certain trending topics, particularly for more abstract keywords/concepts or terms not well-established, using the hash tag and the keyword:

> *If I'm interested in a particular idea or topic I'll search a #tag for whatever I'm interested in…I might be able to search for something a bit more abstract…I haven't quite thought about why I would use specifically Twitter over Google but I think if the word that I'm looking, the key words are a bit more*

> *abstract, Google doesn't really pick them up necessarily. Twitter might pick up those words better because they're so generic. Google doesn't know what exactly I'm looking for, but the way other people use that word like when they make a tweet, it's very specific* [P5]

Twitter was also the preferred platform during conferences, to increase conference visibility, reach non-attendees, and increase delegate participation.

Wikipedia was the only wiki mentioned and reported as a commonly sought source of information, frequently accessed during working hours. Wikipedia was regarded as a convenient initial resource for searching for information because it provided background knowledge for further reading and learning: "If I'm looking for something I know nothing about, that can be a really good starting point" [P17].

Additionally, Wikipedia was also used when well-established common references failed to provide the information needed, and then Wikipedia references were commonly scrutinized in order to expand the information search.

It appeared that participants had their own rules about when it would be appropriate to use Wikipedia in a professional capacity. For instance, it was not used as a reference source to guide their professional actions. However, it was used when a general knowledge about a topic or quick access to information was required. Pharmacy-related information retrieved ranged from brand names, to pharmacological effects, to therapeutic classification and uses of a drug or medicine, to unfamiliar health topics, such as unfamiliar medicines or rare disease states.

Several characteristics of Wikipedia made it a useful SM tool for the participants. These included the speed with which participants could obtain information, knowing the date of the last update, the fact that entries were referenced, its not-for-profit nature, and usability. Despite the frequent use of Wikipedia by most participants, it was not regarded highly by many. Participants generally provided some rationale for not considering Wikipedia as a reputable source of information, such as lack of credentials and anonymity of Wikipedia contributors, and inaccuracy of some information. Therefore, the academic participants were not eager to recommend Wikipedia as a resource/reference for student use. Only 2 participants were engaged in editing Wikipedia entries, though just one contributed to medicine- and pharmacy-related information.

Four participants had their own blogs. Each of them had different purposes and audiences. They were used for pharmacy education (detailed below), entertainment (to publish drawings and cartoons created by the blogger), rural and aboriginal health care, and pharmacy profession in general. All those actively blogging emphasized that blog traffic generally depended on links to them being spread on SNS: "When I post on a blog, I promote that onto Twitter and Facebook" [P26] and:

> *Depending on the news significance, after I publish that on the blog I will provide a link to it on Facebook…most access to the blog comes from Facebook...when I publish a post on Facebook with*

a link to the blog, that news is much more accessed than other news not published on my Facebook account [P24]

Though very few participants mentioned accessing blogs, the pharmacist hosting the general pharmacy blog reported more than 850,000 visits during 2 years of the existence of the blog, with an average daily audience of 1700 in 2014. Topics published in the pharmacy blog included job opportunities (both private and public sectors), pharmacy and pathology laboratories for sale, pharmacy fixtures and fittings for sale, professional events (eg, talks, workshops, and conferences), pharmacy legislation changes and updates, product recalls, new drug and treatment studies, and disease outbreaks (eg, Ebola). The most popular topics were those related to daily pharmacists' working lives (eg, legislation affecting work procedures, community recognition, and level of trust in the profession). However, the most popular topic so far was a product recall in that country, with more than 23,000 visits.

The most used video sharing platform was YouTube. The video sharing site Vimeo was mentioned by a few participants, but its use was not common. Even though entertainment was the major motivation to access YouTube (eg, watching music clips), some professional uses were reported, such as obtaining content on technical information. Some participants expressed a preference for watching a video instead of reading, especially when dealing with practical procedures and "how-to-do things": "Every now and again if I forget how to use a technique and need help, go back to YouTube, usually it's from a professional organization, but it's a YouTube link anyway" [P31]. Professionally made videos and lectures from reputable universities and educational institutions were highly used as self-learning material.

Some participants used YouTube frequently and considered it a good repository, while others were not regular users and did not access any professional-related content. None of the participants uploaded professionally related videos on social media platforms.

Professional Discussions

SM provided an alternative venue for professional discussion and peer collaboration. Discussions on SNS, particularly Facebook and Twitter, were preferred over pharmacy blogs. Pharmacy- or pharmacist-themed private groups on SNS were very common for this purpose. The "Facebook groups" function allows users to create or join groups based on common interests [20]. These professional pharmacy groups were set up either by pharmacy organizations, such as the Pharmaceutical Society of Australia, or individual pharmacists. Membership in these groups was restricted to pharmacists, and the groups functioned as forums where discussion about regulations, provision of services, and professional ethical issues took place. The ability to provide and obtain information quickly was an important driving force in these groups. However, not all professional Facebook groups were constantly active, and activity could range from daily to quarterly. Moreover, within each group, individual members' participation varied from a predominantly observational role to active participation.

Those who contributed more to the pharmacy Facebook forums were perceived to be the ones with stronger opinions. Thus, the moderation task was regarded as crucial.

Pharmacy Jobs, Career, and Social Media

Participants believed that SM had an impact on access to the pharmacy job market by increasing visibility of both job opportunities available and prospective employees' resumes. In terms of recruitment, although no participant mentioned using SM for their own benefit, it (namely pharmacy-themed closed Facebook groups and pharmacy blogs) was widely recognized as a medium for both employers and employees within the profession: "Job advertisements get utilized a lot, so on a local [Facebook] pharmacist's page a lot of people will post up 'job needed' or whether they are looking for a job or whether they are trying to fill that job, so that's quite common" [P31].

Although several participants had LinkedIn profiles to convey their professional qualifications and achievements, the vast majority of participants were not actively engaged with it. Its main purpose, as perceived by participants, was to increase professional visibility among peers and potential employers and be used for job searching purposes. Professional achievements were also posted on other social media outlets, particularly Facebook: "I know that there are colleagues that use Facebook to disseminate their research data so if a publication comes up they'll put on their status update" [P15].

Additionally, some pharmacy practice academics reported having an account on Research Gate, a kind of "Facebook for science" [21]. The reported advantage of this social network was the formation of a clear professional network, the opportunity to promote one's own research (eg, uploading articles), and establishing a network of researchers within their field. Pharmacy academics were aware that their research papers could be accessed by other researchers without access to expensive databases. However, none expressed a great level of activity or enthusiasm about its use.

Social Media Use in Pharmacy Education

Social media also had its place in the delivery of pharmacy education. It was perceived to facilitate the learning experience in the current pharmacy student population that uses these platforms extensively and routinely for most aspects of their lives. Most pharmacy academic participants reported using SM for teaching purposes, though at different intensity and frequency. The teaching activities using SM were performed in class or designed to supplement in-class activities and served as an alternative method for the usual education management systems (eg, Moodle, Blackboard) used by pharmacy schools.

The major platforms used were Facebook and YouTube. It was recognized that information posted on Facebook would reach students faster because they perceived Facebook as the students' favorite way of accessing course information as they were always accessing their personal Facebook pages.

Some of the pharmacy academic participants could be identified as early adopters of technologies. They tended to provide more examples of SM use in their pharmacy teaching. For these early adopters, SM was seen as an integral part of teaching and they

tended to use different types of SNS platforms and groups for different teaching-related activities. Although the use of Facebook was perceived to be beneficial to supplement in-class activities and to improve both academic-student and student-student communication, its use within the classroom without teacher's guidance was not regarded advantageous since it disrupted student concentration, preventing them from paying due attention or engaging in the topic being taught and discussed.

YouTube was perceived as a good resource to better illustrate concepts during lectures, provide supplementary information, serve as alternative learning resources, and increase student understanding especially students who were more visual learners. Wiki platforms were also used, specifically to assign group tasks to students, enabling the educators to gain a more accurate understanding of individual student participation. This "virtual control" was regarded as one of the great advantages of wiki platforms for educational purposes.

Even though pharmacists were using SM websites for self-education and educating students, they were not using it to educate other health care professionals. One interesting exception was the provision of a blog as a reference on how to write prescriptions, which was regularly recommended by a participant to junior physicians in a hospital.

Other Uses of Social Media

Some of the academic participants had also used SM for research, specifically for recruiting participants or collaborators to research projects. SM was also used to collect data through posting questionnaires on Facebook.

As SM is a two-way platform, some participants used it as a medium for sparking discussions about the political issues within the profession and health sector, as well as voicing opinions about changes in professional activities. Consequently, Facebook also served as a venue to put pressure on decision makers and policy makers either for change or to demand more accountability from the profession's leaders:

> I think what's important is to provide an avenue where pharmacist can use it as an outlet for some of their I guess frustrations with policy decision makers...Will it lead to change? I think it just really depends on the numbers...I think it will also force them to respond in a manner…to justify the decisions that have been made simply because of the reaction, the outbursts on these social websites [P18]

While the major aim of this project was to explore how pharmacists used and perceived peers' use of SM from a practitioner's perspective, some additional relevant information on how SM is used by pharmacy stores was also identified. Many community pharmacies were reported to host a Facebook page, with a marketing aim. Most frequently, the Facebook page served as a supplementary or additional form of promotion about products already advertised (eg, in the pharmacies' catalogue), as well as services available. Pharmacy Facebook pages were also reportedly used as a channel to spread general health information to the community.

Discussion

Principal Findings

This paper provides the first insight into the broad professional use of social media by pharmacists. It was believed that SM usage in pharmacy ranged from only social and personal communication (eg, with friends and family) at one end of the spectrum, to professional communication only (eg, with colleagues and clients) at the other end of the spectrum [22]. However, our study has shown that the professional use of SM by pharmacists is common and very diverse. Although many participants initially believed that their use of SM was exclusively for personal reasons, it became apparent during interviews that all participants used social media in a professional capacity. Even those few participants without an SNS account commonly accessed popular social media websites like Wikipedia and YouTube for professional purposes. The lack of awareness of professional use of SM by the participants could be attributed to how social media was initially narrowly defined and perceived.

Peer communication was one of the most common professional activities on social media. Participants preferred to use general SNS, like Facebook and Twitter, instead of pharmacy-only SNS for professional networking (eg, PharmQD). Similarly the use of professional SNS (eg, LinkedIn) was far less common. As reported in a separate publication [23], most participants had a blended approach [24] when using SNS, which means they had both professional and personal activities taking place in the same SNS. Participants' initial motivation to set up a Facebook profile was social and personal in nature, and over time gradually evolved to have a mixed audience (both social and professional contacts) as pharmacy school friends became professional peers and colleagues became closer friends. Not surprisingly, professional online conversations and discussions then took place over a platform initially intended for personal reasons, like Facebook. Since community pharmacists commonly do not work alongside peers [25,26], this increased interaction among pharmacists provided by SNS would be a major advantage because it could diminish pharmacists' isolation helping them to better deal with technical, clinical, and ethical issues that arise at work. However, this increased interaction is not taking place yet during pharmacists' time in the community pharmacy as the use of these platforms was commonly restricted.

A natural development that followed the increased professional interaction afforded by SNS was the increased access and further spread of relevant professional information. In fact, one of the major professional activities performed by pharmacists on SM was to access information. It is widely known that health care professionals frequently use the Internet as a working tool, especially to access information [3]. As the Internet has shifted to be more interactive [1], it would be expected to find pharmacists using SM as an information source for work purposes and as a tool to learn from a variety of sources, such as from organizations, peers, and patients. So obtaining health and medicines information via SM platforms represents a natural step in the communication technology evolution for pharmacists. The ubiquitous nature of SM platforms places them in a

prominent position as a medium for relevant pharmacy professional information dissemination, as demonstrated in this study. The increased access to information and the consequent knowledge expansion can lead not only to professional development, but also to better services and patient care.

As one of the major open sources on the Internet, Wikipedia was highly used by pharmacists during work time to access health and medicines information. As Wikipedia is one of the top results from medical queries in general search engines [27], it is not surprising to see it used commonly. However, most participants did not consider Wikipedia as an authoritative source and several of them voiced concerns about its accuracy and reliability. These concerns matched pharmacy literature restriction recommendations on the use of Wikipedia in both pharmacy education [28] and practice [29]. Nevertheless, it is important to emphasize that other studies have pointed out that Wikipedia entries have a high degree of quality compared to the prestigious Encyclopedia Britannica [30] and can be useful as a reference for health students' assignments [31].

The high use of Wikipedia found in this study is consistent with a high rate of Wikipedia users among pharmacists previously reported [12]. Wikipedia use among pharmacists might be on the rise since the first rate of Wikipedia use published in 2009 had indicated that only 35% of US pharmacists were using it [6]. Wikipedia was commonly used during work time for initial explorations of less known topics related to health and medicines. This is similar to how physicians have been reported using it to get an overview of unfamiliar topics [32]. It is also substantiated by a review that found Wikipedia was widely used as a reference source by health care professionals in general [33]. The Wikipedia editing process was well known by all participants in our study, in contrast with the 2009 US survey where only a third of Wikipedia users knew how it was edited [6].

User-friendliness and ease of access, as well as the fact that Wikipedia often appears in the top 10 results of searches in general search engines such as Google, are important aspects for consideration in developing approaches for conveying information to pharmacists in all areas of the profession [34]. Pharmacists' contribution to the "free encyclopedia" seems to be negligible, despite, as our study asserts, being highly accessed by pharmacists. Although there is a call for health care professionals, their societies, patient groups, and institutions to join the effort in improving Wikipedia's health-related entries [27], it is believed that the community who edits health-related entries is very small and is driven by an intrinsic set of values and beliefs [35]. The rationale underlying collaborative projects like wikis is that the joint effort of many actors leads to a better outcome than any actor could achieve individually [1]. Perhaps the best approach to improve medicine and pharmacy entries on Wikipedia would be via collaboration between academia, health organizations, practitioners, and consumers.

The increased access to up-to-date information provided by SM also led pharmacists to engage in another professional task: sharing information. SM made information sharing easier, reaching a bigger audience. Some very active participants saw the opportunity afforded by SM and launched their own platforms, mainly blogs and Facebook groups. It is important to emphasize that when sharing professional information, participants commonly provided links to original sources of information and tried to post the original source of the material. This significant aspect of information sharing behavior might be caused by the fact that our sample consisted of experienced pharmacists (mean 13.3 years of practice [SD 9.8]) and a higher level of education (many had a post-graduate degree or were getting research degrees).

The use of SM channels to spread health- and pharmacy-related knowledge and spark professional discussions should be considered an additional tool to traditional forms of professional interaction in pharmacy. The SM platforms can be complementary and even increase the effectiveness of these traditional ways. The use of Twitter in conferences is seen as example of how the traditional congregation of peers can have their discussions spread beyond the setting of the conference instantly and at the same time causing an impact on the participants' perceptions and level of interaction within the conference. This is congruent with the assertion that Twitter use during medical conferences was a medical education application [36]. However, the benefits of such application in pharmacy should not be overestimated. A study about Twitter use during a major US pharmacy conference found that Twitter was used by less than 2% of attendees [37]. It has also been reported that almost half of all tweets were made by a tiny group of participants (0.125%) and only a third of tweets were related to educational sessions. Nevertheless, the reach and impact was not verified.

Although our snowball approach led us to recruit key Twitter users, the use of this microblogging platform was less prominent than other SM applications among our participant cohort. This is congruent with findings from surveys among pharmacy preceptors, which revealed that less than 10% of respondents had a Twitter account and that their use was very limited (ie, majority accessed on a monthly bases) [9]. Similarly, a more recent direct observation study on Twitter indicated that less than 1% of practicing US pharmacists had a Twitter account [11]. Most participants who used Twitter in our study used it primarily for professional reasons. This is in contrast to an earlier observation study that evaluated tweet samples and showed that only 10% used Twitter exclusively or predominantly for professional reasons [11]. This cohort of Twitter users might be very similar to our Twitter participants who also used Twitter to both obtain and disseminate pharmacy-related news, recently published pharmacy articles, and other useful and relevant health-related information. Twitter is a suitable application for this purpose since microblogging is designed to offer real-time updates [19], and participants can access important and relevant information up-to-the minute.

Although the use of SNS for recruiting and hiring employees is a new process with limited research [38], this study has found that this use of SNS was perceived to be very common within the pharmacy sector, playing an important role in the pharmacy job market. This seems to be a very beneficial professional use of SNS by pharmacists. It expanded the reach of pharmacists searching for job opportunities. At the same time, it also favored recruiters, who could access a pool of candidates within a closer

XSL•FO

RenderX

network (eg, a local Facebook pharmacists' group). In other words, SNS amplified the "word of mouth" strategy commonly employed in the recruiting of community pharmacists.

The use of SM in pharmacy education described in this study was very similar to previously published approaches [5]. While the Internet, since its first years, facilitated teaching and learning, the need to master the standard mark-up language (HTML) and related concepts (eg, servers, file transfer protocol [FTP] of files, client-side plug-ins, Java applets for interactivity) served as barriers to its use since they required intricate programing [39]. The advent of SM has revolutionized the use of the Internet as an education tool as it has eliminated the interface barriers and information technology constraints for users to create and share content online.

Limitations

The findings presented should be viewed in light of certain limitations. First, as an exploratory qualitative study with a sample size of 31 pharmacists, the study findings are not generalizable to the pharmacist population. Second, the results may have been impacted by a self-selection bias, despite the fact that not all participants were active SM users. However, a wide range of views on the topic was provided, which could also be interpreted as a strength since participants with more

SM expertise can provide more information and insights. Third, as social media is constantly changing due to its dynamic nature, it is advisable to keep in mind that the results represent the situation only at the time of the study.

Conclusions

Although personal and social reasons generally were the main drivers to join SM (namely Facebook), participants frequently progressed to using SM for professional activities, with most actively engaged with SM in a professional capacity. Primary activities included accessing and sharing professionally relevant information among peers, intraprofessional networking, and job announcements. Wikipedia, YouTube, and Facebook were the most commonly used platforms. However, few professional contributions to the first two were noted. Pharmacists as experts on medicine could provide a great contribution by taking the lead in adding and editing medicine information accessible to consumers in highly accessed SM platforms such as Wikipedia. As the use of SM tends to increase among the population in general, a corresponding expansion of its application in the health sector and pharmacy may also be probable. Consequently, it is expected that this research might stimulate debate and professional use of social media among practitioners as well as stimulate further research on this topic.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide.

[PDF File (Adobe PDF File), 16KB - jmir_v18i9e258_app1.pdf]

References

1. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
2. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
3. Lupiáñez-Villanueva F, Mayer MA, Torrent J. Opportunities and challenges of Web 2.0 within the health care systems: an empirical exploration. Inform Health Soc Care 2009 Sep;34(3):117-126. [doi: 10.1080/17538150903102265] [Medline: 19670002]
4. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
5. Benetoli A, Chen TF, Aslani P. The use of social media in pharmacy practice and education. Res Social Adm Pharm 2015 Feb;11(1):1-46. [doi: 10.1016/j.sapharm.2014.04.002] [Medline: 24814268]
6. Brokowski L, Sheehan AH. Evaluation of Pharmacist Use and Perception of Wikipedia as a Drug Information Resource. Annals of Pharmacotherapy 2009 Oct 20;43(11):1912-1913. [doi: 10.1345/aph.1m340]
7. Cain J, Dillon G. Analysis of pharmacy-centric blogs: Types, discourse themes, and issues. Journal of the American Pharmacists Association 2010 Nov;50(6):714-719. [doi: 10.1331/japha.2010.10025]

XSL•FO

RenderX

8.   Clauson KA, Ekins J, Goncz CE. Use of blogs by pharmacists. Am J Health Syst Pharm 2010 Dec 1;67(23):2043-2048. [doi: 10.2146/ajhp100065] [Medline: 21098377]

9.   Kukreja P, Heck SA, Riggins J. Use of social media by pharmacy preceptors. Am J Pharm Educ 2011 Nov 10;75(9):176 [FREE Full text] [doi: 10.5688/ajpe759176] [Medline: 22171104]

10.   Ulbrich T, Metzger AH, Finley KS, McAuley JW. Evaluating the online networking relationships between preceptors and pharmacy students. Currents in Pharmacy Teaching and Learning 2013 Jul;5(4):256-262. [doi: 10.1016/j.cptl.2013.01.009]

11.   Hajar Z, Clauson KA, Jacobs RJ. Analysis of pharmacists' use of Twitter. Am J Health Syst Pharm 2014 Apr 15;71(8):615-619. [doi: 10.2146/ajhp130307] [Medline: 24688034]

12.   Alkhateeb FM, Clauson KA, Latif DA. Pharmacist use of social media. Int J Pharm Pract 2011 Apr;19(2):140-142. [doi: 10.1111/j.2042-7174.2010.00087.x] [Medline: 21385246]

13.   Shcherbakova N, Shepherd M. Community pharmacists, Internet and social media: an empirical investigation. Res Social Adm Pharm 2014 Nov;10(6):e75-e85. [doi: 10.1016/j.sapharm.2013.11.007] [Medline: 24388002]

14.   Smith FJ. Conducting Your Pharmacy Practice Research Project : A step-by-step guide. London: Pharmaceutical Press; 2010.

15.   Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

16.   Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? International Journal of Qualitative Studies on Health and Well-Being 2014 Oct 16;9:26512. [doi: 10.3402/qhw.v9.26152]

17.   Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]

18.   About Moodle. 2015. URL: https://docs.moodle.org/30/en/About_Moodle [accessed 2016-09-15] [WebCite Cache ID 6kYNq9RBB]

19.   Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons 2011 May;54(3):241-251. [doi: 10.1016/j.bushor.2011.01.005]

20.   Park N, Kee KF, Valenzuela S. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. Cyberpsychol Behav 2009 Dec;12(6):729-733. [doi: 10.1089/cpb.2009.0003] [Medline: 19619037]

21.   Noorden R. Nature. Online collaboration: Scientists and the social network URL: http://www.nature.com/news/online-collaboration-scientists-and-the-social-network-1.15711 [accessed 2016-09-15] [WebCite Cache ID 6kYOXqOA3]

22.   Cain J, Fink JL. Legal and Ethical Issues Regarding Social Media and Pharmacy Education. Am J Pharm Educ 2010 Sep;74(10):184. [doi: 10.5688/aj7410184]

23.   Benetoli A, Chen TF, Schaefer M, Chaar B, Aslani P. Pharmacists' perceptions of professionalism on social networking sites. Res Social Adm Pharm 2016:- (forthcoming)(forthcoming). [doi: 10.1016/j.sapharm.2016.05.044]

24.   Childs LM, Martin CY. Social media profiles: striking the right balance. Am J Health Syst Pharm 2012 Dec 1;69(23):2044-2050. [doi: 10.2146/ajhp120115] [Medline: 23172261]

25.   Barber N, Smith F, Anderson S. Improving quality of health care: the role of pharmacists. Quality & Safety in Health Care 1994 Sep 01;3(3):153-158. [doi: 10.1136/qshc.3.3.153]

26.   Cooper RJ, Bissell P, Wingfield J. 'Islands' and 'doctor's tool': the ethical significance of isolation and subordination in UK community pharmacy. Health (London) 2009 May;13(3):297-316. [doi: 10.1177/1363459308101805] [Medline: 19366838]

27.   Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]

28.   Lavsa SM, Corman SL, Culley CM, Pummer TL. Reliability of Wikipedia as a medication information source for pharmacy students. Currents in Pharmacy Teaching and Learning 2011 Apr;3(2):154-158. [doi: 10.1016/j.cptl.2011.01.007]

29.   Clauson KA, Polen HH, Boulos MNK, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]

30.   Giles J. Internet encyclopaedias go head to head. Nature 2005 Dec 15;438(7070):900-901. [doi: 10.1038/438900a] [Medline: 16355180]

31.   Haigh CA. Wikipedia as an evidence source for nursing and healthcare students. Nurse Educ Today 2011 Feb;31(2):135-139. [doi: 10.1016/j.nedt.2010.05.004] [Medline: 20646799]

32.   Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]

33.   von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. Journal of the American Medical Informatics Association 2012;19(5):777-781. [doi: 10.1136/amiajnl-2012-000990]

34.   Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009 Jul;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]

35.   Farič N, Potts HWW. Motivations for contributing to health-related articles on Wikipedia: an interview study. J Med Internet Res 2014 Dec;16(12):e260 [FREE Full text] [doi: 10.2196/jmir.3569] [Medline: 25498308]

36.   Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]

37.   Awad NI, Cocchio C. Use of Twitter at a major national pharmacy conference. Am J Health Syst Pharm 2015 Jan
       1;72(1):65-69. [doi: 10.2146/ajhp140175] [Medline: 25511841]
38.   Davison HK, Maraist C, Bing MN. Friend or Foe? The Promise and Pitfalls of Using Social Networking Sites for HR
       Decisions. J Bus Psychol 2011 May 4;26(2):153-159. [doi: 10.1007/s10869-011-9215-8]
39.   Guzdial M, Rick J, Kehoe C. Beyond Adoption to Invention: Teacher-Created Collaborative Activities in Higher Education.
       Journal of the Learning Sciences 2001 Jul;10(3):265-279. [doi: 10.1207/s15327809jls1003_2]

Abbreviations

SM: social media
SNS: social networking site

Original Paper

Exploring the Relationship Between Online Social Network Site Usage and the Impact on Quality of Life for Older and Younger Users: An Interaction Analysis

Darren Quinn[1], BSc (Hons), PhD; Liming Chen[2], BEng (Hons), MEng, PhD; Maurice D Mulvenna[1], BSc (Hons), MPhil, PhD; Raymond Bond[1], BSc (Hons), PhD

[1]Computer Science Research Institute, Faculty of Computing and Engineering, Ulster University, Newtownabbey, United Kingdom
[2]De Montfort University, School of Computer Science and Informatics, Leicester, United Kingdom

Corresponding Author:
Darren Quinn, BSc (Hons), PhD
Computer Science Research Institute
Faculty of Computing and Engineering
Ulster University
Shore Road, Newtownabbey
United Kingdom
Phone: 44 2890368252
Fax: 44 2890368252
Email: Quinn-D15@email.ulster.ac.uk

Abstract

Background: Analyzing content generated by users of social network sites has been shown to be beneficial across a number of disciplines. Such analysis has revealed the precise behavior of users that details their distinct patterns of engagement. An issue is evident whereby without direct engagement with end users, the reasoning for anomalies can only be the subject of conjecture. Furthermore, the impact of engaging in social network sites on quality of life is an area which has received little attention. Of particular interest is the impact of online social networking on older users, which is a demographic that is specifically vulnerable to social isolation. A review of the literature reveals a lack of knowledge concerning the impact of these technologies on such users and even less is known regarding how this impact varies across different demographics.

Objective: The objective of our study was to analyze user interactions and to survey the attitudes of social network users directly, capturing data in four key areas: (1) functional usage, (2) behavioral patterns, (3) technology, and (4) quality of life.

Methods: An online survey was constructed, comprising 32 questions. Each question directly related to a research question. Respondents were recruited through a variety of methods including email campaigns, Facebook advertisements, and promotion from related organizations.

Results: In total, data was collected from 919 users containing 446 younger and 473 older users. In comparison to younger users, a greater proportion of older users (289/473, 61.1% older vs 218/446, 48.9% younger) ($P<.001$) stated that Facebook had either a positive or huge impact on their quality of life. Furthermore, a greater percentage of older users strongly agreed that Facebook strengthened their relationship with other people (64/473, 13.5% older vs 40/446, 9.0% younger) ($P=.02$). In comparison to younger users, a greater proportion of older users had more positive emotions—classified as slightly better or very good—during their engagement with Facebook (186/473, 39.3% older vs 120/446, 26.9% younger) ($P<.001$).

Conclusions: The results reveal that despite engaging at considerably lower rates with significantly fewer connections, older users gain a greater quality-of-life benefit. Results disclose how both cohorts vary in their use, interactions, and rationale for engaging with Facebook.

KEYWORDS

online social network; social networking; Facebook; quality of life; interaction analysis; younger users; older users

XSL•FO
RenderX

Introduction

Background

The benefits of analyzing user interactions generated from social network sites (SNS) have been well-documented [1-6]. A key finding has been the ability to isolate and report on not just individual behaviors, but also on specific user groups that can be attributed to specific patterns of engagement among variant age cohorts [7,8]. This paper looks to add to the body of evidence that younger and older users interact differently with feature usage and engagement frequency. In addition, we will begin to address if SNS have an impact on the quality of life (QoL) of their users. QoL is understood as being an "individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" as defined by the World Health Organization [9].

Applying this view on QoL, through several works it is clear that social interaction is a key aspect for individuals; however, it is particularly so for older people [10-15]. Evidence reveals that declines in QoL directly relate to the vulnerability felt by older people from social isolation and loneliness [10,11,16-18]. The percentage of the world's older population (ie, 60+ years of age) will increase from 12.3% to 21.5% in 2050 [19]. SNS are an exciting option when we consider potential solutions that have the ability to impact large-scale populations and combat social isolation for the elderly. However, at present little is known about how using SNS impacts the QoL of its users.

Given that younger users are accepted as having much greater volumes of online social activity [7,20-22], does this necessarily imply that SNS have a greater impact upon a younger user's QoL? Or could the reverse be true, that older users with lower volumes of activity to have a higher QoL, with a focus on quality, not quantity, of SNS interactions? Important questions relating to the relationship between SNS usage and QoL remain unanswered. This is largely due to the complexities in the dual process of acquiring SNS user-generated content (UGC) for divergent age groups and then measuring the QoL of both cohorts. However, if achieved, the result would detail what relationships exist between age, usage, and QoL.

Related Work

A review of literature detailing how users perceive the impact of SNS on QoL has been carried out; it details how the views of network users have been assessed to date. Burke et al [23] provided an analysis of the relationship between usage of the popular social network site Facebook and social well-being, whereby 1193 users were recruited via a Facebook ad campaign. Crucially, it investigated the concept of social capital (ie, networks forming together for mutual benefit) in the context of using Facebook, which had been observed in similar works [24,25]. Importantly, however, this work was largely apprehensive of findings by related works [25-28]. Burke et al [23] investigated if findings in related research [25-28] showed whether students who are active do experience higher levels of social capital and whether this can be generalized. The methodology employed a study into the precise interactions to which social capital can be attributed (eg, wall posts and friends'

conversations with other friends). Resulting outcomes demonstrated that directed interaction (eg, a wall post) is akin to greater feelings of bonding and lower levels of loneliness. Results confirmed that the use of SNS increases social capital and reduces loneliness, stating that "engagement with Facebook is correlated with greater overall well-being."

There are a number of further works associating well-being and a user's online interactions [29-34]. Sundar et al [35] explored the issue with regard to older users. Their work focused on evaluating usage by retirees on Facebook, asking if SNS can help alleviate social isolation from aging alone, whereby the QoL of subjects was measured through 33 items adapted using the Life Satisfaction Index (LSI). The work focused on the vulnerability of older users in assessing the potential of SNS, particularly that of Facebook, to positively contribute to the lives of retirees. Results stated that QoL was "not linked to Facebook use, frequency of use, or intensity." However, this fact was stressed as being likely due to a number of factors. First, it observed users as having a small number of online friends and the small amount of time spent on Facebook by each subject. Second, a limitation was that only 34 retired Facebook users were assessed. Moreover, it was observed that the sample of older users already had a high QoL rating, leaving little room for any impact to be observed.

As highlighted by Burke et al [23], an issue within the literature at the time of this study is that studies are largely restricted to observing college students or adolescents only. Sundar et al [35] stress the significant fact that older users are distinct in nature from their younger counterparts (eg, lifestyles and experiences). Moreover, the work states that younger users have varying motivations for engaging with such technologies. It is clear that previous approaches were not explicit and direct in evaluating the experiences of users and the impact these experiences had on their QoL. For example, users were not directly asked if using application "x" or feature "x" contributed to their overall QoL and, if so, to what extent. As a result, this study aims to address these issues by constructing a questionnaire capable of evaluating user experience regardless of age or gender, an approach which can be employed for cross-comparison among varying demographics.

In summary, the literature shows that SNS can have a direct positive impact upon the well-being of users. However, the literature suggests that critical knowledge gaps remain in understanding the impact they have on the well-being of older users, since few works observed older users as a discrete cohort and, more crucially, they have not yet been directly compared and contrasted with a younger cohort.

Study Aim

Since SNS can *potentially* alleviate the burden of social isolation for older users, the extent of this impact upon end users' needs was assessed and quantified. The findings would have many applications, from formal policy decisions to design and usability considerations.

As social interaction is a key contributor to QoL [14], the following presented work investigated the impact of interacting within Facebook upon the QoL of two distinct age cohorts:

those considered *young*, chosen for being the primary user of the technology, versus an *older* cohort, a grouping which may be vulnerable to increasing levels of social isolation and may therefore benefit from a technology designed to increase social interactivity.

In the context of this study and keeping in line with several previous works [5,7,8,20,36] in this area, users between 15 and 30 years of age were labeled as *younger* and those aged 50+ years old were labeled as *older*. This work extends current knowledge by examining the perceptions of users as they engage with SNS. It establishes relationships, if any, between SNS and the individual user's QoL. We hypothesized that, given that interaction and usage have been shown to differ between older and younger users, we would expect SNS to positively contribute to the QoL of older people.

Methods

Overview

Within this section we present the methods applied for identifying and engaging with end users, which principally involved the development of an online survey. We also discuss the methods employed in data gathering and analysis.

Data Collection

The design of the survey was driven by a series of 18 research questions (see Multimedia Appendix 1) based on several previous studies [5,7,8]. This survey aimed to capture the views of both cohorts in four key areas: (1) functional usage, (2) behavioral patterns, (3) technology, and (4) QoL impact. The approach to the phrasing of questions was that of a nontechnical, simplistic approach, being as concise as possible throughout. The online survey comprised 32 questions as presented in Table 1. The survey was implemented and hosted using SurveyGizmo (Widgix, LLC) [37].

Survey Testing

Following the preliminary design, testing was carried out with a range of subjects who varied in terms of both gender and age. The aim of testing was to acquire user feedback in terms of design and coherence for end users. It requested feedback on aesthetics, question style, and phraseology, as well as overall usability. A series of survey iterations occurred following test responses, in advance of an agreed-upon final version.

Given that this research was to involve surveying two distinct age cohorts, an approach was applied to employ separate surveys. Given the depth of literature on barriers for older users, it was viewed that a one-look survey would be unlikely to be satisfactory for both cohorts. Older users were provided with minimal text, with only critical information to minimize cognitive strain, with a larger font size to increase readability. However, for both cohorts every question in each section (from Demographics section onwards) was identical and compulsory, regardless. To reduce survey dropouts, an error message

indicated incomplete questions acting as a control loop. A progress bar was also provided to indicate progression.

Sample Size

Required user numbers were determined through sample-size calculations. A confidence level of 95% was applied. At the time of data collection, the UK Facebook population was 28,940,400, with calculations determining the required sample size to be 474 completed surveys [8]. Consideration was given to the volumes of users recruited within related research. However, as previously noted, only a limited number of works were available, providing a lack of consistency in relation to user numbers in this area. Nevertheless, it is observed in the work of Sundar et al [35] that issues were raised concerning the ability to determine a QoL impact following the evaluation of only a small number of individuals (ie, n=34). Although no older users were sampled, an approach more akin to that of Burke et al [23] (ie, with significantly higher numbers [n=1193]) was viewed as providing greater confidence in terms of reliability and representative analysis.

Recruitment

As a first phase uptake of younger users, an email was circulated to all Ulster University students in April 2012, with an estimated reach of approximately 28,500 registered students. A secondary phase consisted of posting on a series of Facebook accounts to stimulate interest and uptake, along with the use of Twitter to publicize the survey. The application of an all-student email proved extremely successful, quickly acquiring more than the required number of users. Recruitment of older users proved significantly more challenging. First, organizations with direct access to potential older subjects were consulted (see Multimedia Appendix 2) and asked for support in publicizing through their media outlets (eg, websites, blogs, Facebook, and Twitter accounts). A secondary phase promoted the survey during the annual *National Silver Surfer Day 2012*, an event aimed at encouraging the over 50-year-olds to engage with online technologies. In a localized context, the event was promoted in Northern Ireland libraries that hosted open days. Promotional leaflets were distributed throughout centers, providing information and a link to the study to encourage local uptake. Both strategies proved unfruitful. A final phase of Facebook advertising was employed as the core promotional strategy. This was a direct approach to engage end users, enabling advertisements to be displayed on the walls of a highly specific demographic. For this directly targeted audience, advertisements were displayed on only those profiles of Facebook users who were (1) over 50 years of age and (2) had attended university. The result was that two groups were acquired with a comparable socioeconomic status, with knowledge that both populations had entered third-level education with similar education attainment. Furthermore, with knowledge of the groups' educational attainment, it provided an indication of both groups sharing a similar socioeconomic status.

Table 1. Online survey questions and answer styles.

Category	Question	Answer style
Demographics		
1	Gender	Radio buttons
2	Age	Textbox
3	Which of the following do you currently, or have you ever attended?	Radio buttons
4	How would you rate your level of computer literacy?	Likert scale
5	Please indicate which of the following Online Social Networks (OSN) accounts you have.	Check buttons
6	Do you have a Facebook account?	Radio buttons
Functional usage		
7	How important is Facebook for maintaining your real-world social connections?	Likert scale
8	How much consideration goes into adding or accepting new friend connections?	Likert scale
9	A list of Facebook's most popular features has been compiled. Please rate each function in terms of importance.	Table of radio buttons
10	How important do you feel it is to keep your Facebook profile up-to-date, such as changing your profile picture or updating your relationship status, etc?	Likert scale
11	In relation to the above question, please take a moment to state why.	Textbox
12	After you created your Facebook account, rate how easy or difficult you found it to use its applications/functions.	Likert scale
13	In relation to the above question, please take a moment to state why.	Textbox
Patterns of usage		
14	How often do you log into Facebook?	Radio buttons
15	Indicate which you feel is your prime time for activity.	Radio buttons
16	Which day are you most likely to check Facebook for updates?	Radio buttons
17	Which day are you most likely to use Facebook for making plans (eg, social events)?	Radio buttons
18	Do you feel using Facebook is part of your routine?	Radio buttons
19	How important do you feel Facebook is for planning and broadcasting new events?	Likert scale
20	From the three options, rank how you would describe your general behavior when using Facebook.	Rank 1st, 2nd, and 3rd
21	Generally speaking, how much thought would you put into the posting of comments or replies?	Likert scale
22	Do you use Facebook more during the week or at the weekend?	Radio buttons
23	In relation to the above question, please take a moment to state why.	Textbox
24	What category of user would you class yourself as?	Likert scale
Technology section		
25	A list of 12 potential reasons has been compiled for using Facebook. For each, please state whether you agree, disagree, or it is not applicable to using Facebook.	Table of radio buttons
26	Which of the following factors would discourage you from using Facebook?	Table of radio buttons
27	Do current technologies such as smartphones and tablet PCs encourage you to use Facebook more often?	Radio buttons
Social impact		
28	Quality of Life (QoL) is used to evaluate the general well-being of individuals and societies (eg, recreation and leisure time and social belonging). Do you feel using Facebook contributes to your overall QoL?	Radio buttons
29	Do you feel it strengthens the relationship with the people you connect with?	Radio buttons
30	How important do you feel Facebook is for keeping in contact with family and friends?	Likert scale
31	Generally speaking, how do you feel when using Facebook during the following three stages, from feeling very down to feeling very good?	Table of radio buttons

Category	Question	Answer style
32	You have reached the end of the survey. Please provide any observations/criticisms which you feel may improve future survey takers' experience.	Textbox
33	If you would like to receive a copy of the analysis, please provide an email address or contact information.	Textbox

Statistical Analysis

The number of completed surveys is stated along with the responses to each of the 33 questions by each cohort, which are directly contrasted using frequency analysis and horizontal stacked bar charts. Statistical significance between the groups—younger users versus older users—per question was tested using the N-1 chi-square test for comparing independent proportions ($P<.05$).

Multivariate logistic regression was used to assess potential confounders that might bias the hypothesis that Facebook has a positive impact on the QoL for a greater proportion of older users in comparison to younger users. The independent or exposure variables included age group (ie, younger or older cohort), gender, country (ie, 17 different nationalities), computer literacy, and the type of user (eg, frequent user or occasional user). The dependent variable, or response variable, was binary (ie, Facebook has a positive or negative impact, or none at all, on QoL). The model provides odds ratios (ORs) that indicate how each exposure variable contributes to a participant stating that Facebook has an impact on their QoL. The model is described in equation 1:

$$logit(p) = \beta_0 + \Sigma_{i=0}^{n} \beta_i X_i (1)$$

where β_0 is the intercept, β_i is a vector of coefficients (log odds), and X_i a vector of values from each independent/exposure variable. All data analysis was carried out using the R programming language in combination with RStudio (RStudio, Boston, MA).

Results

Overview

Excluding ineligible participants or those with partial surveys, a total of 919 completed surveys were collected, with 446 younger users and 473 older users. Demographics for the two cohorts are provided in Table 2. This table shows an equal distribution of gender in both cohorts, which eliminates gender bias, a common confounding factor. While computer literacy was slightly higher among the younger group, this was to be expected; however, the difference was minimal. Figure 1 shows the popularity of SNS among the cohorts. Younger users preferred Twitter and YouTube whereas older users preferred LinkedIn and Google+.

Table 2. Demographics of older and younger participants.

Demographics	Older users (n=473)	Younger users (n=446)
Male, n (%)	173 (36.6)	140 (31.4)
Female, n (%)	300 (63.4)	306 (68.6)
Age in years, mean (SD)	59.35 (6.92)	21.81 (2.48)
Age in years, range	50-87	18-30
Computer literacy[a], mean (SD)	3.27 (0.84)	3.64 (0.64)

[a]Computer literacy ranges from 1 (novice) to 5 (expert).

Figure 1. Online social network (OSN) accounts among older and younger users.

XSL•FO
RenderX
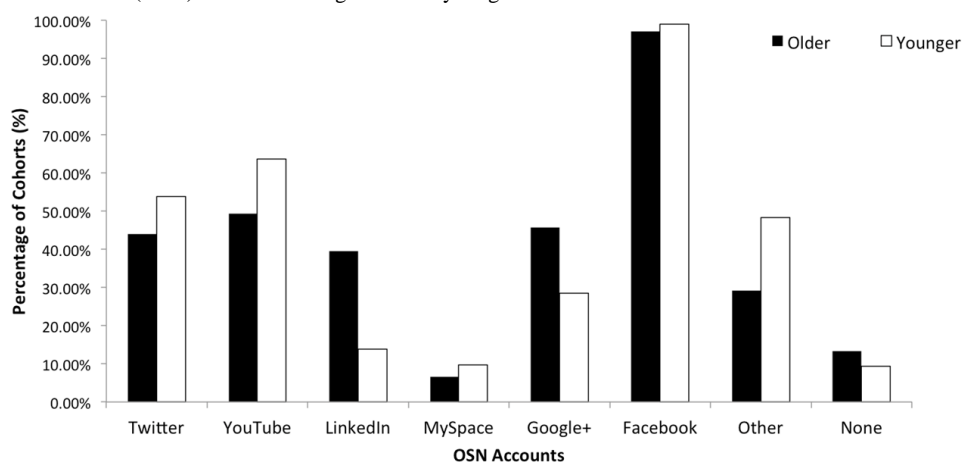
Functional Usage

Figure 2 (A) shows how both cohorts considered how important Facebook is for maintaining real-world social connections, and Figure 2 (B) shows how much consideration went into adding or accepting new friends. Results show older users put more consideration into who they connect with; 249 out of 473 (52.6%) older users consider who they connect with *a lot* versus 157 out of 446 (35.2%) younger users consider who they connect with *a lot* (*P*<.001).

Figure 3 presents the most important Facebook features. Applying statistical significance, younger users gave greater importance to creating groups, tagging, instant messaging,

notifications, news feeds, status updates, and photos. However, older users gave greater importance to questions (*P*<.001) and surveys (*P*<.001). In reference to question 10 (see Table 1), only 7.6% (36/473 older and 34/446 younger) of subjects in both cohorts agreed that maintaining an up-to-date Facebook profile (eg, profile picture and relationship status) is *Very Important*. When asked how difficult it is to use Facebook's features, older users encountered more usability problems (160/473, 33.8%) than younger users (64/446, 14.3%) (*P*<.001). For all users, notifications and news feeds were considered important, but photos and posting were the most important features.

Figure 2. (A) The importance of Facebook for maintaining connections among older and younger users and (B) the consideration taken by both cohorts when adding new friends.
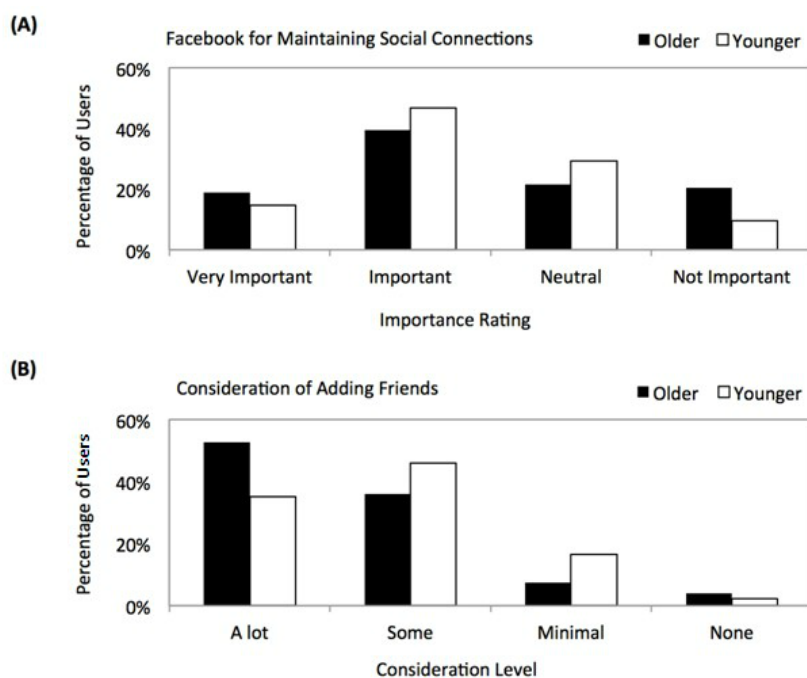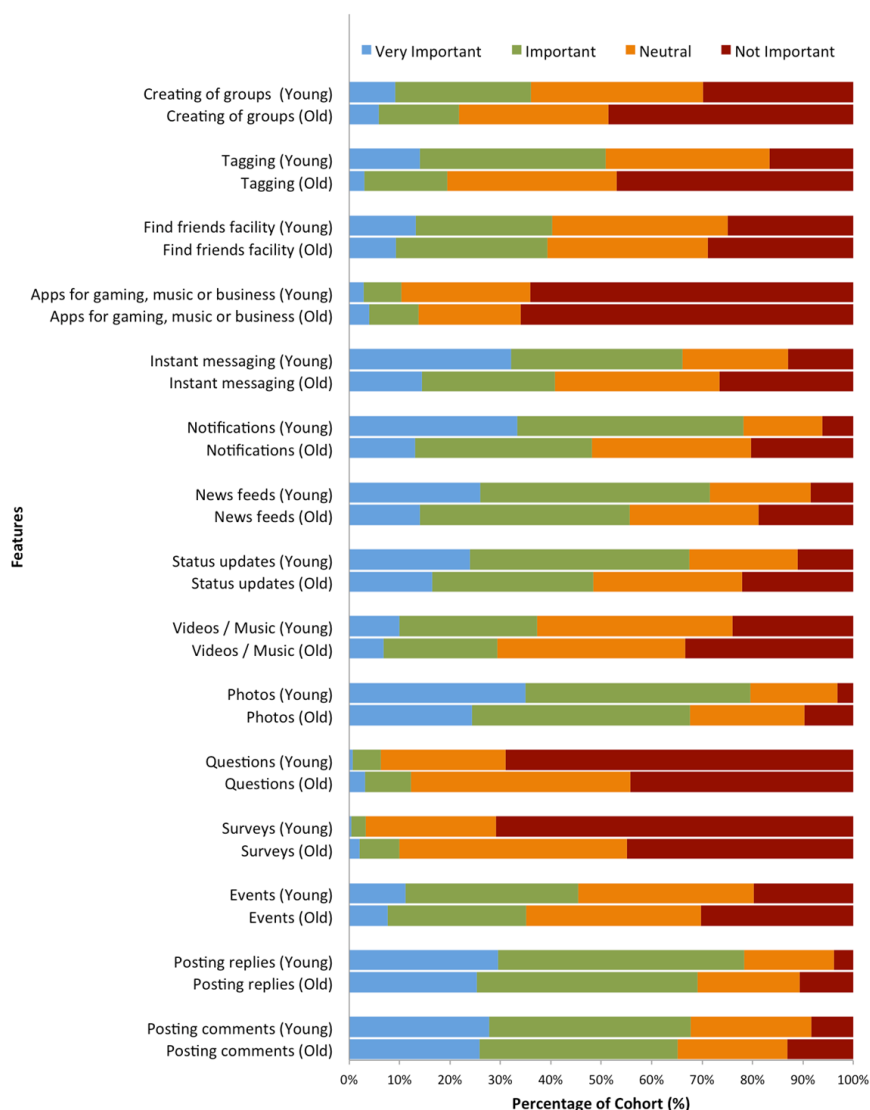
Figure 3. Ratings of Facebook features by older and younger users.

Patterns of Usage

Figure 4 shows graphs that indicate when both cohorts are likely to engage with Facebook. Most subjects stated they log on *daily*—362 out of 473 (76.5%) for older users and 286 out of 446 (64.1%) for younger users, as seen in Figure 4 (A). However, more younger users (105/446, 23.5%) log in *hourly* when compared to older users (38/473, 8.0%) (*P*<.001). In addition, more younger users are active in the late evening (135/446, 28.5%) versus in the morning (222/446, 49.8%) (*P*<.001), whereas more older users are active in the morning (103/473, 21.8%) versus in the late evening (32/473, 6.8%) (*P*<.001), as seen in Figure 4 (B). Also, regarding question 22 (see Table 1), 281 of 446 younger users (63.0%) use Facebook during weekdays, whereas the majority of older users (253/473, 53.5%) use Facebook independent of whether it is a weekday or the weekend. Figure 4 (C) indicated that older users do not feel they have a specific day for checking updates. It is also evident from Figure 4 (D) that older people do not use Facebook to

make plans or arrange social events, whereas younger users do. In answering question 19 (see Table 1), more younger users ranked Facebook as being *Important* or *Very Important* for planning and broadcasting events (206/446, 46.2%) versus 274 of 473 older users (57.9%) (*P*<.001).

Figure 5 shows the amount of consideration given to posting comments to a friend. A higher percentage of older subjects (232/473, 49.0%) provide *a lot* of consideration to posting comments compared to the younger cohort (94/446, 21.1%) (*P*<.001).

Answers to question 20 from Table 1 (see Table 3) asked users to consider their overall behavior and then to rank it within three options— *Responder*, *Observer*, or *Instigator*. Both cohorts primarily declared their behavior to be *Respond to others*. However, younger users ranked this higher when compared to older users. Answers to question 24 from Table 1 (see Figure 6) show that the majority of users in both cohorts classify themselves as a *high frequency user* or a *moderate user*.

XSL•FO
RenderX

Table 3. Behavior rankings.

Descriptor	Older users		Younger users	
	Total score[a]	Overall rank[a]	Total score	Overall rank
Respond to others' activity	859	1	936	1
Observe others' activity	685	2	800	2
Instigate activity (eg, posting comments, videos, and pictures)	636	3	673	3

[a]Score is a weighted calculation. Items ranked first are valued higher than the following items; the score is the sum of all weighted rank counts.

Figure 4. (A) The frequency of how often both cohorts log into Facebook, (B) where each cohort designated their prime time for activity on Facebook, (C) which days the two cohorts are most likely to check Facebook for updates, and (D) which days the cohorts stated to be their most likely to use Facebook for making plans and arranging social events. N/A: not applicable.
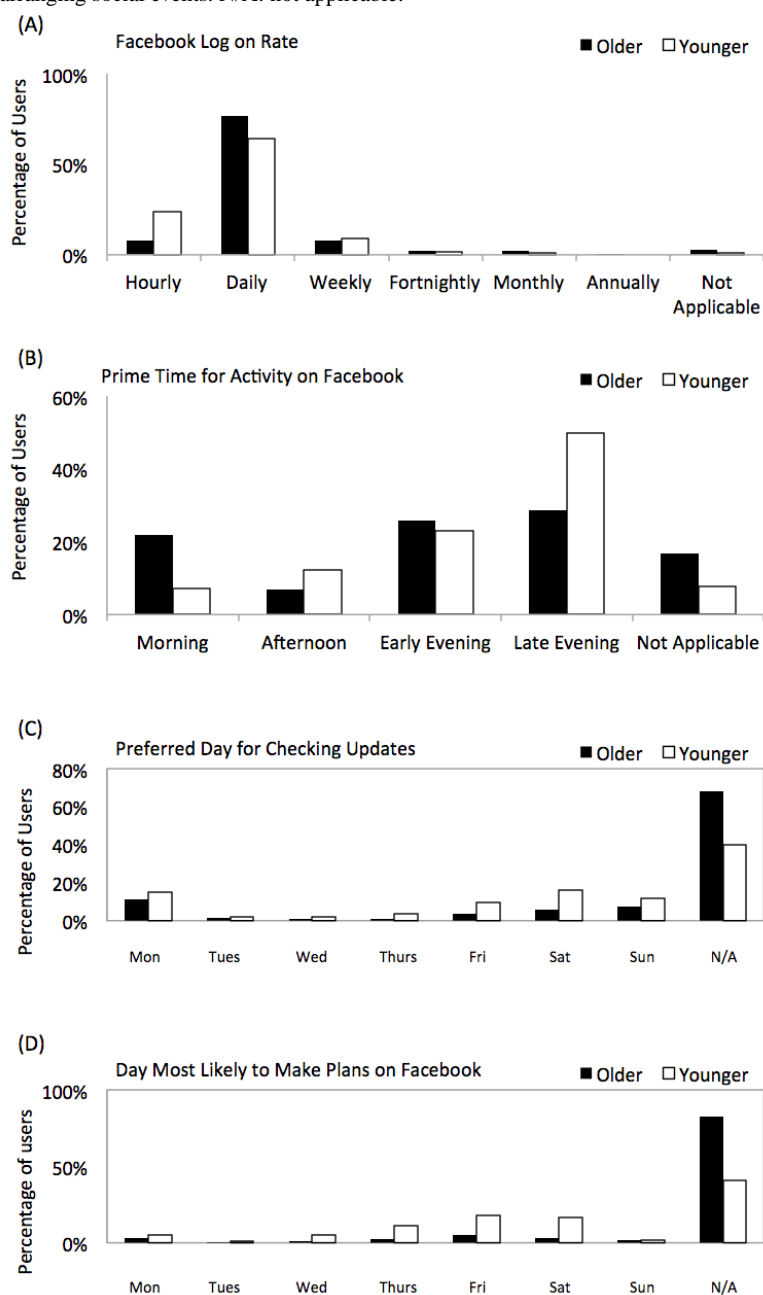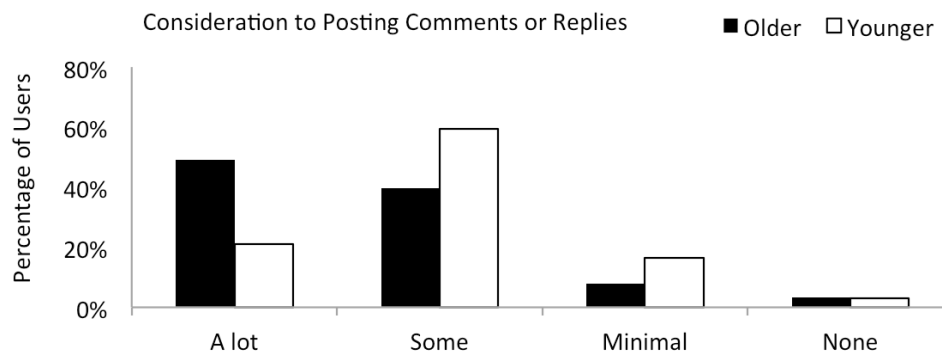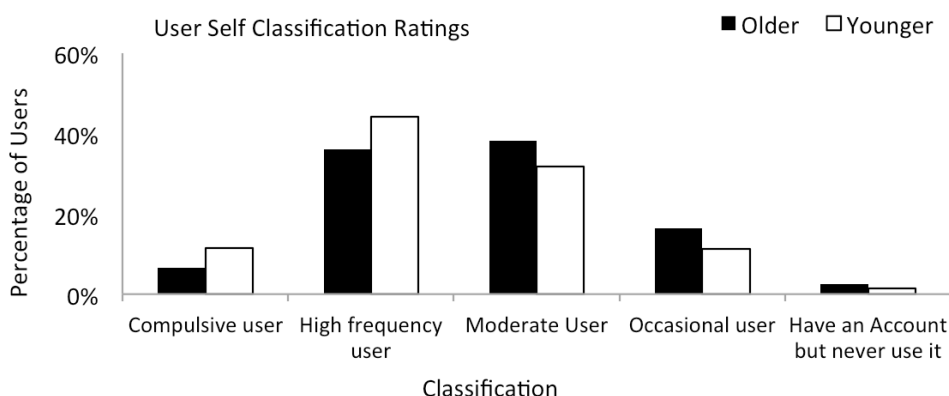
Figure 5. Amount of consideration given to posting comments or a reply for both cohorts.

Figure 6. Behavior classification in both cohorts.

Technology

Figure 7 shows 12 reasons for using Facebook. The main reasons for the younger cohort to use Facebook, that are all statistically significant (*P*<.001) in comparison to older users, is that it helps them in making new plans (278/446, 62.3% younger vs 123/473, 26.0% older), they can view other profiles (372/446, 83.4% younger vs 256/473, 54.1% older), and the fact that everybody else uses it (363/446, 81.4% younger vs 198/473, 41.9% older). Interestingly, 387 of 446 younger users (86.8%) engage with Facebook due to boredom, which is in contrast to 139 of 473 older users (29.4%) who do so. Conversely, a greater proportion of older users identified that they engage with Facebook because they can debate with like-minded people, which is distinct from the younger cohort (239/473, 50.5% older vs 128/446, 28.7% younger).

Figure 8 shows the factors that could discourage users from using Facebook. A greater percentage of older users agree that such factors would include the following: (1) Facebook is too technically demanding (217/473, 45.9% older vs 152/446, 34.1% younger) (*P*<.001) and (2) the continual format changes would discourage users from engaging (362/473, 76.5% older vs 267/446, 59.9% younger) (*P*<.001). Conversely, a greater percentage of younger users agree that profile viewing from potential employers is a key reason not to engage (330/446, 74.0% younger vs 206/473, 43.6% older) (*P*<.001). Interestingly, 354 of 446 younger users (79.4%) agreed that technologies such as mobile phones and tablet personal computers (PCs) encourage them to use Facebook more frequently, whereas only 146 of 473 older users (30.9%) agreed with this statement (*P*<.001).
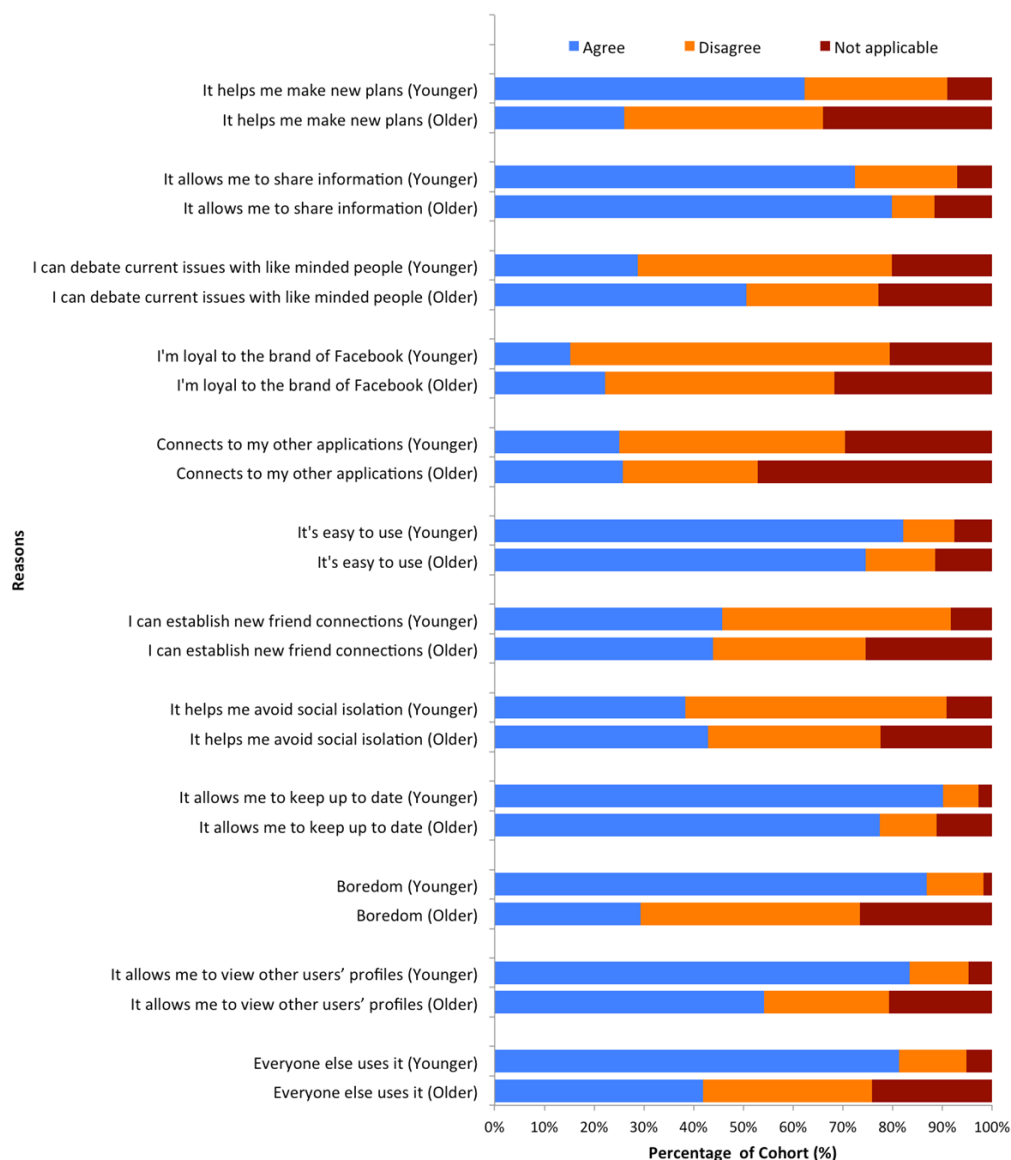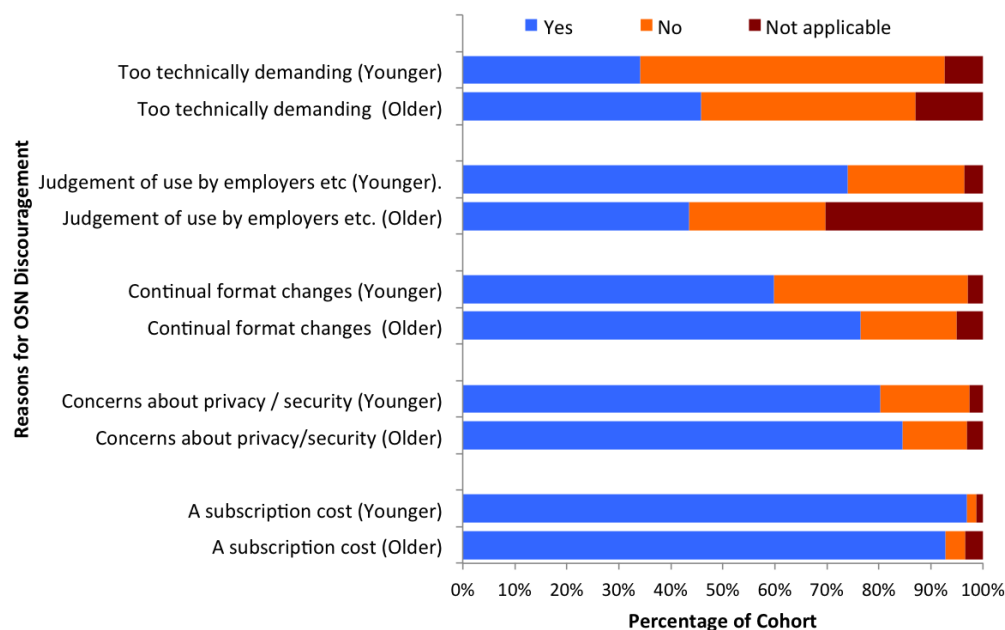
Figure 7. A list of 12 reasons for using Facebook according to both cohorts.

Figure 8. Factors that discourage the use of Facebook for younger and older users. OSN: online social networks.

Social Impact

Figure 9 presents graphs that show how Facebook contributes to the users' QoL and their relationships with people. In comparison to younger users, a greater proportion of older users (289/473, 61.1% older vs 218/446, 48.9% younger) (*P*<.001) stated that Facebook has either a *positive impact* or a *huge impact* on their QoL, as seen in Figure 9 (A). There are few differences between the cohorts in Figure 9, (B) and (C). However, a greater percentage of older users *strongly agree* that Facebook strengthens their relationships with other people (64/473, 13.5% older vs 40/446, 9.0% younger) (*P*=.02).

Table 4 indicates the ORs for each exposure variable that may contribute to a participant indicating that Facebook contributes to their QoL. The table confirms that younger users are less likely (OR 0.45, *P*<.001) to indicate that Facebook improves their QoL. Conversely, there are greater odds that Facebook will have a positive impact on the QoL of older users when compared to younger users. However, Table 4 indicates that

there are two other statistically significant ORs related to the frequency of using Facebook. Thus, those participants who ranked themselves as a *moderate* or *occasional* user were less likely to state that Facebook has an impact on their QoL. However, on inspection, this is not a confounding factor since the variable is considerably proportionately split between the younger and older cohort (ie, 26% of younger users are moderate/occasional users who said Facebook has no positive impact on QoL and, similarly, 29% of older users are moderate/occasional users who said Facebook has no positive impact on QoL). While 17 countries were represented in the dataset, this variable was also not confounding. Being German was almost significant (OR 9.25, *P*=.08); however, the confidence interval has a significant range and there were only 6 German participants: an equal split of 3 German participants in each group. Interestingly, the per unit increase in computer literacy, which could be associated with education and socioeconomic status, was not statistically significant (OR 0.92, *P*=.46) in contributing to stating whether Facebook has an impact on QoL.

Table 4. Odds ratios for each of the independent (exposure) variables where the response variable is whether Facebook has or has not made an impact on the user's quality of life.

Exposure variable	Odds ratio	95% CI	SE	Z	P
Age group (younger)	0.45	0.32-0.62	0.17	-4.74	<.001
Gender (male)	1.19	0.87-1.65	0.17	1.07	.28
Computer literacy (per unit increase)	0.92	0.74-1.14	0.11	-0.75	.46
Country (Germany)	9.25	0.96-22.04	1.29	1.73	.08
Type of user (high-frequency user)	1.12	0.66-1.88	0.27	0.43	.67
Type of user (moderate user)	0.55	0.32-0.93	0.27	-2.20	.02
Type of user (occasional user)	0.05	0.02-0.09	0.38	-8.04	<.001

Figure 9. (A) Both cohorts feel Facebook contributes to their quality of life (QoL), (B) users feel it strengthens the relationship with the people they are connected to, and (C) users feel Facebook is important for keeping in contact with family/friends.
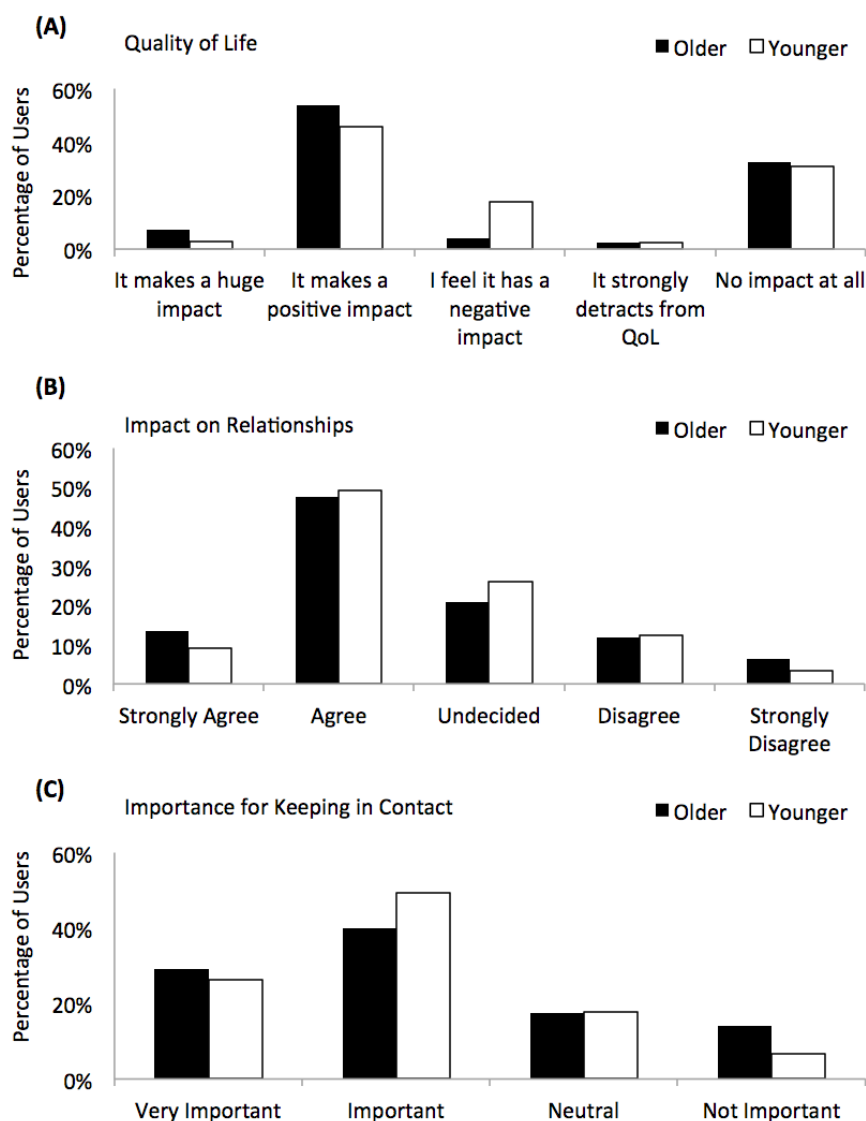
[Figure 10](#) shows two graphs that illustrate the changes in positive emotion before, during, and after using Facebook. Before using Facebook, 45 of 473 older users (9.5%) felt *good* or *very good*. However, when engaged with Facebook, this statistic increased to 186 out of 473 (39.3%) ($P<.001$). Likewise, before using Facebook, 31 of 446 younger users (7.0%) felt *good* or *very good*, however, when engaged with Facebook this statistic increased to 120 out of 446 (26.9%) ($P<.001$). This also indicates that in comparison to younger users, a greater proportion of older users have more positive emotions—classified as *slightly better* or *very good* —during their engagement with Facebook (186/473, 39.3% older vs 120/446, 26.9% younger) ($P<.001$).

Figure 10. Emotional state before, during, and after using Facebook for (A) older users and (B) younger users. The graphs also present the means and standard deviations for the ratings of the state in each phase (where 1=very down and 5=very good). The P values compare the proportion of users who feel slightly better or very good before and during Facebook engagement for both cohorts.
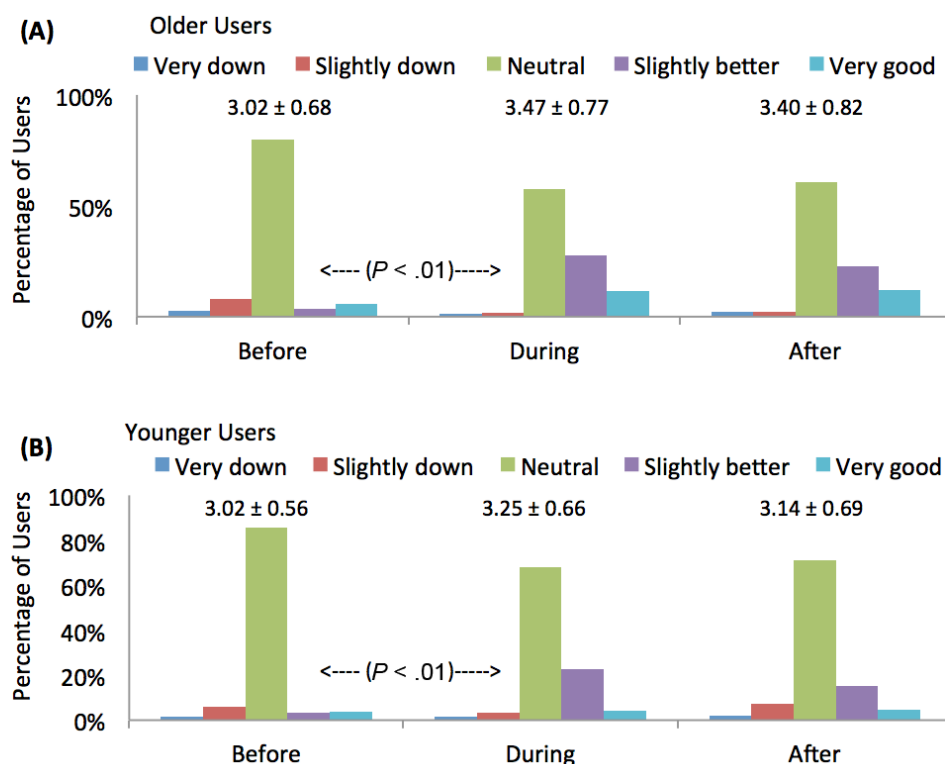
Discussion

The aim of this study was to disclose the impact of online social networking on QoL, examining the perceptions of users as they engage. The hypothesis was that SNS positively contribute to the QoL of older people. Results disclosed why younger users have 11 times more Facebook "friends" (observed by Quinn et al [7]). This is due to the fact that younger users create new linkages without much consideration. Relating to functional usage, a hypothesis was proposed that anomalies, identified by Quinn et al [8], were due to the fact that each cohort attached different values to different functionalities. Results of this study reveal that differing Facebook features are clearly identifiable in terms of their importance for each of the age groups. In asking about the *prime time* for activity, interpretation of results leads to the conclusion that younger users integrate SNS as a part of their daily life. For these users, online social networking activity occupies a dedicated time within each day, which has been clearly evidenced by the volume of users recorded in this study. Facebook is now an accepted communication modality adjudged to be part of the user's routine. In this regard, younger users gave greater importance for using SNS for planning and broadcasting new events. However, the view of older users was more divided, with only a slim majority recording *Neutral*, followed by *Important* (171/473, 36.2% and 156/473, 33.0%, respectively). Results therefore disclose older users to be the more reflective users.

Questions concerning patterns of usage evaluated if users had a bias for *Weekday* or *Weekend* usage, following the user metric results contained within Quinn et al [7]. Results demonstrated

a strong preference for weekend use among younger respondents when compared to older users. Twice as many older users selected *Weekday* as their preference. Results indicated that given that the majority of younger users were attending university, they accessed their accounts during weekdays. Older users are not bound by such restrictions and results are therefore reflective of such facts. Based on this new evidence for both cohorts, Facebook is now shown to be important for maintaining real-world connections. Results demonstrated that age cohorts are identifiable with particular functionalities. Discouraging factors united users in opposing subscription costs, with concerns relating to privacy/security, and continual platform changes. SNS were shown to strengthen relations, regardless of age, as both cohorts agreed upon its importance for keeping in contact with family and friends. In terms of emotional state, a definite shift was observed when a greater volume of users were recorded among the more positive emotions *During* and *After* usage. It is now shown that usage directly affects the emotional state of users as they engage, the limitation being that emotional state was self-reported. Given the content that users will frequently observe, such as pictures of friends and family or messages from friends, it is often emotive content that will directly stimulate the emotions of a user. As a social platform, it is clearly an established mode of communication. However, results demonstrated no loyalty to the brand, indicating a willingness by both cohorts to potentially switch to an alternative.

Given that older users are engaging extensively in online social media [38], a key aim of this study was to investigate the impact of such users who interact with a social networking application, namely Facebook. A particular focus was to establish what, if

any, relationship existed between using social networking technologies and their impact upon users' QoL. Although a number of works have emerged in the area, it was identified that (1) results were the subject of conjecture and (2) no works addressed the real-world impact upon the QoL of end users and, more specifically, the impact upon older users. This is important since social technology has the potential to alleviate the burden of social isolation. Although generic traits could be shared across both cohorts, there were many characteristics which were identifiable to specific age ranges, as described in Textbox 1.

Textbox 1. Evidence-based personas of older and younger users.

Younger persona

- They create connections with less consideration, largely due to their real-world social structures; subsequently, they will update profiles frequently due to their life updates (eg, relationships, jobs, and attainment).

- They engage very frequently, due to the fact that the majority of their friends use the same application and the source, therefore, of the majority of social content; subsequently, they check for updates more frequently.

- They prefer weekday activity.

- They provide dedicated time to keep up-to-date with their peers and the news, which is facilitated by the fact that they find the interface easy to use.

- They frequently share information on social network sites (SNS) for making plans.

- They feel that the SNS positively impacts on their quality of life (QoL).

Older persona

- They give a lot of consideration to creating new connections and crucially operate a quality approach to posting; with fewer connections, there are fewer reasons to engage.

- They are more purposeful in their reason for engagement; they only log on for a direct purpose.

- As a cohort, their failure to continue engaging is explicable due to difficulties in using features on a multifaceted platform.

- Older users encounter more usability problems with the user interface.

- Contemporary technologies do not encourage older users to engage, most likely due to the low adoption rates of other technologies.

- They feel that the SNS positively impacts on their QoL.

In conjunction with previous research, results from 919 surveyed users—446 younger users (18-25 years) and 473 older users (50+ years)—form a new body of knowledge applicable to many domains, from policy makers to SNS designers. Results showed that older users have a quality, rather than a quantity, approach to SNS usage. This was directly in contrast to that of younger users. Although older users interacted with the SNS less frequently, they gained a significantly greater QoL and emotional benefit to using Facebook when compared to younger users. Future work could explore how SNS effectiveness can help users avoid social isolation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Research questions.

[PDF File (Adobe PDF File), 23KB - jmir_v18i9e245_app1.pdf]

Multimedia Appendix 2

Older people's organizations.

[PDF File (Adobe PDF File), 15KB - jmir_v18i9e245_app2.pdf]

References

1. Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, Settle JE, et al. A 61-million-person experiment in social influence and political mobilization. Nature 2012 Sep 13;489(7415):295-298 [FREE Full text] [doi: 10.1038/nature11421] [Medline: 22972300]
2. Cheung C, Chiu P, Lee M. Online social networks: Why do students use Facebook? Comput Human Behav 2011 Jul;27(4):1337-1343. [doi: 10.1016/j.chb.2010.07.028]

XSL•FO

RenderX

3.   Chu S, Kim Y. Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. Int J Advert 2011;30(1):47-75. [doi: 10.2501/IJA-30-1-047-075]

4.   Newman M, Lauterbach D, Munson S, Resnick P, Morris M. 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. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work. New York, NY: ACM; 2011 Presented at: ACM Conference on Computer Supported Cooperative Work; March 19–23, 2011; Hangzhou, China p. 341-350.

5.   Quinn D, Chen L, Mulvenna M. Discovering the social interaction patterns of younger and older Facebook users. In: Bravo J, Hervás R, Rodriguez M, editors. Ambient Assisted Living and Home Care. Berlin, Germany: Springer; 2012:295-303.

6.   Quinn D, Chen L, Mulvenna M. Contemporary gold rush or scientific advancement: A review of social network analysis. In: Recent Advances in Ambient Intelligence and Context-Aware Computing. Hershey, PA: IGI Global; 2014:210-226.

7.   Quinn D, Mulvenna M, Chen L. Discovering social behaviour variances of younger and older users through social interaction analysis. Int J Web Sci 2013;2(1-2):44-65. [doi: 10.1504/IJWS.2013.056574]

8.   Quinn D. Analysis of How Young and Older People Interact Using Online Social Networks [PhD dissertation]. Newtownabbey, UK: University of Ulster; 2013.

9.   The World Health Organization. The World Health Organization Quality of Life assessment (WHOQOL): Position paper from the World Health Organization. Soc Sci Med 1995 Nov;41(10):1403-1409. [Medline: 8560308]

10. Bond J, Corner L. Quality of Life and Older People. Maidenhead, UK: Open University Press; 2004:86-98.

11. Bowling A. What things are important in people's lives? A survey of the public's judgments to inform scales of health related quality of life. Soc Sci Med 1995 Nov;41(10):1447-1462. [Medline: 8560313]

12. Bowling A, Gabriel Z. Lay theories of quality of life in older age. Ageing Soc 2007;27(6):827-848. [doi: 10.1017/S0144686X07006423]

13. Farquhar M. Elderly people's definitions of quality of life. Soc Sci Med 1995 Nov;41(10):1439-1446. [Medline: 8560312]

14. Gabriel Z, Bowling A. Quality of life from the perspectives of older people. Ageing Soc 2004;24(5):675-691.

15. Glass TA, de Leon CM, Marottoli RA, Berkman LF. Population based study of social and productive activities as predictors of survival among elderly Americans. BMJ 1999 Aug 21;319(7208):478-483 [FREE Full text] [Medline: 10454399]

16. Victor C, Scambler S, Bond J. The Social World of Older People: Understanding Loneliness and Social Isolation in Later Life. Maidenhead, UK: Open University Press; 2009.

17. Victor C, Grenade L, Boldy D. Measuring loneliness in later life: A comparison of differing measures. Rev Clin Gerontol 2006;15(1):63-70.

18. Victor C, Scambler S, Bowling A, Bond J. The prevalence of, and risk factors for, loneliness in later life: A survey of older people in Great Britain. Ageing Soc 2005;25(6):357-375.

19. United Nations. World Population Prospects: The 2015 Revision. Key Findings and Advance Tables. New York, NY: United Nations; 2015. URL: http://esa.un.org/unpd/wpp/Publications/Files/Key_Findings_WPP_2015.pdf [accessed 2016-08-01] [WebCite Cache ID 6jRfTybRy]

20. Quinn D, Chen L, Mulvenna M. Does age make a difference in the behaviour of online social network users? In: Proceedings of the Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing. 2011 Presented at: Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing; October 19-22, 2011; Dalian, China p. 266-272.

21. Boyd D. MacArthur Foundation Series on Digital Learning – Youth, Identity, and Digital Media Volume. Why Youth Heart Social Network Sites: The Role of Networked Publics in Teenage Social Life. Cambridge, MA: MIT Press; 2007. URL: http://www.danah.org/papers/WhyYouthHeart.pdf [accessed 2016-08-11] [WebCite Cache ID 6jgj5BNqX]

22. Jones S, Millermaier S, Goya-Martinez M, Schuler J. First Monday. Whose Space is MySpace? A Content Analysis of MySpace Profiles.. 2008 Sep 01. URL: http://firstmonday.org/article/view/2202/2024 [accessed 2016-08-11] [WebCite Cache ID 6jgjK9pYc]

23. Burke M, Marlow C, Lento T. Social network activity and social well-being. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA p. 1909-1912.

24. Valenzuela S, Park N, Kee K. Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. J Comput Mediat Commun 2009;14(4):875-901.

25. Ellison N, Steinfield C, Lampe C. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. J Comput Mediat Commun 2007;12(4):1143-1168.

26. Coleman J. Social capital in the creation of human capital. Am J Sociol 1988;94:95-120.

27. Putnam RD. Bowling Alone: The Collapse And Revival Of American Community. New York, NY: Simon & Schuster; 2001.

28. Steinfield C, Ellison N, Lampe C. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. J Appl Dev Psychol 2008;6:434-445.

29. Burke M, Kraut R, Marlow C. Social capital on Facebook: Differentiating uses and users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 571-580.

30. Yoder C, Stutzman F. Identifying social capital in the Facebook interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 585-588.

31. Vitak J, Ellison N, Steinfield C. The ties that bond: Re-examining the relationship between Facebook use and bonding social capital. In: Proceedings of 44th Hawaii International Conference on System Sciences (HICSS). 2011 Presented at: The 44th Hawaii International Conference on System Sciences (HICSS); Jan 4-7, 2011; Kauai, HI p. 1-10.

32. Kim J, Lee J. The Facebook paths to happiness: Effects of the number of Facebook friends and self-presentation on subjective well-being. Cyberpsychol Behav Soc Netw 2011 Jun;14(6):359-364. [doi: 10.1089/cyber.2010.0374] [Medline: 21117983]

33. Lee G, Lee J, Kwon S. Use of social-networking sites and subjective well-being: A study in South Korea. Cyberpsychol Behav Soc Netw 2011 Mar;14(3):151-155. [doi: 10.1089/cyber.2009.0382] [Medline: 20649450]

34. Zhang F, Zhu T, Li A, Li Y, Xu X. A survey of Web behaviour and mental health. In: Proceedings of the 6th International Conference on Pervasive Computing and Applications (ICPCA). 2011 Presented at: The 6th International Conference on Pervasive Computing and Applications (ICPCA); October 26-28, 2011; Port Elizabeth, South Africa p. 189-195. [doi: 10.1109/ICPCA.2011.6106503]

35. Sundar S, Oeldorf-Hirsch A, Nussbaum J, Behr R. Retirees on Facebook: Can online social networking enhance their health and wellness? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Extended Abstracts). 2011 Presented at: The SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 2011-2292.

36. Quinn D, Chen L, Mulvenna M. An examination of the behaviour of young and older users of Facebook. In: Electronic Healthcare. Berlin, Germany: Springer; 2012:9-16.

37. SurveyGizmo. URL: http://www.surveygizmo.com/ [accessed 2016-06-19] [WebCite Cache ID 6iNpF01W1]

38. Kania-Lundholm M, Torres S. The divide within: Older active ICT users position themselves against different 'Others'. J Aging Stud 2015 Dec;35:26-36. [doi: 10.1016/j.jaging.2015.07.008] [Medline: 26568212]

Abbreviations

LSI: Life Satisfaction Index
OR: odds ratio
QoL: quality of life
SNS: social network site(s)

The Implementation of Internet Interventions for Depression: A Scoping Review

Filip Drozd[1], PhD; Linda Vaskinn[2,3], MPhil; Hans Bugge Bergsund[1,3], MPhil; Silje Marie Haga[1], PhD; Kari Slinning[1,3], PhD; Cato Alexander Bjørkli[3], PhD

[1]National Network for Infant Mental Health, Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Oslo, Norway

[2]Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Oslo, Norway

[3]Department of Psychology, University of Oslo, Oslo, Norway

Corresponding Author:
Filip Drozd, PhD
National Network for Infant Mental Health
Centre for Child and Adolescent Mental Health, Eastern and Southern Norway
PO Box 4623 Nydalen
Oslo, 0405
Norway
Phone: 47 975 16 188
Fax: 47 22 58 60 01
Email: filip.drozd@r-bup.no

Abstract

Background: Depression is one of the most common mental health problems among adults, but effective treatments are not widely accessible. The Internet holds promise as a cost-effective and convenient delivery platform of interventions for depression. However, studies suggest that Internet interventions are not widely available in routine settings.

Objective: The aim of this study was to review the literature and examine whether there are systematic differences in reporting of the various implementation components on Internet interventions for depression, and then to examine what is known about and is characteristic of the implementation of these Internet interventions in regular care settings.

Methods: We performed a scoping review, drawing upon a broad range of the literature on Internet interventions for depression in regular care, and used the active implementation framework to extract data.

Results: Overall, the results suggested that knowledge about the implementation of Internet interventions for depression in regular care is limited. However, guided support from health professionals emphasizing program adherence and recruitment of end users to the interventions emerged as 2 main themes. We identified 3 additional themes among practitioners, including their qualifications, training, and supervision, but these were scarcely described in the literature. The competency drivers (ie, staff and user selection, training, and supervision) have received the most attention, while little attention has been given to organizational (ie, decision support, administration, and system intervention) and leadership drivers.

Conclusions: Research has placed little emphasis on reporting on the implementation of interventions in practice. Leadership and organizational drivers, in particular, have been largely neglected. The results of this scoping review have implications for future research and efforts to successfully implement Internet interventions for depression in regular care.

Introduction

According to the World Health Organization [1,2], about 350 million people have depression worldwide every year. However, less than 50% (in some countries, less than 10%) of those affected have access to or seek professional help [3]. Barriers to treatment include limited access to effective treatments, stigma, undertreatment, and lack of trained providers [4]. Internet interventions have been proposed as one innovate way to overcome such barriers, and systematic reviews show that people can benefit from both unguided and therapist-supported interventions [5-8]. Despite these findings, Internet interventions

are not commonly used in practice, and their uptake and actual reach, among both practitioners and users, appears to be low (eg, see [9-11]). It is, therefore, necessary to understand factors that can explain the gap between our knowledge about the effects of Internet interventions and how to translate these findings into practice.

Barriers to Uptake of Internet Interventions

A lack of availability of Internet interventions and, thus, experience with such a treatment modality has been identified as one important barrier to their uptake and use in practice [12]. Thus, it makes sense that a lack of training in the delivery of Internet interventions is another barrier [9,13]. However, recent studies showed that discrete implementation strategies, such as availability or training, would not necessarily translate an innovation into sustainable changes in practice. In a study by Friesen et al [14], 12 graduate students completed a training workshop in the delivery of Internet-based cognitive behavioral therapy (iCBT) for depression, anxiety, and panic disorder; 1 year after the workshop, each student had treated, on average, only 3 clients. In another study, conducted by Wilhelmsen and colleagues [15], only 1 of 11 general practitioners deployed the guided MoodGYM program as prescribed after training, even though the general practitioners had expressed a desire to acquire iCBT as a tool in their treatment of depression. Thus, more multifaceted strategies for implementing Internet interventions seem necessary, especially for more complex interventions and large-scale implementation.

Earlier experiences from the Improving Access to Psychological Therapies program in the United Kingdom showed that iCBT can be implemented effectively using various configurations and setups (ie, implementation strategies; [16]). However, there were insufficient data to clearly demonstrate that any one configuration was superior to another, which necessitated setups based on the needs of the local population and services (ie, bottom-up implementation). A more recent examination of iCBT in primary care trusts across England [17] identified more-specific barriers to national implementation: (1) availability of alternative interventions, (2) supporter attitudes, and (3) organizational issues such as management support, funding, and intraorganizational communication. These findings allow for a more top-down implementation and emphasize the need for multifaceted, multilevel approaches to implementation (eg, see [18]); that is, a need to deploy several implementation strategies across different levels in a health care service, to ensure successful integration of an intervention in practice. This also means having models of service delivery that describe the practical implementation of Internet interventions as a part of health care services.

Service delivery models would not only describe the practical implementation of Internet interventions, but also provide the infrastructural, legal, managerial, and institutional frameworks needed to operate and maintain Internet interventions as a service. However, service delivery models have not been adequately described in the literature [19-22], although variations of stepped-care models have been proposed (eg, see eg [23]). In stepped care, Internet interventions are suggested as a first step, while reserving more intensive and resource-demanding treatment for those who do not respond and for the most severe cases. One study in the United Kingdom found that the implementation of computer-based cognitive behavioral therapy as stepped care within a specialist service actually increased service capacity by approximately 50% [24].

There are also alternatives to stepped care, such as the centralized unit model [25]. The centralized unit, which is responsible for the Web app, training and supervision of therapists, and screening and referral of patients to practitioners, is considered to be a cost-efficient model, providing a high degree of oversight and quality control. A similar model has been used in Sweden and has helped achieve desirable results [26], albeit in a small number of clinics. However, a centralized unit model with a high degree of control may not be viable for large-scale implementation, where a decreasing degree of control and more variability in performance may be expected (eg, see [27]). Thus, more work is needed to enable integration and dissemination of Internet interventions in practice. In this regard, a first step is to map the state-of-the-art of the implementation of Internet interventions in routine practice and to identify any knowledge gaps in the literature.

Aims of This Study

The overall aim of this study was to review what is known about the implementation of Internet interventions for depression in regular care, based on the scientific literature. More specifically, the goal was, first, to examine whether there are any systematic differences in the reporting of different aspects of the implementation of Internet interventions, and thereby identify any gaps in the literature. Second was to examine what characterizes the literature on implementation of Internet interventions for depression in terms of core implementation components.

Methods

Study Design and Search Strategy

We conducted a scoping review, which has the purpose of identifying gaps in the literature by systematically assessing the breadth of a body of literature in a particular area, rather than the narrow and specific research questions typical of systematic reviews such as meta-analyses [28,29]. The search was conducted by a medical librarian, using the following scientific databases: (1) ISRCTN registry, (2) OpenGrey, (3) Ovid MEDLINE, (4) PsycINFO, (5) PubMed, (6) Web of Science, (7) World Health Organization International Clinical Trials Registry Platform, (8) CINAHL, (9) ClinicalTrials.gov, (10) Cochrane, (11) Embase, and (12) Google Scholar. Google Scholar was only used for additional searches, as it is not a traditional scientific database and has been found unsuitable for systematic literature search [30].

Search terms consisted of the combination of (1) internet, (2) intervention, and (3) depression, including synonyms for all terms (for the complete search strategy, see Multimedia Appendix 1). The inclusion of the term implementation and its synonyms (eg, adoption, integration, and dissemination) often used in the literature produced a large and unmanageable number of irrelevant search results initially (ie, 35,000–45,000 articles

per database) due to its inconsistent use and different definitions in various disciplines (also, see the Study Selection subsection below). Thus, it was not feasible to include implementation in the search strategy. The final search included references published between 1946 and March 24, 2014. After running an initial duplicate check, we imported the search results to Mendeley Desktop v1.13.8 (Mendeley Ltd). After the initial screening process, we also hand searched reference lists in identified reviews and meta-analyses, as well as relevant journals (for a list of journals, see Multimedia Appendix 2). We also contacted researchers involved in the European and international societies for research on Internet interventions [31,32].

Study Selection

Raters (HBB and LV) independently reviewed all references for eligibility based on their title, abstract, and author-provided keywords. Included references had to study (1) an Internet-based (2) intervention for (3) depression in (4) a regular care setting or (5) clearly indicate examining concepts relevant for implementation (eg, dissemination, fidelity, acceptability, and effectiveness). Systematic reviews and nonempirical references such as trial protocols, book reviews, editorials, magazine articles, and theoretical or methodological articles were excluded. Studies clearly identified as efficacy trials and offline interventions, such as desktop-based, computer-based, and CD-ROM interventions, were also excluded from this review. Efficacy trials are conducted in highly controlled settings and outside of regular care, and we did not expected them to contribute to our research questions. We included only references in English and Scandinavian languages in the coding process. In case of disagreements between the 2 reviewers, agreement was reached through discussion. Agreement between the 2 coders was estimated using Cohen kappa, resulting in a coefficient of 0.72 (95% CI 0.64–0.81), which is considered to be good [33].

Implementation Components

In order to systematically extract data, we applied the active implementation framework (AIF) developed by Fixsen et al [34]. In their comprehensive review, they identified a set of core implementation components, which they described as "the most essential and indispensable components of an implementation practice or program" [34]. These are (1) staff and client selection, (2) training, (3) supervision, (4) performance assessment, (5) decision support, (6) administrative support, (7) system intervention, and (8) leadership. These core components are considered universal and apply to all efforts of implementing an intervention in practice. However, they are also considered compensatory, such that weaknesses in one of the components may be overcome by the strengths in other components (eg, high-quality coaching and performance assessments may compensate for poor training). Thus, it is not the applied number of components (ie, the more, the better) that determines the quality of implementation, but rather the quality of how these components are carried out. We coded all references as either containing information (1 = yes) on the respective implementation components or not (0 = no). For each reference coded on an implementation component, we extracted corresponding information for the qualitative synthesis.

According to the AIF, selection, training, supervision, and performance assessments (ie, treatment fidelity) are referred to as the *competency drivers* [35]. These are concerned with the development, improvement, and sustainment of, most often, the practitioners and supervisors' abilities to work with an intervention in a competent manner. Implementation requires essentially a behavior change by means of training and coaching carefully selected staff in the initial stages of implementation whose performance is assessed (eg, how well practitioners work with the intervention). The context of an intervention also includes a clear definition of the population for whom the program is intended, and the application of inclusion and exclusion criteria to provide safer and better health services to end users. Thus, the competency drivers in this study may also pertain to selection, training, supervision, and assessment of end users.

Decision support, administrative support, and system intervention are the *organizational drivers* [35]. Organizational drivers are concerned with the planning and establishment of support systems, such that new interventions can be implemented effectively. This entails collecting data for continuous quality assurance and improvement (ie, decision support); reducing obstacles by establishing or making changes to internal policies, rules, procedures, routines, organizational culture, and climate (ie, administrative support); and developing strategies to cooperate with external systems to assure the availability of the financial, organizational, and human resources required to support and continue the intervention (ie, system intervention). Finally, the *leadership driver* is the final core component that is important in terms of setting priorities, establishing consensus, offering incentives, and managing the overall process of implementation [36].

We also coded the extracted analysis units on different organizational levels to account for multilevel approaches to implementation, based on the individual, group, leadership, and organizational levels (IGLO) framework [37]. To differentiate between individuals at the receiving end of the intervention and individuals delivering the intervention or providing supervision, we subdivided the individual level into users, practitioners, and supervisors. Certain components and organizational levels seemingly overlap (eg, supervision and leadership). However, adding the IGLO framework to the coding contributes to specifying whether the devised strategies are targeted at formal organizational structures, management, groups, or individuals, and who is the target of those strategies. So including the IGLO framework may help to distinguish units pertaining to, for example, the supervision component that describe the scope of supervision for *supervisors* and units describing who is being supervised (eg, practitioners).

Data Analysis

We used descriptive statistics to summarize the included studies. To examine whether there were any systematic differences in the reporting between implementation components, we used Cochran Q tests to account for pairwise data (ie, the same or dependent references), while we analyzed qualitative data according to the template approach to examine what

characterizes the implementation of Internet interventions in the literature [38].

A template analysis allows researchers to identify a priori themes, which are subsequently revised by theoretical concepts or perspectives that emerge during the analysis and, hence, inform the research question(s). The analysis in our study was consistent with template analysis described by King [39]: (1) defining a priori themes (ie, AIF), (2) extracting analysis units (by LV and HBB), (3) coding on a relevant theme, modifying an existing theme, or devising a new theme (LV and HBB), (4) producing the initial template (LV and FD), (5) developing the template (LV and FD), and (6) interpreting the final template (FD). Finally, each time we modified the template, we reanalyzed preceding units according to the modified template and, in order for a theme to be included in the final template, each theme had to include information from a minimum of 10 references. There are no formal procedures for determining the amount of information necessary to constitute a theme. However, it is generally important to avoid producing very narrow thematic structures and becoming too concerned with fine distinctions at lower levels of the coding hierarchy, which may not help to make sense of the data or re-present data in a disproportionate way. There is no perfect, final template, but a law of diminishing returns applies, and theme saturation is reached when continuing (re-)coding does not enrich the data. For each identified theme, we report the frequencies and percentages, and provide definitions and examples of the themes.

Results

As Figure 1 shows, the final list included 164 publications (for a complete list, see Multimedia Appendix 3). The main reason for excluding full-text articles was that they did not meet our inclusion criteria (ie, effectiveness or implementation study of an Internet intervention for depression in a regular setting). We also excluded 43 (11.9%) efficacy trials because they were conducted in a university clinic or laboratory or a research context. Hence, the results would have been explainable by the aim of these studies, since these were focused on assessing the effectiveness of interventions in a research context and would not contribute to identifying relevant information about implementation. In addition, we excluded full-text articles on the basis that they were theoretical or methodological articles (K=57, 15.8%) or unavailable in English or any of the Scandinavian languages (K=28, 7.8%).

Studies of Internet interventions for depression meeting our inclusion criteria were published from 2002 and up to the date of our final search on March 24, 2014. There was a modest increase in the number of publications during this period, with a marked increase and peak occurring in 2013 (K=51, 31.1%; Figure 2). Based on the first authors' affiliation, the majority of publications originated from Australia and the United States (Figure 3). If categorized based on geographical region, Europe (K=69, 42.1%) generated most publications, followed by North America (K=52, 31.7%), Australia (K=41, 25.0%), and Asia (K=2, 1.2%).

Figure 1. Flow diagram of selection of studies on Internet interventions for depression in regular care settings, 2002 to March 24, 2014.

XSL•FO
RenderX
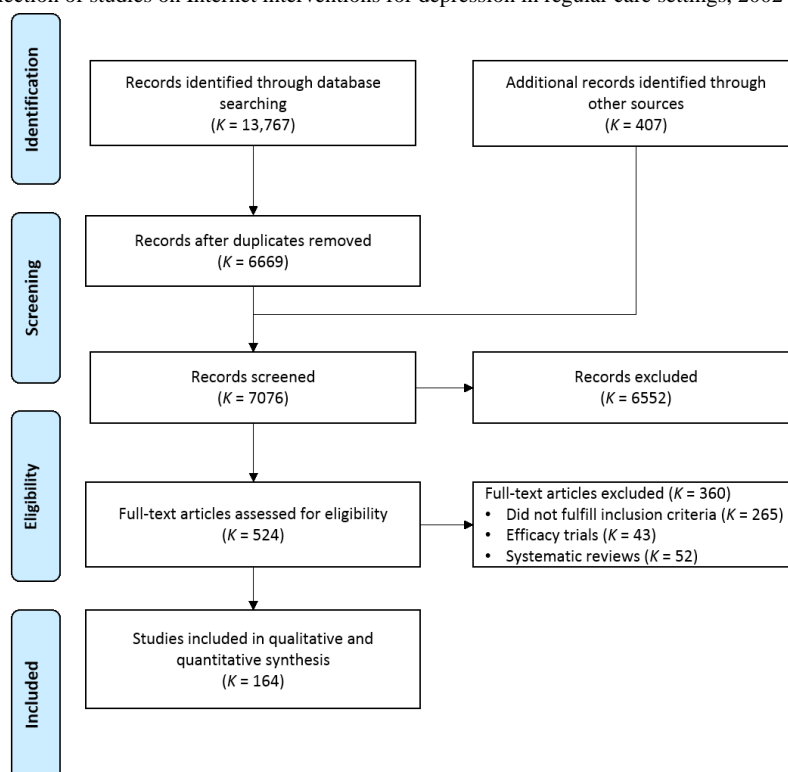
Figure 2. Number of publications per year on Internet interventions for depression in regular care settings. *The number for 2014 is up to March 24 only.
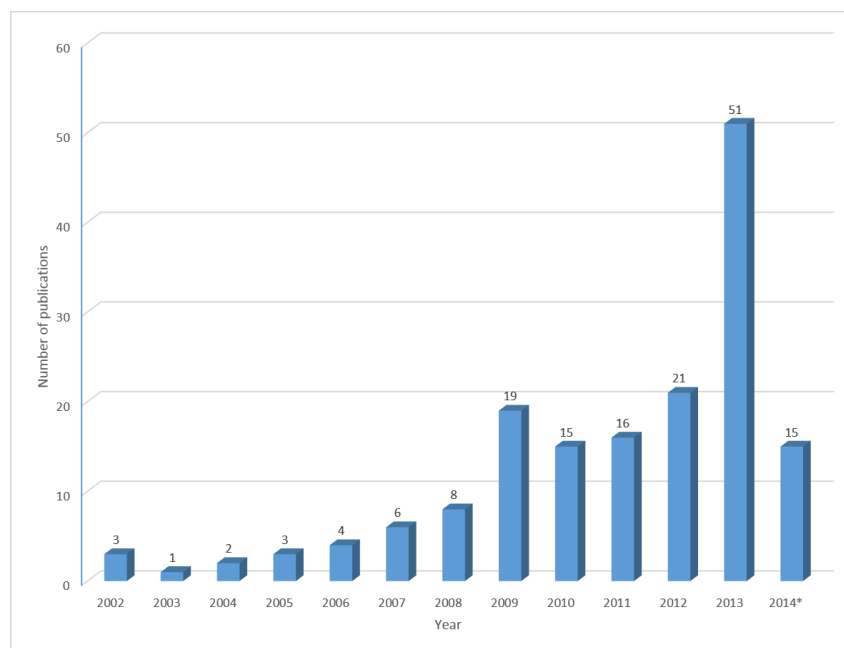
Figure 3. Number and percentage of publications on Internet interventions for depression in regular care settings, by country, 2002 to March 24, 2014.
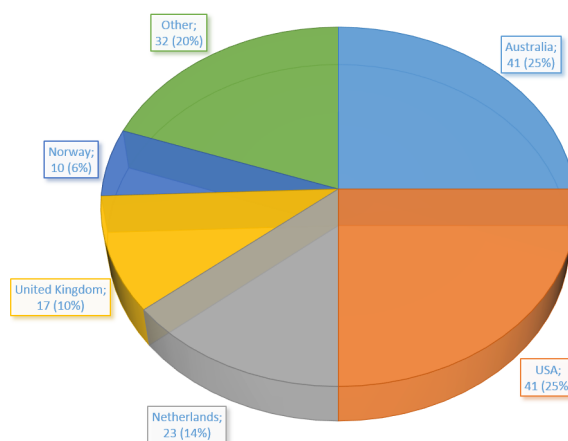
Implementation Components

Of the 164 included references, 122 (74.4%) were coded onto one or several of the implementation components, although none explicitly reported using the AIF. A Cochran Q test indicated a significant difference in the reporting of implementation components ($\chi^2_{7,N=164}$= 484.56, P<.001; see Table 1). Of the 122 references, 120 (98.4%) were coded on the competency drivers (ie, selection, training, coaching, and performance) and 13 (10.7%) on the organizational drivers (ie, administrative support, system intervention, and decision support); none of the references reported any information on aspects of leadership (Table 1). These results were also reflected in the total of 302 analysis units that were extracted, of which 281 (93.0%) units concerned the competency drivers and 21 (7.0%) units concerned the organizational drivers (Table 1).

Table 1. Number and percentage of references and units coded on the initial implementation components.

Implementation components	K [a]	%	k [b]	%
Selection	114	69.5	164	54.3
Training	28	17.1	44	14.6
Supervision	36	22.0	61	20.2
Performance	9	5.5	12	4.0
Decision support	2	1.2	4	1.3
Administrative support	7	4.3	8	2.6
System intervention	8	4.9	9	3.0
Leadership	0	0.0	0	0.0

[a]Unique references coded onto the various implementation components.

[b]Analysis units extracted from the references.

As Table 2 shows, of the 164 included references, most contained information pertaining to the selection of users for an intervention, few reported information relevant for practitioners, and almost none reported on higher levels. However, it is interesting to note that almost all information was related to the competency drivers (ie, selection, training, supervision, and performance assessments), while there was barely any information reported on the organizational drivers (ie, decision support, administration, system intervention, and leadership). This may explain the lack of information about the implementation at higher or across organizational levels (ie, beyond the practitioner level).

Table 2. Number and percentage of references (K) coded onto the initial implementation components across organizational levels.

Level	Selection, K (%)	Training, K (%)	Supervision, K (%)	Performance, K (%)	Decision support, K (%)	Administrative support, K (%)	System intervention, K (%)	Leadership, K (%)
User	100 (61.0)	8 (4.9)	28 (17.1)	2 (1.2)	0 (0.0)	1 (0.6)	1 (0.6)	0 (0.0)
Practitioner	13 (7.9)	14 (8.5)	15 (9.1)	8 (4.9)	1 (0.6)	1 (0.6)	1 (0.6)	0 (0.0)
Supervisor	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Group	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)
Leadership	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Organization	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.2)	4 (2.4)	3 (1.8)	0 (0.0)
Residual[a]	1 (0.6)	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)	4 (2.4)	0 (0.0)

[a]References not accounted for by any of the organizational levels.

Similar to the results in Table 2, the extracted analysis units show that most of the reported information pertained to the user level (Table 3). In addition, it became clear that 283 (93.7%) of the analysis units were coded either on the user or the practitioner level (ie, the individual level). This shows that the included references did not take into account a multilevel perspective on Internet interventions or examined Internet interventions from an organizational perspective. Furthermore, 281 (93.0%) of the analysis units were also coded on one of the competency drivers. This suggests that key aspects of the overall performance of the organization itself, to support and assure the continuing implementation of an intervention, and the work of practitioners and supervisors has not been adequately addressed in the literature.

XSL•FO

RenderX

Table 3. Number and percentage of analysis units (*k*) coded onto the initial implementation components across organizational levels.

Level	Selection, *k* (%)	Training, *k* (%)	Supervision, *k* (%)	Performance, *k* (%)	Decision support, *k* (%)	Administrative support, *k* (%)	System intervention, *k* (%)	Leadership, *k* (%)	Total, *k* (%)
User	122 (40.4)	8 (2.6)	33 (10.9)	4 (1.3)	0 (0.0)	1 (0.3)	1 (0.3)	0 (0.0)	168 (55.6)
Practitioner	40 (13.2)	36 (11.9)	28 (9.3)	8 (2.6)	2 (0.7)	1 (0.3)	0 (0.0)	0 (0.0)	115 (38.1)
Supervisor	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Group	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.3)
Leadership	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Organization	1 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.7)	4 (1.3)	4 (1.3)	0 (0.0)	11 (3.6)
Residual[a]	2 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	4 (1.3)	0 (0.0)	7 (2.3)
Total	164 (54.3)	44 (14.6)	61 (20.2)	12 (4.0)	4 (1.3)	8 (2.6)	9 (3.0)	0 (0.0)	302

[a]Analysis units not accounted for by any of the organizational levels.

Qualitative Synthesis

The implementation components served as a priori themes and were applied throughout the analysis to examine the characteristics of the implementation of Internet interventions for depression. After coding and revising the a priori themes, we broke the initial template down into 5 levels: (1) user, (2) practitioner, (3) group, (4) organization, and (5) residual. However, it was clear that many themes counted fewer than 10 references, indicating narrow thematic structures. Thus, we reduced the development of the final template to 2 levels and produced 5 main themes: (1) guided support and (2) user recruitment, both subsumed under the user level, and practitioner (3) qualifications, (4) training, and (5) supervision (Table 4) [16,40-52]. It is important to note that, despite the emergence of these themes, the units that these themes comprise were typically global and scarcely described in the original text. For example, Clarke et al [53] stated that they "employed the HMO's [health maintenance organization's] electronic medical record," with no further information on how they were given access to the health maintenance organization's medical records (ie, important system knowledge on access to the target population).

Table 4. The final template with meaningful themes, corresponding codes, definitions, and examples.

Level	Theme	1st-level code	2nd-level code	K [a]	%	Definition	Example
1. User				110	67.07		
	1.1. Guided support			26	15.85	An Internet-based self-help program including minimal, but regular, human involvement and support.	"Program coaches...provided motivational support to participants and clarified information contained within the program" [40].
		1.1.1. Program usage		15	9.15	Human support with guidance and direction on how to work through the intervention and its activities.	"…a weekly 10-minute telephone call from a telephone counselor. The purpose of these calls was to address any issues associated with the participants' use of the intervention" [41].
	1.2. Recruitment			101	61.59	Activities related to promoting and advertising the intervention to potential end users.	"Individuals were spontaneous visitors from around the world to an automated internet-delivered program (e-couch)" pg 344 in [42]
		1.2.1. Direct-to-consumer marketing		88	53.66	Efforts to promote the intervention directly to end users.	"Callers to Lifeline's 24 hour telephone counselling service in four major Australian cities were invited to participate in the trial by a telephone counsellor either during or at the conclusion of a counselling call" pg 2 in [43]
			1.2.1.1. Multi-channel market-ing	23	14.02	Efforts to promote the intervention directly via multiple platforms or communication channels.	"…recruited through press releases, banners and advertisements on the Internet, advertisements in magazines, referral by school doctors, through brochures and posters in schools, and through information to parents who are treated in mental health care" pg 2 in [44].
			1.2.1.2. Online	28	17.07	Efforts to promote the intervention directly using digital technologies mainly via the Internet.	"Paid advertising with Google was directed at people who had searched for a short questionnaire online to find out whether they had depression" pg 412 in [45].
			1.2.1.3. Print me-dia	21	12.80	Efforts to promote the intervention directly by means of physical publications.	"…recruited...via a screening survey posted to 70,000 adults randomly selected from the electoral rolls of eight Australian electoral divisions (4 rural, 4 metropolitan)" pg 61 in [46].
		1.2.2. Professional referral		22	13.41	Transfer or direction of end users for an intervention, both directly and indirectly, by health professionals or treatment providers.	"Participants were recruited from 11 General Practices...A member of each of the eleven participating GP [general practitioner] Practices identified searched their patient record system to identify patients…Following identification of appropriate patients, a study information pack was sent by the practice" pg 642 in [47]
			1.2.2.1. In-person	15	9.15	Direct transfer of end users for an intervention by health professionals or treatment providers.	"Two general practitioners and two psychologists, all in Sydney (Australia), referred individuals with symptoms of depression to the first author" pg 2 in [48].
2. Practitioner				46	28.05		

Level	Theme	1st-level code	2nd-level code	K [a]	%	Definition	Example
	2.1. Qualifications			37	22.56	Formal and informal background education or training, or both, among practitioners delivering Internet interventions.	"Psychological support can be provided by the following: • Graduate mental health worker • Practice nurse • GP • Assistant psychologist • Care worker • Other mental health professionals (Administrators/receptionists can offer some support to ensure that users are set up correctly on the programme but not psychological support)" pg 21 in [16].
		2.1.1. Therapists		20	12.20	A practitioner formally qualified and trained in psychological treatment methods.	"Coaches differed in their level of formal training, ranging from master's level psychology students (n=1) and psychotherapists-in-training (n=1) to experienced CBT [cognitive behavioral therapy]-trained psychotherapists with more than 10 years of professional experience (n=3)" [49].
	2.2. Training			21	12.80	Acquisition of new knowledge, skills, and abilities required to work with Internet interventions.	"Therapists were given training and a treatment manual containing a broad guide of how to respond in their reviews (supporting progress, giving encouragement, specific feedback on activities shared)" [50].
		2.2.1. Method		11	6.71	Prescribed practice or process of acquiring new knowledge, skills, and abilities needed to produce desired outcomes with the Internet intervention.	"Training involved a mix of didactics and roleplay around conducting functional analysis in perinatal-specific domains with the chief investigator (H.O.), a clinical psychologist with specialty expertise in BA [behavioral activation] and perinatal depression, and an IAPT [Improving Access to Psychological Therapies] trainer (J.W.)" pg 3 in [51].
	2.3. Supervision			11	6.71	Coaching of practitioners working with users through some form of on-the-job training.	"Comments made by the health care staff to participants in the e-mail sessions were discussed beforehand … with the CBT specialists." pg 497 in [52].

[a]Number of references coded on a theme or subtheme.

Guided Support (1.1)

Guided support emerged as 1 of 2 main themes among users derived from the a priori coaching theme, where users were provided support online or by telephone, most often by a therapist. These contact points were typically regular (eg, weekly), therapist initiated, and brief (eg, 10–20 minutes). Guided support usually involved varying forms of nonclinical supervision, such as technical support (K=1), preparations for general practitioner visits (K=1), and help with homework assignments (K=1). Most often, however, it was not possible to determine the exact purpose, methods, or contents of guided support.

Program Usage (1.1.1)

Program usage emerged as the only subtheme under guided support. This specific type of support was related to direction and guidance of users on how to work through the intervention and its activities. In other words, program usage was concerned with assuring fidelity to the intervention. There were two ways of providing support for program usage: either by (1) attending an introductory course or being briefed by practitioners early on about the operations of the intervention, or (2) receiving ongoing help when facing any problems with the intervention or certain tasks.

XSL•FO
RenderX

Recruitment (1.2)

Recruitment was derived from the a priori selection theme that was concerned with activities related to selecting end users, practitioners, and organizations to use or work with an intervention. However, Table 3 shows, 122 (74.4%) of 164 units were related to end users. Thus, information initially pertaining to selection across multiple organizational levels was conceptualized as user recruitment and divided in 2 first-level subthemes: (1) direct-to-consumer marketing (DTC) and (2) professional referral.

Direct-to-Consumer Marketing (1.2.1)

DTC marketing refers to recruitment activities aimed directly at end users (ie, the consumer). There was a wide range of DTC recruitment strategies, from counseling services (K=4) to organizations (K=2) and school settings (K=2). However, a multichannel strategy, online recruitment, and recruitment using print media were most common and emerged as second-level subthemes.

Multichannel Marketing (1.2.1.1)

Multichannel DTC marketing involved the combination of two or more recruitment strategies. The studies using multichannel marketing often recruited users from a wider population and used more targeted marketing efforts. For example, in a study by Haga and colleagues [54], the researchers recruited pregnant women through midwives and public health nurses in well-baby clinics and hospitals, who, in turn, handed out brochures about the study and intervention. At the same time, pregnant women were also recruited through social media (ie, Facebook).

Online (1.2.1.2)

Online DTC recruitment strategies have mainly consisted of using and testing ads and banners, such as in the study of Barrera et al [55]. They examined the impact of Spanish and English keywords for a Google AdWords campaign to recruit pregnant women and found that broad descriptive words related to pregnancy, health, and distress resulted in higher international enrollment rates. For most recruitment strategies, however, more geographically targeted online advertisements may be necessary, as investigated by Jones and colleagues [56]. Interestingly, they found that between one-third and half of the ads were wrongly targeted by AdWords to nearby postal code areas. In a follow-up study [57], AdWords location targeting was still found to be more effective than posting ads at local organization websites, despite the misdirected ads. Organization websites may still be effective but need to be advertised through trustworthy, relevant, and familiar mental health organizations [58].

Print Media (1.2.1.3)

Use of print media consisted of ads and articles in national and local newspapers, and invitation letters by postal mail (eg, questionnaires, brochures, or study information). In contrast to multichannel marketing, marketing through print media mostly recruited users from the general public. Only 3 studies using print media appear to have used a more targeted approach [47,59,60]. For example, in the study of Woodford et al [47], general practices searched their patient records to identify patients with a diagnosis of depression or who may have experienced mild to moderate depression over the last 6 months.

Professional Referral (1.2.2)

Professional referral is the second first-level subtheme that emerged under the main theme selection. Referrals were either a part of a multichannel marketing strategy or, most often, direct in-person referrals to an intervention. According to the AIF, routines for referral are usually related to system intervention because they entail collaboration with external agencies such as general practitioners. However, none of the articles using referrals contained any information on how interorganizational agreements and routines for referrals were established or evaluated, or how these were embedded in the larger system. Thus, we coded and analyzed these units only as a form of recruitment procedure.

In-Person (1.2.2.1)

Direct in-person referrals were most common where, for instance, patients were prescribed an Internet intervention directly by their general physician or mental health specialist (eg, see [61]).

Qualifications (2.1)

We identified qualifications as the first main theme among practitioners, derived from the selection theme. Of the 38 studies, 36 (94.7%) involved practitioners that either had a completed college or university degree or were in their last year of a formal training program, most of whom were therapists (see below). In 11 (28.9%) of the 38 studies, interventions were delivered by medical staff consisting of either nurses or general practitioners, or both, and, in 7 (18.4%) studies, interventions were delivered by various practitioners such as school teachers, mental health workers, or occupational health staff. Interestingly, according to 2 (5.3%) studies, such formal qualifications may not be necessary, and it appears that laypersons may administer Internet interventions just as effectively as therapists or mental health workers [62,63].

Therapists (2.1.1)

Of the 38 studies, 20 (52.6%) reported the use of therapists to deliver the interventions. Psychologists participated in 13 (34.2%) studies and an additional 2 (5.3%) studies involved psychologists in combination with other mental health professionals. The remaining 5 (13.5%) studies either involved mental health workers or did not specify the therapists' formal training background.

Training (2.2)

Training emerged as the second main theme among practitioners, with 1 subtheme relating to how practitioners were trained in the administration of the Internet intervention (ie, method; Table 4). Of the remaining studies, 9 (23.7%) of 21 reported on the scope of training (ie, ranging from brief 1-hour training sessions to 5 days of training); 4 (17.1%) reported providing special training in the skills required to administer the Internet intervention, such as electronic, text-based communication (eg, see pg 210 in [64]); 3 (14.3%) mentioned that training was provided by either the intervention developers or the principal study investigator; and 1 (4.8%) arranged educational sessions.

XSL•FO

RenderX

In addition, 4 (17.1%) studies noted in passing that practitioners received training, but without providing any further information.

Method (2.2.1)

A variety of methods were used to train practitioners. Of 11 studies, 4 (36.4%) reported the use of video demonstrations (eg, see pg 186 in [65]). In 3 (27.2%) other studies, practitioners reviewed the contents of the intervention in order to adequately address participants' questions and provide assistance with tasks and activities. Of 11 studies, 2 (18.1%) used a mix of didactics and practice in, for example, cognitive behavioral skills or responding to clinical emergencies (eg, pg 743 in [63]). Also, 1 (9.1%) study had a specialist in cognitive behavioral therapy provide a 3-hour lecture about it to practitioners, while 1 (9.1%) study provided newly educated psychologists with additional training in delivering the specific treatment manuals.

Supervision (2.3)

Supervision was the third main theme among practitioners that we derived from the initial coaching component. Information was mostly concerned with *the who* (K=8, 44.5%) and *the extent* (K=6, 33.3%) of supervision. That is, supervision was mostly provided regularly (eg, weekly) by psychologists or therapists. Beyond that, supervision was used for case management in 3 (16.7%) of the 18 studies (eg, discuss practitioners' response to users in email sessions; see pg 497 in [52]), and 1 study reported using supervision to develop interview scripts for users [66].

Discussion

Principal Findings

The aim of this study was twofold: first, to examine whether there are any systematic differences in the implementation of Internet interventions for depression in the literature in terms of core implementation components and, thereby, identify any knowledge gaps; and second, to examine what characterizes the implementation in the literature. In total, we identified 164 references, of which 122 (74.4%) were coded onto the AIF. Overall, the results show that no studies had any hard data about which components are critical for implementation and which components may be adapted without compromising intervention outcomes in regular practice. Information related to the competency drivers (ie, selection, training, and supervision) was most frequently reported; however, in terms of the organizational drivers, fewer than 10 references were coded onto decision support, administrative support, and system intervention. No studies contained any information related to leadership.

Competency Drivers

Our results revealed that studies concerned with selection were focused on the recruitment of users for the intervention or study, rather than on finding the right personnel or organizations to carry out or support the new intervention [34]. This likely reflects a common practice of reporting on participant recruitment in studies (eg, see the CONSORT statement [67]), rather than that research has been genuinely concerned with investigating various recruitment strategies. This is supported by the few studies identified to actually investigate recruitment processes (eg, see [55,56]). Nevertheless, print media and online recruitment strategies were typically used, although a multichannel marketing strategy was most common. The predominance of DTC strategies (ie, self-referrals) supports the notion that Internet interventions for depression have yet to become an integral part of routine care.

Our review found that staff selection has not been studied, but that almost all practitioners had higher education in psychological, medical, or other health sciences. We also found that formal qualifications may not be necessary to administer Internet interventions effectively. However, regardless of qualifications, a strong and active implementation strategy, which integrates and addresses all of the implementation components, is important to maintain their quality and effectiveness [68]. As such, a lack of formal qualifications and practitioner heterogeneity may be compensated for by, for example, receiving high-quality training and supervision from highly competent and experienced practitioners. Training was, however, typically brief and consisted of lectures, videos, and written materials (eg, program review or treatment manuals). This does not tell much about the quality of training, and it can be argued that the technology does most of the work of delivering the interventions and, thus, extensive and complex training may not be required. However, a few small-scale studies have shown that brief training is insufficient to sustain changes in practice over time [14,15], while previous studies have demonstrated that frequently used training methods, such as reviewing treatment manuals, are not necessarily efficient for acquiring a new set of skills [69,70].

Supervision may compensate for brief training and has been shown to increase practitioner behavior change [71]. This may be particularly true when regular, ongoing supervision is provided by highly educated and experienced supervisors, which makes it possible for practitioners to embed new clinical skills into their existing repertoire and ongoing work. This may, however, depend on highly qualified and skilled supervisors and the methods that are used during supervision. There are many various labels for supervision, such as consultation [71], coaching [72], and auditing [73]. However, such implementation strategies were rarely defined, and often inadequately described in the literature. This also seemed to apply to guided support, which emerged as a theme among end users. Thus, it remains unclear what therapist support consisted of, and it appears that the contents of the therapist support that is provided are heterogeneous. Of 31 (18.9%) studies, only 1 subtheme—program usage—emerged. We could not determine whether this is because the purpose and clinical guidelines for therapist support are inadequately described in the literature, or whether the heterogeneity in therapist support is real.

Organizational Drivers

It is important to acknowledge that competency drivers do not exist in a vacuum, but rely on and are supported by an organization that provides management and administrative structures, and relates to external systems (ie, service delivery models), all of which can affect the implementation. However, there were no emerging themes among any of the organizational

drivers, despite the importance of components, such as decision-support systems and leadership, for improving clinical practice [74,75]. This does not mean that systems for decision making do not exist or that aspects of leadership have not been addressed in practice. It simply reflects that no studies have properly assessed these implementation components or implementation processes and quality more generally. Some studies did, however, report information related to administrative support (*K*=7) and system intervention (*K*=8), although there was insufficient information for any themes to emerge. However, Andersson and Hedman [19] identified several issues related to the implementation of iCBT in practice, and which are highly relevant for the administration of Internet interventions and system-related work with external agencies: (1) data security, (2) robust Web solutions, (3) online assessment procedures (including diagnostic interviews), (4) referral routes, patient management, and outcome monitoring, (5) the role of professional organizations, and (6) development of clearly formulated policies, procedures, and practice guidelines (see also [14]). In addition, Titov and colleagues [22] suggested technical support and legislation across federal, state, and international laws as barriers that become more actualized with Internet interventions. Most of these issues, however, have not been studied, except for referrals, in particular self-referrals, which are likely to affect the uptake of and, possibly, adherence to iCBT [76,77].

General Discussion

It is important to establish a robust evidence base for Internet interventions, but it is equally important to establish a robust evidence base for the delivery of Internet interventions in practice, by moving beyond studies of efficacy and effectiveness to implementation. This is a necessary step to scale up the dissemination and integration of Internet interventions in routine practice, and ultimately to provide better and safer health care services. Limited reporting on the different implementation components limits the value of these studies for decision makers and other stakeholders, as most of the studies did not include sufficiently relevant information to understand how to translate these results into practice. The lack of emphasis on organizational drivers, in particular, may impede effective implementation. Thus, stakeholders have to rely on and become dependent on the know-how of the relatively few communities working with Internet-based prevention and treatment of depression (ie, Australia, United States, the Netherlands, Sweden, United Kingdom, Canada, and Norway).

Implementation of Internet interventions in routine care is still in its infancy, and there is no strong evidence or methods for transferring Internet-based prevention and treatment to service delivery settings (see also [78]). Service delivery models have not been adequately developed or tested, even though governments and professional societies in several countries are recommending Internet interventions in their national guidelines (eg, Australia [79], United Kingdom [80], Sweden [81], and Norway [82]). Yet the obvious question is how to integrate Internet interventions into new or existing health care services for large-scale implementation [22]. Stepped-care models and other service delivery models have been proposed [23,25] but should be appraised critically. Technological advances and novel practices may not fit with existing models of health care service delivery and may need to be redefined [21]. Already in 2009, Bennett and Glasgow noted [20] that there has been relatively little discussion of contextual issues in eHealth. The results of our review suggest that this has not changed much since then. To date, studies have largely focused on testing the effects of Internet interventions on users, while more work remains to understand the organizational, systemwide, and contextual features of implementation. Thus, the important future lessons for Internet interventions are really those concerning the knowledge transfer from science to practice. This will support governments, researchers, and other stakeholders in implementing effective Internet interventions in practice, replicating studies, conducting independent research, building competence, and driving development of Internet interventions. Currently, however, Governments, researchers, and others are dependent on the few existing experts and research milieus in this field for the implementation of Internet interventions.

Strengths and Limitations

This study has several strengths and limitations. First, we used the scoping review methodology, which is effective in mapping the state-of-the-art and identifying gaps in the literature. However, in line with the methodology, we emphasized the breadth rather than depth of knowledge and did not assess the quality of the included studies [83].

Second, we constructed a comprehensive search strategy, but Internet interventions and implementation are relatively new areas of scientific inquiry. Thus, there is a tremendously wide range of terminology, which prevented us from combining search terms for Internet interventions and depression with implementation, which produced a large and unmanageable number of search results. Furthermore, the terms efficacy and effectiveness are sometimes used interchangeably, and it may be difficult to distinguish whether unguided interventions should be classified as efficacy or effectiveness trials. In contrast to guided interventions, unguided interventions may be offered directly to users from university clinics, private companies, or online, without being implemented in a health care setting (eg, see [84]). Consequently, we may have missed some relevant articles or included some efficacy trials.

Third, the extensive number of included studies and text material in this scoping review also means that we, most likely, have missed information relevant for implementation. This is mainly because none of the studies were de facto implementation studies, which means that much of the reported information has been scarce and thereby ambiguous. This has occasionally made it difficult to identify, assess, and code information from the studies, and has, most certainly, left some information unidentified or incorrectly classified and analyzed. However, the extensive number of studies and text material also means that a larger number of studies and data would be needed for substantial changes to occur in the results.

Fourth, we did not consider that publications are nested within authors. Different authors may emphasize and report on different aspects of the implementation of Internet interventions, and publish several articles based on 1 study. An author may also

vary in his or her influence on publications depending on their role and contribution (eg, principal investigator versus supervisor).

Fifth, we used Fixsen and colleagues' [34] AIF. Other implementation theories and models may have identified other types of relevant information and thus provided different results. According to Tabak et al [85], the AIF is more concerned with integrating evidence-based practices within a setting than with disseminating them to the target audience via determined channels using a planned strategy. One of the strengths of this study, however, is that we applied the AIF in a flexible manner and revised the model through the template analysis. As our results show, dissemination of evidence-based practices to target audiences is not adequately addressed in the AIF, which resulted in recruitment emerging as a main theme at the user level. Furthermore, the AIF does not operate at the policy level or higher socioecological levels (eg, government). The conclusions in this review must be interpreted in the light of these limitations.

Future Research Directions

This scoping review has highlighted several important issues for future research. First, for Internet interventions for depression to become more widespread and embedded in regular practice, it is necessary to move from studies of efficacy and effectiveness to implementation. There is a clear need for more primary implementation studies that are based on clearly defined models and theories of implementation (for overview, see [85]), and preferably that link implementation outcomes with intervention outcomes. This will help distinguish between what is known about an effective treatment (ie, studies of effectiveness) and what is actually benefiting clients and providing safer and better health care services (ie, implementation [86]).

Second, there is an urgent need to improve reporting guidelines such as the eHealth CONSORT statement [87]. The lessons learned about implementation in randomized trials and other studies can increase sharply, simply by requiring that future studies regularly and systematically describe the implementation of the intervention.

Third, there is a need for experimental studies on the effects of specific implementation strategies—for example, which formal and informal qualifications are important for administration of Internet interventions among practitioners or what recruitment strategies are likely to be more efficient. However, probably the most striking gap in the literature is the lack of investigation of the interaction *between* the different implementation components and their relative influence *over time* [34]—for example, the relationship between the amount of training and duration of supervision necessary to administer an intervention in a competent and skillful manner.

Fourth, more research is needed at different organizational levels of implementation, including leadership and management practices. Much of the available information on implementation pertains to end users, and much less is known about the practitioners, organizations, and systems within which interventions are embedded.

Fifth, we would encourage authors and journals to routinely publish implementation protocols similar to study or intervention protocols. This would provide a greater understanding of *what* activities are necessary and *how* these activities need to be carried out to (re-)produce the achieved results from any given trial. It is, however, also important that implementation protocols use standardized reporting guidelines such as the assessment of transferability and adaptation of health promotion interventions (ASTAIRE) [88] to, among other things, ensure that they are directly applicable in practice. Implementation protocols may help explain why some intervention trials succeed and others do not and, most important, they would support independent research and knowledge transfer between different research communities and contexts.

Conclusions

This review aimed at investigating what is known about the implementation of Internet interventions for depression in the literature. Overall, the results showed that limited emphasis has been given to their implementation in practice and that leadership and organizational drivers have been largely neglected. Recruiting users for the interventions was, by far, most commonly reported and typically carried out by the use of print media, online recruitment, or multichannel marketing strategies and, to some extent, professional referrals. Therapist support to ensure program usage was also characteristic of Internet interventions and, although brief training and regular supervision may be sufficient for administering Internet interventions, more research is needed.

The Internet holds promise as an effective platform for the delivery of interventions for depression. However, to progress and make Internet interventions more widely available, it is of utmost importance that the field prioritize implementation practice and research, and move beyond studies of efficacy and effectiveness. Only by allocating research efforts to implementation may the field be able to provide stakeholders and decision makers with knowledge of which strategies promote effective implementation and, consequently, provide better health services for the target population.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 28KB - jmir_v18i9e236_app1.pdf]

Multimedia Appendix 2

Complete list of hand-searched journals.

[PDF File (Adobe PDF File), 323KB - jmir_v18i9e236_app2.pdf]

Multimedia Appendix 3

List of references included for analysis.

[PDF File (Adobe PDF File), 361KB - jmir_v18i9e236_app3.pdf]

References

1. Marcus M, Yasamy M, van Ommeren O, Chisholm D, Saxena S. Depression: A Global Public Health Concern. Geneva: World Health Organization; 2012. URL: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf [accessed 2016-08-26] [WebCite Cache ID 6k3NyiqUb]
2. World Health Organization. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. Geneva: World Health Organization; 2009. URL: http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf [accessed 2016-08-26] [WebCite Cache ID 6k3O10miT]
3. World Health Organization. Depression. 2010. URL: http://www.who.int/mental_health/management/depression/en/ [accessed 2016-08-26] [WebCite Cache ID 6k2ta0zjW]
4. Overland S, Glozier N, Krokstad S, Mykletun A. Undertreatment before the award of a disability pension for mental illness: The HUNT Study. Psychiatr Serv 2007;58:82.
5. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
6. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
7. Cuijpers P, Donker T, Johansson R, Mohr D, van Straten A, Andersson G. Self-guided psychological treatment for depressive symptoms: a meta-analysis. PLoS One 2011;6(6):e21274. [Medline: 21712998]
8. 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]
9. Nordgreen T, Havik OE. Use of self-help materials for anxiety and depression in mental health services: a national survey of psychologists in Norway. Prof Psychol Res Pract 2011;42(2):185-191. [doi: 10.1037/a0022729]
10. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
11. Whitfield G, Williams C. If the evidence is so good, why doesn't anyone use them? A national survey of the use of computerized cognitive behaviour therapy. Behav Cogn Psychother 2004;32:57-65.
12. Carper MM, McHugh RK, Barlow DH. The dissemination of computer-based psychological treatment: a preliminary analysis of patient and clinician perceptions. Adm Policy Ment Health 2013 Mar;40(2):87-95. [doi: 10.1007/s10488-011-0377-5] [Medline: 22001968]
13. Kurki M, Koivunen M, Anttila M, Hätönen H, Välimäki M. Usefulness of internet in adolescent mental health outpatient care. J Psychiatr Ment Health Nurs 2011 Apr;18(3):265-273. [doi: 10.1111/j.1365-2850.2010.01661.x] [Medline: 21395918]
14. Friesen LN, Hadjistavropoulos HD, Pugh NE. A qualitative examination of psychology graduate students' experiences with guided Internet-delivered cognitive behaviour therapy. Internet Intervent 2014 Apr;1(2):41-48. [doi: 10.1016/j.invent.2014.04.001]
15. Wilhelmsen M, Høifødt RS, Kolstrup N, Waterloo K, Eisemann M, Chenhall R, et al. Norwegian general practitioners' perspectives on implementation of a guided web-based cognitive behavioral therapy for depression: a qualitative study. J Med Internet Res 2014 Sep;16(9):e208 [FREE Full text] [doi: 10.2196/jmir.3556] [Medline: 25208886]
16. Improving Access to Psychological Therapies (IAPT) Programme. Computerised Cognitive Behavioural Therapy (cCBT) Implementation Guidance. London, UK: Department of Health; 2007.

XSL•FO
RenderX

17. Palmili L. Factors Influencing National Implementation of Online Computer-Aided Cognitive Behaviour Therpay in Primary Care Trusts Across England [doctoral thesis]. London, UK: King's College London; 2013.

18. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Q 2001 May;79(2):281-315. [doi: 10.1111/1468-0009.00206]

19. Andersson G, Hedman E. Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013;23(3):140-148.

20. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009 Apr;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

21. 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]

22. Titov N, Andrews G, Sachdev P. Computer-delivered cognitive behavioural therapy: effective and getting ready for dissemination. F1000 Med Rep 2010 Jul;2:49 [FREE Full text] [doi: 10.3410/M2-49] [Medline: 20948835]

23. Green KE, Iverson KM. Computerized cognitive-behavioral therapy in a stepped care model of treatment. Prof Psychol Res Pract 2009;40(1):96-103. [doi: 10.1037/a0012847]

24. Learmonth D, Trosh J, Rai S, Sewell J, Cavanagh K. The role of computer-aided psychotherapy within an NHS CBT specialist service. Couns Psychother Res 2008 Jun;8(2):117-123. [doi: 10.1080/14733140801976290]

25. Hadjistavropoulos HD, Pugh NE, Nugent MM, Hesser H, Andersson G, Ivanov M, et al. Therapist-assisted Internet-delivered cognitive behavior therapy for depression and anxiety: translating evidence into clinical practice. J Anxiety Disord 2014 Dec;28(8):884-893 [FREE Full text] [doi: 10.1016/j.janxdis.2014.09.018] [Medline: 25445078]

26. Andersson G. Using the Internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180. [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]

27. Eisen J, Marko-Holguin M, Fogel J, Cardenas A, Bahn M, Bradford N, et al. Pilot study of implementation of an internet-based depression prevention intervention (CATCH-IT) for adolescents in 12 US primary care practices: clinical and management/organizational behavioral perspectives. Prim Care Companion CNS Disord 2013;15(6):e1-e11. [doi: 10.4088/PCC.10m01065] [Medline: 24800110]

28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

29. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

30. Boeker M, Vach W, Motschall E. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Med Res Methodol 2013;13:131 [FREE Full text] [doi: 10.1186/1471-2288-13-131] [Medline: 24160679]

31. European Society for Research on Internet Interventions. Stockholm, Sweden: ESRII; 2015. URL: http://esrii.org/ [accessed 2016-08-26] [WebCite Cache ID 6k2vc3k8U]

32. International Society for Research on Internet Interventions. Sanford, NC: ISRII; 2015. URL: http://isrii.org/ [accessed 2016-08-26] [WebCite Cache ID 6k2vk4s3p]

33. Peat J. Health Science Research: A Handbook of Quantitative Methods. Sydney: Allen & Unwin; 2001.

34. Fixsen D, Naoom S, Blase K, Friedman RM, Wallace F. Implementation Research: A Synthesis of the Literature (FMHI Publication #231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005. URL: http://cfs.cbcs.usf.edu/_docs/publications/NIRN_Monograph_Full.pdf [accessed 2016-08-26] [WebCite Cache ID 6k3QgqyXe]

35. Fixsen D, Blase K, Naoom S, Duda M. Implementation Drivers: Assessing Best Practices. Chapel Hill, NC: National Implementation Science Network; 2013.

36. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008 Jun;41(3-4):327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]

37. Nielsen K, Stage M, Abildgaard J, Brauer CV. Participatory intervention from an organizational perspectiveemployees as active agents in creating a healthy work environment. In: Bauer GF, Jenny GJ, editors. Salutogenic Organizations and Change: The Concepts Behind Organizational Health Intervention Research. Dordrecht, NL: Springer Science; 2013:327-350.

38. King N. Template analysis. In: Symon G, Cassell C, editors. Qualitative Methods and Analysis in Organizational Research: A Practical Guide. Thousand Oaks, CA: Sage Publications; 1998:118-134.

39. King N. Template Analysis Technique. URL: http://www.hud.ac.uk/hhs/research/template-analysis/technique/ [accessed 2016-08-26] [WebCite Cache ID 6k3Qr2UgC]

40. Wojtowicz M, Day V, McGrath PJ. Predictors of participant retention in a guided online self-help program for university students: prospective cohort study. J Med Internet Res 2013;15(5):e96 [FREE Full text] [doi: 10.2196/jmir.2323] [Medline: 23697614]

41. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Web-based cognitive behavior therapy for depression with and without telephone tracking in a national helpline: secondary outcomes from a randomized controlled trial. J Med Internet Res 2012;14(3):e68 [FREE Full text] [doi: 10.2196/jmir.1859] [Medline: 22738715]

42. Donker T, Batterham P, 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. [Medline: 23953024]

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

44. Hoek W, Schuurmans J, Koot H, Cuijpers P. Effects of Internet-based guided self-help problem-solving therapy for adolescents with depression and anxiety: a randomized controlled trial. PLoS One 2012;7(8):e43485 [FREE Full text] [doi: 10.1371/journal.pone.0043485] [Medline: 22952691]

45. Morgan AJ, Jorm AF, Mackinnon AJ. Email-based promotion of self-help for subthreshold depression: Mood Memos randomised controlled trial. Br J Psychiatry 2012 May;200(5):412-418 [FREE Full text] [doi: 10.1192/bjp.bp.111.101394] [Medline: 22442102]

46. Crisp D, Griffiths K, Mackinnon A, Bennett K, Christensen H. An online intervention for reducing depressive symptoms: secondary benefits for self-esteem, empowerment and quality of life. Psychiatry Res 2014 Apr 30;216(1):60-66. [doi: 10.1016/j.psychres.2014.01.041] [Medline: 24534125]

47. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011 Sep;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]

48. Jacmon J, Malouff J, Taylor N. Treatment of major depression: effectiveness of cognitive-behavioural therapy with an internet course as a central component. E-Journal Appl Psychol 2009;5(2):1-8.

49. Ebert DD, Gollwitzer M, Riper H, Cuijpers P, Baumeister H, Berking M. For whom does it work? Moderators of outcome on the effect of a transdiagnostic internet-based maintenance treatment after inpatient psychotherapy: randomized controlled trial. J Med Internet Res 2013 Aug;15(10):e191 [FREE Full text] [doi: 10.2196/jmir.2511] [Medline: 24113764]

50. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. J Med Internet Res 2013;15(6):e121 [FREE Full text] [doi: 10.2196/jmir.2248] [Medline: 23807565]

51. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]

52. Kojima R, Fujisawa D, Tajima M. Efficacy of cognitive behavioral therapy training using brief e-mail sessions in the workplace: a controlled clinical trial. Ind Health 2010;48(4):495-502.

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

54. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]

55. Barrera AZ, Kelman AR, Muñoz RF. Keywords to recruit Spanish- and English-speaking participants: evidence from an online postpartum depression randomized controlled trial. J Med Internet Res 2014 Jan;16(1):e6 [FREE Full text] [doi: 10.2196/jmir.2999] [Medline: 24407163]

56. Jones RB, Goldsmith L, Williams CJ, Kamel Boulos MN. Accuracy of geographically targeted internet advertisements on Google AdWords for recruitment in a randomized trial. J Med Internet Res 2012;14(3):e84 [FREE Full text] [doi: 10.2196/jmir.1991] [Medline: 22718043]

57. Jones RB, Goldsmith L, Hewson P, Williams CJ. Recruitment to online therapies for depression: pilot cluster randomized controlled trial. J Med Internet Res 2013;15(3):e45 [FREE Full text] [doi: 10.2196/jmir.2367] [Medline: 23462072]

58. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]

59. Hunkeler EM, Hargreaves WA, Fireman B, Terdiman J, Meresman JF, Porterfield Y, et al. A web-delivered care management and patient self-management program for recurrent depression: a randomized trial. Psychiatr Serv 2012 Nov;63(11):1063-1071. [doi: 10.1176/appi.ps.005332011] [Medline: 22983558]

60. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59. [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]

61. Sunderland M, Wong N, Hilvert-Bruce Z, Andrews G. Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. Behav Res Ther 2012 Jun;50(6):374-380. [doi: 10.1016/j.brat.2012.03.005] [Medline: 22498311]

62. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]

63. Sheeber LB, Seeley JR, Feil EG, Davis B, Sorensen E, Kosty DB, et al. Development and pilot evaluation of an Internet-facilitated cognitive-behavioral intervention for maternal depression. J Consult Clin Psychol 2012 Oct;80(5):739-749 [FREE Full text] [doi: 10.1037/a0028820] [Medline: 22663903]

64. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]

65. Van Voorhees BW, Vanderplough-Booth K, Fogel J, Gladstone T, Bell C, Stuart S, et al. Integrative internet-based depression prevention for adolescents: a randomized clinical trial in primary care for vulnerability and protective factors. J Can Acad Child Adolesc Psychiatry 2008 Nov;17(4):184-196 [FREE Full text] [Medline: 19018321]

66. Van Voorhees BW, Ellis J, Stuart S, Fogel J, Ford DE. Pilot study of a primary care internet-based depression prevention intervention for late adolescents. Can Child Adolesc Psychiatr Rev 2005 May;14(2):40-43 [FREE Full text] [Medline: 19030514]

67. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;63(8):e1-e37. [doi: 10.1016/j.jclinepi.2010.03.004]

68. Tommeraas T, Ogden T. Is there a scale-up penalty? Testing behavioral change in the scaling up of Parent Management Training in Norway. Adm Policy Ment Health 2015 Dec 29:1-14. [doi: 10.1007/s10488-015-0712-3] [Medline: 26715496]

69. Kendall PC, Beidas RS. Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. Prof Psychol Res Pract 2007;38(1):13-20. [doi: 10.1037/0735-7028.38.1.13]

70. Miller WR, Sorensen JL, Selzer JA, Brigham GS. Disseminating evidence-based practices in substance abuse treatment: a review with suggestions. J Subst Abuse Treat 2006 Jul;31(1):25-39. [doi: 10.1016/j.jsat.2006.03.005] [Medline: 16814008]

71. Beidas RS, Edmunds JM, Marcus SC, Kendall PC. Training and consultation to promote implementation of an empirically supported treatment: a randomized trial. Psychiatr Serv 2012;63(7):660-665.

72. Vale MJ, Jelinek MV, Best JD, Santamaria JD. Coaching patients with coronary heart disease to achieve the target cholesterol:A method to bridge the gap between evidence-based medicine and the "real world"—randomized controlled trial. J Clin Epidemiol 2002 Mar;55(3):245-252. [doi: 10.1016/S0895-4356(01)00460-7]

73. Dulko D. Audit and feedback as a clinical practice guideline implementation strategy: a model for acute care nurse practitioners. Worldviews Evid Based Nurs 2007;4(4):200-209. [doi: 10.1111/j.1741-6787.2007.00098.x] [Medline: 18076463]

74. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]

75. Klein KJ, Conn AB, Sorra JS. Implementing computerized technology: an organizational analysis. J Appl Psychol 2001 Oct;86(5):811-824. [Medline: 11596799]

76. Andersson G, Bergström J, Buhrman M, Carlbring P, Holländare F, Kaldo V, et al. Development of a new approach to guided self-help via the Internet: the Swedish experience. J Technol Hum Serv 2008;26(2-4):161-181.

77. Cavanagh K, Seccombe N, Lidbetter N. The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self help clinic. Behav Cogn Psychother 2011;39(4):427-442.

78. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: A model for dissemination into primary care. Clin Psychol Rev 2014 Jun 11:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]

79. Australian Psychological Society. Guidelines for Providing Psychological Services and Products Using the Internet and telecommunications technologies. Melbourne, Australia: Australian Psychological Society; 2011.

80. National Institute for Health and Care Excellence. Depression in Adults: Recognition and Management. London, UK: National Institute for Health and Care Excellence; 2009.

81. The National Board of Health and Welfare. Nationella riktlinjer för vård vid depression och ångestsyndrom 2010 - stöd för styrning och ledning [National guidelines for care in cases of depression and anxiety disorders]. Stockholm, Sweden: The National Board of Health and Welfare; 2010.

82. Norwegian Directorate of Health. Nasjonale retningslinjer for diagnostisering og behandling av voksne med depresjon i primær- og spesialisthelsetjenesten [National Guidelines for Diagnosis and Treatment of Adults With Depression in Primary and Secondary Care]. Oslo, Norway: Norwegian Directorate of Health; 2009.

83. Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]

84. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]

85. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]

86. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
87. 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]
88. Cambon L, Minary L, Ridde V, Alla F. A tool to analyze the transferability of health promotion interventions. BMC Public Health 2013;13:1184 [FREE Full text] [doi: 10.1186/1471-2458-13-1184] [Medline: 24341441]

Abbreviations

AIF: active implementation framework
ASTAIRE: assessment of transferability and adaptation of health promotion interventions
DTC: direct-to-consumer marketing
iCBT: Internet-based cognitive behavior therapy
IGLO: individual, group, leadership, and organizational levels

XSL•FO
RenderX

Original Paper

Tracking Dabbing Using Search Query Surveillance: A Case Study in the United States

Zhu Zhang[1], PhD; Xiaolong Zheng[2], PhD; Daniel Dajun Zeng[1,2], PhD; Scott J Leischow[3], PhD

[1]Department of Management Information Systems, The University of Arizona, Tucson, AZ, United States

[2]State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[3]Department of Health Sciences Research, Mayo Clinic Arizona, Scottsdale, AZ, United States

Corresponding Author:
Xiaolong Zheng, PhD
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun E Rd
Beijing, 100190
China
Phone: 86 10 82544537
Fax: 86 10 82544550
Email: xiaolong.zheng@ia.ac.cn

Abstract

Background: Dabbing is an emerging method of marijuana ingestion. However, little is known about dabbing owing to limited surveillance data on dabbing.

Objective: The aim of the study was to analyze Google search data to assess the scope and breadth of information seeking on dabbing.

Methods: Google Trends data about dabbing and related topics (eg, electronic nicotine delivery system [ENDS], also known as e-cigarettes) in the United States between January 2004 and December 2015 were collected by using relevant search terms such as "dab rig." The correlation between dabbing (including topics: dab and hash oil) and ENDS (including topics: vaping and e-cigarette) searches, the regional distribution of dabbing searches, and the impact of cannabis legalization policies on geographical location in 2015 were analyzed.

Results: Searches regarding dabbing increased in the United States over time, with 1,526,280 estimated searches during 2015. Searches for dab and vaping have very similar temporal patterns, where the Pearson correlation coefficient (PCC) is .992 ($P<.001$). Similar phenomena were also obtained in searches for hash oil and e-cigarette, in which the corresponding PCC is .931 ($P<.001$). Dabbing information was searched more in some western states than other regions. The average dabbing searches were significantly higher in the states with medical and recreational marijuana legalization than in the states with only medical marijuana legalization ($P=.02$) or the states without medical and recreational marijuana legalization ($P=.01$).

Conclusions: Public interest in dabbing is increasing in the United States. There are close associations between dabbing and ENDS searches. The findings suggest greater popularity of dabs in the states that legalized medical and recreational marijuana use. This study proposes a novel and timely way of cannabis surveillance, and these findings can help enhance the understanding of the popularity of dabbing and provide insights for future research and informed policy making on dabbing.

Introduction

"Dabbing" is a colloquial term referring to the inhalation of vaporized marijuana concentrates and is an increasingly popular method of marijuana ingestion [1]. Marijuana concentrates contain high levels of delta-9-tetrahydrocannabinol (THC), which is the main psychoactive ingredient in marijuana. Butane hash oil (BHO), one of the major marijuana concentrates, is often produced by extracting THC from marijuana plants with liquid butane as the solvent. The resulting BHO products are

often called "shatter," "honeycomb," "crumble wax," "budder," and "earwax" according to their form and quality [2]. Generally, a "dab" is used to describe a small amount of marijuana extract that is vaporized and inhaled using an "oil rig" (a specific dabbing device), vaporizer, or electronic nicotine delivery system (ENDS, also known as e-cigarettes) [3,4]. Figure 1 illustrates dabbing with a screenshot of a YouTube video.

Although dabbing can reduce the ingestion of smoke-related toxins and carcinogens that are typically inhaled when smoking cannabis, there are potential risks of dabbing that have not been studied sufficiently [5]. First, the high THC concentration and novel means of administration might result in some psychological and physical problems [1,4,6,7]. Second, researchers have found that burn injuries associated with BHO manufacture have increased in recent years [8,9]. Finally, concentrated cannabis extract for dabbing may be contaminated by residual solvent and pesticide during commercial or homemade production.

Existing studies about dabbing are scarce. The earliest study on dabbing found that the use of BHO had been outside the medical marijuana user community and it is viewed as significantly more dangerous compared with other forms of cannabis use [1]. Another study discovered that many baby boomers were exploring alternative cannabis products including cannabis concentrates to improve well-being and to reduce the potential risks of traditional marijuana smoking, as they got older and less healthy [10]. A recent paper investigated the contamination concerns of cannabis concentrates and cannabinoid transfer efficiency during dabbing [11]. Additionally, several studies investigated some problems associated with using ENDS to vape cannabis [5,12,13].

Because current national surveys in the United States do not track the use of marijuana extracts [4,14], some studies collected data on dabbing from social media, such as Twitter and YouTube, and obtained several significant findings. For instance, a study based on Twitter data suggested the popularity of dabs was greater in the states that legalized recreational and medical use or only medical use of cannabis [4]. Another study analyzed the content of 116 dabbing videos on YouTube and found that dabbing-related videos on YouTube can be easily accessed [14]. However, little is known about the temporal evolution and regional distribution of public perceptions and interest about dabbing, as well as the association between public interest in dabbing and the interest in ENDS across the United States.

Internet data such as Google searches have filled many public health data gaps [15-17], especially in behavioral outcomes, where traditional data such as telephone surveys are rare and expensive to generate [18]. Hundreds of studies exploited Google search data to yield valid insights in public health research [19]. For instance, Google search data have been used to estimate influenza epidemics [20-25], to track tobacco or emerging products such as ENDS [26-30], to study psychology-related problems [31-33], and to analyze cancer-related information seeking [34-36].

Given the great value of Google search data in digital surveillance systems for public health, this study aimed to fill some of the aforementioned knowledge gaps about dabbing using Google search data. In particular, this study characterizes (1) the popularity, on the Web, of dabbing across time compared with other forms of cannabis use; (2) the popularity of dabbing across the US states, including comparisons of searches across states with varying marijuana legalization policies; and (3) the correlations between dabbing-related searches and ENDS-related searches.

Figure 1. A YouTube video screenshot illustrating dabbing.

Methods

Data Collection

Data were collected from Google Trends [37], which provides a public and timely way to analyze the trends of certain search query terms by time, geographic location, and category. Each Google Trends curve consists of many weekly aggregated data points, and each data point indicates the fraction of searches for the chosen terms (or categories) in a geographic location at a particular time relative to the total number of searches at that time. Note that the value of each data point represents relative search volume (RSV). In a trend curve, all indices are scaled to the highest search week (RSV=100). Extremely low searches are normalized and scaled to be a zero volume (RSV=0). The study qualifies as nonhuman subjects research because this study does not involve data through intervention or interaction with a living individual or his or her identifiable private information.

To understand the popularity of dabbing on the Web, the search topic "dabbing" was compared with the topics relating to other forms of cannabis use: "cannabis smoking" and "cannabis edibles" (see Figure 2) [38]. Note that all topics were on the same RSV scale. To examine the variations of dabbing search terms, "dabbing" was roughly divided into 2 smaller topics: "dab" and "hash oil" (see Figure 2). As a supplement to RSV data, Google AdWords [39] was used to estimate the raw search volume of the topic "dabbing" during 2015, which was similar

to some previous studies exploiting advertisement keywords for infodemiology studies [30,40]. Furthermore, "dab" and "hash oil" were compared with 2 topics relating to ENDS (ie, "vaping" and "e-cigarette") in terms of Google search queries (see Figure 3). The comparison was done to compare their individual temporal patterns, so each of the topics was on its own RSV scale. The division of ENDS into 2 topics was similar to that in a previous study [30], and the reason was also to examine the variations of search terms. Note that the time interval covered by the aforementioned data was between January 2004 (when data were first available) and December 2015. Finally, the search data relating to "dabbing" during 2015 were gathered to understand the popularity of dabbing across the US states (see Figure 4).

During data collection, basic query terms were initially identified according to related literature (eg, "dab rig," "marijuana smoking," "cannabis edibles"), and then related terms suggested by Google Trends were added to form candidate terms. Candidate terms were sorted by RSV, and terms with higher RSV were chosen because Google Trends limits the maximum number of words in query terms for a topic to 30. Unclear terms (eg, the single term "dab" can refer to the name of a bank) were omitted. The chosen terms were combined with "+" to form a composite term to collect the search data for a topic. For example, the composite term "dab rig+dab rigs+make dabs..." was used for the collection of searches relating to the topic "dabbing." The specific terms used in data collection are provided in Multimedia Appendix 1.

Figure 2. National trends for dabbing-related Google searches in the United States, 2004-2015. Panel A compared dabbing searches with searches for cannabis smoking and cannabis edibles, and panel B compared dabbing searches that included terms indicative of dab and hash oil.
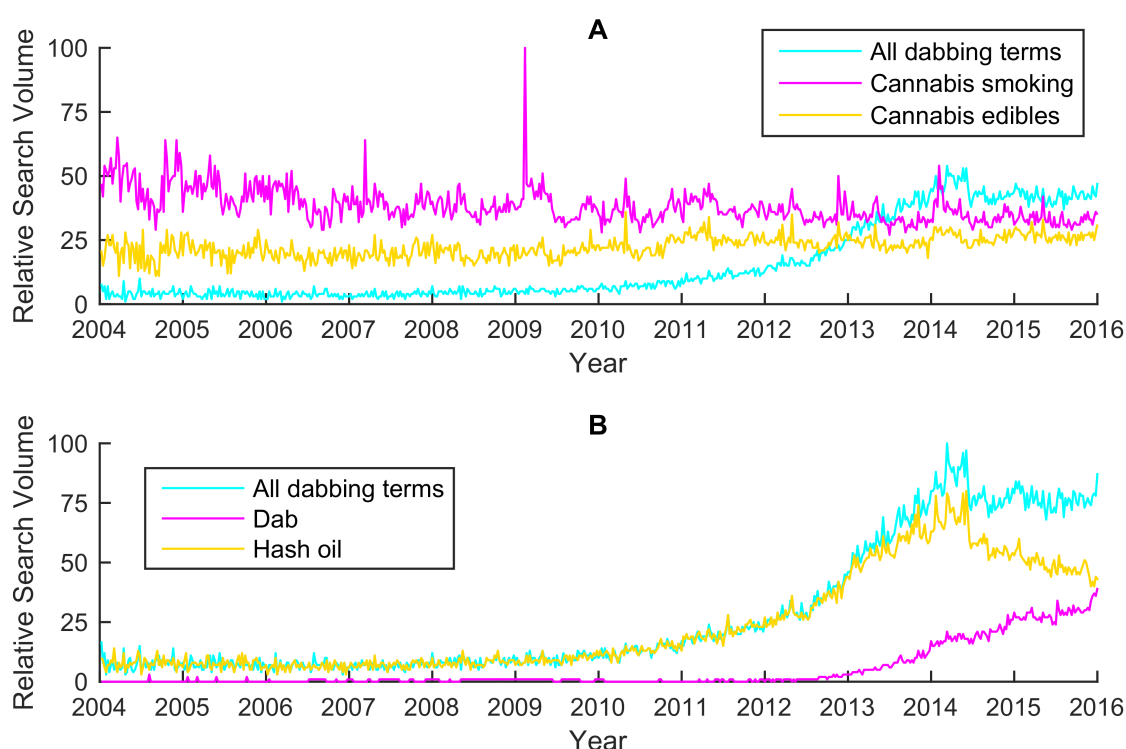
Figure 3. Temporal pattern comparison of searches for the dabbing and ENDS topics (ie, dab, hash oil, vaping, and e-cigarette). Each time series is on its own scale (ie, not applicable for relative search volume comparison between different time series).
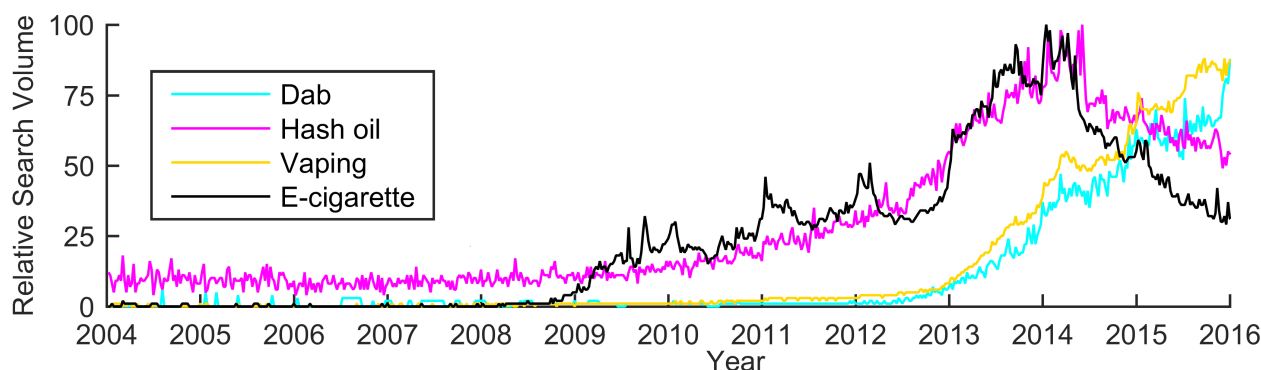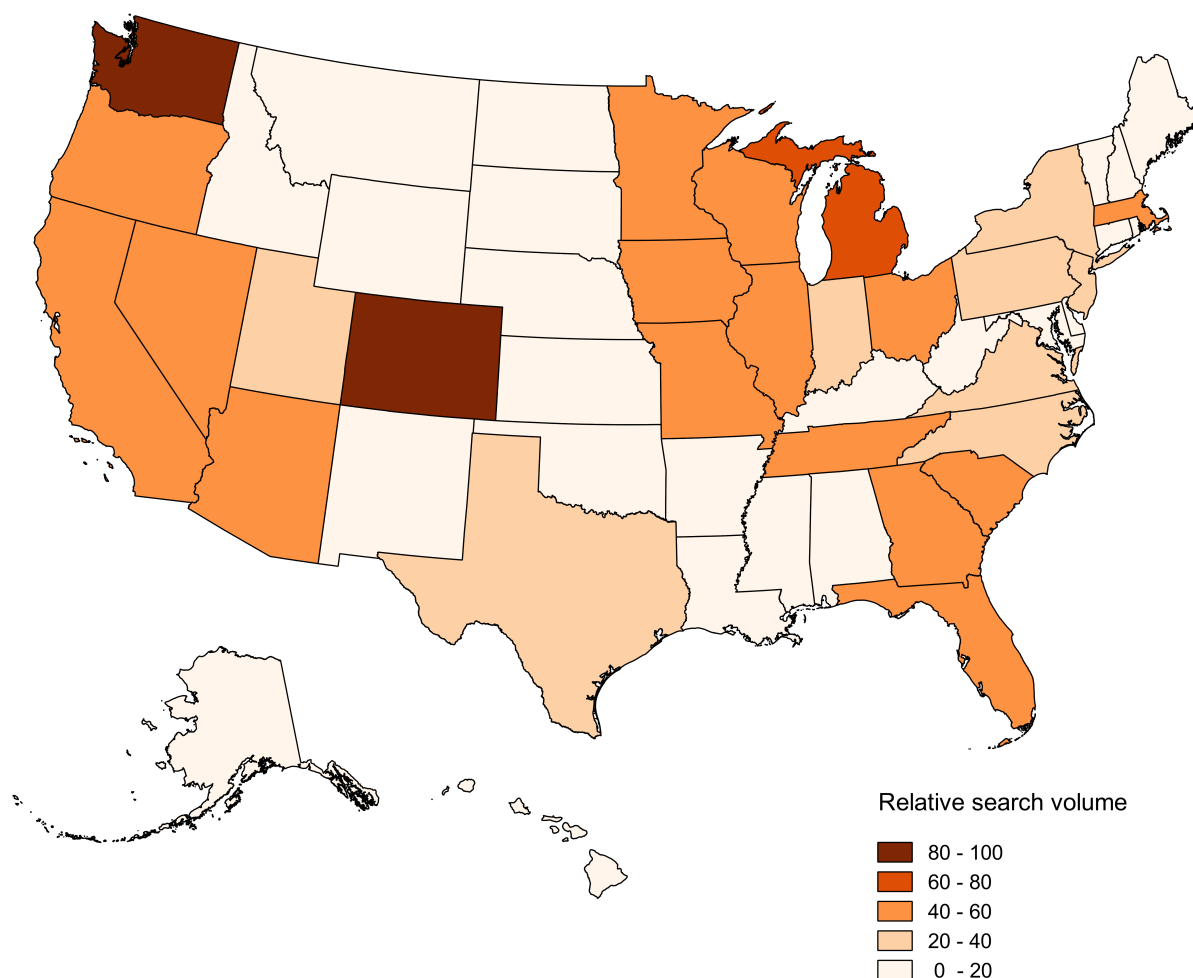

Figure 4. Choropleth map of raw searches regarding dabbing during 2015.

Data Analysis

The RSV of the topic "dabbing" for 2016 was predicted by using the autoregressive integrated moving average model and the R package called forecast [41]. To detect latent association between dabbing searches and ENDS searches, 2-tailed Pearson correlation coefficient (PCC) was adopted to analyze the pairwise correlations of the topics "dab," "hash oil," "vaping," and "e-cigarette."

The top 10 states with the highest RSV were obtained by sorting the raw search data relating to "dabbing" during 2015. The raw data for the 50 US states and the District of Columbia are provided in Multimedia Appendix 2. Differences in the 2015 raw dabbing data across US states with varying marijuana legalization policies were examined by 1-way ANOVA (analysis of variance) with 95% confidence interval. Similar to a prior study [4], states' legal statuses before January 1, 2016, were grouped into 3 types: (1) type 1 includes 4 states and the District

of Columbia that passed laws legalizing medical and recreational use of cannabis (Colorado, Washington, Alaska, Oregon, and District of Columbia); (2) type 2 includes 19 states that have legalized medical but not recreational use of cannabis (Arizona, California, Connecticut, Delaware, Hawaii, Illinois, Maryland, Maine, Massachusetts, Michigan, Minnesota, Montana, New York, Nevada, New Hampshire, New Jersey, New Mexico, Rhode Island, and Vermont); (3) type 3 includes 27 states that have not yet passed laws legalizing medical use of cannabis (Alabama, Arkansas, Florida, Georgia, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Mississippi, Missouri, Nebraska, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, Wisconsin, and Wyoming). Then the pairwise difference between the estimated group means was tested. Note that the time series plotting, correlation analysis, and ANOVA in this study were performed using MATLAB (Mathworks) [42].

Results

As seen in panel A of Figure 2, searches relating to dabbing increased over time in the United States before 2014. The estimated dabbing searches during 2015 were 1,526,280. The predicted dabbing searches for 2016 were 22% (95% CI 19%-24%) more than dabbing searches during 2015. Before 2013, dabbing was less often searched than traditional cannabis smoking or cannabis edibles, but dabbing searches surpassed

searches for cannabis smoking or cannabis edibles after the middle of 2013. For instance, searches regarding dabbing during 2015 were 28% (95% CI 25%-32%) more than searches regarding cannabis smoking and 58% (95% CI 54%-62%) more than searches regarding cannabis edibles.

As seen in panel B of Figure 2, hash oil searches occurred more often than searches for dab since 2004. Hash oil searches began increasing starting in 2008, and dab searches increased in the latter part of 2012. By the middle of 2014, searches regarding dab continued to increase, whereas searches for hash oil terms decreased. However, hash oil searches during 2015 were still 63% (95% CI 58%-68%) more than searches for dab.

In Figure 3, the search trends for hash oil and e-cigarettes show similar temporal patterns. Both had a wave peak in the early part of 2014. Similar results were also obtained between dab searches and vaping searches. The searches for dab and vaping showed a rapid increase beginning in 2013 and continued to show an increase.

Table 1 summarizes the temporal correlations between searches regarding dab or hash oil and searches regarding vaping or e-cigarettes. Searches for dab and hash oil have high correlations with searches of vaping and e-cigarette terms. For instance, the PCC for searches regarding dab and vaping is .992. The PCC for searches regarding hash oil and e-cigarettes is .931. Note that all P values for the abovementioned PCC values are less than .001.

Table 1. Temporal correlation of searches regarding dab and related topics. Note that the value in each cell is the Pearson correlation coefficient and all P values are less than .001.

Electronic nicotine delivery system	Dab	Hash oil
Vaping	.992	.783
E-cigarette	.600	.931

The 12 states with the highest raw RSV in 2015 were Colorado, Washington, Michigan, South Carolina, Nevada, Arizona, California, Oregon, Florida, Georgia, Missouri, and Massachusetts (see Figure 4). Note that Florida, Georgia, and Missouri had the same RSV. Among the Census Regions and Divisions of the United States, dabbing searches in the Pacific, East North Central, Middle Atlantic, and South Atlantic divisions were relatively higher than those in the other divisions.

During 2015, the means of searches for type 1, type 2, and type 3 were 55.800, 21.316, and 20.889, respectively. The group means of dabbing searches for type 1 and type 2 were significantly different at the 5% significance level (P=.02), and dabbing searches for type 1 and type 3 had similar results (P=.01). However, there was no significant difference at the 5% significance level between the group means of dabbing searches for type 2 and type 3 (P>.99; see Figure 5).

Figure 5. Raw dabbing Google searches by predictor for cannabis legal status of the United States. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers. Two medians are significantly different at the 5% significance level if their intervals of the notches do not overlap. Refer to the help document of [42] for more details of the box plots. Note that the details about the statuses are given in the Methods section.
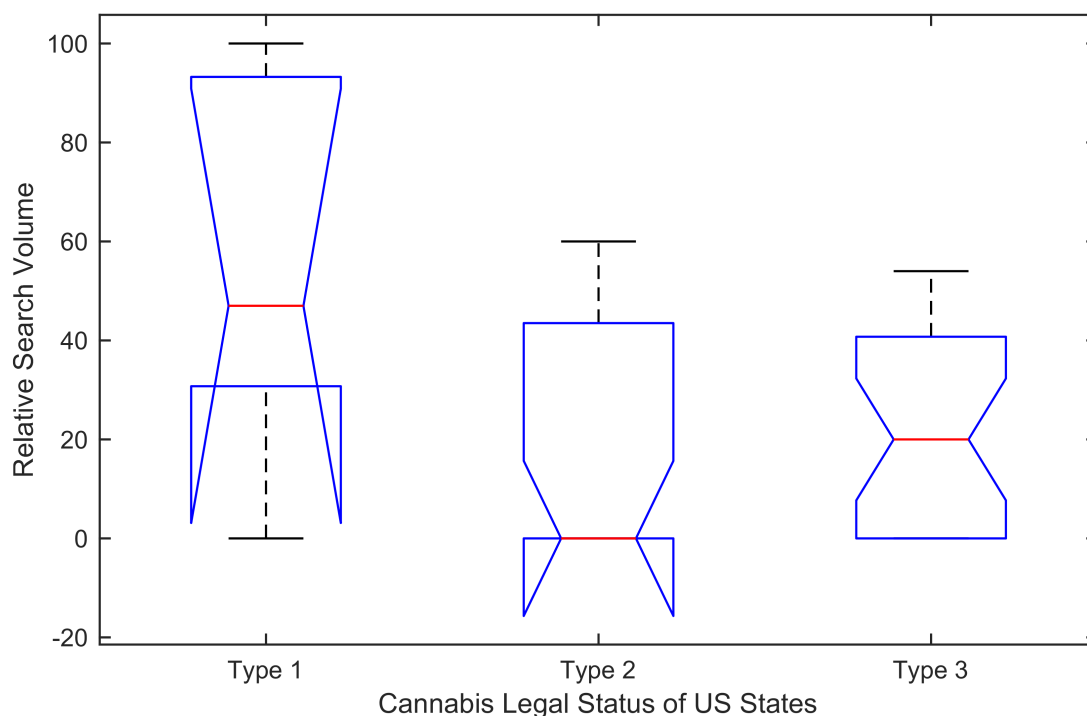
Discussion

Principal Findings

Dabbing searches are very common in the United States, and they have increased rapidly over time. Similar temporal patterns are found between searches for dab and vaping searches as well as searches for hash oil and e-cigarettes. Overall, dabbing was more frequently searched in the western states than other regions. The average dabbing searches were significantly higher in the states with recreational marijuana legalization than in the states without recreational marijuana legalization.

These findings fill some of the knowledge gaps regarding dabbing surveillance, but improved cannabis surveillance systems are needed to more fully understand the breadth and scope of variations in marijuana use. This study is the first to address the temporal associations between dabbing and ENDS. In general, the results based on search query monitoring can provide novel insights for further research and policy making.

Dabbing is becoming a popular alternative form of cannabis use. It became more popular than cannabis edibles and traditional cannabis smoking after the middle of 2013. When searches for dabbing are grouped into 2 categories, dab and hash oil, the searches demonstrate different developmental patterns over time. After 2014, searches for hash oil decreased, whereas dab searches increased. This suggests that the impact of the variations of dabbing-related terms used by cannabis users should be considered when designing survey questionnaires [30]. In addition, these temporal differences very

likely reflect the changing technology opportunities afforded by the increased use of ENDS.

Previous studies have found that some dabbing users use ENDS for dabbing [5,12,13], which is a kind of emerging ENDS misuse. Its health risks and impact on cannabis control are still unknown. This study found that searches for dab and vaping searches have very similar temporal patterns, as do searches for hash oil and e-cigarettes. This finding suggests that there is a certain association between dabbing searches and ENDS searches. One possible reason is that a large number of people use ENDS for dabbing in the United States, but it still needs to be investigated further. In particular, searches for dab and vaping increased rapidly since 2013. Our results are almost consistent with the observations by a leading, popular marijuana magazine, High Times [43], which did a cover story on dabbing in July 2013. A senior editor stated that dabbing was an underground activity 5 months before the cover story [1].

On the basis of the editor's claim, we can infer that the popularity of dabbing increased from an unobvious state before February 2013 to a relatively significant state in July 2013, which attracted the editor's attention. However, the dabbing searches had begun to increase earlier than the time mentioned by the editor. One explanation is that the wisdom of crowds in Google Trends is more sensitive than individuals in terms of perceiving emerging phenomena, but the true reason still needs to be investigated. Considering the close associations between dabbing and ENDS, addressing dabbing issues together with ENDS may be an effective approach for a better understanding of how a variety of drugs are or can be delivered to the lungs using similar technology [5]. Further spatiotemporal analysis

methods are needed to characterize their additional associations in the future [44,45].

Among the top 12 states with the largest raw RSV, all states legalized medical marijuana use except for South Carolina, Florida, Georgia, and Missouri. Some of these states (ie, Colorado, Washington, and Oregon) have already legalized recreational cannabis use. A recent study claims dabbing is more popular in states that have legalized medical marijuana use, which could be related to the emergence of vaporizer use among patients using medical marijuana and the recent increased availability of marijuana concentrates at medical marijuana dispensaries [4]. The claim was partially supported by the abovementioned results showing the temporal correlations between dabbing and ENDS.

This study found that dabbing searches are more prevalent on the West Coast of the United States, which is consistent with a prior study [4]. Previous analysis of Twitter data suggested that higher dabbing searches in western United States might be partially related to medical marijuana use laws that were passed much earlier there [4]. Another explanation was that the states on the West Coast have older and less strictly controlled medical marijuana programs. Besides, recreational marijuana legalization took effect in Washington after December 2012 and in Oregon after July 2015, so it is easier for people in these two states to do dabbing.

The average dabbing searches were significantly higher in the states with medical and recreational marijuana legalization than in the states with only medical marijuana legalization or the states without medical and recreational marijuana legalization. However, there was no significant difference in the means of dabbing searches between the states with only medical marijuana legalization and the states without medical and recreational marijuana legalization. The findings suggest greater popularity of dabs in the states that legalized medical and recreational marijuana use, which is partly consistent with the previous findings [4]. The comparison between this study and previous studies suggests that selecting suitable data and developing analytic standards are needed by those analyzing Web search data, social media, etc, just as those conducting surveillance and epidemiologic research have developed some standardized approaches.

Future research on dabbing and similar new technologies for drug delivery is needed to more fully understand use patterns that are tied to demographic characteristics, policy changes, drug availability, changes in technology, and other variables. In addition, there is a need to assess whether searches on topics such as dabbing are associated with actual use patterns (eg, as

determined by sales patterns) and reports of adverse events (eg, via poison control or Food and Drug Administration reporting). If clear temporal relationships can be demonstrated between dabbing search changes over time and specific measurable behaviors or health outcomes, it will be possible to more fully characterize the public health value of tracking dabbing and other similar search outcomes as an early warning system of emerging substance use and abuse.

Limitations

Some limitations need to be taken into account when interpreting this study. First, the Google Trends data in this study are the adjusted relative search values, because Google Trends does not provide actual search volume data. Data for only popular terms are analyzed by Google Trends, which causes the RSV of search terms with extremely low volume to appear as 0. Google Trends data lack demographic information compared with survey data. Although Google Trends data include only search activities using Google, Google accounts for an estimated 65% of the market share of search engines in the United States between January 2008 and October 2015 [46]. Second, the data in this study are limited to the United States and English search terms, so caution is needed when generalizing our results to other countries. The selection of search term keywords might lead to some minor biases; however, nearly all of our search term keywords have almost reached the number limit of Google Trends search terms. Third, the search data have not been assessed relative to year-by-year trends in actual use of products, so we cannot at this time suggest that search trends on dabbing predicted eventual increases in actual use of those products, but future research will assess those relationships. Finally, our results demonstrate high correlations between dabbing and ENDS in terms of Google searches. The reasons for the positive correlations between dabbing and ENDS searches require additional research across a variety of disciplines, although we provided some possible explanations.

Conclusions

In recent years, the general public has increasingly accepted marijuana legalization, but the potential adverse effects and health risks of dabbing are still being researched. This study provides a novel and timely way of conducting cannabis-related surveillance that may complement the current but limited epidemiologic data on dabbing. In the future, fusing Web-based data, such as Google searches and Web-based surveys, and offline community-recruited samples may help enhance the understanding of dabbing and similar substance use and provide insights for relevant research and informed policy making.

XSL•FO

RenderX

Authors' Contributions

ZZ provided substantial contributions to the conception and design of this work; provided acquisition, analysis, and interpretation of data for the work; and drafted the manuscript and revised it critically. XZ provided substantial contributions to the conception and design of this work, provided analysis and interpretation of data for the work, and revised the manuscript critically. DDZ provided substantial contributions to the conception and design of this work, provided study supervision, and contributed to the editing of the manuscript. SJL revised the manuscript critically and contributed to the editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The search terms for collecting Google Trends data.

[PDF File (Adobe PDF File), 12KB - jmir_v18i9e252_app1.pdf]

Multimedia Appendix 2

The raw Google searches for the 50 US states and the District of Columbia.

[XLS File (Microsoft Excel File), 29KB - jmir_v18i9e252_app2.xls]

References

1. Loflin M, Earleywine M. A new method of cannabis ingestion: the dangers of dabs? Addict Behav 2014 Oct;39(10):1430-1433. [doi: 10.1016/j.addbeh.2014.05.013] [Medline: 24930049]
2. Stogner JM, Miller BL. The Dabbing Dilemma: A Call for Research on Butane Hash Oil and Other Alternate Forms of Cannabis Use. Subst Abus 2015;36(4):393-395. [doi: 10.1080/08897077.2015.1071724] [Medline: 26241175]
3. Stogner JM, Miller BL. Assessing the Dangers of "Dabbing": Mere Marijuana or Harmful New Trend? Pediatrics 2015 Jul;136(1):1-3. [doi: 10.1542/peds.2015-0454] [Medline: 26077476]
4. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": Analyzing Twitter data on marijuana concentrates across the U.S. Drug Alcohol Depend 2015 Oct 1;155:307-311. [doi: 10.1016/j.drugalcdep.2015.07.1199] [Medline: 26338481]
5. Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction 2015 Nov;110(11):1699-1704. [doi: 10.1111/add.13036] [Medline: 26264448]
6. Whiting PF, Wolff RF, Deshpande S, Di NM, Duffy S, Hernandez AV, et al. Cannabinoids for Medical Use: A Systematic Review and Meta-analysis. JAMA 2015;313(24):2456-2473. [doi: 10.1001/jama.2015.6358] [Medline: 26103030]
7. Wilkinson ST, D'Souza DC. Problems with the medicalization of marijuana. JAMA 2014 Jun 18;311(23):2377-2378. [doi: 10.1001/jama.2014.6175] [Medline: 24845238]
8. Bell C, Slim J, Flaten HK, Lindberg G, Arek W, Monte AA. Butane Hash Oil Burns Associated with Marijuana Liberalization in Colorado. J Med Toxicol 2015 Dec;11(4):422-425. [doi: 10.1007/s13181-015-0501-0] [Medline: 26289652]
9. Jensen G, Bertelotti R, Greenhalgh D, Palmieri T, Maguina P. Honey oil burns: a growing problem. J Burn Care Res 2015;36(2):e34-e37. [doi: 10.1097/BCR.0000000000000067] [Medline: 24823328]
10. Murphy F, Sales P, Murphy S, Averill S, Lau N, Sato S. Baby Boomers and Cannabis Delivery Systems. Journal of Drug Issues 2015 Apr 10;45(3):293-313. [doi: 10.1177/0022042615580991]
11. Raber JC, Elzinga S, Kaplan C. Understanding dabs: contamination concerns of cannabis concentrates and cannabinoid transfer during the act of dabbing. J Toxicol Sci 2015;40(6):797-803 [FREE Full text] [doi: 10.2131/jts.40.797] [Medline: 26558460]
12. Giroud C, de Cesare M, Berthet A, Varlet V, Concha-Lozano N, Favrat B. E-Cigarettes: A Review of New Trends in Cannabis Use. Int J Environ Res Public Health 2015 Aug 21;12(8):9988-10008. [doi: 10.3390/ijerph120809988]
13. Morean ME, Kong G, Camenga DR, Cavallo DA, Krishnan-Sarin S. High School Students' Use of Electronic Cigarettes to Vaporize Cannabis. Pediatrics 2015 Oct;136(4):611-616. [doi: 10.1542/peds.2015-1727] [Medline: 26347431]
14. Krauss MJ, Sowles SJ, Mylvaganam S, Zewdie K, Bierut LJ, Cavazos-Rehg PA. Displays of dabbing marijuana extracts on YouTube. Drug Alcohol Depend 2015 Oct 1;155:45-51. [doi: 10.1016/j.drugalcdep.2015.08.020] [Medline: 26347408]
15. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
16. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
17. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]

XSL•FO
RenderX

18.   Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr
       9;311(14):1399-1400. [doi: 10.1001/jama.2014.1505] [Medline: 24577162]

19.   Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a
       systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline:
       25337815]

20.   Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search
       engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]

21.   Pervaiz F, Pervaiz M, Abdur RN, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res
       2012;14(5):e125 [FREE Full text] [doi: 10.2196/jmir.2102] [Medline: 23037553]

22.   Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to)
       google flu trends? Am J Prev Med 2014 Sep;47(3):341-347. [doi: 10.1016/j.amepre.2014.05.020] [Medline: 24997572]

23.   Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl
       Acad Sci U S A 2015 Nov 24;112(47):14473-14478. [doi: 10.1073/pnas.1515373112] [Medline: 26553980]

24.   Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE, et al. Google Flu Trends Spatial Variability
       Validated Against Emergency Department Influenza-Related Visits. J Med Internet Res 2016;18(6):e175 [FREE Full text]
       [doi: 10.2196/jmir.5585] [Medline: 27354313]

25.   Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating Influenza Outbreaks Using Both Search Engine Query Data
       and Social Media Data in South Korea. J Med Internet Res 2016;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955]
       [Medline: 27377323]

26.   Zhang Z, Zheng X, Zeng DD, Cui K, Luo C, He S, et al. Discovering seasonal patterns of smoking behavior using online
       search information. 2013 Presented at: Proc IEEE Int Conf Intell Secur Inform ISI; 2013; Seattle, WA p. 371-373. [doi:
       10.1109/ISI.2013.6578861]

27.   Zhang Z, Zheng X, Zeng DD, Leischow SJ. Information seeking regarding tobacco and lung cancer: effects of seasonality.
       PLoS One 2015;10(3):e0117938 [FREE Full text] [doi: 10.1371/journal.pone.0117938] [Medline: 25781020]

28.   Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, Chaloupka FJ, et al. Monitoring of non-cigarette
       tobacco use using Google Trends. Tob Control 2015 May;24(3):249-255 [FREE Full text] [doi:
       10.1136/tobaccocontrol-2013-051276] [Medline: 24500269]

29.   Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic
       cigarettes) using search query surveillance. Am J Prev Med 2011 Apr;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007]
       [Medline: 21406279]

30.   Ayers JW, Althouse BM, Allem J, Leas EC, Dredze M, Williams RS. Revisiting the Rise of Electronic Nicotine Delivery
       Systems Using Search Query Surveillance. Am J Prev Med 2016 Jun;50(6):e173-e181. [doi: 10.1016/j.amepre.2015.12.008]
       [Medline: 26876772]

31.   Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With
       Google Trends: Evidence From Taiwan. J Med Internet Res 2015;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516]
       [Medline: 26586281]

32.   Lee D, Lee H, Choi M. Examining the Relationship Between Past Orientation and US Suicide Rates: An Analysis Using
       Big Data-Driven Google Search Queries. J Med Internet Res 2016;18(2):e35 [FREE Full text] [doi: 10.2196/jmir.4981]
       [Medline: 26868917]

33.   Jankowski W, Hoffmann M. Can Google Searches Predict the Popularity and Harm of Psychoactive Agents? J Med Internet
       Res 2016;18(2):e38 [FREE Full text] [doi: 10.2196/jmir.4033] [Medline: 26916984]

34.   Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine,
       United States 2001-2003. J Med Internet Res 2005 Jul 1;7(3):e36 [FREE Full text] [doi: 10.2196/jmir.7.3.e36] [Medline:
       15998627]

35.   Bloom R, Amber KT, Hu S, Kirsner R. Google Search Trends and Skin Cancer: Evaluating the US Population's Interest
       in Skin Cancer and Its Association With Melanoma Outcomes. JAMA Dermatol 2015 Aug;151(8):903-905. [doi:
       10.1001/jamadermatol.2015.1216] [Medline: 26061357]

36.   Foroughi F, Lam AK, Lim MS, Saremi N, Ahmadvand A. "Googling" for Cancer: An Infodemiological Assessment of
       Online Search Interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016
       May 04;2(1):e5. [doi: 10.2196/cancer.5212]

37.   Google Trends. URL: https://www.google.com/trends/ [accessed 2016-09-08] [WebCite Cache ID 6kNOiMNsz]

38.   Schauer GL, King BA, Bunnell RE, Promoff G, McAfee TA. Toking, Vaping, and Eating for Health or Fun: Marijuana
       Use Patterns in Adults, U.S., 2014. Am J Prev Med 2016 Jan;50(1):1-8. [doi: 10.1016/j.amepre.2015.05.027] [Medline:
       26277652]

39.   Google AdWords. URL: https://www.google.com/adwords/?zd=1 [accessed 2016-09-08] [WebCite Cache ID 6kNOnzOFd]

40.   Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc
       2006:244-248 [FREE Full text] [Medline: 17238340]

41.   Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. J. Stat. Soft 2008;27(3):1-22.
       [doi: 10.18637/jss.v027.i03]

42.   MATLAB - The Language of Technical Computing.: MathWorks; 2016. URL: http://www.mathworks.com/products/matlab/?requestedDomain=www.mathworks.com [accessed 2016-09-08] [WebCite Cache ID 6kNPj5hOI]
43.   High Times. URL: http://www.hightimes.com/ [accessed 2016-03-25] [WebCite Cache ID 6gGQoaso0]
44.   Chang W, Zeng D, Chen H. A stack-based prospective spatio-temporal data analysis approach. Decision Support Systems 2008 Nov;45(4):697-713. [doi: 10.1016/j.dss.2007.12.008]
45.   Ma J, Zeng D, Zhao H, Liu C. Cross-Correlation Measure for Mining Spatio-Temporal Patterns. J Database Manag 2013;24(2):13-34. [doi: 10.4018/jdm.2013040102]
46.   Statista. Search engine market share in the United States 2015 URL: http://www.statista.com/statistics/267161/market-share-of-search-engines-in-the-united-states/ [accessed 2016-03-24] [WebCite Cache ID 6gG3YvlZh]

Abbreviations

ANOVA: analysis of variance
BHO: butane hash oil
ENDS: electronic nicotine delivery system
PCC: Pearson correlation coefficient
RSV: relative search volume
THC: delta-9-tetrahydrocannabinol

Original Paper

Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis

Vibhu Agarwal[1], MS; Liangliang Zhang[2,3], PhD; Josh Zhu[2], MS; Shiyuan Fang[2], MS; Tim Cheng[2], PhD; Chloe Hong[2], AM; Nigam H Shah[4], MBBS, PhD

[1]Biomedical Informatics Training Program, Stanford University, Stanford, CA, United States

[2]Baidu Inc, Sunnyvale, CA, United States

[3]Stanford University, Department of Electrical Engineering, Stanford University, Stanford, CA, United States

[4]Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States

Corresponding Author:
Vibhu Agarwal, MS
Biomedical Informatics Training Program
Stanford University
1265 Welch Road
Medical School Office Building
Stanford, CA, 94305
United States
Phone: 1 650 723 1398
Fax: 1 650 725 7944
Email: vibhua@stanford.edu

Abstract

Background: By recent estimates, the steady rise in health care costs has deprived more than 45 million Americans of health care services and has encouraged health care providers to better understand the key drivers of health care utilization from a population health management perspective. Prior studies suggest the feasibility of mining population-level patterns of health care resource utilization from observational analysis of Internet search logs; however, the utility of the endeavor to the various stakeholders in a health ecosystem remains unclear.

Objective: The aim was to carry out a closed-loop evaluation of the utility of health care use predictions using the conversion rates of advertisements that were displayed to the predicted future utilizers as a surrogate. The statistical models to predict the probability of user's future visit to a medical facility were built using effective predictors of health care resource utilization, extracted from a deidentified dataset of geotagged mobile Internet search logs representing searches made by users of the Baidu search engine between March 2015 and May 2015.

Methods: We inferred presence within the geofence of a medical facility from location and duration information from users' search logs and putatively assigned medical facility visit labels to qualifying search logs. We constructed a matrix of general, semantic, and location-based features from search logs of users that had 42 or more search days preceding a medical facility visit as well as from search logs of users that had no medical visits and trained statistical learners for predicting future medical visits. We then carried out a closed-loop evaluation of the utility of health care use predictions using the show conversion rates of advertisements displayed to the predicted future utilizers. In the context of behaviorally targeted advertising, wherein health care providers are interested in minimizing their cost per conversion, the association between show conversion rate and predicted utilization score, served as a surrogate measure of the model's utility.

Results: We obtained the highest area under the curve (0.796) in medical visit prediction with our random forests model and daywise features. Ablating feature categories one at a time showed that the model performance worsened the most when location features were dropped. An online evaluation in which advertisements were served to users who had a high predicted probability of a future medical visit showed a 3.96% increase in the show conversion rate.

Conclusions: Results from our experiments done in a research setting suggest that it is possible to accurately predict future patient visits from geotagged mobile search logs. Results from the offline and online experiments on the utility of health utilization predictions suggest that such prediction can have utility for health care providers.

XSL•FO

RenderX

Introduction

Over the past years, Internet search engines have changed the way people report health outcomes and/or seek information regarding symptoms, diseases, and treatments, resulting in a parallel growth of a large amount of medical information. The potential of addressing public health challenges and advancing medical research through the analysis of such information repositories, as well as the challenges inherent in working with such sources, are being recognized [1-4]. In a novel study, the use of click-through statistics generated by a commercial Web advertising service was shown to be an effective strategy for influenza surveillance [5]. As a Web-scale repository of patient-generated information, Internet search logs have been mined for diverse applications, such as screening patients with pancreatic adenocarcinoma [6] and discovering adverse drug events [7,8]. The use of Internet searches as an indicator of individuals' interests and concerns related to health care has been studied for understanding the relationship between health anxiety and its effect on information-seeking behavior [9]. Notably, through the analysis of search logs collected from consenting users via a browser toolbar and complementary surveys, it has been shown that the analysis of long-term search behavior reveals patterns that may serve as markers for a transition to health care utilization [3].

Viewed against the backdrop of the recent tectonic shifts in the health care landscape in the Unites States, search log repositories present an opportunity to understand the nature of interactions between health care organizations and users—specifically interactions that result in utilization of health care resources. Gaining such an understanding is crucial for improving efficiencies and eventually, the accessibility of health care services [10,11]. Because search log data closely mirrors users' daily concerns and activities, they contain embedded clues about imminent health episodes. As a result, predictions of health care utilization based on geotagged search histories provide a snapshot of future health care demand that is based on an aggregation of personalized health trajectories.

Logs of Internet searches initiated from mobile devices contain search text and time stamp information, as well as the location of where the search was initiated. The location information in search logs contains clues about the searcher's interactions with the real world. For instance, consecutive searches from approximately the same location that are separated by a significant span of time could indicate an engagement at the particular location. The information utility of an individual's approximate location within a virtual geographical boundary (referred to as a "geofence") has been studied extensively within the ubiquitous computing community and forms the basis of several location-based services [7,9,12].

Based on deidentified searches and distances of searches from medical facilities, White and Horvitz [2] have argued that the evidence of health care utilization is related to the acuity of symptom searches. Although such observational analysis of Internet search logs has shown the feasibility of the approach for discovering population-level patterns, the utility of the endeavor to the various stakeholders in a health ecosystem remains unclear. Among other things, assessment of the utility of a predicted health outcome depends on the cost of making the prediction, the "actionability" of the prediction [13], as well as individual goals and value perception [14]. In general, the determinants of predictive utility are hard to measure and possibly subjective. As of this writing, we are not aware of a study that evaluates data mining experiments on Internet search logs from a utility standpoint.

In this study, we assess the utility of predicting health care utilization from the health care provider's perspective in the context of behaviorally targeted advertising. Recent studies on Internet consumer behavior have sought to model the publisher's and the advertiser's payoffs in terms of the performance metrics of a behaviorally targeted advertising campaign [15,16]. The advantage of using such a framework is that it allows for a fine level of control on experimental parameters and produces a concrete measure of the impact of health care utilization predictions within a specific setting. We compute features representing different aspects of search behavior along with surrogate measures of health care resource use directly from users' search logs and train statistical models to predict future health care utilization from search logs. Our models incorporate features that summarize temporal trends in the semantic and location patterns of searches and allow us to investigate the differential effect of various classes of features on utilization prediction. We then evaluate the impact on ad show conversion rates for search users who were shown ads and whose utilization prediction scores were computed from historical search logs. Our overall study design is depicted in Figure 1.

Figure 1. Overall study design: (A) generation of search logs based on searches within geofences, (B) identification of searches proximal to medical facilities and selection of patients and controls based on filtering criterion, (C) learning statistical model for predicting health, (D) evaluation of impact on ad show conversion rates for search users with high utilization prediction scores prior to ad views.
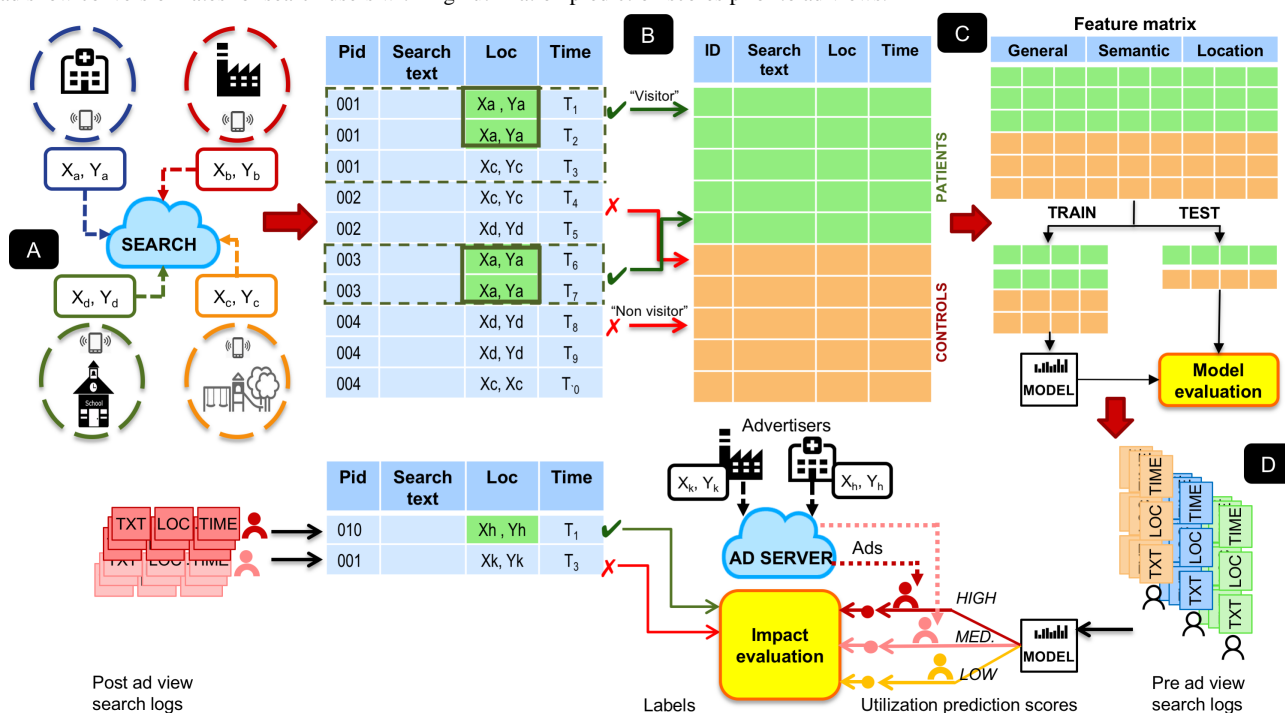
Methods

Data

Our dataset consisted of deidentified mobile Internet search logs representing more than 1 billion searches from 9.5 million search users of the Baidu search engine from March 2015 to May 2015, made accessible to the authors under a collaborative research program. The search logs contain search text (Chinese), the time stamp, and the location (latitude and longitude) of the search, and represent searches made from locations within China. Identifying health care utilization based on evidence of searches on mobile devices that were made close to hospitals is prone to false positives and negatives. Search users may work inside or close to hospital locations or may be passing by one, and may not be consumers of health care resources at the time they make the search from locations proximal to a hospital. Similarly, search users may visit a hospital as patients but not search during their visit. We acknowledge that it is not possible to completely eliminate false positives and false negatives when labels are assigned purely on the basis of searches made within the geofences of medical facilities. However, we have successfully reduced the number of false positives in our data by explicitly filtering out "weak" labels.

Inclusion and Exclusion Criteria

From all searches that took place between the aforementioned dates, we discarded search users who searched from within 200 meters of a hospital but the evidence of their presence in that location was less than 900 seconds. We also discarded search users who had more than 15 searches in a month from locations in the proximity of a medical facility because these may be individuals who lived or worked close by, or they may be health care professionals. Finally, we excluded users who searched

more than five times in a month in the vicinity of different medical facilities. Search logs for the remaining 4 million search users were deemed to have been evidence of visiting a medical facility and 1.5 million distinct search users were randomly sampled from them. We drew a proportionate random sample from all searches, which did not originate within the geofence of a known medical facility, to obtain 8 million distinct search users with no evidence of visiting a medical facility, thus giving a total set of 9.5 million search users. In the absence of relevant user information, we matched our controls on the number of days of available search logs.

Because we were interested in studying the temporal characteristics of search logs that culminated in the visit to a medical facility, we selected those who had search logs for each of 42 or more days preceding their last visit to a medical facility (a higher threshold, we discovered, would reduce our cohort size significantly and adversely impact statistical power). In the remainder of our paper, we refer to this cohort as "patients."

Longitudinal Partitioning

We partitioned the search logs for patients and controls by search days, in which search day n is the n th day in the sequence of days for which logs are available for a search user, preceding an endpoint. For patients, we defined the endpoint as the date of their first visit to a medical facility. For controls, the endpoint was picked randomly from each user's search log. After excluding the last day of visiting a medical facility, we could define an analysis window comprising 41 consecutive search days for patients. We defined a similar analysis window with 41 consecutive search days for controls with an endpoint coterminus with the first search day. Figure 2 illustrates the longitudinal partitioning of our search log data as previously described.

Feature Engineering

We chose three classes of features to study the discriminatory patterns in search logs across patients and controls over a succession of search days. The classes, as described in Table 1, represent general attributes of search logs, semantic properties of the search text, and the location attributes of search logs for each search day in the analysis window. We also created aggregated features that were based on counts of search log properties aggregated over the entire analysis window.

Figure 2. Longitudinal partitioning of search log data with analysis windows covering 41 search days. Control end points were selected randomly.
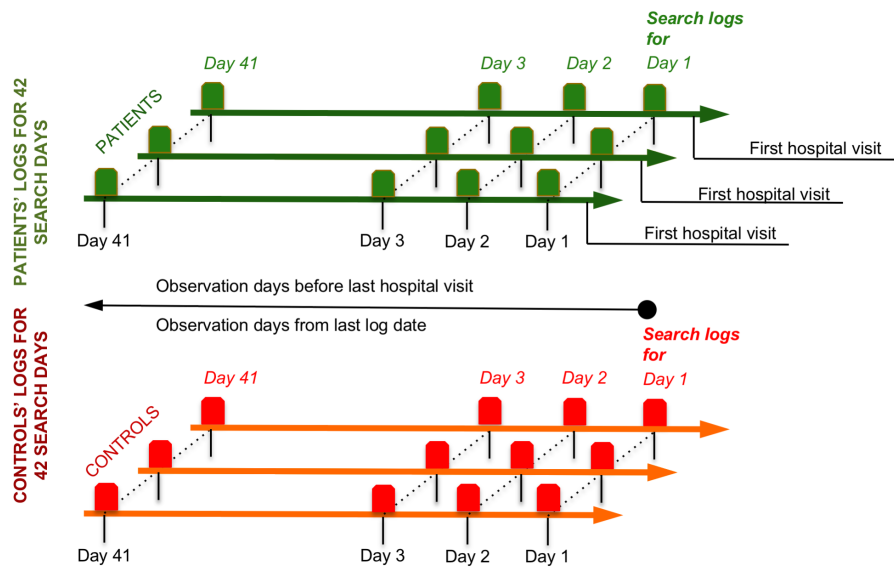
Table 1. Description of feature categories.

Feature categories and description	Aggregate	Daywise
General		
Number of searches	Yes	Yes
Number of health care-related searches	Yes	Yes
Mean session duration	Yes	Yes
Mean length of search text	Yes	Yes
Session interval reduction score	No	Yes
Semantic		
Number of searches for a disease	Yes	Yes
Number of searches for a drug	Yes	Yes
Number of searches for a medical device	Yes	Yes
Number of searches for a medical procedure	Yes	Yes
Number of searches containing one of 100 enriched (Chinese) words	Yes	Yes
Location		
Number of searches mapped to one of 53 enriched location categories	Yes	Yes
Number of searches whose location labels contain one of 113 words	Yes	No

General Features

It has been shown that the linguistic structure of Internet searches influences information retrieval from Web search engines [17]. In a study comparing search characteristics that originate from mobile devices to searches that originate from PCs, Jadhav et al [18] showed that health search queries tend to be longer than general search queries. We chose to include attributes related to the length of the text search and the duration of a search session among our general features. A search session represents several searches linked together by the most prominent themes returned in search results and longer sessions are likely to suggest the evolution of a search user's interest from general to more specific concepts. We also included the number of searches and number of health care-related searches in both our aggregate and daywise features because a user's level of concern about a health care issue is likely correlated to

the number of times they search for information for reassurance or remedy.

Interval Reduction Score

Earlier studies on health information-seeking behavior of Internet users have studied the relationship between search behavior and health concerns of search users. For instance, it has been shown that physician information-seeking behavior, as assessed by the mean page view duration for specific websites, is distinct from general online media activity [19]. Health anxiety in patients has also been shown to be associated

with specific search patterns, such as intense search activity punctuated by periods of calm [20]. With the intention of capturing differences in such patterns between utilizers and nonutilizers within our data, we computed an interval reduction score (please see Multimedia Appendix 1) for each search day for each user (Figure 3).

For search days with two or fewer searches, the interval reduction score was 1. We experimented with different values of w and obtained best results with small values close to 0.1 and by only considering health care-related searches in a search day.

Figure 3. Equation for the Interval Reduction Score (IRS).

We define IRS, $s(j, d)$ for search user j on search day d as

$$s(j, d) = \frac{1}{(n_{j,d} - 1)} \left[\sum_{i=2}^{n_{j,d}-1} 1 + w\mathbb{I} \cdot (2t_i - t_{i+1} - t_{i-1}) \right]$$

$n_{j,d}$ is the number of search logs for search user j on search day d

t_i is the starting time stamp for the ith search session for search user j or search day d

$\mathbb{I} \cdot$ is the indicator function whose value is 1 if $2t_i - t_{i+1} - t_{i-1} > 0$ and 0 otherwise

w is a weight parameter such that $0 \leq w \leq 1$.

Semantic Features

The language of the search queries in our dataset posed a unique challenge to our analysis. On one hand, we could leverage advantageous aspects of the Chinese language, such as the lack of verb conjugation and plural forms. In addition, this allowed us to capture the meaning behind idiomatic expressions that could be challenging to translate. On the other hand, English tokens would enable us to use a wider variety of existing language analytic tools. Thus, we balanced the two approaches by analyzing our tokens in Chinese, and performing token translation and carrying out further analysis in English.

For our Chinese semantic analysis, we identified enriched tokens that were used by patients and controls using a Fisher test with Bonferroni correction. We also evaluated the number of patients and controls that searched for each token on any given day, and compared the term frequency between patients and controls. We took the union of the tokens from these two analyses, and the best performing 100 tokens were included as features in subsequent analysis, after manually inspecting the features for procedure artifacts. Following this analysis, all tokens from health care queries were translated from Chinese to English for downstream analyses (Figure 4).

To model the variation in search content across search days for patients and controls, we further chose to explicitly characterize the medical content within the search text by using an approach that has been validated in clinical text mining research.

Although the form and structure of search text is fundamentally different from the free text in patients' records, we noted that certain aspects of the two were strikingly similar from a linguistic standpoint. Intuitively, one may take advantage of

this similarity by employing proven tools and techniques that have been used to characterize lexical coverage in the former for achieving similar goals with the latter. In particular, the use of ill-formed sentences, abbreviations, and spelling errors are common to both search text and clinical text, and motivated our choice of a biomedical terminology for identifying and delineating the use of medical terms in search text. We decided to use an extensive terminology of terms drawn from 22 clinically relevant ontologies from the Unified Medical Language System (UMLS) and BioPortal [21]. The lexicon represents more than 3.1 million terms that map to nearly 1.2 million concepts and a functional evaluation of annotations of clinical text based on the same have shown equivalence with more sophisticated natural language processing-based approaches [22]. Because the UMLS provides a mapping from each concept to one or more semantic types, by coalescing relevant semantic types into groups that represent diseases, drugs, devices, or procedures, one may achieve a fine-grained characterization of search terms by determining their group membership. However, because the semantic types as defined by the UMLS semantic network represent key relationships between biomedical concepts, inferring the medical semantics of search text using these semantic types is likely to be noisy. Results from our initial experiments showed many examples of misattributions on account of the domain specificity of our lexicon. For example, commonly occurring terms in search texts (eg, dame, gift, blade) are mapped to medical concepts that are grouped as a drug (D-Ala2-methionine enkephalinamide), a procedure (gamete in fallopian transfer), and a device, respectively. To address this issue, we investigated whether such misattributions were more pronounced for certain semantic groups than others. We segregated search texts in our training

XSL•FO

RenderX

set into health care and non-health care categories on the basis of whether a token in the translated text mapped to any one of the four high-level semantic groups mentioned previously. We then compared the segregation with an independent categorization based on Baidu's knowledge graph. We then repeated the experiment, each time leaving out one of the four semantic groups for making the health care categorization of search texts.

We observed that leaving out the device subgroup results in an overall improved agreement between the two segregations (from 15.2% with all four groups to 40% when only using drugs,

diseases, and procedures), whereas leaving out other groups did not show improvement. Therefore, we used membership counts of individual searches into the groups drugs, diseases, and procedures for features indicating the nature of the medical content in queries on a given search day. We noted that the semantic groups for drugs, diseases, and procedures represented the largest grouping of concepts within the UMLS [23-25] and, hence, are likely to be the most relevant groups in a lexicon-based semantic analysis approach. Similarly, membership counts across the full analysis window yielded the aggregated semantic features related to medical content.

Figure 4. Framework for search text translation and mapping.
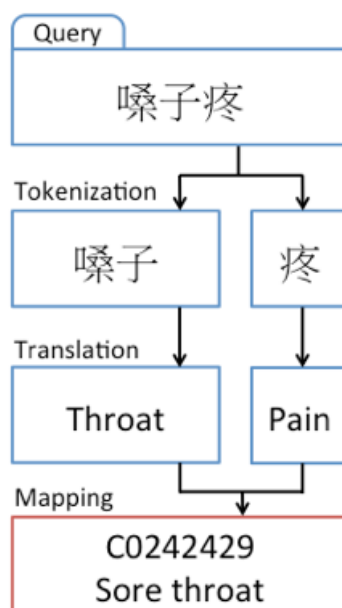
Search Specificity

Information returned by medical searches is known to influence concerns related to health, which in turn modulates subsequent search behavior [20]. Concerns about common symptomatology have been shown to escalate into searches for serious and rare diseases [26] and anxiety regarding one's health is likely to influence health care utilization intent [3], possibly precipitating a visit to a medical facility. Thus, we were interested in modeling the evolution of searches that progressed from a general inquiry regarding symptoms into a specific inquiry regarding a serious health condition. We chose to use the information content score of the most specific term in a search as an indicator of the generality of the subject matter of the search. The information content score takes advantage of the hierarchical structure of a medical ontology to ensure a monotonically nondecreasing measure of specificity and may be computed based on document-level frequency of a term in a corpus. For search terms that mapped to our medical terminology, we computed session-level information content scores from the segmented search texts. As with other semantic features, we chose the highest daywise score for all searches from a given search day for a daywise measure of search specificity.

Location Features

To build location features, we attached location labels to the latitude and longitude coordinates of searches. We used the Gecko Landmarks (Gecko Landmarks Ltd, Espoo, Finland) application program interface (API), which takes the latitude and longitude as input and outputs the 10 closest landmarks to this referenced location ranked by distance along with name and category labels for each landmark. For example, for a reference location given by latitude 39.903651 E, 116.415505 N, the Gecko API returns Beijing Hospital as the closest landmark with a category label of "hospital."

We rounded each geographic coordinate to four decimal places and then filtered for uniqueness. This resulted in less than 10 million coordinate pairs without a significant loss of precision. Coordinates with accuracy up to four decimal places represented an accuracy of approximately 11 meters, which we considered adequate for our location features. We then obtained access to a rate-limited instance of a Gecko server and carried out a batchwise conversion of our unique, binned coordinates to obtain the respective location labels (Figure 5). The batchwise lists of landmarks were merged and mapped back to the original location coordinates from which we built our feature matrix. Specifically, the features were the number of searches made by a user from a given landmark category (eg, health, restaurant).

We also created features based on individual words in the location names. For example, searches from an educational building could have the words "elementary" or "university" in the location names that were identified by the Gecko API. Although both universities and elementary schools have a category label of "education," they have opposite effects on predicting hospitalization; the former term is enriched whereas the latter is slightly depleted in patient's search logs. We indicated the presence or absence of a word token in a search location name using a binary (0,1) variable. Features based on individual tokens in location names captured additional granularity without imposing a structure on the location names a priori.

Figure 5. Extraction of place category labels.
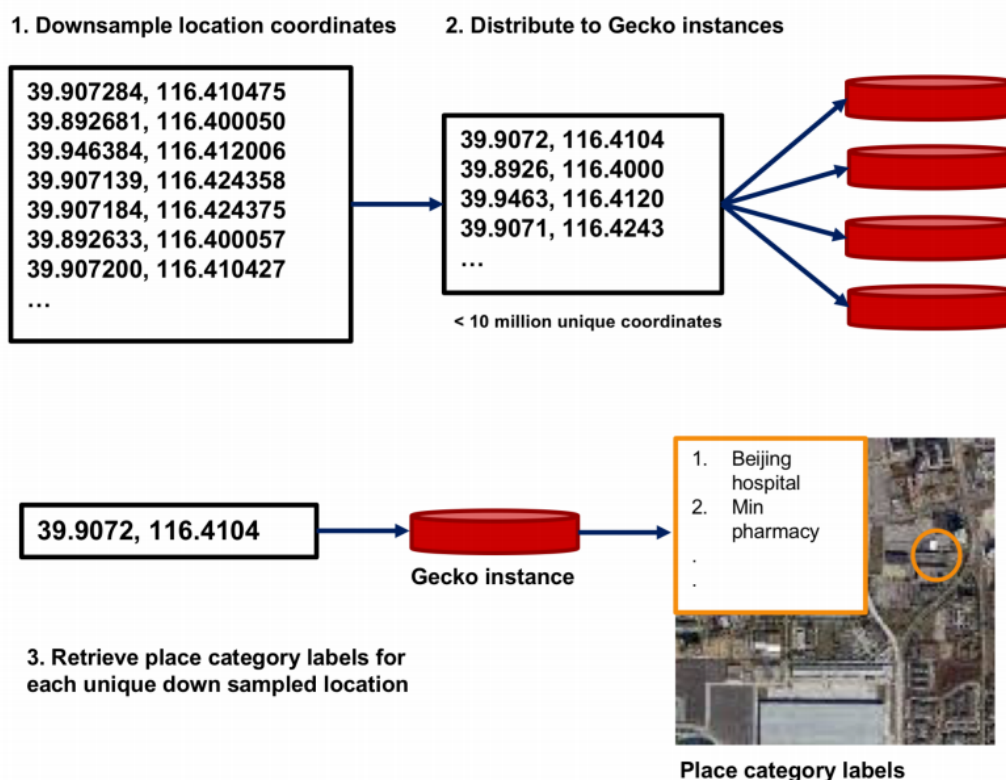
Building Prediction Models

We constructed a variety of supervised machine-learning models based on our aggregate and daywise features. In fitting our models, the aggregate feature set and the daywise feature sets were divided into 80% for training and the remaining 20% for testing. Given the sparsity and correlations within our features, we focused primarily on using regularized models to reduce the dimensionality of our feature set and to avoid overfitting. All machine-learning analyses were performed using R 3.2.0 (R Development Core Team, Vienna, Austria). We selected linear, nonparametric, and ensemble methods to evaluate the best fit to our data. For linear models, we built lasso, ridge, and elastic net models using the "glmnet" package. Five-fold cross-validation was applied to the training set to determine the optimal tuning parameter lambda for lasso and ridge classification. For elastic net, a grid search was performed to determine lambda and alpha. The lambda that was within 1 standard error of the lambda that produced the minimum cross-validation error was selected for these models to prevent overfitting. In addition to linear models, we built support vector machine (SVM) models with a Gaussian kernel (using e1071) and random forest models (using the "randomForest" package). For our SVM models, gamma was set to 1 divided by the number of features, and the cost was chosen via cross-validation.

To evaluate the performance of our models, we constructed receiver operating characteristic (ROC) curves. We used the area under the curve (AUC) of the ROC curve to compare the performance of our classifiers in the held-out test sets.

Choosing the Most Informative Features

As described in the section on feature engineering, our initial feature design choices were guided by insights generated in prior work as well as by our experience of mining medical content. Our three feature categories attempt to discriminate between utilizers and nonutilizers based general search use, search contents, and search locations, respectively. Learning spatial trajectory patterns from location tags and learning linguistic patterns from embedded search texts requires the use methods from different subfields within machine learning, each an area of active research by itself. To guide further work on feature design and improvement, we measured the individual contribution of each of three feature categories on prediction performance. We trained three models on three different feature matrices, each containing features from only two of the three feature categories. Each model was tested against the held-out test data, featured in the same way.

Measuring the Impact of Utilization Prediction

We validated our prediction model through an experiment conducted in Baidu's mobile search ads system, which charges advertisers based on user clicks. The model used in online evaluation is modified to comply with commercial limitations (such as the use of location APIs). However, the model contains similar category of features, reconstructed from the raw search log data. Medical facilities that advertise via the system are interested in a low cost per conversion (CPC) and a higher show conversion rate, which implies more efficient use of the advertising budget. Our objective was to measure the relationship between ad conversions and the prediction of health care utilization. In particular, we wished to evaluate if displaying medical facility advertisements to predicted health care utilizers results in higher show conversion rate, "conversion" being defined as a single health care utilization by a search user, satisfying the following two conditions: (1) the search user had no utilization of the same medical facility within 1 month before the conversion, and (2) the utilization took place within 2 weeks after the search user was shown the ads from this specific hospital.

Intuitively, the first condition restricts the conversion to be a new hospital utilization rather than a readmission, while the second encourages a relationship between ads display and hospitalization. A rigorous argument establishing the linkage

between condition 2 and causality is beyond the scope of this paper. Instead, we refer the interested reader to relevant work in this area [27,28] and the references therein.

Offline Evaluation

We carried out an offline evaluation prior to deploying the model in a production setting. Our goal was to ascertain the efficacy of the predicted health care utilization scores on real posterior conversion data. We first collected logs of all search users who had been shown medical facility advertisements between June 1 and June 7, 2015. Using our prediction model, we generated health care utilization scores for these users based on their search logs from May 2 through May 31, 2015. For each search user A, we obtained a predicted health care score (a real number between 0 and 1) and a conversion label (0 or 1). We adopted the following metric, which is considered to be more interpretable for the evaluation task at hand compared to the conventional AUC. We ranked these search users by their predicted scores and plotted the show conversion rate versus risk score. The local show conversion rate for a search user is defined as the conversion proportion within a T-person window centered at the risk score for the user (Figure 6). This local show conversion rate at a given risk score represents the approximate conversion probability of people with scores close to the given risk score and is aligned with the way our model was adopted in the online experiment (described in the following sections).

Figure 6. Equation for local show conversion rate.

$$\text{Prediction score for a search user } A_i \text{ is denoted } \hat{p}_{A_i}$$

$$\hat{p}_{A_1} \leq \hat{p}_{A_2} \leq \cdots \leq \hat{p}_{A_n}$$

The local conversion rate for a search user A_i is denoted as $r(\hat{p}_{A_i})$

Formally

$$r(\hat{p}_{A_i}) = \sum_{B \in W_T(A_i)} L_B \ / \ |W_T(A_i)| \, ,$$

$$\text{where } W_T(A_i) = \left\{ A_j \colon\ i - \frac{T}{2} \leq j \leq i + \frac{T}{2} \right\}$$

Online Evaluation

Health care utilization prediction scores were generated on a daily basis from search users' previous day's logs. Feature generation and prediction was carried out using Baidu's MapReduce framework. MapReduce jobs running on a cluster with 50 HPC machines, composed of 550 physical cores and 6.4 terabytes of memory in total typically take 3 hours for feature and score generation. The predicted scores are stored in Baidu's online k-v servers with 70 gigabytes of memory, where the online ads system can use them to modify the ads bidding coefficients. The online evaluation was carried out on approximately 10 million mobile users of Baidu apps. The user pool for this experiment is large enough to avoid user-to-user variations. We split these users into two parts to conduct the A/B test. For each user in the treatment group B, the predicted score was mapped to a coefficient in the interval (0.7, 1.3), which was then used to adjust the probability that a health care

ad was displayed, whereas the control group A was left unchanged.

Results

Predictive Model for Health Utilization

The AUCs for our feature sets across all models are shown in Table 2. Our models showed the poorest performance with the aggregate features, with the highest AUC at 0.627 from the lasso model. On the daywise features, we observed the best performance with random forest, reaching an AUC of 0.796. The AUC when a single feature category was omitted was 0.781, 0.789, and 0.779 when the omitted categories were semantic, general, and location, respectively. Omission of location features caused the largest drop in AUC, followed by the omission of semantic and general features. We used a random forests classifier with daywise features for measuring the individual contribution of the feature categories.

Table 2. Areas under the curve for aggregate and daywise feature sets.

Model	Aggregate		Daywise	
	training	Test	training	Test
Lasso	0.899	0.627	0.899	0.589
Ridge	0.890	0.598	0.920	0.639
Elastic net	0.901	0.601	0.920	0.621
Support vector machine, radial kernel	0.983	0.590	—	—

Predicting Utilization by Specialty

Many of the geofenced locations allow sublocalization of searches via network access points that served as sublocation indicators. For medical facilities where this was feasible, the user search logs could be tagged with access point identifiers that were, in turn, mapped to hospital departments. We collapsed department tags into four broad categories covering treatments related to male utilizers, treatments related to female utilizers, beauty treatments, and other specialty treatments, and investigated the feasibility of learning search log predictors that may classify a health care utilization by visit type. For a random forests classifier trained on our daywise features, we obtained a held-out test accuracy of 0.632 and AUC of 0.592, where AUC for the multiclass classification was computed as a mean of all pairwise comparisons as discussed by Hand and Till [29].

Feature Importance

An examination of the top 30 positive and negative features from our daywise models (tabulated in Multimedia Appendix 2) indicates a high degree of overlap among the top ranking features. The overlap increased if we disregarded the day for which the feature was selected. All features pertained to search log events within 15 search days preceding the outcome (health care utilization). The number of search sessions per search day was among the top features for all models except the elastic net. The interval reduction score feature, which summarizes the lengthening/shortening of search intervals for all searches made by a user per day, appeared among the top 30 positively associated features in two of our models. The number of visits to place categories labeled as accommodation or health care facility appeared as important features in three of the daywise models.

Offline Evaluation

Figure 7 shows the expected change in the show conversion rate with changes in the prediction risk score. Note that the horizontal axis has been normalized and represents the percentile of population with increasing predicted scores. Therefore, the AUC approximates the proportion of all converted ad views. Our offline experiment using data from May 2 through May 31, 2015, resulted in a monotonically increasing curve confirming that a high predicted score implies a high show conversion rate; thus, justifying our online experiment.

Figure 7. Show conversion rate versus utilization prediction score.
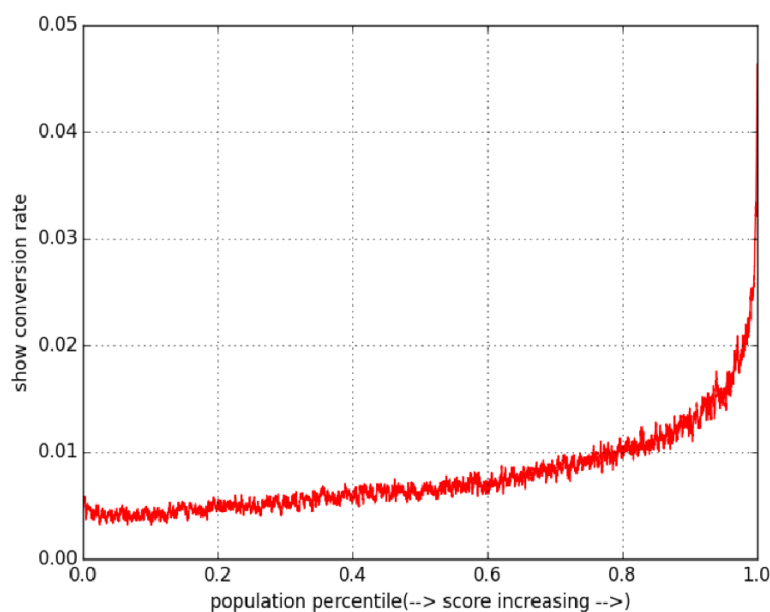
Online Evaluation

Several metrics including show conversion (the ratio between number of ads shown and hospital visits) and the cost per action (CPA) for shown conversions were tracked in the online experiment for the time period between August 24 and September 27, 2015. The reported numbers are the relative change between the experiment group and the control group. For reasons of confidentiality, we do not disclose the revenue of the advertisers. Two models were applied separately on two experiment groups, one had location-based features and the other did not. As expected, the model with location features showed a higher conversion (3.96% vs 0.67%) and lower CPA (–1.77% vs 1.61%).

Discussion

The value of Internet search logs as valuable repositories of patient-generated biomedical information is widely appreciated. Because location data offer an opportunity to link search users' virtual behavior with their real-world activities, it holds promise for many areas of medical research that have heretofore remained impregnable to conventional research techniques.

Principal Results

The experiments described in our study demonstrate the relationship between predicted health care utilization and show conversion rate of targeted advertising (our surrogate measure of utility). The show conversion evaluates how many hospital visits happen after we show health care-related ads to the users and demonstrates that our online evaluation results are in good agreement with the model. Both models (with or without location-based features) have higher show conversion, suggesting users in experiment groups are prone to be affected by those advertisements. The model with location-based features has higher change of show conversion, which is consistent with results that the performance drops most suddenly when omitting the location features. The CPA is a similar metric to the CPC, which considers the cost per show conversions. It is observed that the location-based features helped lower CPA.

Limitations

With its high sparsity, missingness, and vulnerability to contamination, geotagged search data also pose unique challenges for informatics research. A key limitation is the vulnerability to noise from searches close to the observation boundaries, as well as from false negatives that arise from our labeling approach. We expect that search data that span larger observation windows will allow creation of clear feature sets. Methods that attempt to learn from only positive labeled data may also be explored to control for false negative labels. Although feature sparsity in longitudinal patient histories is an active area of research in clinical informatics, search log texts add a new dimension to the challenge in at least two ways. First, given the "consumer" nature of the search texts, identifying biomedical content requires the use of a consumer health lexicon as well as session-level context detection, unlike patient notes where a term within the context of the note unambiguously represents a concept. Secondly, the "Internet scale" of the data and the approximated truth labels result in long-tailed

distributions of potential health care concepts (eg, the occurrence of the token "gift" in health care utilizers' searches versus in nonutilizers' searches).

The relationship between the performance of prediction algorithms and the impact on utility—analogous to net reclassification in clinical studies [30]—has not been examined in our work. Specifically, the decision rules mapping a health care utilization prediction score to the probability of a health care advertisement being served impacts the overall show conversion rate. For example, a prediction model with a high AUC (capable of a high true positive rate and a low false positive rate simultaneously) coupled to a decision rule that results in advertisement displays to searcher users with low prediction scores would be inefficient. In general, this holds true in many health care prediction tasks where a high AUC may be a misleading indicator of the utility of the prediction model as it effectively leads to no net reclassification [30,31]. A full investigation into dynamic calibration of decision rules for mapping prediction scores to advertisement views and the impact on CPC metrics is beyond the scope of this study.

Comparison With Prior Work

We note that mining of search texts for the purposes of syndromic surveillance has been actively studied in the past [32-34]. The study by Wang et al [35] in forecasting new outpatient visits related to dementia based on predictive tokens in search texts focuses on the problem of predicting health care utilization from the provider's perspective. Nagar et al [36] have constructed spatiotemporal models based on tweets localized to New York for influenza surveillance. Our experiments suggest that in addition to using tokens as semantic predictors, features based on the location of the search improve the performance of utilization prediction models. Further, we were able to link click-throughs with subsequent geolocation search data to improvise a metric for assessing prediction impact.

Our work on predicting health care utilization from geotagged search logs is conceptually similar to the privacy-sensitive analysis of geotagged data from mobile devices by White and Horvitz [2] in that we make use of biomedical lexical resources to characterize the medical content in a search. However, our approach is novel in the use of a stack of temporal features based on a fixed analysis window of search days. The higher resolution of our approach is able to capture the progression of a wide variety of search attributes. Among our top daywise features, the number of queries on the search day preceding the day of a medical visit is selected by both the random forests and the ridge regression models and is in agreement with the results from White and Horvitz.

The association between health care utilization and temporal news trends has been examined previously [37]. A key finding in earlier studies on search log-characterized patient behavior is the escalation of health-related anxiety in the period leading to a health care utilization episode [8,21]. We believe that characterizing the temporal progression of relevant search features can reveal markers of anxiety escalation that precipitate health care utilization. In this work, we relied on the information content of search text tokens as a measure of the specificity of the text and compute the mean of the highest information content

score (most specific token) of all searches made by a user within a search day. The mean information content score a few weeks prior to the hospital visit is selected by our L1 penalized logistic regression model as one of the top predictors. Given the sparse feature space (consisting of more than 1 million tokens), the L1 penalty results in the shrinkage of all but a small set of predictors. The mean information content score is unable to survive the selection criterion in models that do not have such sparsity-handling mechanisms.

In addition, we define another measure of anxiety—the interval reduction score. The interval reduction score, as defined in the Feature Engineering section, is the mean (daywise) shortening or lengthening of the interval between consecutive health care-related searches. For two of our daywise models, the interval reduction score in the 2 weeks prior to the hospital visit is among the top predictors. The models trained on daywise features for predicting health care utilization performed better compared to the models trained on aggregate features suggesting that daywise progression of search attributes better represents search behavior that signals health care resource utilization.

Conclusions

Overall, the results of the two sets of experiments—first the proof-of-concept done in a research setting and second the offline as well as online experiments on the utility of health utilization predictions—support the claim that it is possible to accurately predict future patient visits from geotagged mobile search logs and that such prediction can have utility for health care providers.

Authors' Contributions

NHS, VA, and CH envisioned the study. VA performed the feature engineering and statistical modeling experiments at Stanford. LZ, JZ, SF, and TC performed the offline and online experiments at Baidu. VA and LZ compiled the results. NHS, VA, and LZ participated in the editing of the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interval Reduction Score.

[PDF File (Adobe PDF File), 30KB - jmir_v18i9e251_app1.pdf]

Multimedia Appendix 2

Top 30 daywise features.

[PDF File (Adobe PDF File), 193KB - jmir_v18i9e251_app2.pdf]

References

1. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
2. White R, Horvitz E. From web search to healthcare utilization: privacy-sensitive studies from mobile data. J Am Med Inform Assoc 2013 Jan 1;20(1):61-68 [FREE Full text] [doi: 10.1136/amiajnl-2011-000765] [Medline: 22661560]
3. White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc 2014;21(1):49-55 [FREE Full text] [doi: 10.1136/amiajnl-2012-001473] [Medline: 23666794]
4. Horvitz E, Mulligan D. Machine learning and inference makes it increasingly difficult for individuals to understand what others can. Science 2015;349(6245):253-255 [FREE Full text]
5. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
6. Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J Oncol Pract 2016 Jun 7:737-744. [doi: 10.1200/JOP.2015.010504] [Medline: 27271506]
7. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013 May 1;20(3):404-408 [FREE Full text] [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]

8.   Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]

9.   Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]

10. Hu Z, Hao S, Jin B, Shin A, Zhu C, Huang M, et al. Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. J Med Internet Res 2015;17(9):e219 [FREE Full text] [doi: 10.2196/jmir.4976] [Medline: 26395541]

11. Rakovski C, Rosen A, Wang F, Berlowitz D. Predicting elderly at risk of increased future healthcare use: How much does diagnostic information add to prior utilization? Heal Serv Outcomes Res Methodol 2002;3(3-4):267-277. [doi: 10.1023/A:1025866331616]

12. Phithakkitnukoon S, Smoreda Z, Olivier P. Socio-geography of human mobility: a study using longitudinal mobile phone data. PLoS One 2012;7(6):1-9 [FREE Full text] [doi: 10.1371/journal.pone.0039253] [Medline: 22761748]

13. Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. Circulation 2011 Mar 15;123(10):1116-1124 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.943860] [Medline: 21403122]

14. Wilkinson D. Knowing when to stop: futility in the intensive care unit. Curr Opin Anaesthesiol 2011;24(2):160-165 [FREE Full text] [doi: 10.1097/ACO.0b013e328343c5af]

15. Yan J, Liu N, Wang G, Zhang W. How much can behavioral targeting help online advertising? In: Proceedings of the 18th International Conference on World Wide Web. 2009 Presented at: 18th International Conference on World Wide Web; Apr 20-24, 2009; Madrid p. 261-270. [doi: 10.1145/1526709.1526745]

16. Chen J, Stallaert J. An economic analysis of online advertising using behavioral targeting. MIS Quarterly 2010;38(2):429-449 [FREE Full text]

17. Barr C, Jones R, Regelson M. The linguistic structure of English web-search queries. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: Empirical Methods in Natural Language Processing; Oct 25-27, 2008; Honolulu p. 1021-1030. [doi: 10.3115/1613715.1613848]

18. Jadhav A, Andrews D, Fiksdal A, Kumbamu A, McCormick JB, Misitano A, et al. Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. J Med Internet Res 2014;16(7):e160 [FREE Full text] [doi: 10.2196/jmir.3186] [Medline: 25000537]

19. Callahan A, Pernek I, Stiglic G, Leskovec J, Strasberg HR, Shah NH. Analyzing information seeking and drug-safety alert response by health care professionals as new methods for surveillance. J Med Internet Res 2015;17(8):e204 [FREE Full text] [doi: 10.2196/jmir.4427] [Medline: 26293444]

20. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]

21. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011 Jul;39(Web Server issue):W541-W545 [FREE Full text] [doi: 10.1093/nar/gkr469] [Medline: 21672956]

22. Jung K, LePendu P, Iyer S, Bauer-Mehren A, Percha B, Shah NH. Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. J Am Med Inform Assoc 2015 Jan;22(1):121-131 [FREE Full text] [doi: 10.1136/amiajnl-2014-002902] [Medline: 25336595]

23. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. J Biomed Inform 2003 Dec;36(6):414-432 [FREE Full text] [doi: 10.1016/j.jbi.2003.11.002] [Medline: 14759816]

24. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform 2001;84(Pt 1):216-220 [FREE Full text] [Medline: 11604736]

25. Wu ST, Liu H, Li D, Tao C, Musen MA, Chute CG, et al. Unified Medical Language System term occurrences in clinical notes: a large-scale corpus analysis. J Am Med Inform Assoc 2012 Jun;19(e1):e149-e156 [FREE Full text] [doi: 10.1136/amiajnl-2011-000744] [Medline: 22493050]

26. Muse K, McManus F, Leung C, Meghreblian B, Williams JM. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. J Anxiety Disord 2012 Jan;26(1):189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]

27. Barajas J, Kwon J, Akella R, Flores A, Holtan M, Andrei V. Marketing campaign evaluation in targeted display advertising. In: Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy. 2012 Presented at: ADKDD '12 Sixth International Workshop on Data Mining for Online Advertising and Internet Economy; Aug 12-16, 2012; Beijing p. 1-7. [doi: 10.1145/2351356.2351361]

28. Barajas J, Akella R, Holtan M, Kwon J, Flores A, Andrei V. Dynamic effects of ad impressions on commercial actions in display advertising. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. 2012 Presented at: CIKM '12 21st ACM International Conference on Information and Knowledge Management; Oct 29-Nov 2, 2012; Maui. [doi: 10.1145/2396761.2398510]

29.   Hand D, Till R. A simple generalisation of the area under the roc curve for multiple class classification problems. Mach Learn 2001;45(2):171-186 [FREE Full text] [doi: 10.1023/A:1010920819831]

30.   Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007 Feb 20;115(7):928-935 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.106.672402] [Medline: 17309939]

31.   Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem 2008 Jan;54(1):17-23 [FREE Full text] [doi: 10.1373/clinchem.2007.096529] [Medline: 18024533]

32.   Majumder M, Santillana M, Mekaru S, McGinnis D, Khan K, Brownstein J. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveill 2016;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]

33.   Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. J Med Internet Res 2014;16(12):e289 [FREE Full text] [doi: 10.2196/jmir.3680] [Medline: 25517353]

34.   Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. J Med Internet Res 2016;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]

35.   Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with Google trends: evidence from Taiwan. J Med Internet Res 2015;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]

36.   Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014;16(10):e236 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]

37.   Ghosh S, Chakraborty P, Nsoesie EO, Cohn E, Mekaru SR, Brownstein JS. Temporal Topic Modeling to Assess Associations between News Trends and Infectious Disease Outbreaks. arXiv:1606.00411 2016:1-21.

Abbreviations

API: application program interface
AUC: area under the curve
CPA: cost per action
CPC: cost per conversion
ROC: receiver operating characteristic
SVM: support vector machine
UMLS: Unified Medical Language System

XSL•FO
RenderX

Original Paper

Incentive and Reminder Strategies to Improve Response Rate for Internet-Based Physician Surveys: A Randomized Experiment

David A Cook[1,2,3], MHPE, MD; Christopher M Wittich[3], PharmD, MD; Wendlyn L Daniels[4]; Colin P West[3,5], MD, PhD; Ann M Harris[4], BA; Timothy J Beebe[4], PhD

[1]Mayo Clinic Online Learning, Mayo Clinic College of Medicine, Rochester, MN, United States

[2]Knowledge Delivery Center, Mayo Clinic College of Medicine, Rochester, MN, United States

[3]Division of General Internal Medicine, Mayo Clinic, Rochester, MN, United States

[4]Survey Research Center, Mayo Clinic, Rochester, MN, United States

[5]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
David A Cook, MHPE, MD
Division of General Internal Medicine
Mayo Clinic
200 First St SW
Rochester, MN, 55905
United States
Phone: 1 5072842269
Fax: 1 5072845370
Email: cook.david33@mayo.edu

Abstract

Background: Most research on how to enhance response rates in physician surveys has been done using paper surveys. Uncertainties remain regarding how to enhance response rates in Internet-based surveys.

Objective: To evaluate the impact of a low-cost nonmonetary incentive and paper mail reminders (formal letter and postcard) on response rates in Internet-based physician surveys.

Methods: We executed a factorial-design randomized experiment while conducting a nationally representative Internet-based physician survey. We invited 3966 physicians (randomly selected from a commercial database of all licensed US physicians) via email to complete an Internet-based survey. We used 2 randomly assigned email messages: one message offered a book upon survey completion, whereas the other did not mention the book but was otherwise identical. All nonrespondents received several email reminders. Some physicians were further assigned at random to receive 1 reminder via paper mail (either a postcard or a letter) or no paper reminder. The primary outcome of this study was the survey response rate.

Results: Of the 3966 physicians who were invited, 451 (11.4%) responded to at least one survey question and 336 (8.5%) completed the entire survey. Of those who were offered a book, 345/2973 (11.6%) responded compared with 106/993 (10.7%) who were not offered a book (odds ratio 1.10, 95% CI 0.87-1.38, *P*=.42). Regarding the paper mail reminder, 168/1572 (10.7%) letter recipients, 148/1561 (9.5%) postcard recipients, and 69/767 (9.0%) email-only recipients responded (*P*=.35). The response rate for those receiving letters or postcards was similar (odds ratio 1.14, 95% CI 0.91-1.44, *P*=.26).

Conclusions: Offering a modest nonmonetary incentive and sending a paper reminder did not improve survey response rate. Further research on how to enhance response rates in Internet-based physician surveys is needed.

KEYWORDS

surveys and questionnaires; survey methods; questionnaire design

Introduction

Surveys remain a widely used, cost-effective means of assessing the attitudes, beliefs, and practices of physicians. However, conducting physician surveys is challenging. Response rate, a benchmark commonly used in appraising survey quality, is typically 10% lower in physician surveys than in nonphysician surveys [1,2] and appears to be decreasing for many physician

subgroups [3,4]. Although a high response rate may not be the best predictor of overall survey quality [5,6], low response rates raise concerns about study precision, nonresponse bias, and the overall inferential value of study findings.

Recent reviews of the literature on physician survey response behavior have identified the method of initial contact, mode of administration, and approach to nonrespondent follow-up as important determinants of response rate [4,7-10]. Electronic survey methods, and Internet-based surveys in particular, are becoming increasingly popular owing to their potential for reduced time spent in collecting data, immediate availability of data once collected, improved data quality, and overall cost savings [11-16]. However, electronic modalities often have lower response rates than paper mailed surveys [4,7,11,15-24].

Although uncertainties remain regarding how to enhance physician response rate in Internet-based surveys, incentives constitute one possible support. Incentives increase the perception of value, trust, reciprocity, and appreciation on the part of the respondent [4,25]. Research indicates that incentives improve response rates and that monetary incentives are more effective than nonmonetary incentives [3,7,8,26,27], yet most studies investigating the impact of incentives on physician survey response have done so in the context of mailed, paper-based forms [11]. There remains a paucity of evidence concerning how incentives work in the context of electronic surveys of physicians.

Physician survey response can also be improved through multiple follow-up contacts, although the optimum number of contact attempts remains unclear [27]. In the context of an Internet-based survey, sending follow-up reminders via email to invitees who fail to complete an Internet-based survey after the initial invitation costs virtually nothing [11]. However, repeated email reminders may backfire, hardening prospective respondents' resistance to future requests. Thus, email reminders may be less desirable than reminders via other methods (eg, mail or telephone). In one study, even after sending the nonresponding physicians 4 emailed reminders, the overall response rate improved substantially after investigators sent nonresponding physicians a paper letter with a printed URL that the physicians used to access the Internet-based survey [12]. Little consensus exists regarding what reminder approaches are most efficient and effective in increasing response rates among physicians, regardless of data collection mode. This

investigation is a response to specific calls for further testing in this area [28,29].

Despite the importance of physician surveys and ongoing concerns over participation, few randomized trials have examined potential strategies to address nonresponse [7,25]. Moreover, even when physician surveys are optimally executed (and enjoy high levels of participation) the investigators infrequently describe the challenges they encountered [3,28]. As such, significant gaps remain in our understanding of best practices in physician survey research. We sought to address these knowledge gaps.

We hypothesized that a low-cost nonmonetary incentive and a paper mail reminder would improve response rates in an Internet-based survey of physicians. We also hypothesized that a personalized formal letter would have only a slightly greater benefit than a personalized postcard. We tested these hypotheses using a factorial-design randomized experiment in the context of a nationally representative Internet-based physician survey. To our knowledge, no one has applied such a design to study the process of conducting physician surveys.

Methods

Survey Procedures

From September to December 2015, we conducted a self-administered Internet survey among a random sample of licensed physicians in the United States. The survey sought physician opinions regarding 2 topics that we believed they would perceive as personally important, namely, maintenance of certification and continuing medical education. Maintenance of certification is presently a matter of controversy and debate [30,31], and both this and continuing medical education [32,33] directly affect the professional lives of nearly all physicians. Details of survey development and results will be published separately.

The survey was administered using Qualtrics (Qualtrics LLC, Provo, UT, USA). All invitees received an email (Textbox 1) inviting them to complete the survey, accessible via an embedded URL link unique to each invitee. Consistent with the protocol used by Braithwaite and colleagues [15] in which it took 5 reminder emails to achieve a 52% response rate, we sent 6 reminder emails on days 11, 20, 31, 36, 48, and 58. A subset of participants also received 1 paper reminder as outlined below.

Textbox 1. Initial invitation email.

[Email subject line: CME and MOC – Survey of physicians' needs]

Dear Dr. [last name],

Continuous professional development (CPD), including maintenance of certification (MOC), affects all physicians. Our group is trying to promote changes to make CPD easier and more effective.

We invite you to participate in a nation-wide survey, and make your voice heard on these important issues! *Those who complete the survey can request a free copy of the "Mayo Clinic Handbook for Happiness: A Four-Step Plan for Resilient Living."*

This survey asks physicians' opinions about MOC, continuing medical education, and online learning. We want to understand what you do to maintain your professional knowledge and skills, what challenges you face, and—most importantly—what needs to change.

We anticipate this will take about 7 minutes to complete. Your responses will be anonymous. We would be grateful to receive responses by October 21, 2015.

Click here to start the survey.

We thank you in advance for your participation. Feel free to contact any of us if you have questions.

Sincerely,

[study investigators]

This study has been approved by the Mayo Clinic Institutional Review Board. All responses will be anonymous (unless otherwise explicitly noted) and strictly confidential.

> *Editorial note: The italicized text was deleted in the "no book offer" group for the initial email and all subsequent reminders. The email text varied slightly for each follow-up email, but the incentive text remained unchanged. CME = continuing medical education.*

Interventions

This factorial-design study tested 2 interventions (Figure 1). First, we systematically altered the emails inviting physicians to participate, with one version offering a book (the *Mayo Clinic Handbook for Happiness: A Four-Step Plan for Resilient Living* nominal cost about US $10) upon completion of the survey and the other version making no mention of any tangible incentive (Textbox 1). Although the wording of the reminder emails varied slightly, the presence or absence of the book offer remained constant. We randomly assigned 75% of invitees to receive emails with the book offer. All respondents were offered the book upon completion of the survey; the intervention in this case was the *up-front notification* regarding the book.

Second, 7 days after the first email we sent a reminder via paper mail to a subset of invitees using two formats. We timed this mailing to arrive at approximately the same time as the first email reminder. The invitees in one group received a personalized letter (Multimedia Appendix 1, Box 1), printed on institution letterhead bonded paper and sealed in an envelope, asking them to complete the survey using the link they had received via email, or to contact the study investigators if they had not received or had deleted the email. A second group received a similar message via a personalized postcard (Multimedia Appendix 1, Box 2) that included the institution logo on both sides and was signed by one of the investigators. A third group received no paper reminder. We randomly assigned 40% of the initial sample to the letter group, 40% to the postcard group, and 20% to the no paper reminder group. Only 1 paper reminder was sent.

Figure 1. Participant flow.
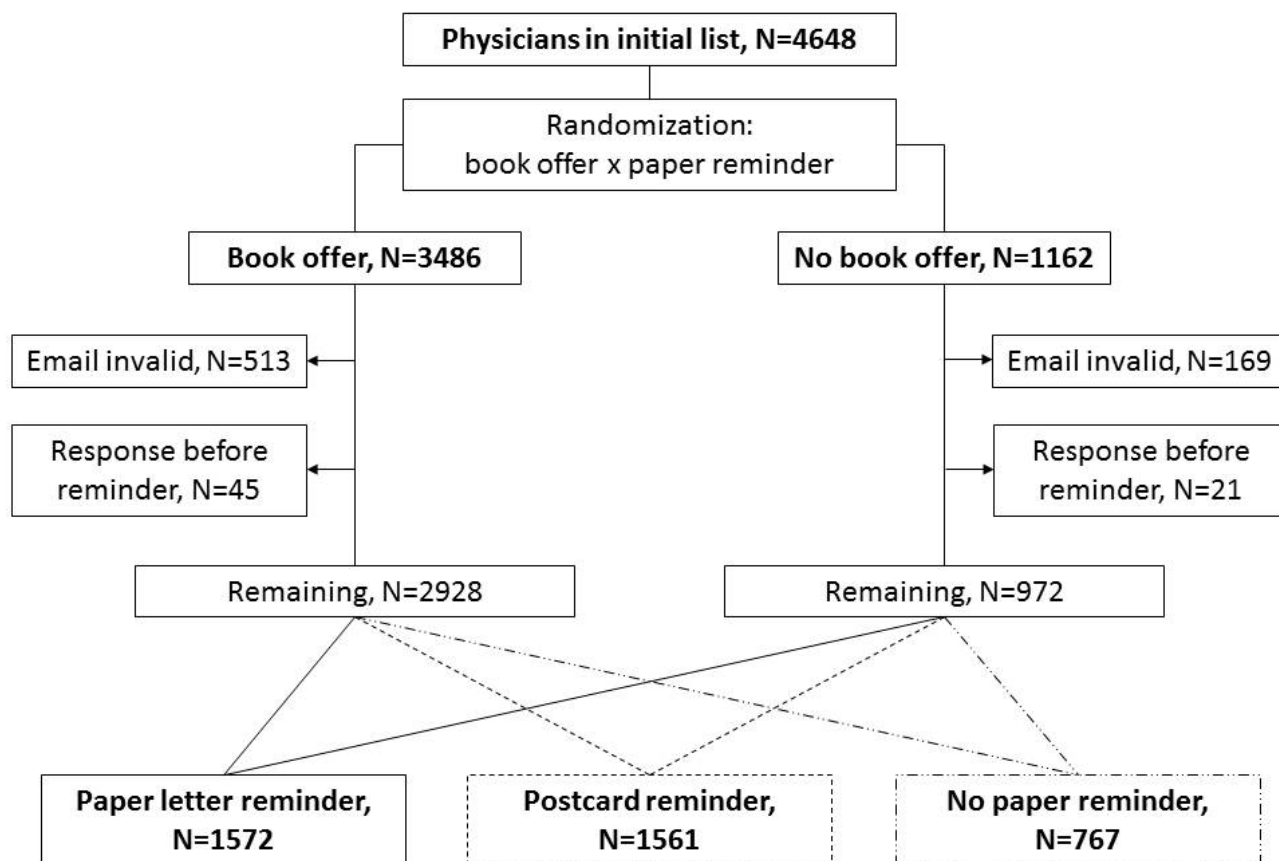
Human Subjects

We obtained a sample of 4648 names and email addresses selected at random from the LexisNexis Provider Data Management and Services database of all licensed US physicians (LexisNexis Risk Solutions, Alpharetta, GA). This study was approved by the Mayo Clinic Institutional Review Board. Survey response was tracked but responses were deidentified before analysis. Randomization for both interventions was done at the same time before the survey began by a research assistant using a random number generator (Microsoft Excel).

Outcomes and Data Collection

The primary outcome of this study was the survey response calculated using the American Association for Public Opinion Research formula RR2 (Response Rate 2) [34], that is, the number answering at least one survey question divided by the number of surveys sent less those returned as undeliverable. Secondary outcomes included the number of respondents who completed the entire survey, those who actually claimed a book, and time to response (defined as the number of days between the initial survey mailing and the response date). We obtained demographic information on physician sex, age, degree, specialty, and practice location from the LexisNexis database. For analysis purposes we dichotomized physician specialty as generalist (family medicine, general internal medicine, or general pediatrics) and nongeneralist.

Data Analysis

We compared response rates between interventions and across physician subgroups using the chi-squared test and report odds ratios with 95% confidence intervals for dichotomous variables. We used logistic regression to explore potential interactions between the interventions and physician demographic subgroups. We graphed time to response to visually explore its association with reminder emails. We performed statistical analyses using SAS version 9.4 (SAS Institute, Inc, Cary, NC, USA), using two-sided tests and $P \leq .05$ as the threshold of statistical significance. To achieve 80% power to detect an absolute improvement of 5%, we estimated we would need to invite 4000 physicians (providing 81% power assuming baseline response rate of 10%, 87% power assuming baseline rate of 25%). All participants were analyzed in the groups to which they were originally assigned ("intent to treat").

Results

Survey Response

We obtained the email addresses of 4648 licensed physicians, randomly assigned these to intervention arms (Figure 1), and sent them an email invitation. We received notification that 682 emails were undeliverable, leaving 3966 potential respondents (Table 1). Of these 3966 physicians, 451 (11.4%) responded to at least one survey question during the study period and 336 (8.5%) completed the entire survey. Twenty-six physicians clicked the survey link but did not answer any questions, and

84 opted out of the survey. We received no response during the study period for the remaining 3405 emails.

Of the undeliverable emails, for 673 we received notice that the email address was invalid and for 9 we received automated messages indicating that the physician would only accept emails from approved senders. One representative message read, "To control spam, I now allow incoming messages only from senders I have approved beforehand. If you would like to be added to my list of approved senders, please fill out the short request form (see link below)." We did not fill out such forms.

Impact of Book Incentive Offer

Of 2973 physicians who were offered a book in the email invitation, 345 (11.6%) responded to at least one survey question compared with 106 of 993 (10.7%) physicians who were not offered a book (odds ratio 1.10, 95% CI 0.87-1.38, P=.42); see Figure 2. Converting this odds ratio to absolute rates relative to the baseline rate for those not offered a book, the 95% CI for response among those offered a book ranged from 9.5% to 14.2%.

Among the 336 respondents who completed the entire survey, 224 (66.7%) requested a book. Those who had been offered a book ultimately requested one slightly more often (177/257, 68.9%) than those who had not been offered a book (47/79, 59.5%; odds ratio 1.51, 95% CI 0.89-2.54, P=.12).

Table 1. Demographics of invitees and respondents.

| Demographic feature	Invitees (N=3966)[a]	Respondents (N=451)[a]
Age in years, mean (SD; interquartile range)	54.9 (11.5; 46-63)	55.3 (10.5; 47-62)
Male, n (%)	2637 (66.5)	290 (64.3)
Degree, MD[b], n (%)	3791 (95.6)	442 (98.0)
Specialty[c], n (%)		
General medicine or pediatrics	1244 (31.5)	117 (25.9)
Surgery or anesthesia	1041 (26.4)	128 (28.4)
Nongeneral clinical	1412 (35.8)	174 (38.6)
Diagnostic	253 (6.4)	32 (7.1)
Region[d], n (%)		
Northeast	852 (21.5)	98 (21.7)
Midwest	803 (20.3)	97 (21.5)
South	1364 (34.4)	151 (33.5)
West	941 (23.8)	105 (23.3)

[a] Invitees are physicians who were sent an email that was not returned as undeliverable; respondents are those who answered one or more survey questions. Responses may not sum to column total N owing to missing data.

[b] MD: doctor of medicine. Non-MD degrees were doctor of osteopathy (DO).

[c] General medicine includes family medicine and internal medicine; surgery includes ophthalmology and obstetrics/gynecology; nongeneral clinical includes medical subspecialties (cardiology, pulmonology, etc), dermatology, emergency medicine, neurology, and physical medicine, among others; diagnostic includes radiology and pathology and their subspecialties.

[d] Regions are classified according to US Census Bureau definitions.

Figure 2. Survey response rate with different incentives and reminders. Dark gray bars show response rate by incentive (book offer or none), *P*=.42. Light gray bars show response rate by paper reminder (letter, postcard, or none), *P*=.35. Error bars indicate the 95% confidence interval for the binomial proportion.
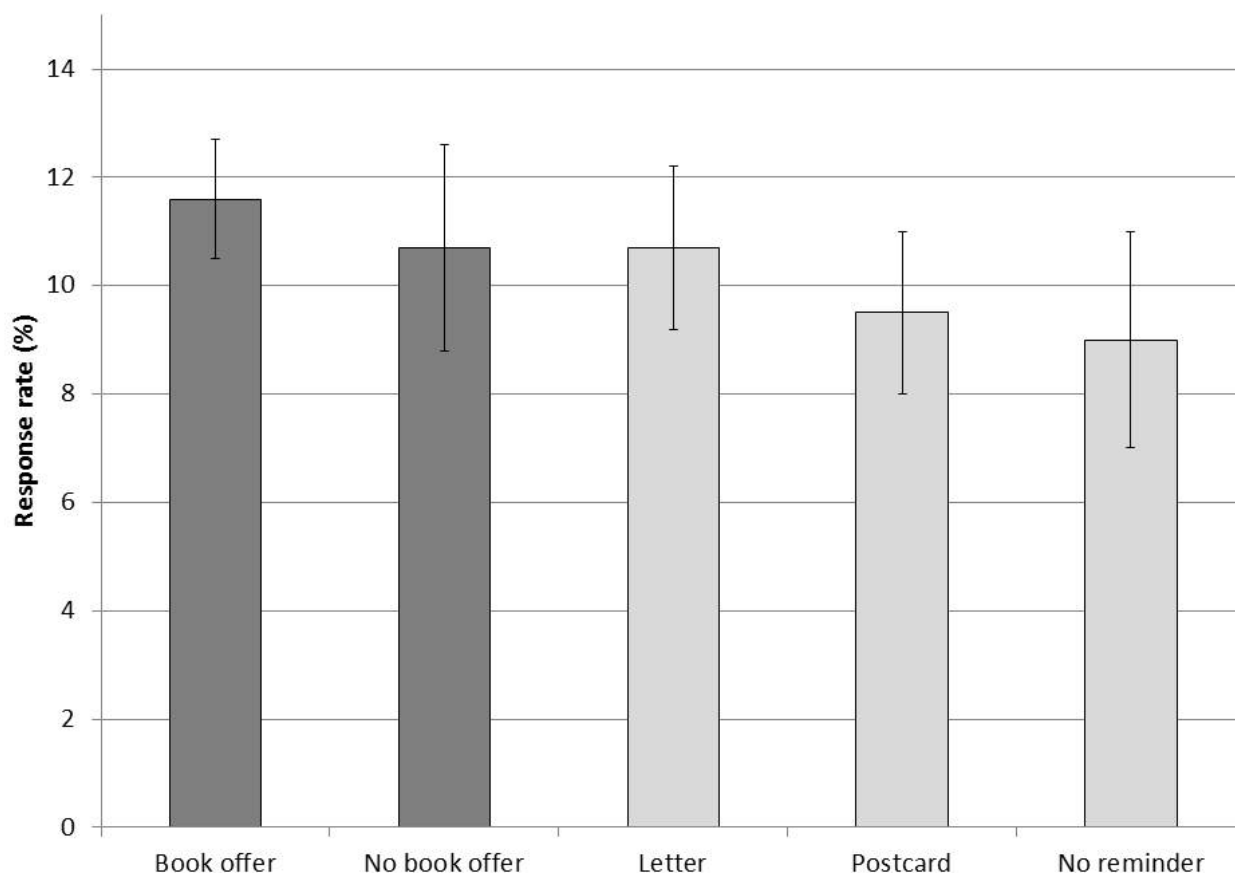
Impact of Paper Reminders

Sixty-six physicians responded before receiving the mailed paper reminders. Among the remaining 3900 physicians, the mailed paper reminders had no significant impact on response rates, with 168/1572 (10.7%) letter recipients, 148/1561 (9.5%) postcard recipients, and 69/767 (9.0%) email-only recipients responding to at least one survey question (*P*=.35). When dichotomized as any paper reminder (316/3133, 10.1%) versus none, the results were again not statistically significant (odds ratio 1.13, 95% CI 0.86-1.49, *P*=.36). Converting this odds ratio to absolute rates relative to the baseline rate for those not receiving a paper reminder, the 95% CI for response among those receiving any paper reminder ranged from 7.9% to 12.8%. Comparison of the response rate for those receiving letters or postcards was likewise not statistically significant (odds ratio 1.14, 95% CI 0.91-1.44, *P*=.26).

Results for both the book incentive and paper reminder were similar when only those who completed the entire survey were counted as respondents (data not shown).

Associations of Response Rate With Demographic Features

We explored associations between response rate and several demographic features. Response rate varied significantly by age

(*P*<.001), with both younger (response rate 8.8% for age < 45 years) and older (9.0% for age ≥65 years) physicians responding less frequently than others (12.8% for age group 45-54 years, 14.9% for age group 55-64 years). We found no significant difference in response rates across specialties as classified in Table 1 (*P*=.06); however, generalist (family medicine, internal medicine, and pediatric) physicians responded less often (9.4%) than nongeneralist physicians (12.3%; *P*=.007). Physicians with a doctor of medicine (MD) degree were twice as likely (442/3791, 11.7%) to respond as those with a doctor of osteopathy (DO) degree (9/174, 5.2%, *P*=.008). We found no significant difference by sex (*P*=.30) or geographic region (*P*=.90).

In multivariate analysis simultaneously accounting for all 5 demographic variables, namely, age, specialty (generalist vs nongeneralist), degree, sex, and region, statistically significant predictors of response were age (*P*<.001), specialty (*P*=.003), and sex (161/1329, 12.1% female, 290/2637, 11.0% male; *P*=.02).

## Interaction Between Interventions and Demographic Features

We explored interactions between the study interventions and demographic features as well. Independent adjustment for age, specialty (generalist or nongeneralist), sex, and geographic

region did not change the finding of no effect from the book incentive or the reminder. We found no significant interaction between any intervention and any of these demographic features ($P_{interaction} \geq .09$).

Influence of Email Reminders

Figure 3 shows a timeline of cumulative responses and illustrates the continued favorable impact of repeated reminder emails.

Figure 3. Timeline of responses for completed surveys.The survey started on day 1. Vertical dashed lines indicate the dates of reminder emails. The paper reminder was mailed on day 7, with the intent that it would arrive on the same day as the first reminder email. Intervals with no dots indicate periods in which no additional responses were received.
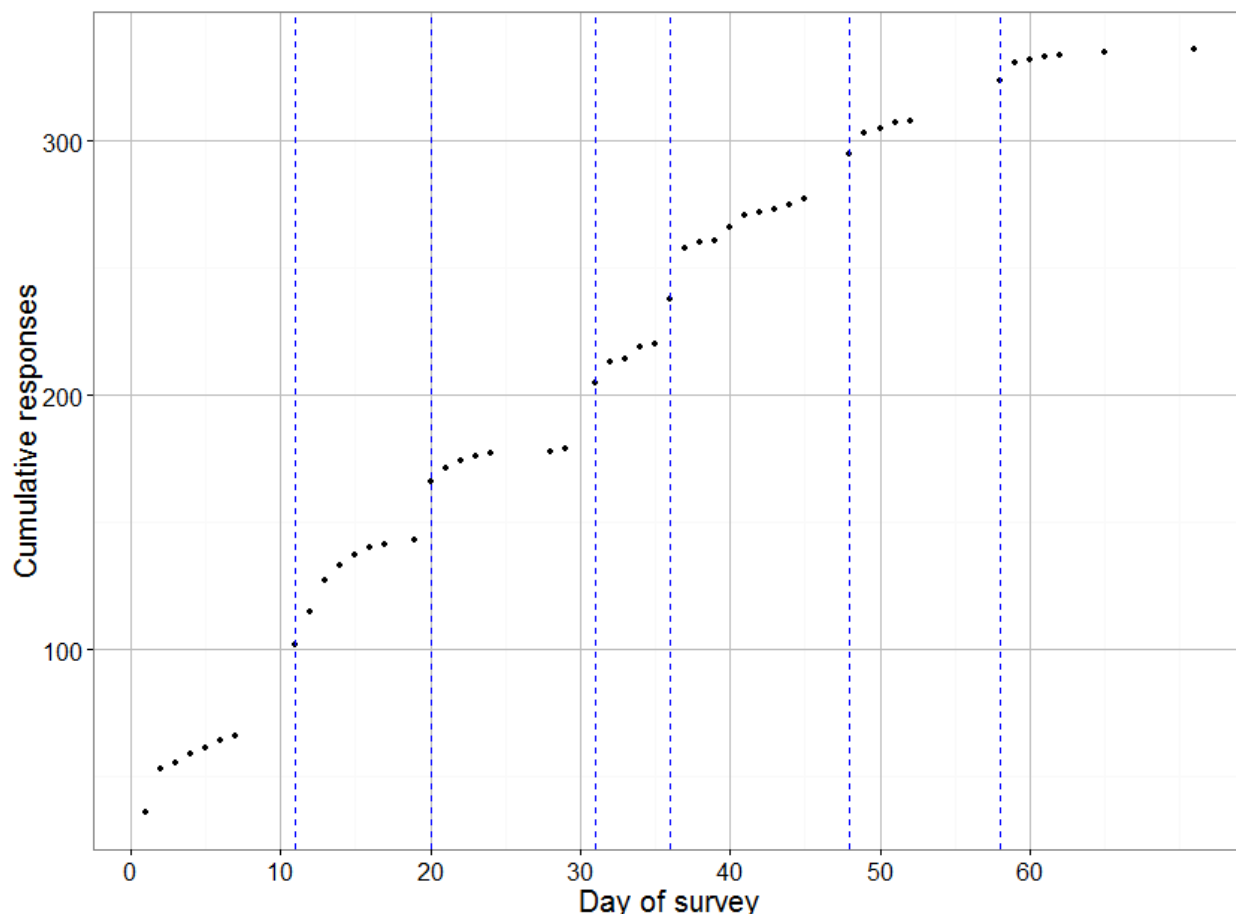
Discussion

Summary of Findings

Optimizing the response rate of Internet-based surveys is a matter of great importance to survey researchers. In this randomized experiment, we found that offering a modest incentive had no impact on response rate in an Internet-based survey. We also found that paper reminders had no impact on response rate, whether in the form of a formal letter or a postcard. Generalist (family medicine, internal medicine, and pediatric) physicians responded less often than nongeneralist physicians, and middle-aged physicians responded more often than younger or older physicians, but we found no significant interaction between the study interventions and specialty or age. We observed a meaningful number of additional responses after each of the 6 reminder emails.

Limitations

The greatest limitation of this study is the overall low response rate. A recent meta-analysis [4] estimated the average survey response rate among health professionals at 53%; however, a

review focused on Internet surveys of physicians found that response rates <20% are not uncommon [11]. Although the low response rate does not affect the internal validity of the study findings, it could affect the generalizability of these results to other survey contexts. We note that despite the low response rate, the confidence intervals exclude what would be considered by most researchers to be a meaningful difference in response rate.

One possible explanation for the overall low response rate is that email addresses were invalid or not monitored, but we were not notified. If true, this would not affect our conclusions unless there were a systematic bias in the distribution of invalid emails (which seems unlikely). Another possibility is that physicians perceived relatively low value in this incentive—perhaps they simply were not interested in this particular book. Research suggests that an unappealing incentive in some situations may be worse than no incentive at all. For example, one study found that offering free continuing medical education credit to Internet survey respondents actually resulted in lower response rates compared with no incentive [35]. Alternatively, the promise that the book would be delivered at a later date might have been

less motivating than an immediately present, tangible incentive. The logistical difficulty of providing an immediate meaningful incentive reflects a limitation of Internet-based surveys generally [36,37] and among physicians specifically [11]. A third possibility is that the invitees did not find the survey topic interesting. Like other professionals, physicians will not complete a survey if the topical importance is unclear or perceived as low [25]. Although maintenance of certification and continuing medical education appear prominently in current physician discussion venues, it is possible that the invited physicians perceived low personal salience in these topics.

Strengths of our study include the randomized design and adequate power to detect even small intervention effects, if present.

Integration With Prior Work

Prior investigations of methods for physician surveys, conducted primarily in the context of paper surveys, indicate that incentives increase participation. Our findings suggest that in Internet-based surveys the incentive may play a smaller role, perhaps because the promise of a future reward prompts less motivation than would a tangible gift. This suggestion is supported by prior research [38-40] indicating that noncontingent prepaid monetary incentives (eg, including US $5 with an initial paper survey invitation) yield higher response rates than contingent postcompletion offers, such as that used in our investigation.

Research also suggests that sending paper reminders after an initial email invitation or offering the opportunity to complete the survey via the Internet or on paper increases physician survey response rates [19,41]. We hoped that a single paper reminder might have a similar effect, but unfortunately, such was not the case.

The field of electronic surveys continues to evolve. Tools to conduct such surveys have become more ubiquitous, more powerful, and more user-friendly in recent years. While these improvements make it easier for health service researchers to conduct Internet-based survey research, it also makes it easier for everyone else to do the same. Thus, physicians (along with many other people) often feel overwhelmed by survey invitations ("surveyed to death") and other unsolicited emails ("spam") and may indiscriminately ignore, delete, or opt out of all survey requests. In addition, the increasing prevalence and awareness of electronic threats (eg, computer viruses, phishing, and identity theft) lead people to protect themselves from emails arising from unknown senders. A small percentage of the physicians invited to participate in our survey blocked all emails from unknown senders, a solution that may become more popular in coming years. The evolving landscape of Internet-based surveys will continue to challenge those using this mode of administration.

Conclusions and Implications for Research Methods

Our study has several implications for best practices in survey research. First, the book incentive did not improve response rates. Because this contradicts prior work on incentives in paper surveys, it remains to be seen whether this finding was simply related to the particular incentive or survey topic or whether it indicates a general lack of benefit from contingent nonmonetary incentives in the context of Internet-based physician surveys.

Second, the paper reminder also did not help, whether in the form of a postcard or a formal letter. Assuming this finding replicates in future work, it suggests that researchers using electronic surveys will need to find other ways to encourage response. Mixed-modality surveys using both paper and Internet will likely achieve the best results.

Third, repeated email reminders continued to work, at least through 6 reminders spaced at approximately 10-day intervals. The long-term impact of such reminders (eg, potential negative impact on physicians' response to other surveys) could not be evaluated in this study.

The methods of Internet-based survey research continue to evolve. What was once a powerful and easily accessible tool may no longer continue to be such, given the current state of electronic threats and excessive surveys from marketing, customer satisfaction, and political pollsters as well as health care service researchers. Extra effort may be required to follow other approaches known to improve survey response, such as highlighting the salience of the topic, having someone known to the invitee champion the survey, broadening the options for survey completion to include additional modes (eg, fax, telephone, interactive voice response telephone, or face-to-face), and decreasing the burden of survey completion by utilizing the briefest forms possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of paper reminder letters.

[PDF File (Adobe PDF File), 46KB - jmir_v18i9e244_app1.pdf]

References

1. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. J Clin Epidemiol 1997 Oct;50(10):1129-1136. [Medline: 9368521]

2. Asch S, Connor SE, Hamilton EG, Fox SA. Problems in recruiting community-based physicians for health services research. J Gen Intern Med 2000 Aug;15(8):591-599 [FREE Full text] [Medline: 10940152]

3. McLeod CC, Klabunde CN, Willis GB, Stark D. Health care provider surveys in the United States, 2000-2010: a review. Eval Health Prof 2013 Mar;36(1):106-126. [doi: 10.1177/0163278712474001] [Medline: 23378504]

4. Cho YI, Johnson TP, Vangeest JB. Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response. Eval Health Prof 2013 Sep;36(3):382-407. [doi: 10.1177/0163278713496425] [Medline: 23975761]

5. Davern M. Nonresponse rates are a problematic indicator of nonresponse bias in survey research. Health Serv Res 2013 Jun;48(3):905-912 [FREE Full text] [doi: 10.1111/1475-6773.12070] [Medline: 23656501]

6. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA 2012 May 2;307(17):1805-1806. [doi: 10.1001/jama.2012.3532] [Medline: 22550194]

7. VanGeest JB, Johnson TP, Welch VL. Methodologies for improving response rates in surveys of physicians: a systematic review. Eval Health Prof 2007 Dec;30(4):303-321. [doi: 10.1177/0163278707307899] [Medline: 17986667]

8. Kellerman SE, Herold J. Physician response to surveys. A review of the literature. Am J Prev Med 2001 Jan;20(1):61-67. [Medline: 11137777]

9. Thorpe C, Ryan B, McLean SL, Burt A, Stewart M, Brown JB, et al. How to obtain excellent response rates when surveying physicians. Fam Pract 2009 Feb;26(1):65-68 [FREE Full text] [doi: 10.1093/fampra/cmn097] [Medline: 19074758]

10. Klabunde CN, Willis GB, Casalino LP. Facilitators and barriers to survey participation by physicians: a call to action for researchers. Eval Health Prof 2013 Sep;36(3):279-295. [doi: 10.1177/0163278713496426] [Medline: 23975758]

11. Dykema J, Jones NR, Piché T, Stevenson J. Surveying clinicians by web: current issues in design and administration. Eval Health Prof 2013 Sep;36(3):352-381. [doi: 10.1177/0163278713496630] [Medline: 23975760]

12. Dykema J, Stevenson J, Day B, Sellers SL, Bonham VL. Effects of incentives and prenotification on response rates and costs in a national web survey of physicians. Eval Health Prof 2011 Dec;34(4):434-447 [FREE Full text] [doi: 10.1177/0163278711406113] [Medline: 21613242]

13. Minniear TD, McIntosh EB, Alexander N, Weidle PJ, Fulton J. Using electronic surveys to gather information on physician practices during a response to a local epidemic--Rhode Island, 2011. Ann Epidemiol 2013 Aug;23(8):521-523. [doi: 10.1016/j.annepidem.2013.05.003] [Medline: 23830934]

14. Schleyer TK, Forrest JL. Methods for the design and administration of web-based surveys. J Am Med Inform Assoc 2000;7(4):416-425 [FREE Full text] [Medline: 10887169]

15. Braithwaite D, Emery J, De LS, Sutton S. Using the Internet to conduct surveys of health professionals: a valid alternative? Fam Pract 2003 Oct;20(5):545-551 [FREE Full text] [Medline: 14507796]

16. Akl EA, Maroun N, Klocke RA, Montori V, Schünemann HJ. Electronic mail was not better than postal mail for surveying residents and faculty. J Clin Epidemiol 2005 Apr;58(4):425-429. [doi: 10.1016/j.jclinepi.2004.10.006] [Medline: 15862729]

17. Crouch S, Robinson P, Pitts M. A comparison of general practitioner response rates to electronic and postal surveys in the setting of the National STI Prevention Program. Aust N Z J Public Health 2011 Apr;35(2):187-189. [doi: 10.1111/j.1753-6405.2011.00687.x] [Medline: 21463418]

18. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a controlled comparison (2). J Med Internet Res 2004 Oct 29;6(4):e39 [FREE Full text] [doi: 10.2196/jmir.6.4.e39] [Medline: 15631963]

19. McMahon SR, Iwamoto M, Massoudi MS, Yusuf HR, Stevenson JM, David F, et al. Comparison of e-mail, fax, and postal surveys of pediatricians. Pediatrics 2003 Apr;111(4 Pt 1):e299-e303. [Medline: 12671142]

20. Nicholls K, Chapman K, Shaw T, Perkins A, Sullivan MM, Crutchfield S, et al. Enhancing response rates in physician surveys: the limited utility of electronic options. Health Serv Res 2011 Oct;46(5):1675-1682 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01261.x] [Medline: 21492157]

21. Raziano DB, Jayadevappa R, Valenzula D, Weiner M, Lavizzo-Mourey R. E-mail versus conventional postal mail survey of geriatric chiefs. Gerontologist 2001 Dec;41(6):799-804. [Medline: 11723348]

22. Seguin R, Godwin M, MacDonald S, McCall M. E-mail or snail mail? Randomized controlled trial on which works better for surveys. Can Fam Physician 2004 Mar;50:414-419 [FREE Full text] [Medline: 15318679]

23. Thomas M, Rogers R, Maclean R. Collecting data from physicians via web-based surveys: recommendations for improving response rates. The Internet Journal of Medical Informatics 2002;1(1):1-6.

24. VanDenKerkhof EG, Parlow JL, Goldstein DH, Milne B. In Canada, anesthesiologists are less likely to respond to an electronic, compared to a paper questionnaire. Can J Anaesth 2004 May;51(5):449-454. [Medline: 15128630]

25. VanGeest J, Beebe T, Johnson T. Surveys of Physicians. In: Johnson T, editor. Handbook of Survey Methods. Hoboken, NJ: John Wiley & Sons, Inc; 2014:515-543.

26. Burns KE, Duffett M, Kho ME, Meade MO, Adhikari NK, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ 2008 Jul 29;179(3):245-252 [FREE Full text] [doi: 10.1503/cmaj.080372] [Medline: 18663204]

27.   Field TS, Cadoret CA, Brown ML, Ford M, Greene SM, Hill D, et al. Surveying physicians: do components of the "Total Design Approach" to optimizing survey response rates apply to physicians? Med Care 2002 Jul;40(7):596-605. [Medline: 12142775]

28.   Klabunde CN, Willis GB, McLeod CC, Dillman DA, Johnson TP, Greene SM, et al. Improving the quality of surveys of physicians and medical groups: a research agenda. Eval Health Prof 2012 Dec;35(4):477-506. [doi: 10.1177/0163278712458283] [Medline: 22947596]

29.   Willis GB, Smith T, Lee HJ. Do additional recontacts to increase response rate improve physician survey data quality? Med Care 2013 Oct;51(10):945-948 [FREE Full text] [doi: 10.1097/MLR.0b013e3182a5023d] [Medline: 23969583]

30.   Iglehart JK, Baron RB. Ensuring physicians' competence--is maintenance of certification the answer? N Engl J Med 2012 Dec 27;367(26):2543-2549. [doi: 10.1056/NEJMhpr1211043] [Medline: 23268670]

31.   Cook DA, Holmboe ES, Sorensen KJ, Berger RA, Wilkinson JM. Getting maintenance of certification to work: a grounded theory study of physicians' perceptions. JAMA Intern Med 2015 Jan;175(1):35-42. [doi: 10.1001/jamainternmed.2014.5437] [Medline: 25365596]

32.   Nissen SE. Reforming the continuing medical education system. JAMA 2015 May 12;313(18):1813-1814. [doi: 10.1001/jama.2015.4138] [Medline: 25965221]

33.   Committee on Planning a Continuing Health Care Professional Education Institute, Board on Health Care Services, Institute of Medicine. Redesigning Continuing Education in the Health Professions. Washington, DC: The National Academies Press; 2010.

34.   American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys URL: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2016-08-25] [WebCite Cache ID 6k1nHsZV6]

35.   Viera AJ, Edwards T. Does an offer for a free on-line continuing medical education (CME) activity increase physician survey response rate? A randomized trial. BMC Res Notes 2012;5:129 [FREE Full text] [doi: 10.1186/1756-0500-5-129] [Medline: 22397624]

36.   Göritz A. Using lotteries, loyalty points,other incentives to increase participant responsecompletion. In: Gosling SD, Johnson JA, editors. Advanced methods for conducting online behavioral research. Washington, DC: American Psychological Association; 2010:219-233.

37.   Göritz AS. Incentives in Web Studies: Methodological Issues and a Review. International Journal of Internet Science 2006;1:58-70.

38.   James KM, Ziegenfuss JY, Tilburt JC, Harris AM, Beebe TJ. Getting physicians to respond: the impact of incentive type and timing on physician survey response rates. Health Serv Res 2011 Feb;46(1 Pt 1):232-242 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01181.x] [Medline: 20880042]

39.   Kanaan RA, Wessely SC, Armstrong D. Differential effects of pre and post-payment on neurologists' response rates to a postal survey. BMC Neurol 2010;10:100 [FREE Full text] [doi: 10.1186/1471-2377-10-100] [Medline: 20973984]

40.   Leung GM, Johnston JM, Saing H, Tin KY, Wong IO, Ho L. Prepayment was superior to postpayment cash incentives in a randomized postal survey among physicians. J Clin Epidemiol 2004 Aug;57(8):777-784. [doi: 10.1016/j.jclinepi.2003.12.021] [Medline: 15485729]

41.   Beebe TJ, Locke GR, Barnes SA, Davern ME, Anderson KJ. Mixing web and mail methods in a survey of physicians. Health Serv Res 2007 Jun;42(3 Pt 1):1219-1234 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00652.x] [Medline: 17489911]

[XSL•FO]

RenderX

Original Paper

Linked Patient-Reported Outcomes Data From Patients With Multiple Sclerosis Recruited on an Open Internet Platform to Health Care Claims Databases Identifies a Representative Population for Real-Life Data Analysis in Multiple Sclerosis

Valery Risson[1], MBA, PhD; Bhaskar Ghodge[2], MCA; Ian C Bonzani[3], PhD; Jonathan R Korn[4], BA; Jennie Medin[1], MPH, PhD; Tanmay Saraykar[5], MBA; Souvik Sengupta[2], BE; Deepanshu Saini[6], MBA; Melvin Olson[1], PhD

[1]Novartis Pharma AG, Basel, Switzerland

[2]IMS Health, Basel, Switzerland

[3]IMS Health, London, United Kingdom

[4]IMS Health, Burlington, MA, United States

[5]IMS Health, Parsippany, NJ, United States

[6]IMS Health, Gurgaon, Delhi, India

Corresponding Author:
Valery Risson, MBA, PhD
Novartis Pharma AG
Postfach
Basel, 4002
Switzerland
Phone: 41 61 324 6036
Fax: 41 61 324 8001
Email: valery.risson@novartis.com

Abstract

Background: An enormous amount of information relevant to public health is being generated directly by online communities.

Objective: To explore the feasibility of creating a dataset that links patient-reported outcomes data, from a Web-based survey of US patients with multiple sclerosis (MS) recruited on open Internet platforms, to health care utilization information from health care claims databases. The dataset was generated by linkage analysis to a broader MS population in the United States using both pharmacy and medical claims data sources.

Methods: US Facebook users with an interest in MS were alerted to a patient-reported survey by targeted advertisements. Eligibility criteria were diagnosis of MS by a specialist (primary progressive, relapsing-remitting, or secondary progressive), ≥12-month history of disease, age 18-65 years, and commercial health insurance. Participants completed a questionnaire including data on demographic and disease characteristics, current and earlier therapies, relapses, disability, health-related quality of life, and employment status and productivity. A unique anonymous profile was generated for each survey respondent. Each anonymous profile was linked to a number of medical and pharmacy claims datasets in the United States. Linkage rates were assessed and survey respondents' representativeness was evaluated based on differences in the distribution of characteristics between the linked survey population and the general MS population in the claims databases.

Results: The advertisement was placed on 1,063,973 Facebook users' pages generating 68,674 clicks, 3719 survey attempts, and 651 successfully completed surveys, of which 440 could be linked to any of the claims databases for 2014 or 2015 (67.6% linkage rate). Overall, no significant differences were found between patients who were linked and not linked for educational status, ethnicity, current or prior disease-modifying therapy (DMT) treatment, or presence of a relapse in the last 12 months. The frequencies of the most common MS symptoms did not differ significantly between linked patients and the general MS population in the databases. Linked patients were slightly younger and less likely to be men than those who were not linkable.

Conclusions: Linking patient-reported outcomes data, from a Web-based survey of US patients with MS recruited on open Internet platforms, to health care utilization information from claims databases may enable rapid generation of a large population of representative patients with MS suitable for outcomes analysis.

Introduction

The Internet and social media are driving a revolution in communication and information sharing, with a fundamental impact on health care. Patients' voices have become more influential through the exchange of information in the form of conversations, blogs, tweets, and other postings on social media. This development is changing the power balance in decisions regarding health care, requiring traditional stakeholders to recognize patients' perspectives in the provision and evaluation of treatments [1-3].

An enormous amount of information relevant to public health is being generated directly by online communities [4,5]. Epidemiology [6,7], pharmacovigilance [8,9], identification of malpractice [10], and the support of health behavior changes [11] are only a few examples of areas where informal data have been successfully applied. Moreover, the Internet and particularly social networks represent a large number of individuals with shared interests, nationalities, or characteristics that can be reached by relatively modest financial or human resources.

We have previously reported on the feasibility of applying social media listening (defined as the mining and analysis of information gathered from social media) to retrospective analyses in outcomes research, specifically the use of patient-reported reasons for switching between different treatment modalities for multiple sclerosis (MS) [12]. The ability to include patient-reported information to enhance prospective analyses of sources such as claims databases would appear to have great promise in outcomes research. We present here an approach to create a dataset that contains both patient outcomes data (from a Web-based survey of US patients with MS recruited on an open Internet platform) and health care utilization information from claims databases. A linkage analysis has recently been performed on data from dedicated patient platforms and invited patients [13]. We hypothesized that linking patient data from the social media survey with those from the claims databases could identify a representative population that can be used for real-life data analysis in MS. The initial analysis focused on verifying the method by demonstrating that the characteristics of the linked population recruited on the open Internet platform are representative of the MS population in the United States.

Methods

Study Aim and Design

The primary aim of this pilot study was to explore the feasibility of creating a dataset that links patient-reported outcomes data, from a Web-based survey of US patients with MS recruited on open Internet platforms, to health care utilization information from health care claims databases. The representativeness of the linked populations was validated by a comparison with the characteristics of known MS populations in the United States.

This study was designed, implemented, and reported in accordance with the Guidelines for Good Pharmacoepidemiology Practices of the International Society for Pharmacoepidemiology [14], the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [15], and the ethical principles laid down in the Declaration of Helsinki [16]. The secondary data source used for the analysis meets all of the US Health Insurance Portability and Accountability Act (HIPAA) compliance standards, ensuring patient anonymity. As such, approval from an institutional review board was not necessary.

The defined target population was a broad, US-based, commercially insured population with MS diagnosed and treated by a specialist. All participants took part in the survey entirely of their own volition, and complete information regarding how the data would be used was provided before patients agreed to take part. Full anonymity was guaranteed at all points of the process of running the survey and performing the linkage and subsequent analysis.

Recruitment and Survey

The survey process is shown schematically in Figure 1. US Facebook users with an interest in MS were alerted to a patient-reported survey by targeted advertisements. The identification of users with a high interest in MS for the placement of advertisements was performed by Facebook as a commercial service and was beyond the control of the researchers. Users clicking on the survey advertisement were provided a disclaimer on the study and how data would be used, followed by options to decline or consent to proceed to the survey. Users were anonymous until the time at which they consented to taking the survey and passed the screening criteria. No identifiable information was collected from users who declined to take the survey. Users who agreed to participate were redirected away from the Facebook domain to the survey, which was hosted on a secure third-party site accessed using an https (hypertext transfer protocol, secure) protocol. Neither the advertisements nor the survey was branded by any commercial entity.

Before completing the survey, patients were screened for eligibility. Screening questions are presented in Textbox 1. The predefined criteria were a diagnosis of MS by a specialist; ≥12 months of history of diagnosed disease before taking the survey; 18-65 years of age; and commercial health insurance, both current and in the ≥12 months before study entry. The disease could be primary progressive, relapsing-remitting, or secondary progressive MS.

Textbox 1. Screening questions.

Screener questions

In which country do you currently live?

(*In the US* / Outside of the US)

Are you currently between the ages of 18 and 65 years of age?

(*Yes* / No)

Have you been diagnosed with Multiple Sclerosis by a specialist?

(*Yes* / No / Not Sure)

How long have you had diagnosed Multiple Sclerosis?

(< 1 year, *> 1 year*, *> 5 years*, *> 10 years*)

What type of Multiple Sclerosis do you have?

(*Relapsing-remitting Multiple Sclerosis (RRMS)*, *Primary Progressive Multiple Sclerosis (PPMS)*, *Secondary Progressive Multiple Sclerosis (SPMS)*, I don't know)

Do you have health insurance through a commercial health plan?

(*Yes*, No—I currently do not have health insurance, No—I am on a Medicare or Medicaid health plan)

How long have you been with your current health insurance provider?

(< 1 year, *> 1 year*, *> 5 years*, *> 10 years*)

If the defined criteria (italics) are not met, the patient will be excluded from the survey.

The survey was designed to provide information on demographics and disease characteristics, current and earlier therapy use, relapses, disease severity and disability, health-related quality of life (on the EQ-5D-3L and EQ-VAS scales), and employment status and productivity. The survey is included in Multimedia Appendix 1. Survey participants were informed about data handling, anonymity, and the right to revoke consent in a disclaimer, provided in Multimedia Appendix 2.

All data were hosted in a secure data enclave using network firewalls. Access was provided only to named users who had

successfully completed training on the handling of health information and signed data nondisclosure agreements. Before the linkage testing, all data collected from the Web-based survey were deidentified by a trusted third party (Management Science Associates Inc, Pittsburgh, PA), using patient deidentification software, encrypting patients' protected health information data elements in accordance with the Expert Determination De-Identification methodology of the HIPAA Privacy Rule law. The linkage to claims data and subsequent analyses were performed on the deidentified survey population.

Figure 1. Survey and linkage analysis process. MS: multiple sclerosis.
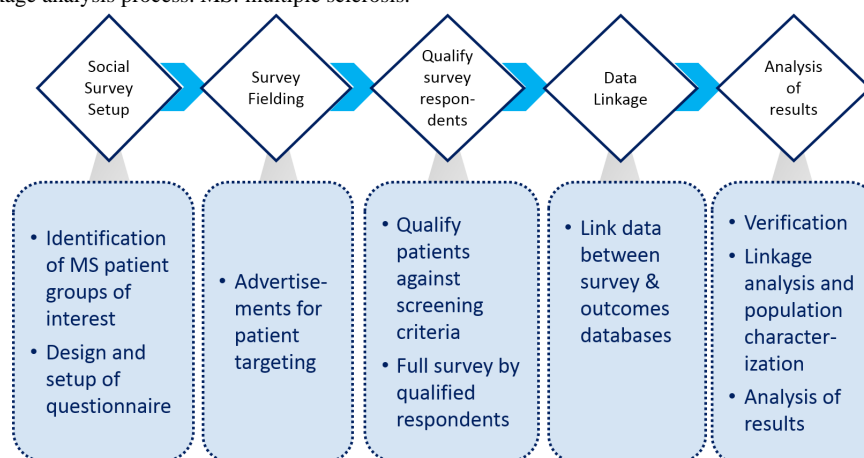
Linkage Analysis

We included 3 claims databases in the linkage analysis: the preadjudicated provider- and pharmacy-level (open-source) medical claims (Dx) and prescription claims (Rx) databases and the commercial health plan PharMetrics Plus database of adjudicated medical and pharmacy claims. The databases are characterized in Multimedia Appendix 3. The Dx and Rx

databases were merged for the linkage analysis (forming the Dx/Rx database). The quality of the records in the adjudicated, plan-level PharMetrics Plus database is overall higher than the open-source Rx pharmacy and Dx medical claims databases. The latter has the advantage of covering a larger number of patients. An overview of the different types of information captured in the survey and in the individual databases is shown in Figure 2.

To link a completed survey to an entry in any of the claims databases, survey respondents' data (eg, name, address, zip code, date of birth, sex) were deidentified by means of a multilevel encryption process that was combined with administrative, physical, and technical safeguards to generate unique, encrypted, deidentified tokens that could not be reidentified. The tokens were used in a deterministic matching process to similarly anonymized patients with claims in the IMS Health database. A detailed review of the anonymization and linking methodologies is beyond the scope of this paper, but the methodologies have been used extensively, including in the study cited above [13]. For details, the reader is referred to earlier publications available on the Web [17]. Successful linkage was defined as a match in the claims database to the same anonymous profile in a deterministic process.

Figure 2. Overview of the data available in the different sources included in the linkage analysis. The cohorts identified in the medical claims (Dx) and prescription claims (Rx) databases were merged for the linkage analysis. *Via International Classification of Diseases, Ninth Revision, codes. **All claims may not have been captured owing to the possibility of patients using providers or pharmacies not in the database. PMTX+: PharMetrics Plus; QoL: quality of life.

Data of interest	Social Survey	PMTX+ Claims (closed-source)	Medical Claims (Dx) (open source)	Pharmacy Claims (Rx) (open source)	Analysis:
Demographics	✓	✓	✓	✓	① Validation
Geography	✓	✓	✓	✓	
Diagnosis	✓	✓	✓**	✓**	
Medication use	✓	✓	✓**	✓	
Symptoms *		✓	✓**		② Generalizability
Comorbidities *		✓	✓**		
Outpatient visits & procedures		✓	✓		
In-patient visits & procedures		✓	✓**		③ Analysis of Results
Costs		✓			
Disease characteristics	✓				
Disease Severity	✓				
QoL	✓				
Work Productivity	✓				

Validation

A total of 4 cohorts were generated for the linkage analysis: cohort 1, all survey participants; cohort 2, survey population successfully linked to the PharMetrics Plus or Dx/Rx claims; cohort 3, survey patients not linkable to claims sources; and cohort 4, patients with MS in PharMetrics Plus. Cohort 4 was made up of all patients with MS in the PharMetrics Plus database aged 18-65 years with ≥1 MS diagnosis and ≥1 month of health plan enrollment between January 1, 2013, and March 31, 2015. For further characterization of linked populations, a subset of linked patients was selected with claims from 2014 or 2015. This was in order to take into account availability of newer therapies and to reduce the discrepancy between the dates of the survey and those of historical claims.

An index date of March 31, 2015 was assigned for all survey patients. For the overall MS population in PharMetrics Plus, the index date was the last month of enrollment between January 1, 2013 and March 31, 2015. For all patients, demographic and clinical characteristics were analyzed and compared for 12 months before the index date.

To validate the representativeness of the cohort identified in the linkage analysis, the degree of concordance between the characteristics of the survey population and those identified in the claims databases was analyzed. Concordance was estimated by calculating the positive predictive values (PPVs; the probability that a claimed characteristic in the survey corresponded to the presence of same characteristic in the linked data). Positive predictive values were calculated as follows: $a/(a+b) \times 100$, where a=the number of survey respondents with a specific claim also found in the linked database and b=the total number of survey respondents with the claim. Positive predictive values were calculated for the variables MS diagnosis, current use of disease-modifying therapies (DMTs), prior DMT use, and relapses.

Furthermore, the means and distribution profiles of disease characteristics of all included cohorts were analyzed and compared between the cohorts as follows: cohorts 1 and 4 were compared on demographic characteristics. Cohorts 2 and 4 were compared on clinical characteristics before the index date: use of DMTs, dalfampridine, and corticosteroids (all databases), and comorbidity profiles and Charlson score, use of magnetic resonance imaging of the brain and spine, and relapse rates (PharMetrics Plus only). Cohorts 2 and 3 were compared on all survey results.

Sample Size

Data on around 302,000 patients with MS were available from the open-source Dx/Rx MS database. The PharMetrics Plus MS database includes data on >100,000 patients in a given year. On the basis of preliminary data and pilot study experience, a 10%-15% linkage rate was expected between the survey and PharMetrics Plus MS cohorts and a linkage rate of >50% to the Dx/Rx cohort. On the basis of these assumptions, a survey sample size of 1000 participants was targeted.

Statistical Methods

Demographic data were analyzed descriptively. Categorical variables are presented as frequency and percentage (%) of total patients observed in each category. Continuous variables are presented as mean (SD) as well as the median. Statistical significance testing used the chi-square test for categorical variables and the Wilcoxon rank sum test for continuous variables. A *P* value of <.05 was considered statistically significant.

Results

Facebook Survey Participants

The flow of respondents to the Web-based advertisements and the survey is shown in Figure 3. The Web-based survey was run between July 21, 2015 and September 15, 2015. During this time, the advertisement was placed on 1,063,973 Facebook users' pages. The advertisements generated a total of 68,674 clicks leading to 3719 attempts at the survey. After filtering out respondents who did not meet the criteria for the survey, 685 respondents completed the survey successfully. The characteristics of 34 respondents were indicative of duplications; thus, 651 unique surveys were included in the linkage analysis (651/1040; 62.60% response rate among eligible respondents who passed screening).

Figure 3. Flow of respondents.
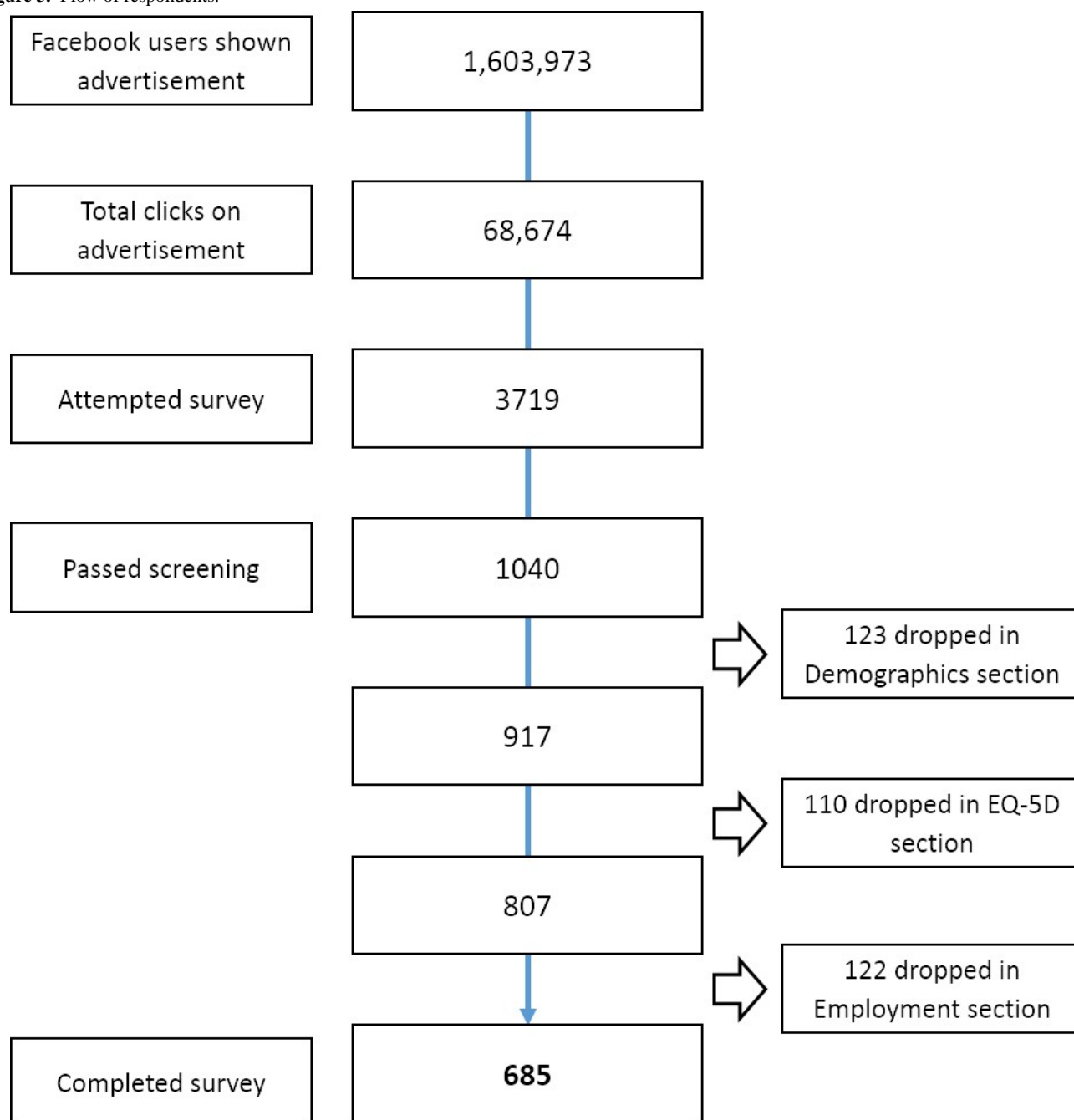
Data Linkage and Validation

Of the 651 unique patients completing the survey, 453 (69.6%) could be linked with the Dx/Rx database and 73 (11.2%) were linkable to the PharMetrics Plus MS database. A total of 198 survey participants could not be linked; a major reason for this was incorrectly entered data, mostly dates of birth that were missing or incorrect for 67 respondents.

The subset of linked patients with claims from 2014 or 2015 used in further characterization consisted of 440/651 patients (67.6%), 387 of whom (88.0%) were linked to the Dx/Rx database only and 53 (12.0%) to the PharMetrics Plus MS database.

There was a high degree of concordance between the linked patients and the PharMetrics Plus database (Table 1), whereas concordance with Dx/Rx plus PharMetrics Plus was moderate. The PPV for MS diagnosis in the linked patients was 98.1% with PharMetrics Plus (88.0% for all 3 databases), that for current DMT use was 86.5% (51.7%), and that for prior DMT use was 70.0% (47.7%). The PPV for relapses, 34.6%, was lower than that for the other variables.

Table 1. Concordance between data from the Web-based survey and the PharMetrics Plus and Dx/Rx + PharMetrics Plus databases.

Variable	PPV[a] with PharMetrics Plus, %	PPV with Dx/Rx[b] or PharMetrics Plus, %
Multiple sclerosis diagnosis	98.1	88.0
Any current DMT[c]	86.5	51.7
No current DMT	68.8	86.4
Any prior DMT	70.0	47.7
Relapse in past 12 months	34.6	34.6

[a]PPV: positive predictive value.

[b]Dx/Rx: merged medical claims (Dx) and prescription claims (Rx) databases.

[c]DMT: disease-modifying therapy.

Generalizability of Survey Data

A comparison of those linkable to the PharMetrics Plus or the Rx/Dx databases in 2014 or 2015 and those not linkable can be found in Table 2. Overall, patients linkable to the open-source databases had slightly greater mean and median age and were more likely to be men than those linkable to PharMetrics Plus. Patients not linkable to the PharMetrics Plus database were more evenly distributed geographically across the United States, a consequence of underrepresentation of this database in the western states (IMS Health internal data). No significant differences were found for educational status, ethnicity, current or prior DMT treatment, or presence of a relapse in the last 12 months between linkable and not linkable individuals.

Because of the complete coverage of health care claims (eg, low likelihood of missing claims compared with open-source databases) captured in the PharMetrics Plus, the additional analysis presented below focuses on the survey patients linked to the PharMetrics Plus MS database.

Among the most common MS symptoms, the frequencies of a majority of symptoms did not differ significantly between the linked patients and the general PharMetrics Plus MS population (Figure 4). The rate of gait, balance, and coordination problems was higher in the linked population (16/53 or 30% vs 14,500/82,845 or 17.50%; P=.015) as was rate of bladder dysfunction (12/53 or 23% vs 9421/82,845 or 11.37%; P=.0098). The rate of numbness was lower in the linked population, but this result was not statistically significant (5/53 or 9% vs 16,452/82,845 or 19.86%; P=.057). Among comorbidities, only rates of depression differed significantly between the groups, with 21/53 (40%) linked patients reporting depression compared with 22.21% (18,402/82,845) of the overall PharMetrics Plus MS population (P=.0023). The proportion of patients with relapses based on a claims-based algorithm in the linked cohort was comparable to that in the overall PharMetrics Plus MS cohort: 11/53 (21%) in the linked cohort versus 15,723/82,845 (18.98%) in the claims database (P=.7417).

Medication use was analyzed for the different populations, displayed graphically in Figure 5. The use of DMTs and corticosteroids was highly similar in the linked cohorts and the overall PharMetrics Plus MS population, whereas more patients in the linked cohort than in the overall population reported dalfampridine use. However, dalfampridine was low in all study populations.

XSL•FO
RenderX

Table 2. Demographic and clinical characteristics of the population included in the linkage analysis and the general multiple sclerosis population in the PharMetrics Plus and the Dx/Rx databases, respectively.

Characteristic	Not linkable to PharMet-rics Plus or Dx/Rx[a] (N=211)	Linkable to PharMet-rics Plus (N=53)	P value	Linkable to Dx/Rx (N=387)	P value
Age in years, mean (SD)	46.0 (14.7)	48.9 (8.6)		51.2 (8.8)	
Median age, years	50	49	.79	52.0	.004
Female sex, n (%)	178 (84.4)	46 (87)	.66	318 (82.2)	.50
Region, n (%)					
Northeast	37 (17.5)	18 (34)		75 (19.4)	
Midwest	51 (24.2)	16 (30)		114 (29.5)	
South	64 (30.3)	17 (32)		128 (33.1)	
West	59 (28.0)	2 (4)	<.001	70 (18.1)	.0435
Ethnicity: white, n (%)	186 (88.2)	50 (94)		361 (93.3)	
Educational status, n (%)					
Less than high school	5 (2.4)	1 (2)	.92	7 (1.8)	.25
Completed high school	43 (20.4)	11 (21)		71 (18.1)	
Some college	78 (37.0)	17 (32)		141 (36.4)	
Completed college	65 (30.8)	17 (32)		106 (27.4)	
Graduate school	20 (9.5)	7 (13)		62 (16.0)	
MS[b] subtype, n (%)					
RRMS[c]	173 (82.0)	46 (87)	.22	317 (81.9)	>.99
PPMS[d]	14 (6.6)	5 (9)		25 (6.5)	
SPMS[e]	24 (11.4)	2 (4)		45 (11.6)	
Time since diagnosis, n (%)					
>1 year	47 (22.3)	9 (17)	.04	119 (30.7)	.06
>5 years	79 (37.4)	11 (21)		118 (30.5)	
>10 years	85 (40.3)	33 (62)		150 (38.3)	
DMT[f] treatment, n (%)					
No DMT	45 (21.3)	16 (30)		72 (18.6)	
Copaxone and interferons	84 (39.8)	12 (23)		140 (36.2)	
Oral DMT[g]	57 (27.0)	17 (32)		130 (33.6)	
Infused DMT[h]	21 (10.0)	8 (15)		41 (10.6)	
Duration of current treatment, n (%)					
1-5 years	145 (68.7)	41 (77)		295 (76.2)	
6-10 years	40 (19.0)	7 (13)		50 (12.9)	
11-15 years	12 (5.7)	2 (4)		24 (6.2)	
16-20 years	10 (4.7)	1 (2)		14 (3.6)	
>20 years	4 (1.9)	2 (4)		4 (1.0)	
Prior DMT treatment, n (%)					
No DMT	67 (31.8)	16 (30)	.83	99 (25.6)	.11
Copaxone and interferons	148 (70.1)	43 (60)		296 (76.5)	
Oral DMT	21 (10.0)	11 (21)		32 (8.3)	

XSL•FO
RenderX

Characteristic	Not linkable to PharMetrics Plus or Dx/Rx[a] (N=211)	Linkable to PharMetrics Plus (N=53)	P value	Linkable to Dx/Rx (N=387)	P value
Infused DMT	22 (10.4)	5 (9)		34 (8.8)	
≥1 Relapse, n (%)	117 (55.5)	26 (49)		230 (59.4)	

[a]Dx/Rx: merged medical claims (Dx) and prescription claims (Rx) database.

[b]MS: multiple sclerosis.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]PPMS: primary progressive multiple sclerosis.

[e]SPMS: secondary progressive multiple sclerosis.

[f]DMT: disease-modifying therapy.

[g]Oral DMTs include Gilenya, Tecfidera, and Aubagio.

[h]Infused DMTs include Tysabri and Lemtrada.

Figure 4. Frequencies of the most common multiple sclerosis (MS) symptoms based on International Classification of Diseases codes on paid claims in the cohort linked to the PharMetrics Plus MS database (blue bars) and the overall PharMetrics Plus MS population (red bars). Only symptoms with prevalence >10% are shown. Asterisk indicates $P<.05$.
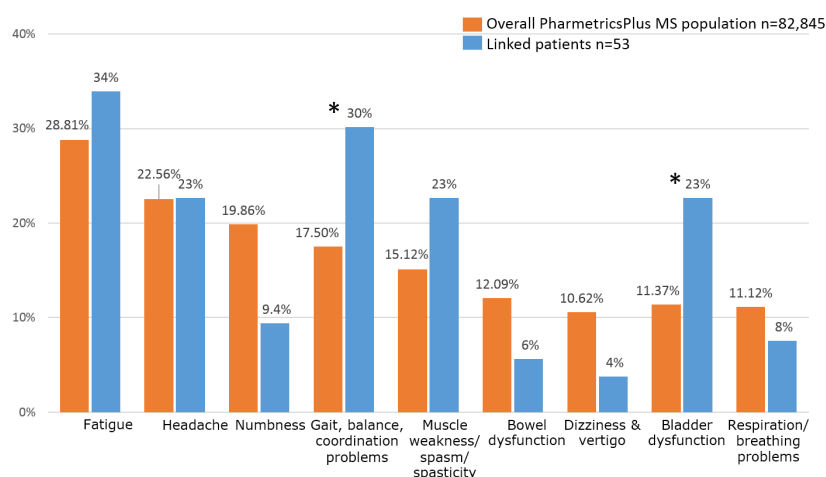
Figure 5. Use of multiple sclerosis (MS)–specific medication in the 1-year index period in the overall PharMetrics Plus MS cohort (red bars) and in the cohorts linked to the PharMetrics Plus MS (light blue bars) and Rx/Dx (dark blue bars) databases. DMT: disease-modifying therapy; Rx/Dx: merged prescription claims (Rx) and medical claims (Dx) database.
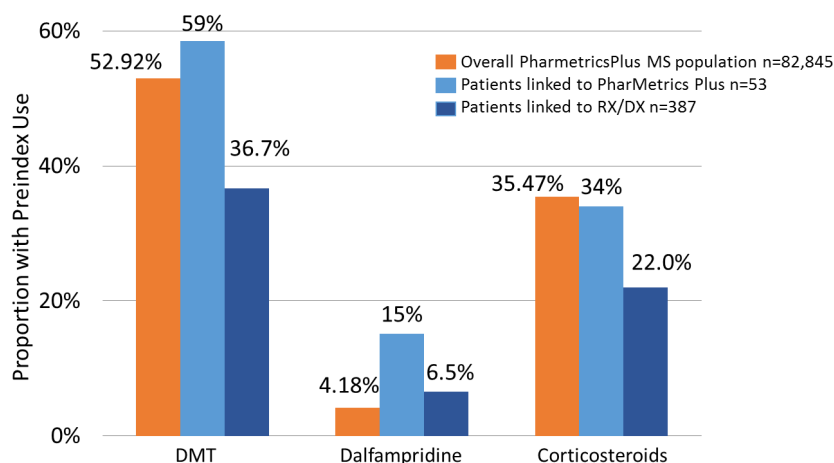
Discussion

Principal Findings

The aim of this study was to explore the feasibility of creating a dataset that contains both patient outcomes data (from a Web-based survey of US patients with MS recruited on an open Internet platform) and health care utilization information from pharmacy and medical claims databases. The initial results presented here indicate that this aim was fulfilled and that the survey population is broadly representative of the general MS patient population in the United States.

People with MS are highly active on social media [18,19]. We have previously shown that the demographics of cohorts of patients with MS identified by their activities on social media correspond well with cohorts identified in other sources, indicating that social media data analysis can be usefully applied to outcomes research [12]. Facebook, forums, and blogs have been used previously to recruit participants into surveys of health outcomes and lifestyle interventions in MS [20-22]. Our work differs in a number of important aspects from previous reports, however. First, patients were recruited through unbranded advertisements placed on Facebook pages, not by active, personal invitations sent to specific target groups. Although this approach led to a markedly lower rate of successfully completed surveys than that reported with patients invited directly, the absence of active targeting of participants can be expected to reduce the scope for bias and generate a more representative population.

The same differences apply to a very recent report that described successful linking of data from invited patients on a dedicated online patient community (PatientsLikeMe) with administrative claims data [13]. In contrast to our approach, targeted invitations were sent to eligible patients identified on the social network by email and private messages. Both approaches have their merits, but the population from the online patient community may well have been less diverse than patients who can be reached by untargeted advertisements on open Internet platforms such as Facebook.

The growing interest in linking health care–related social media content to information obtained by traditional means, for example, claims databases, reflects the realization that such linked data may provide a rapid, cost-effective, and credible method to capture patient outcomes, behavioral data, and health care claims. Our use of a standardized survey allowed us to overcome a common limitation of data from open Internet platforms: a lack of structured clinical, socioeconomic, and demographic data necessary for observational research [23]. The survey allowed us to obtain structured, disease-specific information on disease duration, medications, disabilities, and impact of the disease on quality of life as well as work productivity. The linkage to claims databases generated a dataset that combines patient-reported MS-related information and standardized data driven by the claims classification system. The linked database thus includes a wealth of information that is typically only available in separate databases. This could enable deeper and more rounded insights into burden of illness and other outcomes beyond what is possible with conventional database approaches. For example, complementing claims databases with patient-derived information on MS disease type would enable analyses of the impact of different types of MS on productivity or disability. Such information cannot be obtained from either data source in isolation. The potential to derive outcomes data was not assessed in this study but is currently being explored in future analysis of the database.

In our analysis, the concordance of the data, measured as PPV, between survey responses and both the claims databases used was high for diagnosis and current and prior DMT but lower for relapses. In the open-source database, relapse was based on a claims-based algorithm [24,25]. The survey was based on patient recall, which is typically less exact than data entered into claims databases.

The size of the cohort and the percentage of survey respondents linkable to the PharMetrics Plus database were relatively modest. There are several reasons for this, none of which invalidates the approach. The odds of successful linkage depend on the size and population coverage of the databases selected. Although it has a lower likelihood of missing claims, PharMetrics Plus is an order of magnitude smaller than the open-source databases and underrepresents patients from the western United States, which reduced the potential to link survey respondents from this region. These are weaknesses specific to the specific database, not to the method.

When expanding the linkage methodology to include databases such as electronic medical records [26], it is important to take the privacy aspect into account. It has been strongly argued that important privacy concerns must be interpreted alongside the social good that can come from this kind of health research [27-30]. In this study, the patient survey required active opt-in from the participants and no data were obtained from participants' Facebook accounts.

Limitations

There are weaknesses in the analysis. The sample size of 53 patients linked to the PharMetrics Plus database is too small to read much into the data, and the main value of this particular dataset is in demonstrating the feasibility of the method. The high linkage rates but low concordance with the Dx/Rx databases are noted, attributable to the open-source nature of the databases. As social media content is user driven, there is no independent verification of the correctness of the data, although the concordance analysis indicates that the social media information reflected patients' actual situations. This limitation applies to all social media analyses [31]. There is a wide range of trust in Web-based information [19], and differences in attitudes toward social media among patients with MS may produce a certain bias in the survey population toward those more willing to use and trust social media. All data are for US populations and the generalizability of the methods has yet to be established. There is a possibility of selection bias as participants were recruited by Web-based advertisements. From the survey results, the Facebook-recruited survey participants may have been somewhat more severely affected by their illness than the overall population. More symptomatic patients may be more motivated to complete a Web-based survey. Such bias toward more severely affected patients would also affect an

analysis of MS-related costs. It should also be underlined that the survey relied on patient recall, which is less than 100% reliable [32]. The date for an event recorded by a patient in the survey may not correspond to the time point for the same event in the PharMetrics Plus database. However, given the large number of participants and the relatively high degree of linkage, these two risks do not seem to have invalidated the collected data.

With these limitations in mind, this study shows that the combination of advertisements on open Internet platforms and Web-based surveys may enable rapid gathering of real-life data on a large US population of representative patients with MS. The applicability of the approach to diseases other than MS would need independent verification.

Acknowledgments

Authors' Contributions

Data were collected and analyzed by IMS Health and interpreted by all authors. The manuscript was drafted by VR. All authors provided comments and additional interpretations at the writing stage. All authors approved the final manuscript.

Conflicts of Interest

VR, JM, and MO are employees of Novartis.

Multimedia Appendix 1

Survey questionnaire. The EQ-5D-3L and EQ-VAS instruments are proprietary and cannot be displayed in publicly available materials.

[PDF File (Adobe PDF File), 94KB - jmir_v18i9e249_app1.pdf]

Multimedia Appendix 2

Disclaimer presented to survey participants explaining confidentiality and handling of the data.

[PDF File (Adobe PDF File), 47KB - jmir_v18i9e249_app2.pdf]

Multimedia Appendix 3

Description of the databases used in the analysis.

[PDF File (Adobe PDF File), 29KB - jmir_v18i9e249_app3.pdf]

References

1.   European Medicines Agency. 2013 Oct 18. The patient's voice in the evaluation of medicines Internet URL: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2013/10/WC500153276.pdf [accessed 2016-08-31] [WebCite Cache ID 6kAs2dIbD]

2.   US Food & Drug Administration. United States Government Printing Offices. 2012. Food and Drug Administration Safety and Innovation Act (FDASIA) Internet URL: https://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf [accessed 2016-08-31] [WebCite Cache ID 6kAsJm4cq]

3.   Stafinski T, Menon D, Nardelli A, Bakal J, Ezekowitz J, Tymchak W, et al. Incorporating patient preferences into clinical trial design: results of the opinions of patients on treatment implications of new studies (OPTIONS) project. Am Heart J 2015 Jan;169(1):122-31.e22. [doi: 10.1016/j.ahj.2014.10.002] [Medline: 25497257]

4.   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]

5.   Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]

6.   Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]

7.   Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]

8. European Medicines Agency. 2014 Sep 08. Guideline on good pharmacovigilance practices (GVP): Module VI - Management and reporting of adverse reactions to medicinal products (Rev 1) Internet URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/09/WC500172402.pdf [accessed 2016-08-31] [WebCite Cache ID 6kAsgj1Iq]

9. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]

10. Nakhasi A, Passarella R, Bell S, Paul M, Dredze M, Pronovost P. Malpractice and malcontent: Analyzing medical complaints in twitter. AAAI Tech Rep 2012 Apr;FS-12-05:84-85.

11. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. 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. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of Treatment Switching in Multiple Sclerosis Therapies in US Patients Active on Social Media: Application of Social Media Content Analysis to Health Outcomes Research. J Med Internet Res 2016;18(3):e62 [FREE Full text] [doi: 10.2196/jmir.5409] [Medline: 26987964]

13. Eichler GS, Cochin E, Han J, Hu S, Vaughan TE, Wicks P, et al. Exploring Concordance of Patient-Reported Information on PatientsLikeMe and Medical Claims Data at the Patient Level. J Med Internet Res 2016;18(5):e110 [FREE Full text] [doi: 10.2196/jmir.5130] [Medline: 27174602]

14. Epstein M, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf 2005 Aug;14(8):589-595. [doi: 10.1002/pds.1082] [Medline: 15918159]

15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]

16. World Medical Association. 2008. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects, amended by the 59th WMA General Assembly, Seoul, October 2008 Internet URL: http://http//:www.wma.net/en/30publications/10policies/b3/index.html.pdf [accessed 2016-08-31] [WebCite Cache ID 6kAtEm5Rs]

17. IMS Health. 2016. HEOR Bibliographies URL: http://www.imsheorbibliography.com/ [accessed 2016-08-05] [WebCite Cache ID 6jWvOKQ6k]

18. Shirazipour CH, Baillie CP, Galaviz K, Jarvis JW, Latimer-Cheung AE. Evaluating the theoretical content of online physical activity information for people with multiple sclerosis. Int J MS Care 2015;17(2):91-100 [FREE Full text] [doi: 10.7224/1537-2073.2013-048] [Medline: 25892979]

19. Colombo C, Mosconi P, Confalonieri P, Baroni I, Traversa S, Hill SJ, et al. Web search behavior and information needs of people with multiple sclerosis: focus group study and analysis of online postings. Interact J Med Res 2014;3(3):e12 [FREE Full text] [doi: 10.2196/ijmr.3034] [Medline: 25093374]

20. Hadgkiss E, Jelinek G, Weiland T, Pereira N, Marck C, van der Meer D. Methodology of an International Study of People with Multiple Sclerosis Recruited through Web 2.0 Platforms: Demographics, Lifestyle, and Disease Characteristics. Neurol Res Int 2013;2013:1-12 Article ID 580596 [FREE Full text] [doi: 10.1155/2013/580596] [Medline: 23691313]

21. Jelinek GA, Weiland TJ, Hadgkiss EJ, Marck CH, Pereira N, van der Meer D. Medication use in a large international sample of people with multiple sclerosis: associations with quality of life, relapse rate and disability. Neurol Res 2015 Aug;37(8):662-673 [FREE Full text] [doi: 10.1179/1743132815Y.0000000036] [Medline: 25905471]

22. Weiland T, Jelinek G, Marck C, Hadgkiss E, van der Meer D, Pereira N, et al. Clinically significant fatigue: prevalence and associated factors in an international sample of adults with multiple sclerosis recruited via the internet. PLoS One 2015;10(2):e0115541 [FREE Full text] [doi: 10.1371/journal.pone.0115541] [Medline: 25692993]

23. Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. J Med Toxicol 2013 Jun;9(2):184-191 [FREE Full text] [doi: 10.1007/s13181-013-0299-6] [Medline: 23619711]

24. Chastek BJ, Oleen-Burkey M, Lopez-Bresnahan MV. Medical chart validation of an algorithm for identifying multiple sclerosis relapse in healthcare claims. J Med Econ 2010;13(4):618-625. [doi: 10.3111/13696998.2010.523670] [Medline: 20883151]

25. Ollendorf DA, Jilinskaia E, Oleen-Burkey M. Clinical and economic impact of glatiramer acetate versus beta interferon therapy among patients with multiple sclerosis in a managed care population. J Manag Care Pharm 2002;8(6):469-476 [FREE Full text] [doi: 10.18553/jmcp.2002.8.6.469] [Medline: 14740608]

26. Cascade E, Marr P, Winslow M, Burgess A, Nixon M. Conducting research on the Internet: medical record data integration with patient-reported outcomes. J Med Internet Res 2012;14(5):e137 [FREE Full text] [doi: 10.2196/jmir.2202] [Medline: 23060427]

27. Verma IM. Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jul 22;111(29):10779 [FREE Full text] [doi: 10.1073/pnas.1412469111] [Medline: 24994898]

28. Goel V. NY Times Internet. 2014 Aug 12. As Data Overflows Online, Researchers Grapple With Ethics URL: http://www.nytimes.com/2014/08/13/technology/the-boon-of-online-data-puts-social-science-in-a-quandary.html?_r=2 [accessed 2016-08-22] [WebCite Cache ID 6SVvgOJ3k]

XSL•FO
RenderX

29.   Padrez KA, Ungar L, Schwartz HA, Smith RJ, Hill S, Antanavicius T, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. BMJ Qual Saf 2015 Oct 13:1-10. [doi: 10.1136/bmjqs-2015-004489] [Medline: 26464519]

30.   Kahn JP, Vayena E, Mastroianni AC. Opinion: Learning as we go: lessons from the publication of Facebook's social-computing research. Proc Natl Acad Sci U S A 2014 Sep 23;111(38):13677-13679 [FREE Full text] [doi: 10.1073/pnas.1416405111] [Medline: 25217568]

31.   Goff DA, Kullar R, Newland JG. Review of Twitter for infectious diseases clinicians: useful or a waste of time? Clin Infect Dis 2015 May 15;60(10):1533-1540. [doi: 10.1093/cid/civ071] [Medline: 25652087]

32.   Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]

Abbreviations

DMT: disease-modifying therapy
HIPAA: Health Insurance Portability and Accountability Act
https: hypertext transfer protocol, secure
MS: multiple sclerosis
PPV: positive predictive value
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Original Paper

Australian Gay Men Describe the Details of Their HIV Infection Through a Cross-Sectional Web-Based Survey

Ian Down[1], MPH; Garrett Prestage[1], PhD; Jeanne Ellard[2], PhD; Kathy Triffitt[1], PhD; Graham Brown[2], PhD; Denton Callander[1,3], PhD

[1]Kirby Institute, University of New South Wales Australia, Sydney, Australia

[2]Australian Research Centre in Sex, Health and Society, La Trobe University, Melbourne, Australia

[3]Centre for Social Research in Health, University of New South Wales Australia, Sydney, Australia

Corresponding Author:
Ian Down, MPH
Kirby Institute
University of New South Wales Australia
Wallace Wurth Building
Sydney, 2052
Australia
Phone: 61 2 9385 9954
Fax: 61 2 9385 0920
Email: idown@kirby.unsw.edu.au

Abstract

Background: With emerging opportunities for preventing human immunodeficiency virus (HIV) transmission, it remains important to identify those at greatest risk of infection and to describe and understand the contexts in which transmissions occur. Some gay and bisexual men with recently diagnosed HIV infection are initially unable to identify high-risk behaviors that would explain their HIV infection. We explored whether Web-based data collection could assist them in identifying the circumstances of their infection.

Objective: To assess the capacity of a Web-based survey to collect reliable self-report data on the event to which gay and bisexual men ascribe their HIV infection.

Methods: The HIV Seroconversion Study included a Web-based survey of gay and bisexual men with recently diagnosed HIV infection in Australia. Participants were asked if they could identify and describe the event they believe led to their infection. Men were also asked about their sexual and other risk practices during the 6 months before their diagnosis.

Results: Most (403/506, 79.6%) gay and bisexual men with newly diagnosed HIV infection were able to identify and describe the circumstances that likely led to their infection. Among those who were initially unable to identify possible exposure events, many could nonetheless provide sensible information that ostensibly explained their seroconversion. Free-text responses allowed men to provide more detailed and contextual information, whereas questions about the totality of their sexual behavior before diagnosis provided opportunities for men to describe their sexual risk behavior in general. Overall, 84.0% indicated having engaged in condomless anal intercourse before their HIV diagnosis, including 71.8% in the receptive position.

Conclusions: This study demonstrates the effectiveness of using Internet-based technologies to capture sensitive information about the circumstances in which HIV infection occurs among gay and bisexual men. By providing a range of opportunities for relaying experience, this research reveals some of the complexity in how individuals come to understand and explain their HIV infection. These findings may assist in obtaining detailed sexual history in the clinical setting.

KEYWORDS

HIV; transmission; sexual behavior; surveys and questionnaires

Introduction

The diagnoses of human immunodeficiency virus (HIV) infection among gay and bisexual men continue to increase in Australia, and elsewhere [1,2]. Epidemiological studies have provided measures of per-contact probability of HIV transmission through various transmission modes [3,4]. Among gay and bisexual men, condomless anal intercourse presents the highest HIV risk exposure, particularly for those men who take the receptive position [4] and if their partner ejaculates inside the rectum [3]. Trends in condomless anal intercourse with casual partners have been found to be a strong predictor of trends in HIV infections among gay and bisexual men [5]. Furthermore, the presence of some sexually transmissible infections can facilitate HIV transmission [6].

Many gay and bisexual men have adapted their sexual behaviors to minimize the potential of HIV transmission [7-9]. The use of condoms during anal sex continues to be the primary method to reduce HIV transmission among gay and bisexual men [10-12]. In Australia, about a third of gay and bisexual men report no recent condomless anal intercourse (within the previous 6 months), whereas around 1 in 5 restrict condomless anal intercourse to seroconcordant regular partners [10]. However, recent behavioral surveillance data suggest that a sizeable proportion of men, more than a third, are engaging in condomless anal intercourse with casual partners [13].

While injecting drug use remains a common mode of transmission in many other settings [14], the practice has been attributed to fewer than 3% of all new HIV diagnoses in Australia over the past 5 years [2]. Although gay and bisexual men report injecting drug use at rates higher than their heterosexual peers [15], the actual proportion is nonetheless small and has been declining somewhat over the past decade [13]. Moreover, most Australian gay and bisexual men appear to adhere to safer injecting practices and are less likely than their heterosexual peers to share injecting equipment [16].

Studies of recent seroconverters have found HIV risk to be associated with sexual adventurism [17], the use of sex venues to meet partners [18,19], using illicit drugs while having sex [20,21], and the presence of sexually transmissible infections [6]. Data have also provided evidence of men acquiring HIV while attempting, and in the absence of, non–condom-based HIV risk reduction strategies [22]. In Australia, gay and bisexual men tend to have a fairly well-developed sense of the level of risk involved in particular sexual practices [23], and this perceived hierarchy of risks appears to broadly reflect what is known of relative risk [24].

The collection and analysis of reliable behavioral surveillance data is an essential part of an effective response to HIV. Self-report survey responses to measure sexual risk behaviors raise questions about reliability, with the potential effects of recall bias due to participants' inability to accurately recall their sexual behavior, sensitivity about answering questions concerning sexual practices, and the unwillingness of some to acknowledge having engaged in risk behavior [25-27]. In clinic settings, when individuals provide this information to their doctor, what they describe may be no less subject to recall and

social desirability bias. Indeed, there has been some suggestion that when collecting sensitive information, doing so in a clinic setting may inhibit full disclosure [27].

Such limitations may be even more pronounced when the data being collected concern a recent HIV diagnosis experience [28]. Previous research has indicated that between 20% and 36% of men are unable to identify a risk behavior that may have led to their infection [28,29]. One study found that men who were reluctant or unable to disclose information about condomless anal intercourse during surveys were later able to identify such events during face-to-face interviews [30].

Previous studies of HIV seroconversion in Australia have relied on clinical referral, involved clinic-based data collection, and were limited to sites in Sydney and Melbourne. Between 1993 and 2001, a total of 92 participants were interviewed (an average of 11.5 per year), whereas from 2003 to 2006, a total of 158 men were recruited (an average of almost 53 per year) [22,31]. These studies predated, or did not use, Web-based data collection, which may resolve these issues around recruitment and the sensitivity of data collection processes.

We sought to assess whether gay and bisexual men with recently diagnosed HIV infection could provide consistent and reliable information about their sexual risk behavior before their HIV diagnosis, using Web-based data collection techniques to capture information about possible exposures. Here, we report on sexual and other risk behavior data collected from these men, and compare these with known modes of transmission.

Methods

Eligibility Criteria

The HIV Seroconversion Study included a Web-based survey of people in Australia with recently diagnosed HIV infection. Eligibility criteria were being older than 18 years, living in Australia, and having been diagnosed HIV positive within the 2 years before enrollment. Ethics approval was obtained from the University of New South Wales and La Trobe University Human Research Ethics Committees.

Recruitment

A convenience sample was obtained through a range of recruitment strategies: referrals from the staff of state AIDS Councils and organizations for people living with HIV, referrals from clinics (mostly sexual health services), or direct Web-based enrollment by individuals who have found a link to the survey posted on another website. The study was promoted on the websites of gay community organizations and in gay press. The only direct contact to recruit individuals took place either in a clinic setting or via a community-based organization. Eligible participants were directed to a dedicated study website, which provided information about the study, its purpose, and what participation involved. Participants were told that the survey was anonymous and would take around 20 minutes to complete. Participation was voluntary, and no incentives were offered. Those who chose to consent to participate were invited to begin the survey. As the survey was anonymous, participants' consent is implied by their continuation. The study methods have been described in more detail elsewhere [32].

Survey Questionnaire

The survey questionnaire was developed in consultation with community partners and was pilot-tested for usability and reliability. The questionnaire included demographic characteristics, details of participants' HIV diagnosis, and their sexual and drug-taking behaviors both before and since their diagnosis. Respondents were able to navigate back to previous answers, should they wish to review their responses. Only key questions required a response, to ensure participants were only asked questions relevant to their circumstances.

Participants were asked if they could identify one or more events that they believed may have led to them acquiring HIV. Those who were not immediately able to identify an event were provided with information about potentially risky activities, such as anal intercourse without a condom, sharing a needle or other injecting equipment, or any other activity where they may have been exposed to someone's body fluids (such as blood or semen), and then asked again if they could identify an event that may have exposed them to HIV. Those who could identify their most likely highest-risk event were then asked questions about their sexual and other risk behavior at that event.

To capture additional detail, the behavioral survey questions were complemented by optional free-text questions, with men invited to provide further detail in their own words of any factors they believe contributed to their infection. These free-text responses were reviewed for sexual practices that may have facilitated the transmission of HIV and coded accordingly. Sex practices other than anal or oral intercourse were coded as semen-related (for example, ejaculating around the anus and then inserting a finger into the anus using the ejaculate as a lubricant) or blood-related (for example, fisting and sharing sex toys). An in-depth analysis of the free-text data collected through the Web-based survey is reported separately [33].

Finally, and separately, participants were asked about their sexual and injecting drug use behavior in the 6 months before diagnosis. They were asked specifically about anal intercourse, with and without condoms, according to their sexual position, with both regular and casual partners.

Analysis

The quantitative data were analyzed with SPSS version 22 software. We report the range of behaviors that occurred at the highest-risk event as they were provided in participants' survey responses—including those revealed in free-text responses—and other risk practices reported in the 6 months before their diagnosis.

Results

Study Participation

From December 2007 to March 2013, a total of 506 unique male respondents who reported that their HIV infection was due to homosexual contact had enrolled in the study. Most men (449/506, 88.7%) reported the year they received their diagnosis (Table 1). On average, over the time of these analyses, (449/5298) 8.47% of the eligible population enrolled in the study.

Table 1. Survey participant enrollment, by year of diagnosis and as proportion of eligible population.

Year	Survey respondents, n	Number of Australian diagnoses recorded, attributed to sex between men, N	Proportion of eligible population that participated in survey, %
2006	19	678	2.8
2007	53	702	7.5
2008	83	670	12.4
2009	82	691	11.9
2010	83	679	12.2
2011	69	806	8.6
2012	56	867	6.5
2013	4	205[a]	2.0
Total	449	5298	8.5

[a]Survey recruitment occurred only in the first 3 months of 2013, therefore we have provided one-fourth of the number of new diagnoses that occurred in 2013.

Describing Risk Behaviors

Most of the 506 men (403/506, 79.6%) were able to identify one or more high-risk events that they believed may have led to their HIV infection without any prompting. Among those not immediately able to identify such an event, a further 51 men were able to recollect occasions where they may have been exposed to HIV after prompting. Of the 454 who could recall one or more high-risk events, almost half (202/454, 44.5%) identified one such event, whereas for the remaining men there was more than one occasion that may have resulted in their HIV infection (mean 7.32; standard deviation, SD, 12.06). These men were asked to select the event that was the most "risky" based on their own assessment using provided information about relative risk, with consideration to the time since their last HIV-negative test result, as well as a time when they may have experienced seroconversion-like symptoms. On the basis of this assessment, they were then asked to describe the event they felt was most likely to be their seroconversion event. A total of 403 men were able to describe a single highest-risk event. For these

analyses, we included those 403 men, as well as an additional 62 men who provided free-text responses describing risk or responses to questions about their behaviors in the 6 months before diagnosis. Overall, (258/465), 55.5% were unable to describe the HIV status of their sex partner at the time of the highest-risk event, whereas (140/465) 30.1% believed him to be HIV negative and (67/465) 14.4% believed him to be HIV positive.

The 465 men included in these analyses ranged in age, at the time of their HIV diagnosis, from 16 to 73 years, with a mean of 34.5 (SD 9.5) years. Most men (306/465, 65.8%) completed the survey within 12 months of receiving their diagnosis, including (179/465) 38.5% who did so within 3 months of receiving their diagnosis. The vast majority of the men identified as gay (425/465, 91.4%), more than half (249/465, 53.6%) had some university education, and most had been born in Australia (364/465, 71.8%).

Details of the men's HIV risk behaviors are presented in Table 2. In response to the direct survey questions about the sexual practices that occurred at their highest-risk event, most men reported anal intercourse at the time (384/465, 82.6%), with (322/465) 69.2% reporting condomless anal intercourse. Men most commonly (280/465, 60.2% of the men in these analyses) reported that they were the receptive partner, and (177/465) 38.1% reported that their partner ejaculated in their rectum. Among the men, (193/465) 41.5% were exclusively receptive during condomless anal intercourse, whereas (49/465) 10.5% were exclusively insertive. A small proportion of men (63/465, 13.5%) reported only condom-protected anal intercourse as their highest-risk event. Overall, 1 in 6 men (76/465, 16.3%) reported receptive oral intercourse with their partner ejaculating in their mouth. Of the 40 men (40/465, 8.6%) who reported injecting drug use as part of their highest-risk event, 5 indicated that they shared injecting equipment at the time.

In free-text responses, 94 men provided details about engaging in condomless anal intercourse at the highest-risk event. Few men (54/465, 11.6%) attributed their infection to sexual practices other than anal intercourse. Practices that may result in the transfer of blood between partners, such as fisting or sharing sex toys or douching equipment, were described by 21 men (21/465, 2.4%). The possibility of semen or pre-ejaculatory fluid entering the rectum was believed to be the mode of

acquisition for 13 men, who described the brief, partial insertion of their partner's penis into the anus without a condom ("dipping") or their partner ejaculating around their anus and then inserting a finger into the anus. In their text responses, 1 in 10 men commented on abrasions on the skin or mucosal membranes in the mouth, anus, or penis that they believed might have facilitated infection. For 28 men, their responses were sufficient to reasonably imply that they had engaged in condomless anal intercourse; for example, when asked about what may have contributed to their acquiring HIV, one response was "my stupidity and horniness" and another was "too drunk, so didn't take proper care." These sorts of comments were often not accompanied by quantitative survey data or were accompanied by data that did not correspond.

In describing their sexual behavior during the 6 months before their HIV diagnosis, more than half of men (272/465, 58.5%) reported having engaged in condomless anal intercourse during that period. Men were more likely to report having been the receptive partner in condomless anal intercourse (241/465, 51.8%) than the insertive partner (191/465, 41.1%). Most men who reported receptive condomless anal intercourse had also done so to the point of ejaculation inside their rectum. There were 8 men who were unable to identify a highest-risk event and went on to describe occasions of condomless anal intercourse with casual partners in the 6 months before their diagnosis.

Validating Responses

The 3 sources of information in the men's responses were combined and reallocated according to their relative risk. These data are presented in Table 3 as a hierarchy of risk. We identified that (372/465) 84.0% of the men in this sample reported some condomless anal intercourse in the 6 months before their diagnosis. Almost three-fourths of the sample (334/465, 71.8%) reported receptive condomless anal intercourse, and (272/465) 58.5% of men had engaged in receptive condomless anal intercourse with ejaculation. Despite some discrepancies, and the provision of additional information in one set of questions versus other questions, there was broad consistency between the behaviors men reported through the survey responses and the descriptions provided in the free-text responses. Individual responses to the survey questions are compared with the individual free-text responses in Table 4.

Table 2. Risk behaviors reported at the highest-risk event and during the preceding 6 months.

Risk behaviors N=465	Direct response in survey (HRE[a]) n (%)	Text responses (about HRE) n (%)	Behavior in the preceding 6 months n (%)
Anal intercourse			
Any anal intercourse	384 (82.6)	-	-
Any receptive anal intercourse	317 (68.2)	-	-
Any insertive anal intercourse	174 (37.4)	-	-
Receptive anal intercourse only with a condom	39 (8.4)	-	-
Condom slippage or breakage during receptive anal intercourse	25 (5.4)	10 (2.2)	-
Insertive anal intercourse only with a condom	35 (7.5)	-	-
Condom slippage or breakage during insertive anal intercourse	12 (2.6)	-	-
Any condomless anal intercourse	322 (69.2)	-	272 (58.5)
Any receptive condomless anal intercourse	280 (60.2)	-	241 (51.8)
Receptive condomless anal intercourse with ejaculation	177 (38.1)	31 (6.7)	189 (40.6)
Receptive condomless anal intercourse (ejaculation unspecified)	-	13 (2.8)	-
Receptive condomless anal intercourse withdrawal	103 (22.1)	8 (1.7)	-
Any insertive condomless anal intercourse	136 (29.2)	8 (1.7)	191 (41.1)
Unspecified condomless anal intercourse	-	34 (7.3)	-
Other sex activities identified as potentially leading to infection			
Receptive oral sex with ejaculation in mouth	76 (16.3)	5 (1.1)	-
Receptive oral sex (ejaculation unspecified)	-	20 (4.3)	-
Blood-related (eg, fisting, sharing toys)	-	21 (4.5)	-
Semen-related (eg, ejaculate as lubricant, nudging or dipping)	-	13 (2.8)	-
Wound, sore, or infection	109 (23.4)	50 (10.8)	-
Risk implied although not specified	-	28 (6.0)	-
Nonsexual risk			
Injecting drug use—without sharing equipment	36 (7.7)	-	35 (6.5)
Injecting drug use—with shared equipment	4 (0.9)	2 (0.4)	7 (0.2)
Tattoo, medical, or other blood to blood	-	5 (1.1)	-
Could not identify likely risk	59 (12.7)	-	-

[a]HRE: highest-risk event.

Table 3. Highest-risk behavior reported either at the highest-risk event or during the preceding 6 months.

Behavior N=465	n (%)
Receptive condomless sex with ejaculation	272 (58.5)
Receptive condomless sex (ejaculation unspecified)	7 (1.5)
Receptive condomless sex withdrawal	46 (9.9)
Condom slippage or breakage during receptive anal intercourse	9 (1.9)
Insertive condomless sex	54 (11.6)
Condom slippage or breakage during insertive anal intercourse	3 (0.6)
Semen-related sex act (eg, ejaculate as lubricant, nudging, or dipping)	5 (1.1)
Blood-related sex act (eg, fisting, sharing toys)	1 (0.2)
Receptive oral sex with ejaculation	6 (1.3)
Receptive oral sex (ejaculation unspecified)	3 (0.6)
Wound, sore, or infection	12 (2.6)
Injecting drug use—with shared equipment	2 (0.4)
Injecting drug use—without sharing equipment	4 (0.9)
Tattoo, medical, or other blood to blood	5 (1.1)
Only condom-protected anal intercourse	19 (4.1)
No clear evidence of risk	17 (3.7)

XSL•FO
RenderX

Table 4. Risk behaviors described in survey responses and free-text responses.

Risk behaviors	Anal intercourse, n							Other, n	Total, N
	RCLAI[a] ejaculation	RCLAI withdrawal	Condom break RAI[b]	ICLAI[c]	Condom break IAI[d]	RAI with condom	IAI with condom		
RCLAI with ejaculation	20	3	0	2	0	1	1	4	31
RCLAI (ejaculation unspecified)	4	5	1	0	0	0	1	2	13
RCLAI withdrawal	3	3	0	1	0	1	0	0	8
Condom slippage/ breakage during RAI	0	2	5	0	2	1	0	0	10
ICLAI	0	1	0	6	1	0	0	0	8
Unspecified CLAI[e]	18	5	0	3	0	0	0	4	30
Semen-related sex act (eg, nudging)	1	3	0	0	0	2	1	6	13
Blood-related sex act (eg, fisting)	5	4	0	5	0	4	1	2	21
Receptive oral sex with ejaculation	0	0	0	1	0	0	1	3	5
Receptive oral sex (ejaculation unspecified)	3	1	0	0	0	2	2	5	13
Wound, sore, or infection	6	8	2	5	3	2	3	2	31
Injecting drug use-sharing	0	1	0	0	0	0	0	1	2
Tattoo, medical, or other blood to blood	0	0	0	0	0	0	0	5	5
Risk implied, although not specified	11	11	2	2	0	0	0	2	28
Only condom-protected anal intercourse	0	0	0	0	0	1	0	0	1
No clear evidence of risk	2	0	0	0	0	0	0	0	2
Unspecified anal intercourse	2	0	0	0	0	0	0	1	3
Unspecified RAI	0	1	0	0	0	0	0	0	1
No reference to risk in text response	102	55	2	24	2	9	2	44	240
Total	177	103	12	49	8	23	12	81	465

[a]RCLAI: receptive condomless anal intercourse.

[b]RAI: receptive anal intercourse.

[c]ICLAI: insertive condomless anal intercourse.

[d]IAI: insertive anal intercourse.

[e]CLAI: condomless anal intercourse.

Discussion

Principal Findings

We successfully employed Internet-based technologies to recruit a much larger and geographically extensive sample than obtained from earlier Australian seroconverter studies [22,31]. Aside from being a broader sample, the data we collected indicate that most men are both capable of and willing to identify and report their risk behavior in ways that make sense and appear reasonably reliable as a representation of their recollection and understanding of what occurred.

We were able to recruit on average nearly 100 men per year from across Australia, with fewer numbers recruited in later years as recruitment became less intensive because of reduced funding. This compares favorably with previous studies, which collected 15 and 53 participants per year in previous studies, all of those being from Sydney or Melbourne. The size of the sample recruited to this study is much larger than obtained from earlier studies, demonstrating that Internet-based studies can have much wider reach. During recruitment, a common resistance from some sites to refer patients to the study was concern about "burdening" people soon after diagnosis. That almost two-fifths of men completed the survey within 3 months of receiving their diagnosis suggests there is a willingness to participate in this kind of study, even while dealing with the impacts of their recent diagnosis.

The majority of men in this sample identified an occasion of receptive condomless anal intercourse as the mode through which they acquired HIV and, for most of those men, their partner ejaculated in their rectum at the time. This finding accords with what is known about the relative risk for HIV infection of specific practices, which suggests that most men in our sample appear capable of and willing to recall and report risk behavior. Some men identified multiple risk events, but most of them were able to identify what they believed was the highest-risk event. Additionally, some men reported having engaged in condomless anal intercourse during the 6 months before their diagnosis, some of which events may not have been identified by the men as potentially leading to their infection. Furthermore, a small number of men described practices that would otherwise be considered as safe, and these are explored in more detail through the qualitative data collected as part of this study [33]. A small number of men, however, were unable to identify a risk event to explain their infection, which has been noted in previous studies of people with recently diagnosed HIV infection [28,29]. In our study, recall was improved by the use of probing devices and free-text questions, and we were able to identify risk behaviors for many of these same men. In the end, most men reported receptive condomless anal intercourse in the period before their HIV diagnosis. This rate of receptive condomless anal intercourse, and particularly where ejaculation occurs in the rectum, was much higher than what has been found in surveys of Australian gay and bisexual men generally, particularly among HIV-negative gay and bisexual men, and even when compared with HIV-negative gay and bisexual men who engaged in condomless anal intercourse [10].

Although a small number of men reported injecting drug use during their highest-risk event, only a minority of those men reported sharing injecting equipment with others. This is consistent with what has previously been found among gay and bisexual men in Australia [16]. Nonetheless, after our detailed analysis of men's responses, it is clear that the only identifiable risk behavior for a small number of men was injecting drug use.

In the end, most men were both capable of and willing to identify and report their risk behavior in ways that make sense and appear reasonably reliable as a representation of their recollection and understanding of what led to their HIV infection. This study was able to identify rates of condomless anal intercourse at least as high as previous similar studies [28,29].

Limitations

We excluded from our detailed analysis 41 men who could not identify a risk event or provide detail about their sexual behaviors in the 6 months before their diagnosis. We cannot know why they did not complete these sections of the survey, nor can we account for them. Our descriptions here, though, are of men who were at least able to recognize, or describe, something that had happened that put them at risk of infection. Although we did not identify any duplicate entries, it is possible that individuals may have entered their responses more than once.

We cannot determine from these data that the events men believed led to their infection were the actual source of their infection. More than half of the men reported multiple possible events.

Over the last decade, there have been significant changes in how gay men connect with each other [34] and their sense of community [35]. These changes have run parallel to changes in HIV prevention, with focus on emerging biomedical approaches [36-38]. It is therefore understandable that for some men there may be uncertainty about the current relative risk of particular practices [39]. Definitions of what is safe and what is not safe may differ between men and may be changing for individual men. In addition, the ways that the relative risk of specific sex practices are assessed within gay communities appear to be changing, further adding to the possibility of confusion for individual men [40].

Although self-report data on sexual risk behavior are subject to limitations, anonymous Web-based surveys that present culturally appropriate questions have been found to produce reliable data [25]. There are particular challenges to accurately measuring sexual behavior, given that "the object under consideration—sexual practice and its change—is fluid, embedded in specific social formations, and involves the negotiation of meaning" [41,42]. Surveys are discursive and iterative in that they can often imply narrow meanings and interpretation in ways that do not always reflect what was intended, and may miss important mediating factors or contextual influences. Some of the men who did not provide responses to the set survey questions about the activity that occurred at the time they became infected did utilize the free-text component of the survey to describe and explore other events

that may have led to infection, including less common explanations for their transmission, which is the subject of a separate analysis [33].

Given this interpretive variability, even well-developed survey questions can lead to confusion between the researchers' intention and the respondents' interpretation. With opportunity, many men can and will offer further contextual explanations in relation to what they believe led to their HIV infection, demonstrating the value of including free-text opportunities in survey questionnaires where possible. This contextual detail suggests that some men require an opportunity to articulate their story in greater detail, including the reasons they found themselves in a particular situation. Motivation for participants' sexual behavior was not asked in the survey questionnaire.

Conclusions

These findings suggest that seroconverter studies could take advantage of Internet-based technologies to obtain a broad sample of individuals with recently diagnosed HIV infection. Our data demonstrate a willingness for people with recently diagnosed HIV infection to engage in such research and that this method can collect, what appear to be, reasonably reliable data. Individuals with newly diagnosed HIV infection are generally able to identify and describe the likely circumstances that led to their infection. However, for some individuals, this may not always be so straightforward. Providing multiple and accessible methods for people to describe how they believe they were infected may allow some people to explain what they believe happened, even if they could not do so using one particular method. Providing diverse opportunities for relaying experience can sometimes elicit more detailed and contextual information. Given that identifying the circumstances that led to a person's HIV infection is a high priority, providing opportunities for alternative ways of describing the details of what occurred, within survey instruments and through qualitative methods such as narrative interviews, strengthens the evidence base for more effective HIV prevention and support work. In the changing environment, where the circumstances of seroconversion are likely to be more complex, the capacity to reliably collect sensitive information remains important. In the context of limited resources, this appears to be an efficient and reliable method.

In addition to clinic-based data collection, Internet-based technologies should be employed to obtain detailed data about the circumstances in which individuals believe they were infected with HIV.

Authors' Contributions

GP, ID, JE, KT, and GB contributed to the study design. ID took primary responsibility for drafting and redrafting the manuscript, with support from GP, DC, JE, GB and KT. ID conducted the data analyses, with support from GP and DC. All authors read, commented on, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]

2. The Kirby Institute. Australia: The Kirby Institute, UNSW; 2015. 2015 annual surveillance report of HIV viral hepatitis and sexually transmissible infections in Australia URL: https://kirby.unsw.edu.au/sites/default/files/hiv/resources/ASR2015.pdf [accessed 2016-08-14] [WebCite Cache ID 6gAjuQsTj]

3. Jin F, Jansson J, Law M, Prestage GP, Zablotska I, Imrie JC, et al. Per-contact probability of HIV transmission in homosexual men in Sydney in the era of HAART. AIDS 2010 Mar 27;24(6):907-913 [FREE Full text] [doi: 10.1097/QAD.0b013e3283372d90] [Medline: 20139750]

4. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. Am J Epidemiol 1999 Aug 1;150(3):306-311 [FREE Full text] [Medline: 10430236]

5. Zablotska IB, Prestage G, Middleton M, Wilson D, Grulich AE. Contemporary HIV diagnoses trends in Australia can be predicted by trends in unprotected anal intercourse among gay men. AIDS 2010 Jul 31;24(12):1955-1958. [doi: 10.1097/QAD.0b013e32833bf74a] [Medline: 20543653]

XSL•FO

RenderX

6.   Jin F, Prestage GP, Imrie J, Kippax SC, Donovan B, Templeton DJ, et al. Anal sexually transmitted infections and risk of HIV infection in homosexual men. J Acquir Immune Defic Syndr 2010 Jan;53(1):144-149. [doi: 10.1097/QAI.0b013e3181b48f33] [Medline: 19734801]

7.   Dubois-Arber F, Jeannin A, Lociciro S, Balthasar H. Risk reduction practices in men who have sex with men in Switzerland: serosorting, strategic positioning, and withdrawal before ejaculation. Arch Sex Behav 2012 Oct;41(5):1263-1272. [doi: 10.1007/s10508-011-9868-4] [Medline: 22083656]

8.   Prestage G, Van De Ven P, Grulich A, Kippax S, McInnes D, Hendry O. Gay men's casual sex encounters: discussing HIV and using condoms. AIDS Care 2001 Jun;13(3):277-284. [doi: 10.1080/09540120120043928] [Medline: 11397329]

9.   Van de Ven P, Kippax S, Crawford J, Rawstorne P, Prestage G, Grulich A, et al. In a minority of gay men, sexual risk practice indicates strategic positioning for perceived risk reduction rather than unbridled sex. AIDS Care 2002 Aug;14(4):471-480. [doi: 10.1080/09540120208629666] [Medline: 12204150]

10.   Mao L, Kippax SC, Holt M, Prestage GP, Zablotska IB, de Wit JBF. Rates of condom and non-condom-based anal intercourse practices among homosexually active men in Australia: deliberate HIV risk reduction? Sex Transm Infect 2011 Oct;87(6):489-493. [doi: 10.1136/sextrans-2011-050041] [Medline: 21764890]

11.   Saxton PJW, Dickson NP, Hughes AJ. Location-based HIV behavioural surveillance among MSM in Auckland, New Zealand 2002-2011: condom use stable and more HIV testing. Sex Transm Infect 2014 Mar;90(2):133-138. [doi: 10.1136/sextrans-2013-051160] [Medline: 24226099]

12.   Pines HA, Gorbach PM, Weiss RE, Shoptaw S, Landovitz RJ, Javanbakht M, et al. Sexual risk trajectories among MSM in the United States: implications for pre-exposure prophylaxis delivery. J Acquir Immune Defic Syndr 2014 Apr 15;65(5):579-586 [FREE Full text] [doi: 10.1097/QAI.0000000000000101] [Medline: 24378726]

13.   de Wit J, Mao L, Adam P, Treloar C. CSRH. Australia: Centre for Social Research in Health, UNSW; 2014. HIV/AIDS, hepatitis and sexually transmissible infections in Australia: Annual report of trends in behaviour 2014 (monograph 6/2014) URL: https://csrh.arts.unsw.edu.au/media/CSRHFile/Annual_Report_of_Trends_in_Behaviour_2014.pdf [accessed 2016-08-14] [WebCite Cache ID 6jkni0cfp]

14.   Degenhardt L, Mathers BM, Wirtz AL, Wolfe D, Kamarulzaman A, Carrieri MP, et al. What has been achieved in HIV prevention, treatment and care for people who inject drugs, 2010-2012? A review of the six highest burden countries. Int J Drug Policy 2014 Jan;25(1):53-60. [doi: 10.1016/j.drugpo.2013.08.004] [Medline: 24113623]

15.   Lea T, Mao L, Bath N, Prestage G, Zablotska I, de Wit J, et al. Injecting drug use among gay and bisexual men in Sydney: prevalence and associations with sexual risk practices and HIV and hepatitis C infection. AIDS Behav 2013 May;17(4):1344-1351. [doi: 10.1007/s10461-013-0409-0] [Medline: 23321949]

16.   Grulich AE, de Visser RO, Smith AM, Rissel CE, Richters J. Sex in Australia: injecting and sexual risk behaviour in a representative sample of adults. Aust N Z J Public Health 2003;27(2):242-250. [Medline: 14696718]

17.   Kippax S, Campbell D, Van de Ven P, Crawford J, Prestage G, Knox S, et al. Cultures of sexual adventurism as markers of HIV seroconversion: a case control study in a cohort of Sydney gay men. AIDS Care 1998 Dec;10(6):677-688. [doi: 10.1080/09540129848307] [Medline: 9924523]

18.   Thiede H, Jenkins RA, Carey JW, Hutcheson R, Thomas KK, Stall RD, et al. Determinants of recent HIV infection among Seattle-area men who have sex with men. Am J Public Health 2009 Apr;99 Suppl 1:S157-S164. [doi: 10.2105/AJPH.2006.098582] [Medline: 18445808]

19.   Silva AP, Greco M, Fausto MA, Greco DB, Carneiro M. Risk factors associated with HIV infection among male homosexuals and bisexuals followed in an open cohort study: Project Horizonte, Brazil (1994-2010). PLoS One 2014;9(10):e109390 [FREE Full text] [doi: 10.1371/journal.pone.0109390] [Medline: 25279670]

20.   Volk JE, Prestage G, Jin F, Kaldor J, Ellard J, Kippax S, et al. Risk factors for HIV seroconversion in homosexual men in Australia. Sex Health 2006 Mar;3(1):45-51. [Medline: 16607974]

21.   Plankey MW, Ostrow DG, Stall R, Cox C, Li X, Peck JA, et al. The relationship between methamphetamine and popper use and risk of HIV seroconversion in the multicenter AIDS cohort study. J Acquir Immune Defic Syndr 2007 May 1;45(1):85-92 [FREE Full text] [doi: 10.1097/QAI.0b013e3180417c99] [Medline: 17325605]

22.   Jin F, Prestage GP, Ellard J, Kippax SC, Kaldor JM, Grulich AE. How homosexual men believe they became infected with HIV: the role of risk-reduction behaviors. J Acquir Immune Defic Syndr 2007 Oct 1;46(2):245-247. [doi: 10.1097/QAI.0b013e3181565db5] [Medline: 17721400]

23.   Mao L, Adam P, Kippax S, Holt M, Prestage G, Calmette Y, et al. HIV-negative gay men's perceived HIV risk hierarchy: imaginary or real? AIDS Behav 2013 May;17(4):1362-1369. [doi: 10.1007/s10461-012-0406-8] [Medline: 23314802]

24.   Jin F, Crawford J, Prestage GP, Zablotska I, Imrie J, Kippax SC, et al. Unprotected anal intercourse, risk reduction behaviours, and subsequent HIV infection in a cohort of homosexual men. AIDS 2009 Jan 14;23(2):243-252 [FREE Full text] [doi: 10.1097/QAD.0b013e32831fb51a] [Medline: 19098494]

25.   Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. Sex Transm Infect 2001 Apr;77(2):84-92 [FREE Full text] [Medline: 11287683]

26.   Hoff CC, Faigeles B, Wolitski RJ, Purcell DW, Gomez C, Parsons JT. Sexual risk of HIV transmission is missed by traditional methods of data collection. AIDS 2004 Jan 23;18(2):340-342. [Medline: 15075559]

XSL•FO
RenderX

27.   Catania JA. A framework for conceptualizing reporting bias and its antecedents in interviews assessing human sexuality. Journal of Sex Research 1999 Feb;36(1):25-38. [doi: 10.1080/00224499909551964]

28.   Moskowitz DA. Claiming HIV infection from improbable modes as a possible coping strategy. J Appl Soc Psychol 2010;40(2):311-324. [doi: 10.1111/j.1559-1816.2009.00575.x]

29.   Richters J, Grulich A, Ellard J, Hendry O, Kippax S. HIV transmission among gay men through oral sex and other uncommon routes: case series of HIV seroconverters, Sydney. AIDS 2003 Oct 17;17(15):2269-2271. [doi: 10.1097/01.aids.0000088203.77946.99] [Medline: 14523289]

30.   Keet IP, Albrecht van Lent N, Sandfort TG, Coutinho RA, van Griensven GJ. Orogenital sex and the transmission of HIV among homosexual men. AIDS 1992 Feb;6(2):223-226. [Medline: 1558719]

31.   Kippax S, Slavin S, Ellard J, Hendry O, Richters J, Grulich A, et al. Seroconversion in context. AIDS Care 2003 Dec;15(6):839-852. [doi: 10.1080/09540120310001618685] [Medline: 14617505]

32.   Down I, Ellard J, Triffitt K, Brown G, Prestage G. Factors associated with recent previous HIV testing among a sample of recently HIV-diagnosed gay men in Australia: a cross-sectional study. Sex Transm Infect 2015 May;91(3):189-193. [doi: 10.1136/sextrans-2014-051752] [Medline: 25371421]

33.   Callander D, Prestage G, Ellard J, Triffitt K, Brown G, Down I. The Road Less Travelled: Exploring Gay and Bisexual Men's Explanations of 'Uncommon' Routes of HIV Transmission. AIDS Behav 2016 Jan 18:- Epub ahead of print. [doi: 10.1007/s10461-016-1289-x] [Medline: 26781867]

34.   Prestage G, Bavinton B, Grierson J, Down I, Keen P, Bradley J, et al. Online dating among Australian gay and bisexual men: romance or hooking up? AIDS Behav 2015 Oct;19(10):1905-1913. [doi: 10.1007/s10461-015-1032-z] [Medline: 25777506]

35.   Zablotska IB, Holt M, Prestage G. Changes in gay men's participation in gay community life: implications for HIV surveillance and research. AIDS Behav 2012 Apr;16(3):669-675. [doi: 10.1007/s10461-011-9919-9] [Medline: 21424273]

36.   Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]

37.   Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]

38.   Conway DP, Guy R, Davies SC, Couldwell DL, McNulty A, Smith DE, Sydney Rapid HIV Test Study. Rapid HIV Testing Is Highly Acceptable and Preferred among High-Risk Gay And Bisexual Men after Implementation in Sydney Sexual Health Clinics. PLoS One 2015;10(4):e0123814 [FREE Full text] [doi: 10.1371/journal.pone.0123814] [Medline: 25898140]

39.   Prestage G, Brown G, Down IA, Jin F, Hurley M. "It's hard to know what is a risky or not a risky decision": gay men's beliefs about risk during sex. AIDS Behav 2013 May;17(4):1352-1361. [doi: 10.1007/s10461-012-0180-7] [Medline: 22430641]

40.   McKechnie ML, Bavinton BR, Zablotska IB. Understanding of norms regarding sexual practices among gay men: literature review. AIDS Behav 2013 May;17(4):1245-1254. [doi: 10.1007/s10461-012-0309-8] [Medline: 22983537]

41.   Race K. The difference practice makes: Evidence, articulation, and affect in HIV prevention. AIDS Educ Prev 2014 Jun;26(3):256-266. [doi: 10.1521/aeap.2014.26.3.256] [Medline: 24846488]

42.   Kippax S, Stephenson N. Meaningful evaluation of sex and relationship education. Sex Education 2005 Nov;5(4):359-373. [doi: 10.1080/14681810500278436]

Abbreviations

HIV: human immunodeficiency virus
SD: standard deviation

XSL•FO
RenderX

Original Paper

The Inclusion of Ethnic Minority Patients and the Role of Language in Telehealth Trials for Type 2 Diabetes: A Systematic Review

Talia Isaacs[1*], PhD; Daniel Hunt[2*], PhD; Danielle Ward[3], MPH; Leila Rooshenas[4], PhD; Louisa Edwards[4], PhD

[1]University of Bristol, Graduate School of Education, Bristol, United Kingdom

[2]University of Nottingham, School of English, Nottingham, United Kingdom

[3]University of Bristol, School of Social Policy, Bristol, United Kingdom

[4]University of Bristol, School of Social and Community Medicine, Bristol, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Talia Isaacs, PhD
University of Bristol
Graduate School of Education
35 Berkeley Square
Bristol, BS8 1JA
United Kingdom
Phone: 44 1173314312
Fax: 44 1173317391
Email: talia.isaacs@bristol.ac.uk

Abstract

Background: Type 2 diabetes is a serious, pervasive metabolic condition that disproportionately affects ethnic minority patients. Telehealth interventions can facilitate type 2 diabetes monitoring and prevent secondary complications. However, trials designed to test the effectiveness of telehealth interventions may underrecruit or exclude ethnic minority patients, with language a potential barrier to recruitment. The underrepresentation of minorities in trials limits the external validity of the findings for this key patient demographic.

Objective: This systematic review examines (1) the research reporting practices and prevalence of ethnic minority patients included in telehealth randomized controlled trials (RCTs) targeting type 2 diabetes and the trial characteristics associated with recruiting a high proportion of minority patients, and (2) the proportion of included RCTs that report using English language proficiency as a patient screening criterion and how and why they do so.

Methods: Telehealth RCTs published in refereed journals targeting type 2 diabetes as a primary condition for adults in Western majority English-speaking countries were included. Ethnically targeted RCTs were excluded from the main review, but were included in a post hoc subgroup analysis. Abstract and full-text screening, risk of bias assessment, and data extraction were independently conducted by two reviewers.

Results: Of 3358 records identified in the search, 79 articles comprising 58 RCTs were included. Nearly two-thirds of the RCTs (38/58) reported on the ethnic composition of participants, with a median proportion of 23.5% patients (range 0%-97.7%). Fourteen studies (24%) that included at least 30% minority patients were all US-based, predominantly recruited from urban areas, and described the target population as underserved, financially deprived, or uninsured. Eight of these 14 studies (57%) offered intervention materials in a language other than English or employed bilingual staff. Half of all identified RCTs (29/58) included language proficiency as a participant-screening criterion. Language proficiency was operationalized using nonstandardized measures (eg, having sufficient "verbal fluency"), with only three studies providing reasons for excluding patients on language grounds.

Conclusions: There was considerable variability across studies in the inclusion of ethnic minority patients in RCTs, with higher participation rates in countries with legislation to mandate their inclusion (eg, United States) than in those without such legislation (eg, United Kingdom). Less than 25% of the RCTs recruited a sizeable proportion of ethnic minorities, which raises concerns about external validity. The lack of objective measures or common procedures for assessing language proficiency across trials implies that language-related eligibility decisions are often based on trial recruiters' impressionistic judgments, which could be subject to bias. The variability and inconsistent reporting on ethnicity and other socioeconomic factors in descriptions of research participants could be more specifically emphasized in trial reporting guidelines to promote best practice.

KEYWORDS

Introduction

Diabetes, a chronic metabolic condition, is on the rise, placing growing resource pressures on health care systems worldwide [1]. With the risk of developing diabetes increasing with obesity, sedentary behavior, and age, type 2 diabetes accounts for roughly 90% of diabetes cases. Behavioral changes (eg, diet, exercise), coupled with medication aiming to lower blood sugar levels and blood pressure, can reduce the risk of developing diabetes-related complications [2]. Due to its prevalence and the benefits of effective disease management, type 2 diabetes is frequently targeted in studies aiming to promote lifestyle modification through behavioral interventions and education [3].

Telehealth, broadly defined as remote health care delivery using technology [4], can partially alleviate the growing pressures associated with aging populations and rising rates of chronic conditions [5]. Technological advances have led to the availability of wide-ranging telehealth options to support patients as an alternative or supplement to traditional outpatient care, with presumed benefits including cost reduction, increasing convenience and access, and promoting patient self-management [6,7]. For example, diabetes-related telehealth services often involve platforms for measuring and communicating blood glucose information and receiving feedback from an automated system or remote professional. Other services involve structured self-management education or peer or motivational support [8]. Although findings from randomized controlled trials (RCTs) testing the effectiveness of telehealth interventions have been mixed and tend to vary by condition, systematic reviews focusing on type 2 diabetes specifically have generally yielded positive results, including modest but significant improvements for glycemic control and favorable outcomes for patient quality of life and treatment satisfaction [8-15].

Due to its flexibility, telehealth has the potential to reach underserved patients who may experience difficulties accessing traditional health services [16,17]. This extends to ethnic minorities, who are particularly vulnerable to developing type 2 diabetes [18], tending to do so at a younger age and with a lower body mass index than the general population in Western countries [19]. In the United Kingdom, for example, South Asians have up to six times higher prevalence of type 2 diabetes compared to people from Caucasian backgrounds, with incidence up to three times higher in people of African or African-Caribbean descent [20]. Ethnic minority groups also experience poorer long-term outcomes for diabetes, such as worse glycemic control and higher rates of complications [21], even when patients have access to health care at minimal cost

[22] and after adjusting for age and socioeconomic status [19,23]. Ethnic minorities may also experience impaired self-management and underuse services due to a combination of socioeducational factors, including a less developed understanding of diabetes, differing attitudes toward medication, culturally specific dietary practices, and/or language barriers to accessing care [13,24,25].

Despite having a higher incidence of diabetes, ethnic minority patients tend to be underrepresented in trials [26-28]. This means that they cannot receive the potential health benefits from trial participation [29,30], including access to improved treatments and closer monitoring during the trial period. In addition, studies are not able to adequately assess the effectiveness of new treatments for these higher-need groups. For example, telehealth technologies trialed on a disproportionately Caucasian sample may not be generalizable to a diverse patient demographic, with implications for the service-level adoption of such interventions [31,32]. Low participation rates of ethnic minority diabetes patients have been reported in systematic reviews focusing on telehealth interventions, although the ethnic composition of the recruited sample is often unreported. In a 2006 review, only eight of 26 included studies reported on the ethnic composition of trial participants, of which the median proportion of ethnic minority patients was 39% (range 5%-100%) [33]. Similarly, two 2014 reviews found that only half of 16 included studies [34] and four of nine included studies [3] reported on the ethnic makeup of the recruited sample, with variable minority participation across these studies (range 15%-100% and range 24%-100% of total participant population, respectively).

Although the results of these reviews [3,33,34] are revealing in terms of prevalence and research reporting practices, they have several limitations. First, the included studies were restricted to a narrow range of computer-based telehealth interventions—a subset of the wide variety of the telehealth technologies available. For example, they excluded trials that were phone-based or included glucose monitoring, which are pervasive in telehealth diabetes research [15]. Next, two reviews included studies that targeted both type 1 and type 2 diabetes [33,34]. Because ethnic minorities are not disproportionately affected by type 1 diabetes, it is important to identify the prevalence of minority participants in studies that exclusively focus on type 2 diabetes. Third, two of the reviews included trial designs other than RCTs [3,33], with different designs potentially affecting the type of patients who had opted to take part (eg, due to time commitments) [35,36] in addition to not providing the same design safeguards against bias [37]. Finally, both 2014 reviews included studies that specifically recruited from one or more ethnic minority communities [3,34]. The

inclusion of these ethnically targeted studies likely inflated ethnic minority patient prevalence estimates relative to studies recruiting from the general population. This systematic review addresses these limitations by focusing on telehealth RCTs targeting type 2 diabetes in adults recruited from the wider patient population. The telehealth medium used in the RCTs was not restricted in order to more comprehensively survey research reporting practices across the range of technologies. A central aim of this review was to explore the barriers and facilitators to ethnic minority inclusion in telehealth RCTs.

Among the factors that could affect ethnic minority participation in RCTs, language and literacy are often identified as potential obstacles [26,31]. Of foreign-born people living in the United States, 29% reported speaking English "not well" or "not at all" [38] compared to more than 12% in Australia [39]. In England and Wales, nearly 19% of adults from the four largest ethnic minority communities were estimated to speak little or no English [40]. Poor language skills are also related to higher levels of undiagnosed diabetes and to difficulties accessing services [13,41]. Ensuring that patients have the requisite language ability to understand the conditions for trial participation is an ethical imperative in all research (eg, obtaining informed consent). Because communication is a key part of the treatment in telehealth trials, language could act as a barrier to telehealth's ability to provide more accessible, equitable modalities for delivering care. For example, telehealth interventions often necessitate basic literacy skills (eg, understanding and inputting written text) or enhanced communication skills (eg, communicating on the phone with no access to nonverbal cues), potentially barring the participation of patients without adequate language skills to engage with the intervention in the absence of translation or interpretation services [42].

There is a pressing need to examine the role of language in telehealth interventions, especially in countries where a sizeable portion of the population has limited ability in the official language. This is particularly the case for conditions such as type 2 diabetes, to which ethnic minority communities are particularly vulnerable. Yet little is known about whether and how patients are screened for language proficiency or literacy nor the extent to which these factors are cited among the participant inclusion or exclusion criteria in telehealth RCTs targeting type 2 diabetes [30,43]. In light of these gaps, the goals of this systematic review were to investigate (1) the research reporting practices and prevalence of ethnic minority patients included in telehealth RCTs targeting type 2 diabetes and trial characteristics associated with successful minority patient recruitment, and (2) the proportion of included RCTs that report English language proficiency as a participant-screening criterion and how and why proficiency was assessed.

Methods

This review followed the Cochrane Collaboration's handbook on conducting systematic reviews [44] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [45]. Full methodological details,

including MEDLINE search strategy terms, are reported in the published protocol [46].

Search Strategy and Study Screening and Selection

The search, which was conducted in late August 2015, included studies published from January 1, 2000 to July 31, 2015 using MEDLINE, PsycINFO, EMBASE, CINAHL, and CENTRAL. Keywords and inclusion criteria from recent related reviews were examined [15,34,47,48] and a medical librarian was consulted to verify the search strategy (eg, keywords, choice of databases). Abstract and full-text screening were independently performed by two reviewers (LE, KB, or DW), with discrepancies resolved through discussion. Multiple outputs from the same dataset were linked for included studies and imported into Endnote X7.

Included studies were peer-reviewed English-language journal articles on telehealth RCTs recruiting adult (≥18 years) type 2 diabetes patients from Western countries where English is both an official and the majority language (ie, Australia, Canada, Ireland, New Zealand, United Kingdom, United States). All other research designs and publication types were excluded, as were RCTs that included type 1 or gestational diabetes patients, or that had explicitly targeted one or more ethnic minority groups in their recruitment strategy. Interventions could comprise any telehealth medium designed to treat or improve type 2 diabetes as the primary condition. Studies focusing on secondary diabetes-related complications (eg, retinopathy) or mental health were excluded as were telehealth interventions solely targeting health professionals rather than patients.

Data Collection

The data extraction form Multimedia Appendix 1 was developed using Cochrane guidelines [44] to describe study details, participant demographics, and intervention characteristics. Outcome data were extracted using Microsoft Excel and independently checked by two reviewers (DH, DW). The Cochrane Collaboration's risk of bias (ROB) tool [49] was adapted to assess all 79 included articles using Cochrane's Review Manager software (RevMan, The Cochrane Collaboration, Copenhagen, Denmark). Two of three authors (LE, DH, DW) independently evaluated each article for low, unclear, or high ROB, with discrepancies resolved through discussion. The assessment criteria were random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other sources of bias (other bias).

Due to the nature of telehealth interventions, patient and personnel blinding is generally unfeasible. Hence, for the purpose of this review, high risk of performance bias was interpreted as situations where unblinded research personnel interacted with participants across study groups, allowing for differential treatment. Assessment of detection bias focused on primary outcome detection. Studies that used an objective primary outcome measure (eg, laboratory-based blood test, administrative record of number of hospital visits) were assessed as low ROB because knowledge of participants' allocation is

unlikely to seriously affect the outcome. Conversely, studies where the primary outcome was subjectively assessed (eg, self-report measures) were deemed high ROB.

Data Analysis

Data analysis of included studies addressed the following primary outcomes:

1. Proportion of studies that report on the ethnic composition of recruited participants and, where available, the overall prevalence of ethnic minority patients (between-study median and range);

2. Characteristics of studies that recruit a high proportion (≥30%) of ethnic minority participants (eg, telehealth medium, access to translation);

3. Proportion of studies that include English language proficiency or reading and writing literacy as a participant-screening criterion and, where available, the ways in which proficiency/print literacy is operationalized as a screening criterion; and

4. Language-related reasons for patient exclusion, if given (eg, informed consent, lack of resources).

In line with categorizations of race and ethnicity in the United States census [50] and conceptualizations of nonwhite or non-Caucasian respondents in other majority English-speaking countries (eg, United Kingdom [51]), ethnic minorities were defined as those of nonwhite ethnicity, including Hispanics, who may or may not be newcomers to the host country. In the case of studies reporting only the proportion of white patients recruited, all other participants were assumed to be ethnic minorities. In other studies that listed the proportion of participants belonging to an "other" group, "other" was interpreted as patients from an ethnic minority background not specified elsewhere in the study.

A narrative synthesis [52] was conducted to examine the characteristics of RCTs that reported a 30% or greater threshold of ethnic minority participants as a proportion of the total sample (considered relatively high), which is in line with the median prevalence of minority recruitment reported in earlier related reviews [3,33,34]. We would also note that this threshold is close to the proportion of ethnic minorities in the United States population. In the 2015 census, the "white alone, not Hispanic or Latino" category was reported at 61.6%, which implies that the remaining 38.4% are ethnic minorities [53]. After systematically extracting and tabulating the data, groupings and textual descriptions were used to explore heterogeneity within those studies and between studies with higher and lower proportions of ethnic minority participants. Exploration of relationships in the data was iteratively conducted to reveal factors that may promote or impede ethnic minority recruitment.

Finally, a post hoc subgroup analysis was conducted for studies that had explicitly targeted one or more ethnic minority groups as part of their recruitment strategy and that had been excluded from the systematic review for that reason only [54-65]. This additional retrospective analysis of ethnically targeted studies was undertaken to further examine recruitment strategies and features of trials that specifically recruited ethnic minority participants. This analysis involved the same methods as the synthesis described previously, but did not contribute to calculations reported in the main analysis.

Results

Study Selection

The search yielded 2332 records after removing duplicates, which were submitted to abstract screening. After assessing 212 full-length articles for eligibility, 79 articles, consisting of 58 discrete RCTs, met the inclusion criteria (see Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram summarizing the process of selecting eligible studies for the systematic review. RCT: randomized controlled trial, T2DM: type 2 diabetes.
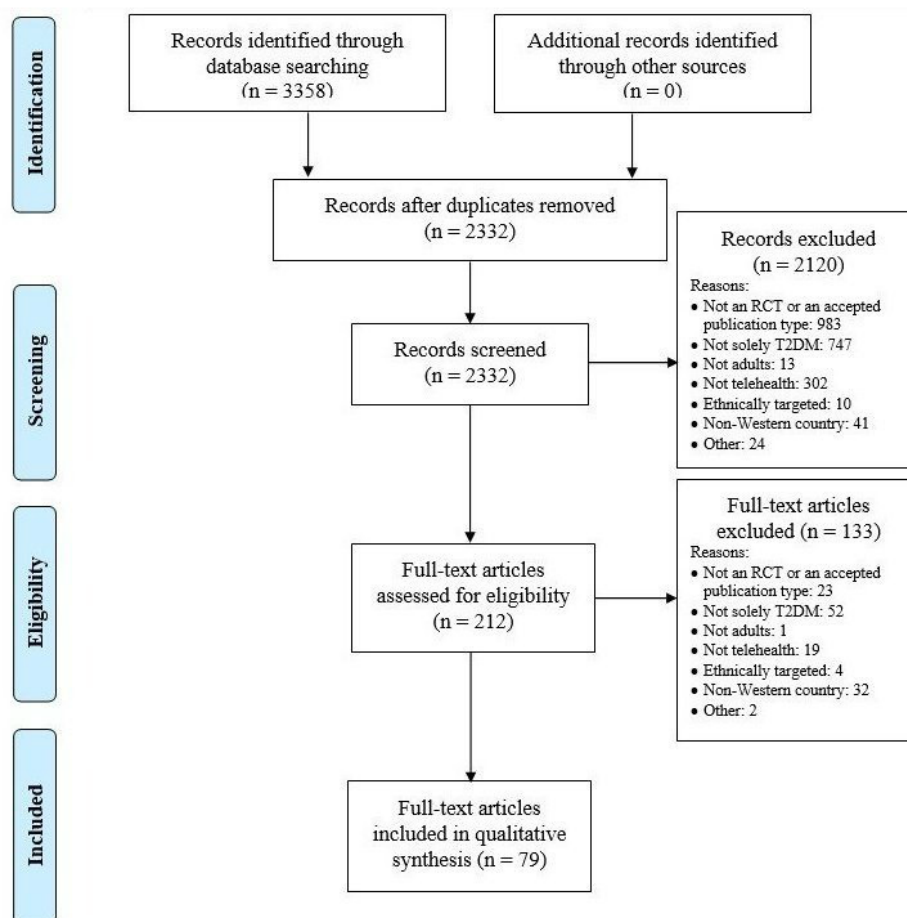
Characteristics of Included Studies

Multimedia Appendix 2 summarizes the characteristics of included studies in the review. The total number of participants across included RCTs was 12,916, with sample sizes ranging from 14 to 1665 (median 160). Nearly three-quarters (43/58) were recruited from the United States, whereas the rest were carried out in Australia (5/58), Canada (5/58), the United Kingdom (4/58), and one in both the United States and Canada. The studies involved a wide range of telehealth media and devices, the most popular of which was phone-based telehealth interventions. Other frequently used telehealth tools included Internet technologies (eg, static and interactive webpages, email, instant messaging), computerized self-management programs, and glucose meters integrated with mobile apps. Several studies combined different media and communication types as part of the intervention, including electronic medical records, educational websites, home monitoring, and videoconferencing (eg, [66]).

Methodological Quality

Parts A and B of Multimedia Appendix 3 show the ROB summary table and graph for included studies. Of the 79 articles, 34 described an appropriate randomization procedure, two [67,68] reported inadequate randomization, and the remainder provided insufficient detail for assessing risk of selection bias (judged unclear). Similarly, allocation concealment reporting

was frequently insufficient to determine the ROB (eg, not clear whether allocation envelopes were opaque and sequentially numbered [69]). Most studies reported designs intended to reduce performance bias (within the parameters of unfeasible patient and interventionist blinding), with studies judged as having a high ROB in cases when nonblinded staff delivered both the telehealth intervention and usual care [69-83] and also collected follow-up measures [84,85]. Included studies employed both subjective and objective outcome measures, with objective glycated hemoglobin (HbA_{1c}) levels being the most common primary outcome. Approximately half of the studies included patient attrition information, with unclear ROB assigned when reasons for attrition were not given, attrition was not broken down by group allocation, or it was unclear how substantial missing data were dealt with in data analysis. Several studies were assessed as having a high ROB due to a large difference in attrition between study arms that was attributed to the intervention [86,87] or that had only presented outcomes for patients who had completed the intervention and/or all follow-up measures [88-91]. There was generally a low risk of reporting bias, with most studies reporting on prespecified outcomes. A high ROB was noted in studies where reported outcomes deviated from the protocol or specified methodologies [70,85,91-97] or insufficiently reported statistical or summary data [67,68,71,72,83,98-105]. Most studies were free of other sources of bias. However, some failed to report treatment dosage (eg, frequency of calls [74,90,106,107]) and its effect on

outcomes even when adherence was low. Others were pilot studies or had small convenience samples, limiting their generalizability [71,85,95,101,108,109].

Ethnic Minority Participation

Thirty-eight of the 58 included RCTs (or 56 of the 79 articles) provided information on the ethnic composition of the sample [66-78,80,85,87,89,91-98,100-130], of which the median proportion of ethnic minority participants was 23.5% (range

0%-97.7%). Two of these recruited no ethnic minority patients [71,77]. The remaining studies (n = 20) provided no ethnicity information, including all five Canadian studies [79,81-84,86,88,90,99,131-144]. Figure 2 shows that the number of included studies in the review markedly increased after 2005, with eight up until that year and 50 thereafter. The proportion of the studies that reported on the ethnicity of recruited patients was 38% (3/8) up to and including 2005 and more than doubled to 70% (35/50) after that date.

Figure 2. Number of included studies (n=58) reporting on the ethnic composition of the recruited sample by year of publication.
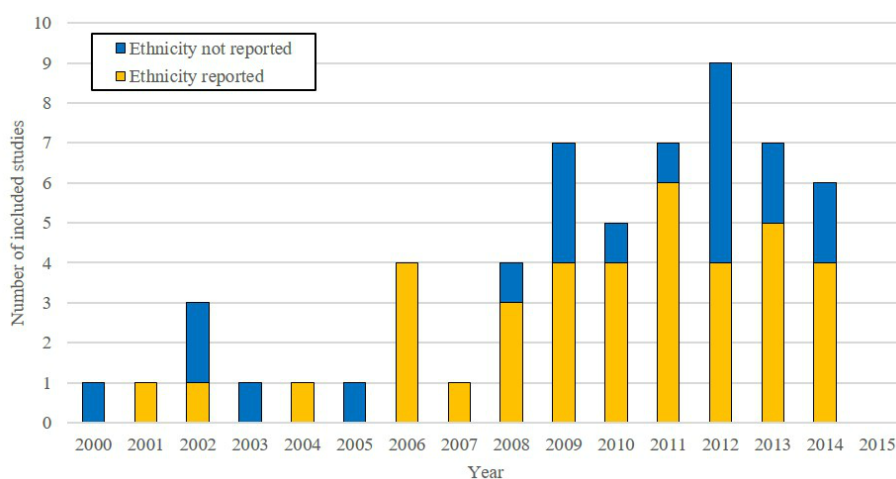
Language Proficiency

Half of the included RCTs (29/58) reported English language proficiency as a patient inclusion or exclusion criterion, with six of these studies alternatively requiring proficiency in Spanish [66,94,105,106,110,111] and one in either Spanish or Cantonese instead of English [118]. In the 29 remaining RCTs, language ability may have been considered in recruitment but not reported in the published article, including one study that did not list any screening criteria at all [72]. Alternatively, language might not have been taken into account in recruitment. Although being able to engage with the intervention may be an implicit reason for including language as an eligibility criterion, only three studies provided explicit explanations for excluding prospective participants on language grounds. In two, this pertained to understanding study information and providing informed consent [99,109]. In the third, this related to language demands required for the intervention, which involved patients receiving tailored feedback through an automated interactive phone service [142].

Of the studies that included language proficiency as an eligibility criterion, there was little consistency in the way that this was defined. More than a third (11/29) emphasized being able to communicate in or fluently speak (and in two cases also understand) English, whereas another specified language without reference to the written medium. Of these, two studies further specified that the context for this was over the phone [91,142], which is more difficult than face-to-face communication [42]. Four other studies referred to participants needing to be able to read and speak English, seven required reading and writing, and two referred to reading and understanding (ie, receptive skills), placing no apparent emphasis on speaking or writing. Finally, five studies emphasized having English (or Spanish) as a main

or primary language, implying that membership to the target language community (ie, native speaker status) was the key criterion.

From these descriptions, there were no indications that any objective measures (eg, temporal measures of fluency or oral comprehension questions) were used to establish whether prospective participants had the necessary skills to meet the specified language criterion for inclusion in the study. Reference to commonly used benchmarks of language proficiency or defined levels (eg, Canadian Language Benchmarks, Common European Framework of Reference for Languages) were not given [145]. The grounds on which a patient was determined to be linguistically eligible to participate were also unspecified. Two studies administered previously validated health literacy instruments to patients [75,118]. However, this was used to assess study outcomes in relation to health literacy and there was no threshold health literacy level required for participation. In sum, there were no explicit or standardized measures of language proficiency across included studies—a variable which could directly affect the ethnic minority patient participation [146].

Narrative Synthesis

Of the 79 articles reporting on 14 distinct RCTs, 28 recruited a high proportion of ethnic minority patients, with the threshold for this set at 30% or greater (median 53.3%, range 30.0%-97.7%; see Table 1) [66,70,75,85, 87,94,103,105,106, 110,111,116, 118,119]. They all took place in the United States, mostly in urban settings, with only one study exclusively recruiting from a rural setting [70]. These studies frequently recruited in medically underserved and financially deprived areas [66,70,106,110,111] and described their target populations

as predominantly uninsured [75,106,111,118]. In addition to including ethnic minorities, these studies also had high numbers of patients with no insurance or who received government-sponsored Medicare or Medicaid benefits. For example, of the three studies recruiting patients from safety-net clinics, which treat uninsured patients, more than 89% were nonwhite [75,110,118]. Two studies required health care coverage as a patient screening criterion [87,103] and one solely recruited low paid health care worker union members insured through their employer [105].

Table 1. Summary of characteristics of 14 parent studies with a high proportion (≥30%) of ethnic minority patients.

Study	Intervention	Setting and patient characteristics[a]			Benefits of participation[b]		Tailoring of intervention[c]		
	Telehealth medium	Recruitment	Urban only	Underserved	Financial	Device	Bilingual	Cultural	Literacy
Anderson et al [106]	Phone	PC	Yes	Yes	Yes		Yes		Yes
Arora et al [110]	Mobile SMS	SN	Yes	Yes	Yes		Yes	Yes	Yes
Davis et al [70]	Videoconference & pedometer	PC	Rural	Yes	Yes			Yes	Yes
Frosch et al [111]	Video & phone	PC Com	Yes	Yes			Yes		
Glasgow et al [94]	Website & phone (automated & from an interventionist)	PC	Yes				Yes	Yes	
Khan et al [75]	Computer multimedia education	SN	Yes	Yes			Yes	Yes	Yes
Krein et al [116]	Phone	PC	Yes						
Quinn et al [85]	Bluetooth blood glucose meter with mobile app	PC Com	Mixed			Yes			
Quinn et al [87]	Mobile SMS, patient portal & phone	PC	Mixed			Yes			
Schillinger et al [118]	Phone	SN	Yes	Yes	Yes		Yes		Yes
Sevick et al [119]	Personal digital assistant	Com	Yes		Yes	Yes			
Shea et al [66]	Website, videoconference, electronic records & email	PC	Mixed	Yes		Yes	Yes		Yes
Tang et al [103]	Bluetooth blood glucose meter & electronic records	PC	Yes			Yes			
Walker et al [105]	Phone	Union	Yes	Yes			Yes		

[a] Recruitment: Description of setting from which participants were recruited; PC: primary care; SN: safety net; Com: community center; Union: health care workers' union. Urban only: refers to recruitment solely from urban areas. All mixed studies recruited from both urban and rural areas, with the exception of [85], which recruited from urban and suburban areas. Underserved: medically underserved/uninsured/low-wage patients or Medicare/Medicaid beneficiaries.

[b] Financial: monetary incentives for research participation, including money, gift cards, or vouchers. Device: provision of technology or equipment including self-monitoring devices used in the intervention.

[c] Bilingual: recruitment or educational materials translated into a minority language or availability of bilingual interventionists. Cultural: materials crafted for or through consultation with members from certain minority communities independent of language. Literacy: materials purposefully written at a particular grade level or staff trained to communicate with low-literacy patients.

High minority-recruiting studies used a mix of telehealth delivery modes, reflecting the breadth of technology used in the wider sample of studies included in the review. This suggests that the medium through which a telehealth intervention is delivered has little bearing on ethnic minority recruitment. In total, 36% (5/14) of the studies offered financial compensation for trial participation, with money or vouchers ranging from US $15 to US $175 per participant [70,106,110,118,119]. By comparison, only 21% (5/24) of low-recruiting studies (<30% ethnic minorities) offered financial incentives. A proportionately higher number of high minority-recruiting trials (35.7% vs 21% in low-recruiting studies) also offered patients equipment to facilitate self-monitoring, such as glucose meters and/or blood testing strips, which can impose a substantial cost to patients

not covered by insurance [70,106,110,118,119], or free mobile phones if they were being used as part of the intervention [85,87,103].

Another key characteristic of high minority-recruiting studies was an emphasis on languages other than English. More than half (8/14) offered the intervention in a language other than English, including the use of bilingual staff for phone interventions [66,75,94,105,106,110,111,118], and two additionally reported using bilingual recruitment staff or translated study information to facilitate the recruitment of nonnative-English-speaking patients [94,118]. Conversely, no study with less than 30% ethnic minority participants offered bilingual interventions (see Multimedia Appendix 2). Four studies reported tailoring intervention materials to the needs and interests of African American and/or Latino patients, such as the use of video testimonials from community members [70,75,94,110], and nearly half reported writing study materials to facilitate low-literacy patients' comprehension [66,70,75,106,110,118].

Post Hoc Subgroup Analysis

A post hoc subgroup analysis was conducted for 12 articles comprising 11 RCTs that had been excluded during full-text screening [54-65]. The findings echoed those of the narrative synthesis (see Table 2). All ethnically targeted studies took place in urban American settings, with nearly half (5/11) reportedly recruiting from economically deprived or medically underserved areas [56,57,60,61,65]. More than half (6/11) used phone calls as part of the intervention, although other telehealth media were also used. Although only one study offered patients intervention-related equipment (laptop and telehealth peripherals) [57], six offered financial remuneration ranging from US $40 to US $60 cash or vouchers [55,59,61-63,65], with one trial additionally offering some patients up to US $200 for reducing their HbA_{1c} levels by a prespecified amount [62]. The ethnically targeted studies frequently reported tailoring interventions to their target demographic, including offering written and spoken aspects of the intervention in a language other than English [56,61,64,65], designing interventions through consultation and feedback from minority groups [55-58,60,65], and using interventionists drawn from the target cultural communities [55,61,64,65]. This latter point included building input from minorities into the recruitment strategy and/or piloting materials and procedures with them.

Table 2. Summary of characteristics of 11 ethnically targeted studies subjected to a post hoc subgroup analysis.

| Study | Intervention | Setting and patient characteristics[a] | | Incentives[b] | | Tailoring of intervention[c] | |
	Telehealth medium	Recruit-ment	Recruited group(s)	Underserved	Financial	Bilingual	Cultural	Literacy
Amoako & Skelly [55]	Phone	PC com	Older African American women		Yes		Yes	Yes
Calderón et al [56]	Video	PC	Spanish speakers	Yes		Yes	Yes	Yes
Carter et al [57]	Laptop with home monitoring, videocon-ference & electronic records	PC	African Americans	Yes			Yes	
Crowley et al [58]	Phone	PC	African Americans				Yes	Yes
Forjuoh et al [59]; Adepoju et al [54][d]	Personal digital assis-tant	PC	African Americans & Hispanics		Yes			
Gary et al [60]	Phone	PC	African Americans	Yes			Yes	
Heisler et al [61]	Tablet computer	Com	African Americans & Hispanics	Yes	Yes	Yes	Yes	Yes
Long et al [62]	Phone	PC	African American veterans		Yes			
Lorig et al [64]	Phone	Com	Spanish speakers			Yes	Yes	
Lorig et al [63]	Website	Web	American Indians/ Alaska Natives		Yes			
Ruggiero et al [65]	Phone	PC	African Americans & Hispanics	Yes	Yes	Yes	Yes	Yes

[a] Recruitment: Description of setting from which participants were recruited; PC: primary care; com: community center; Web: website. Recruited group(s): ethnicities targeted in recruitment and age group, gender, or profession of targeted participants where specified. Underserved: medically underserved/uninsured/low-wage patients or Medicare/Medicaid beneficiaries.

[b] Financial: monetary incentives for research participation, including money or vouchers. Bilingual: recruitment or educational materials translated into a minority language or availability of bilingual interventionists.

[c] Cultural: materials crafted for or through consultation with members from certain minority communities independent of language, or staff undertook cultural sensitivity training. Literacy: materials purposefully written at a particular grade level or staff trained to communicate with low-literacy patients.

[d] These linked articles reported on the same RCT and it was unclear which was the parent study.

Discussion

Principal Findings

This systematic review investigated research reporting practices and prevalence estimates of ethnic minority participation in telehealth diabetes RCTs, extending previous reviews by including a broader range of telehealth technologies targeting type 2 diabetes. Nearly 66% of included studies reported on the ethnic composition of their samples. Although this proportion is higher than in previous reviews [3,33,34], it confirms the underreporting of ethnicity in peer-reviewed journal articles. However, compared to the 2006 review that had also excluded studies recruiting participants from a single ethnic minority background [33], these results yielded lower median participation of ethnic minority patients as a proportion of the total sample (23.5% vs 39%).

All RCTs with 30% or greater ethnic minority recruitment were US-based, mostly recruited from urban areas, and frequently described recruited patients as low-income, socially deprived, or with little or no health care coverage. Participant remuneration or free telehealth monitoring devices incentivized participation in nearly 60% of these studies. Although the United States has a high proportion of ethnic minorities compared to other countries included in the review [147], its dominance as the setting for all studies meeting the 30% or greater cut-off for the narrative review and for all ethnically targeted studies in the post hoc subgroup analysis suggests that the American National Institutes of Health (NIH) *Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research* [148] has been influential in promoting minority representation in trials. Overall, minority participation rates tend to be higher in the United States, which has legislation to mandate their inclusion, than in contexts where no such legislation exists (eg, United Kingdom) [26]. The availability of trial materials in multiple languages was also a recurrent feature of studies recruiting substantial numbers of ethnic minority participants and often occurred in conjunction with trial materials that accounted for cultural factors or presumed literacy levels. This suggests that having a language concordant

interventionist and embedding input from members of target communities into trial materials could have positive effects on the recruitment of groups who may otherwise be underrepresented [149,150]. This supports existing research on strategies to optimize the recruitment of underserved patients, which reveals ways of working holistically to alleviate individual and external factors that impede participation [146,151]. The emphasis on patient and public involvement in research aligns with such practices [29,152].

Half of the included studies listed language proficiency among the patient eligibility criteria, although less than 5% provided a reason for including or excluding patients on this basis. The role of language in participant screening was described in different ways across studies, with emphasis either placed on different combinations of skills (speaking, listening, reading, writing) to reflect the nature of the intervention, or on patients' status as primary speakers of English or another language where offered. There was no evidence of the use of objective measures or instruments or of a common procedure across studies to assess whether patients had the requisite language proficiency to participate and, in individual studies, this level of detail was not given. Thus, it was unclear how language-related determinations about inclusion or exclusion were made—that is, how the language-screening criterion as stated was operationalized in arriving at eligibility decisions.

Limitations

This review has several limitations. First, only English-language articles that recruited patients from Western countries where English is both an official and the dominant language were considered. Findings may be different in reviews focusing on recruitment from other contexts or on other world languages. Second, ethnic minorities are not a monolith, and language barriers to engaging with the intervention are likely to be different for newcomers to a country (eg, migrants), who may have little knowledge of English, than for later generations of native English speakers who are visible minorities (eg, African Americans in the United States) [24]. It is often difficult to disentangle ethnicity from language in secondary data analysis because the language background of participants (ie, proportion of native English speakers and heritage language speakers) is often not reported as a separate category and, in some countries, is not captured in census data [40]. Third, and related to this, we used the ethnicity categories reported in the original studies in data extraction. However, race and ethnicity are complex constructs that frequently conflate social identity with other factors, such as genetic or biological characteristics (particularly as connoted in the former term), geographical origin, cultural practices, or religious persuasion [31]. Categorizations were inconsistent across studies included in this review, making comparisons between and within countries difficult. Fourth, the prevalence of minority patients in the area(s) from which they were recruited was not possible to collate, with several studies recruiting from different geographic sites and not reporting on local population or, in some cases, on the ethnic breakdown of their recruited sample. Finally, peer-reviewed articles were the sole publication type considered in this study, which was consistent with the goal of elucidating research reporting practices in academic journals. That is, the grey literature was not examined and study authors were not contacted to provide further information than was included in their published articles, taking into account linked publications. The brief descriptions of study setting, design, and recruitment in the included telehealth RCTs provided comparatively little qualitative data for the narrative synthesis. This has resulted in a relatively simple theoretical account of the trial features supporting ethnic minority participation, which would need to be corroborated through a more in-depth examination of key variables in subsequent research. Nonetheless, with 79 articles and 58 RCTs included, this review is comprehensive in its account of research reporting and includes more studies than the three previous telehealth diabetes reviews combined [3,33,34].

Concluding Remarks

Despite the link between new technologies and improved outcomes, mixed evidence regarding reducing disparities from the research literature [122,129] suggests that telehealth has yet to fulfill its potential of being truly accessible to and effective for ethnic minority diabetes patients [15]. Cultural and linguistic tailoring to a diverse demographic and offering translation and interpretation services where possible could extend the benefits of telehealth type 2 diabetes interventions to a wider cross-section of patients, thereby promoting more equitable access to health care [43,153].

Findings from this and earlier systematic reviews suggest that between a third and half of telehealth diabetes trials provide no information on the ethnic composition of their samples. Further, other demographic characteristics, such as socioeconomic status, are not consistently reported across studies (see Multimedia Appendix 2). Providing data on participants' gender and age, but no information on ethnicity for either the sample or the local target population is insufficient for assessing the external validity of the findings. Ethnicity information is necessary to evaluate claims about telehealth's accessibility to all patients (and not just a subsection of the population), including its ability to foster social inclusion through the uptake of services.

The Consolidated Standards of Reporting Trials (CONSORT) statement [154], which articulates guidelines for best practice in research reporting and which has been widely adopted, advocates reporting baseline demographic and clinical characteristics for each group in the RCT. However, this is prefaced on the assumption that these will be collected, and the statement does not specify which baseline variables should be captured and described. In the case of clinical baseline variables, further specification of which variables to target would appear to be dependent on the nature of the condition being examined (eg, HbA_{1c} for diabetes). However, this is not the case for sociodemographic variables, which apply regardless of the disease type being investigated. In trials targeting chronic conditions in which prevalence is known to vary by ethnicity and other social factors, further specification of the baseline demographic characteristics that should be reported on could reduce the variability in between-study reporting. This would assist in examining crucial relationships between sociodemographic variables and health outcomes. Journal editors and editorial boards could converge on the most important sociodemographic characteristics that should be reported (eg,

in a checklist) to achieve greater consistency and help streamline such information across studies.

Finally, the fact that telehealth trials seek to recruit patients with an adequate level of language ability is understandable—telehealth interventions rely on effective spoken or written communication as a key part of the "treatment." However, the absence of an objective or standardized measure for assessing whether patients have the requisite language ability to successfully engage with the intervention suggests that such decisions are likely being made on the basis of trial recruiters' subjective judgments. This creates a risk of selection bias, as trial recruiters might exclude ethnic minority participants based

on the subjective view that they will not be able to adequately take part in the intervention, despite actually having sufficient language proficiency [43]. For example, recruiters may misjudge having a perceptible foreign accent as evidence of poor language ability in instances when this does not actually impede communication [155,156]. Conversely, including patients who do not understand the nature of the intervention due to language barriers is problematic, not least for ethical reasons [157]. Future research could focus on the development and validation of a tool to provide trial recruiters with a simple, practical means of assessing language proficiency for trial participation to minimize the possibility of patients being unfairly excluded based on arbitrary judgments.

Acknowledgments

Authors' Contributions

The study was initially conceptualized by TI and LE; carried out by LE, DH, DW, and KB (abstract and full-text screening, ROB assessment, data extraction); with the review jointly drafted by TI and DH, drawing on the published protocol. All authors contributed to and approved this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data extraction form.

[PDF File (Adobe PDF File), 35KB - jmir_v18i9e256_app1.pdf]

Multimedia Appendix 2

Summary of included studies in the review.

[PDF File (Adobe PDF File), 67KB - jmir_v18i9e256_app2.pdf]

Multimedia Appendix 3

A. Risk of bias assessment summary. B. Risk of bias assessment graph.

[PDF File (Adobe PDF File), 104KB - jmir_v18i9e256_app3.pdf]

References

1. World Health Organization. Global Report on Diabetes. Geneva: World Health Organization; 2016. URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [accessed 2016-09-11] [WebCite Cache ID 6kREZjPza]
2. Diabetes UK. Facts and Stats 2015. 2015 Nov. URL: https://www.diabetes.org.uk/Documents/Position%20statements/Diabetes%20UK%20Facts%20and%20Stats_Dec%202015.pdf [accessed 2016-04-08] [WebCite Cache ID 6gcHW2TBL]
3. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
4. Institute of Medicine Committee on Evaluating Clinical Applications of Telemedicine. In: Field MJ, editor. Telemedicine: A Guide to Assessing Telecommunications in Health Care. Washington, DC: National Academies Press; 1996.

5. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]

6. Ackerman MJ, Filart R, Burgess LP, Lee I, Poropatich RK. Developing next-generation telehealth tools and technologies: patients, systems, and data perspectives. Telemed J E Health 2010;16(1):93-95 [FREE Full text] [doi: 10.1089/tmj.2009.0153] [Medline: 20043711]

7. Greenwood DA, Hankins AI, Parise CA, Spier V, Olveda J, Buss KA. A comparison of in-person, telephone, and secure messaging for type 2 diabetes self-management support. Diabetes Educ 2014 Apr 17;40(4):516-525. [doi: 10.1177/0145721714531337] [Medline: 24742540]

8. Bashshur RL, Shannon GW, Smith BR, Woodward MA. The empirical evidence for the telemedicine intervention in diabetes management. Telemed J E Health 2015 May;21(5):321-354 [FREE Full text] [doi: 10.1089/tmj.2015.0029] [Medline: 25806910]

9. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]

10. Worswick J, Wayne SC, Bennett R, Fiander M, Mayhew A, Weir MC, et al. Improving quality of care for persons with diabetes: an overview of systematic reviews - what does the evidence tell us? Syst Rev 2013;2:26 [FREE Full text] [doi: 10.1186/2046-4053-2-26] [Medline: 23647654]

11. Toma T, Athanasiou T, Harling L, Darzi A, Ashrafian H. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]

12. Riazi H, Larijani B, Langarizadeh M, Shahmoradi L. Managing diabetes mellitus using information technology: a systematic review. J Diabetes Metab Disord 2015;14:49 [FREE Full text] [doi: 10.1186/s40200-015-0174-x] [Medline: 26075190]

13. Wilson C, Alam R, Latif S, Knighting K, Williamson S, Beaver K. Patient access to healthcare services and optimisation of self-management for ethnic minority populations living with diabetes: a systematic review. Health Soc Care Community 2012 Jan;20(1):1-19. [doi: 10.1111/j.1365-2524.2011.01017.x] [Medline: 21749529]

14. Long AF, Gambling T, Young RJ, Taylor J, Mason JM. Acceptability and satisfaction with a telecarer approach to the management of type 2 diabetes. Diabetes Care 2005 Feb;28(2):283-289. [Medline: 15677780]

15. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]

16. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]

17. Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. J Med Internet Res 2005;7(5):e50 [FREE Full text] [doi: 10.2196/jmir.7.5.e50] [Medline: 16403714]

18. Khunti K, Morris DH, Weston CL, Gray LJ, Webb DR, Davies MJ. Joint prevalence of diabetes, impaired glucose regulation, cardiovascular disease risk and chronic kidney disease in South Asians and White Europeans. PLoS One 2013;8(1):e55580 [FREE Full text] [doi: 10.1371/journal.pone.0055580] [Medline: 23383233]

19. Negandhi PH, Ghouri N, Colhoun HM, Fischbacher CM, Lindsay RS, McKnight JA, Scottish Diabetes Research Network Epidemiology Group. Ethnic differences in glycaemic control in people with type 2 diabetes mellitus living in Scotland. PLoS One 2013;8(12):e83292 [FREE Full text] [doi: 10.1371/journal.pone.0083292] [Medline: 24358273]

20. Nazroo J. The Health of Britain's Ethnic Minorities: Findings from a National Survey. London: Policy Studies Institute; 1997.

21. Sinha SK, Shaheen M, Rajavashisth TB, Pan D, Norris KC, Nicholas SB. Association of race/ethnicity, inflammation, and albuminuria in patients with diabetes and early chronic kidney disease. Diabetes Care 2014 Apr;37(4):1060-1068 [FREE Full text] [doi: 10.2337/dc13-0013] [Medline: 24550221]

22. Rawshani A, Svensson AM, Rosengren A, Zethelius B, Eliasson B, Gudbjörnsdottir S. Impact of ethnicity on progress of glycaemic control in 131,935 newly diagnosed patients with type 2 diabetes: a nationwide observational study from the Swedish National Diabetes Register. BMJ Open 2015;5(6):e007599 [FREE Full text] [doi: 10.1136/bmjopen-2015-007599] [Medline: 26048210]

23. Abouzeid M, Philpot B, Janus ED, Coates MJ, Dunbar JA. Type 2 diabetes prevalence varies by socio-economic status within and between migrant groups: analysis and implications for Australia. BMC Public Health 2013;13:252 [FREE Full text] [doi: 10.1186/1471-2458-13-252] [Medline: 23517376]

24. Majeed-Ariss R, Jackson C, Knapp P, Cheater FM. A systematic review of research into black and ethnic minority patients' views on self-management of type 2 diabetes. Health Expect 2015 Oct;18(5):625-642. [doi: 10.1111/hex.12080] [Medline: 23710892]

25. Szczepura A. Access to health care for ethnic minority populations. Postgrad Med J 2005 Mar;81(953):141-147 [FREE Full text] [doi: 10.1136/pgmj.2004.026237] [Medline: 15749788]

26. Brown G, Marshall M, Bower P, Woodham A, Waheed W. Barriers to recruiting ethnic minorities to mental health research: a systematic review. Int J Methods Psychiatr Res 2014 Mar;23(1):36-48. [doi: 10.1002/mpr.1434] [Medline: 24474683]

27. Newlin K, Melkus GD, Jefferson V, Langerman S, Womack J, Chyun D. Recruitment of black women with type 2 diabetes into a self-management intervention trial. Ethn Dis 2006;16(4):956-962. [Medline: 17061753]

28. Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242 [FREE Full text] [doi: 10.1002/cncr.23157] [Medline: 18008363]

29. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]

30. Meuter RF, Gallois C, Segalowitz NS, Ryder AG, Hocking J. Overcoming language barriers in healthcare: A protocol for investigating safe and effective communication when patients or clinicians use a second language. BMC Health Serv Res 2015;15:371 [FREE Full text] [doi: 10.1186/s12913-015-1024-8] [Medline: 26357948]

31. Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, et al. The causes and effects of socio-demographic exclusions from clinical trials. Health Technol Assess 2005 Oct;9(38):iii-iv, ix [FREE Full text] [Medline: 16181564]

32. Foster A, Horspool KA, Edwards L, Thomas CL, Salisbury C, Montgomery AA, et al. Who does not participate in telehealth trials and why? A cross-sectional survey. Trials 2015;16:258 [FREE Full text] [doi: 10.1186/s13063-015-0773-3] [Medline: 26044763]

33. Jackson CL, Bolen S, Brancati FL, Batts-Turner ML, Gary TL. A systematic review of interactive computer-assisted technology in diabetes care. Interactive information technology in diabetes care. J Gen Intern Med 2006 Feb;21(2):105-110 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00310.x] [Medline: 16390512]

34. 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]

35. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]

36. Brueton VC, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev 2013 Dec;12:MR000032 [FREE Full text] [doi: 10.1002/14651858.MR000032.pub2] [Medline: 24297482]

37. Jadad A, Enkin M. Randomised Controlled Trials: Questions, Answers and Musings, 2nd Edition. Malden, MA: BMJ Books; 2007.

38. Gambino C, Acosta Y, Grieco E. English-Speaking Ability of the Foreign-Born Population in the United States: 2012. Washington, DC: US Census Bureau; 2014. URL: https://www.census.gov/library/publications/2014/acs/acs-26.html [accessed 2016-09-11] [WebCite Cache ID 6gh56itFY]

39. Department of Immigration Border Protection. The People of Australia: Statistics from the 2011 Census. Canberra, Australia: Commonwealth of Australia; 2014.

40. Gill PS, Shankar A, Quirke T, Freemantle N. Access to interpreting services in England: secondary analysis of national data. BMC Public Health 2009;9:12 [FREE Full text] [doi: 10.1186/1471-2458-9-12] [Medline: 19138392]

41. Mainous AG, Baker R, Majeed A, Koopman RJ, Everett CJ, Saxena S, et al. English language skills and diabetes and hypertension among foreign-born South Asian adults in England. Public Health Rep 2006;121(3):331-336 [FREE Full text] [Medline: 16640158]

42. Isaacs T, Laurier MD, Turner CE, Segalowitz N. Identifying second language speech tasks and ability levels for successful nurse oral interaction with patients in a linguistic minority setting: an instrument development project. Health Commun 2011 Sep;26(6):560-570. [doi: 10.1080/10410236.2011.558336] [Medline: 21512923]

43. Hussain-Gambles M, Atkin K, Leese B. South Asian participation in clinical trials: the views of lay people and health professionals. Health Policy 2006 Jul;77(2):149-165. [doi: 10.1016/j.healthpol.2005.07.022] [Medline: 16216378]

44. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. 2011. URL: http://handbook.cochrane.org/ [accessed 2016-11-09] [WebCite Cache ID 6gsoXE5yk]

45. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]

46. Edwards L, Rooshenas L, Isaacs T. Inclusion of ethnic minorities in telehealth trials for type 2 diabetes: protocol for a systematic review examining prevalence and language issues. JMIR Res Protoc 2016;5(1):e43 [FREE Full text] [doi: 10.2196/resprot.5195] [Medline: 26969340]

47. Huang Z, Tao H, Meng Q, Jing L. Management of endocrine disease. Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur J Endocrinol 2015 Mar;172(3):R93-R101 [FREE Full text] [doi: 10.1530/EJE-14-0441] [Medline: 25227131]

48. Marcolino MS, Maia JX, Alkmim MB, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]

49. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]

50. The Leadership Conference Education Fund. Race and Ethnicity in the 2020 Census: Improving Data to Capture a Multiethnic America. Washington, DC: The Leadership Conference Education Fund; 2014.

51. 2011 Census: Ethnic Group, Local Authorities in the United Kingdom Database on the Internet. 2011. Table KS201UK URL: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census/ key-statistics-and-quick-statistics-for-local-authorities-in-the-united-kingdom---part-1/rft-ks201uk.xls [accessed 2016-04-12] [WebCite Cache ID 6giNGVC94]

52. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme. Lancaster: Lancaster University; 2006.

53. United States Census Bureau. 2016. Quick facts: United States URL: https://www.census.gov/quickfacts/table/PST045215/ 00 [accessed 2016-09-11] [WebCite Cache ID 6kRMXdd6B]

54. Adepoju O, Bolin J, Phillips C, Zhao H, Ohsfeldt R, McMaughan D, et al. Effects of diabetes self-management programs on time-to-hospitalization among patients with type 2 diabetes: a survival analysis model. Patient Educ Couns 2014 Apr;95(1):111-117 [FREE Full text] [doi: 10.1016/j.pec.2014.01.001] [Medline: 24468198]

55. Amoako E, Skelly A. Managing uncertainty in diabetes: an intervention for older African American women. Ethn Dis 2007;17(3):515-521. [Medline: 17985507]

56. 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]

57. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspect Health Inf Manag 2011;8:1b [FREE Full text] [Medline: 21307985]

58. Crowley MJ, Powers BJ, Olsen MK, Grubber JM, Koropchak C, Rose CM, et al. The Cholesterol, Hypertension, And Glucose Education (CHANGE) study: results from a randomized controlled trial in African Americans with diabetes. Am Heart J 2013 Jul;166(1):179-186. [doi: 10.1016/j.ahj.2013.04.004] [Medline: 23816038]

59. Forjuoh SN, Bolin JN, Huber JC, Vuong AM, Adepoju OE, Helduser JW, et al. Behavioral and technological interventions targeting glycemic control in a racially/ethnically diverse population: a randomized controlled trial. BMC Public Health 2014;14:71 [FREE Full text] [doi: 10.1186/1471-2458-14-71] [Medline: 24450992]

60. Gary TL, Batts-Turner M, Yeh H, Hill-Briggs F, Bone LR, Wang N, et al. The effects of a nurse case manager and a community health worker team on diabetic control, emergency department visits, and hospitalizations among urban African Americans with type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2009 Oct 26;169(19):1788-1794. [doi: 10.1001/archinternmed.2009.338] [Medline: 19858437]

61. 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]

62. Long JA, Jahnle EC, Richardson DM, Loewenstein G, Volpp KG. Peer mentoring and financial incentives to improve glucose control in African American veterans: a randomized trial. Ann Intern Med 2012 Mar 20;156(6):416-424 [FREE Full text] [doi: 10.7326/0003-4819-156-6-201203200-00004] [Medline: 22431674]

63. 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]

64. Lorig K, Ritter PL, Villa F, Piette JD. Spanish diabetes self-management with and without automated telephone reinforcement: two randomized trials. Diabetes Care 2008 Mar;31(3):408-414. [doi: 10.2337/dc07-1313] [Medline: 18096810]

65. Ruggiero L, Riley BB, Hernandez R, Quinn LT, Gerber BS, Castillo A, et al. Medical assistant coaching to support diabetes self-care among low-income racial/ethnic minority populations: randomized controlled trial. West J Nurs Res 2014 Oct;36(9):1052-1073 [FREE Full text] [doi: 10.1177/0193945914522862] [Medline: 24569698]

66. Shea S, Weinstock RS, Starren J, Teresi J, Palmas W, Field L, et al. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus. J Am Med Inform Assoc 2006;13(1):40-51 [FREE Full text] [doi: 10.1197/jamia.M1917] [Medline: 16221935]

67. Sacco WP, Bykowski CA, Mayhew LL, White KE. Educational attainment moderates the effect of a brief diabetes self-care intervention. Diabetes Res Clin Pract 2012 Jan;95(1):62-67. [doi: 10.1016/j.diabres.2011.08.027] [Medline: 21992869]

68. Sacco WP, Malone JI, Morrison AD, Friedman A, Wells K. Effect of a brief, regular telephone intervention by paraprofessionals for type 2 diabetes. J Behav Med 2009 Aug;32(4):349-359. [doi: 10.1007/s10865-009-9209-4] [Medline: 19365719]

69. McMahon GT, Fonda SJ, Gomes HE, Alexis G, Conlin PR. A randomized comparison of online- and telephone-based care management with internet training alone in adult patients with poorly controlled type 2 diabetes. Diabetes Technol Ther 2012 Nov;14(11):1060-1067 [FREE Full text] [doi: 10.1089/dia.2012.0137] [Medline: 22953754]

70. Davis RM, Hitch AD, Salaam MM, Herman WH, Zimmer-Galler IE, Mayer-Davis EJ. TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare. Diabetes Care 2010 Aug;33(8):1712-1717 [FREE Full text] [doi: 10.2337/dc09-1919] [Medline: 20484125]

71. Dy P, Morin PC, Weinstock RS. Use of telemedicine to improve glycemic management in a skilled nursing facility: a pilot study. Telemed J E Health 2013 Aug;19(8):643-645. [doi: 10.1089/tmj.2012.0274] [Medline: 23758078]

72. Glasgow RE, Strycker LA, King DK, Toobert DJ, Rahm AK, Jex M, et al. Robustness of a computer-assisted diabetes self-management intervention across patient characteristics, healthcare settings, and intervention staff. Am J Manag Care 2006 Mar;12(3):137-145 [FREE Full text] [Medline: 16524346]

73. Glasgow RE, Toobert DJ. Brief, computer-assisted diabetes dietary self-management counseling: effects on behavior, physiologic outcomes, and quality of life. Med Care 2000 Nov;38(11):1062-1073. [Medline: 11078048]

74. Glasgow RE, Toobert DJ, Hampson SE, Strycker LA. Implementation, generalization and long-term results of the "choosing well" diabetes self-management intervention. Patient Educ Couns 2002;48(2):115-122. [Medline: 12401414]

75. Khan MA, Shah S, Grudzien A, Onyejekwe N, Banskota P, Karim S, et al. A diabetes education multimedia program in the waiting room setting. Diabetes Ther 2011 Sep;2(3):178-188 [FREE Full text] [doi: 10.1007/s13300-011-0007-y] [Medline: 22127826]

76. King DK, Estabrooks PA, Strycker LA, Toobert DJ, Bull SS, Glasgow RE. Outcomes of a multifaceted physical activity regimen as part of a diabetes self-management intervention. Ann Behav Med 2006 Apr;31(2):128-137. [doi: 10.1207/s15324796abm3102_4] [Medline: 16542127]

77. Nagrebetsky A, Larsen M, Craven A, Turner J, McRobert N, Murray E, et al. Stepwise self-titration of oral glucose-lowering medication using a mobile telephone-based telehealth platform in type 2 diabetes: a feasibility trial in primary care. J Diabetes Sci Technol 2013;7(1):123-134 [FREE Full text] [Medline: 23439168]

78. Piette JD, Resnicow K, Choi H, Heisler M. A diabetes peer support intervention that improved glycemic control: mediators and moderators of intervention effectiveness. Chronic Illn 2013 Dec;9(4):258-267 [FREE Full text] [doi: 10.1177/1742395313476522] [Medline: 23585636]

79. Pressman AR, Kinoshita L, Kirk S, Barbosa GM, Chou C, Minkoff J. A novel telemonitoring device for improving diabetes control: protocol and results from a randomized clinical trial. Telemed J E Health 2014 Feb;20(2):109-114. [doi: 10.1089/tmj.2013.0157] [Medline: 24404816]

80. Stone RA, Sevick MA, Rao RH, Macpherson DS, Cheng C, Kim S, et al. The Diabetes Telemonitoring Study Extension: an exploratory randomized comparison of alternative interventions to maintain glycemic control after withdrawal of diabetes home telemonitoring. J Am Med Inform Assoc 2012;19(6):973-979 [FREE Full text] [doi: 10.1136/amiajnl-2012-000815] [Medline: 22610495]

81. Tildesley HD, Mazanderani AB, Ross SA. Effect of Internet therapeutic intervention on A1C levels in patients with type 2 diabetes treated with insulin. Diabetes Care 2010 Aug;33(8):1738-1740 [FREE Full text] [doi: 10.2337/dc09-2256] [Medline: 20668152]

82. Tildesley HD, Wright AM, Chan JH, Mazanderani AB, Ross SA, Tildesley HG, et al. A comparison of internet monitoring with continuous glucose monitoring in insulin-requiring type 2 diabetes mellitus. Can J Diabetes 2013 Oct;37(5):305-308. [doi: 10.1016/j.jcjd.2013.05.006] [Medline: 24500556]

83. Whitlock WL, Brown A, Moore K, Pavliscsak H, Dingbaum A, Lacefield D, et al. Telemedicine improved diabetic management. Mil Med 2000 Aug;165(8):579-584. [Medline: 10957848]

84. Pacaud D, Kelley H, Downey A, Chiasson M. Successful delivery of diabetes self-care education and follow-up through eHealth media. Can J Diabetes 2012 Oct;36(5):257-262. [doi: 10.1016/j.jcjd.2012.08.006]

85. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]

86. Tang TS, Digby EM, Wright AM, Chan JH, Mazanderani AB, Ross SA, et al. Real-time continuous glucose monitoring versus internet-based blood glucose monitoring in adults with type 2 diabetes: a study of treatment satisfaction. Diabetes Res Clin Pract 2014 Dec;106(3):481-486. [doi: 10.1016/j.diabres.2014.09.050] [Medline: 25458333]

87. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]

88. Timmerberg BD, Wurst J, Patterson J, Spaulding RJ, Belz NE. Feasibility of using videoconferencing to provide diabetes education: a pilot study. J Telemed Telecare 2009;15(2):95-97. [doi: 10.1258/jtt.2008.080813] [Medline: 19246610]

89. Shreck E, Gonzalez JS, Cohen HW, Walker EA. Risk perception and self-management in urban, diverse adults with type 2 diabetes: the improving diabetes outcomes study. Int J Behav Med 2014 Feb;21(1):88-98 [FREE Full text] [doi: 10.1007/s12529-013-9291-4] [Medline: 23385488]

XSL•FO
RenderX

90. Long AF, Gambling T, Young RJ, Taylor J, Mason JM. Acceptability and satisfaction with a telecarer approach to the management of type 2 diabetes. Diabetes Care 2005 Feb;28(2):283-289. [Medline: 15677780]

91. Dale J, Caramlau I, Sturt J, Friede T, Walker R. Telephone peer-delivered intervention for diabetes motivation and support: the telecare exploratory RCT. Patient Educ Couns 2009 Apr;75(1):91-98. [doi: 10.1016/j.pec.2008.09.014] [Medline: 19013741]

92. Eakin EG, Reeves MM, Winkler E, Healy GN, Dunstan DW, Owen N, et al. Six-month outcomes from living well with diabetes: A randomized trial of a telephone-delivered weight loss and physical activity intervention to improve glycemic control. Ann Behav Med 2013 Oct;46(2):193-203. [doi: 10.1007/s12160-013-9498-2] [Medline: 23609340]

93. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. J Gen Intern Med 2010 Dec;25(12):1315-1322 [FREE Full text] [doi: 10.1007/s11606-010-1480-0] [Medline: 20714820]

94. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]

95. Hunt CW, Sanderson BK, Ellison KJ. Support for diabetes using technology: a pilot study to improve self-management. Medsurg Nurs 2014;23(4):231-237. [Medline: 25318336]

96. Piette JD, Richardson C, Himle J, Duffy S, Torres T, Vogel M, et al. A randomized trial of telephonic counseling plus walking for depressed diabetes patients. Med Care 2011 Jul;49(7):641-648 [FREE Full text] [doi: 10.1097/MLR.0b013e318215d0c9] [Medline: 21478777]

97. Varney JE, Weiland TJ, Inder WJ, Jelinek GA. Effect of hospital-based telephone coaching on glycaemic control and adherence to management guidelines in type 2 diabetes, a randomised controlled trial. Intern Med J 2014 Sep;44(9):890-897. [doi: 10.1111/imj.12515] [Medline: 24963611]

98. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 8;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]

99. Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ 2009 Jul 7;181(1-2):37-44 [FREE Full text] [doi: 10.1503/cmaj.081272] [Medline: 19581618]

100. Izquierdo R, Meyer S, Starren J, Goland R, Teresi J, Shea S, et al. Detection and remediation of medically urgent situations using telemedicine case management for older patients with diabetes mellitus. Ther Clin Risk Manag 2007 Jun;3(3):485-489 [FREE Full text] [Medline: 18488079]

101. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]

102. Shea S, Kothari D, Teresi JA, Kong J, Eimicke JP, Lantigua RA, et al. Social impact analysis of the effects of a telemedicine intervention to improve diabetes outcomes in an ethnically diverse, medically underserved population: findings from the IDEATel Study. Am J Public Health 2013 Oct;103(10):1888-1894. [doi: 10.2105/AJPH.2012.300909] [Medline: 23488491]

103. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]

104. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hillis SL, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. Telemed J E Health 2011 May;17(4):254-261. [doi: 10.1089/tmj.2010.0176] [Medline: 21476945]

105. Walker EA, Shmukler C, Ullman R, Blanco E, Scollan-Koliopoulus M, Cohen HW. Results of a successful telephonic intervention to improve diabetes control in urban adults: a randomized trial. Diabetes Care 2011 Jan;34(1):2-7 [FREE Full text] [doi: 10.2337/dc10-1005] [Medline: 21193619]

106. Anderson DR, Christison-Lagay J, Villagra V, Liu H, Dziura J. Managing the space between visits: a randomized trial of disease management for diabetes in a community health center. J Gen Intern Med 2010 Oct;25(10):1116-1122 [FREE Full text] [doi: 10.1007/s11606-010-1419-5] [Medline: 20556536]

107. Shea S, Weinstock RS, Teresi JA, Palmas W, Starren J, Cimino JJ, et al. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus: 5 year results of the IDEATel study. J Am Med Inform Assoc 2009;16(4):446-456 [FREE Full text] [doi: 10.1197/jamia.M3157] [Medline: 19390093]

108. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334. [Medline: 11473065]

109. 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]

110. 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]

111. Frosch DL, Uy V, Ochoa S, Mangione CM. Evaluation of a behavior support intervention for patients with poorly controlled diabetes. Arch Intern Med 2011 Dec 12;171(22):2011-2017. [doi: 10.1001/archinternmed.2011.497] [Medline: 21986347]

112. Glasgow RE, Nutting PA, Toobert DJ, King DK, Strycker LA, Jex M, et al. Effects of a brief computer-assisted diabetes self-management intervention on dietary, biological and quality-of-life outcomes. Chronic Illn 2006 Mar;2(1):27-38. [Medline: 17175680]

113. Graziano JA, Gross CR. A randomized controlled trial of an automated telephone intervention to improve glycemic control in type 2 diabetes. ANS Adv Nurs Sci 2009;32(3):e42-e57. [doi: 10.1097/ANS.0b013e3181b117a9] [Medline: 19707086]

114. Huizinga MM, Gebretsadik T, Garcia UC, Shintani AK, Michon SR, Shackleford LO, et al. Preventing glycaemic relapse in recently controlled type 2 diabetes patients: a randomised controlled trial. Diabetologia 2010 May;53(5):832-839 [FREE Full text] [doi: 10.1007/s00125-010-1658-3] [Medline: 20084363]

115. Izquierdo R, Lagua CT, Meyer S, Ploutz-Snyder RJ, Palmas W, Eimicke JP, et al. Telemedicine intervention effects on waist circumference and body mass index in the IDEATel project. Diabetes Technol Ther 2010 Mar;12(3):213-220. [doi: 10.1089/dia.2009.0102] [Medline: 20151772]

116. Krein SL, Klamerus ML, Vijan S, Lee JL, Fitzgerald JT, Pawlow A, et al. Case management for patients with poorly controlled diabetes: a randomized trial. Am J Med 2004 Jun 1;116(11):732-739. [doi: 10.1016/j.amjmed.2003.11.028] [Medline: 15144909]

117. Piette JD, Valenstein M, Himle J, Duffy S, Torres T, Vogel M, et al. Clinical complexity and the effectiveness of an intervention for depressed diabetes patients. Chronic Illn 2011 Dec;7(4):267-278 [FREE Full text] [doi: 10.1177/1742395311409259] [Medline: 21840915]

118. Schillinger D, Hammer H, Wang F, Palacios J, McLean I, Tang A, et al. Seeing in 3-D: examining the reach of diabetes self-management support strategies in a public health care system. Health Educ Behav 2008 Oct;35(5):664-682. [doi: 10.1177/1090198106296772] [Medline: 17513690]

119. Sevick MA, Korytkowski M, Stone RA, Piraino B, Ren D, Sereika S, et al. Biophysiologic outcomes of the Enhancing Adherence in Type 2 Diabetes (ENHANCE) trial. J Acad Nutr Diet 2012 Aug;112(8):1147-1157 [FREE Full text] [doi: 10.1016/j.jand.2012.05.008] [Medline: 22818724]

120. Shea S, IDEATel Consortium. The Informatics for Diabetes and Education Telemedicine (IDEATel) project. Trans Am Clin Climatol Assoc 2007;118:289-304 [FREE Full text] [Medline: 18528511]

121. Stone RA, Rao RH, Sevick MA, Cheng C, Hough LJ, Macpherson DS, et al. Active care management supported by home telemonitoring in veterans with type 2 diabetes: the DiaTel randomized controlled trial. Diabetes Care 2010 Mar;33(3):478-484 [FREE Full text] [doi: 10.2337/dc09-1012] [Medline: 20009091]

122. Trief PM, Izquierdo R, Eimicke JP, Teresi JA, Goland R, Palmas W, et al. Adherence to diabetes self care for white, African-American and Hispanic American telemedicine participants: 5 year results from the IDEATel project. Ethn Health 2013;18(1):83-96. [doi: 10.1080/13557858.2012.700915] [Medline: 22762449]

123. Trief PM, Morin PC, Izquierdo R, Teresi J, Eimicke JP, Goland R, et al. Depression and glycemic control in elderly ethnically diverse patients with diabetes: the IDEATel project. Diabetes Care 2006 Apr;29(4):830-835. [Medline: 16567823]

124. Trief PM, Teresi JA, Eimicke JP, Shea S, Weinstock RS. Improvement in diabetes self-efficacy and glycaemic control using telemedicine in a sample of older, ethnically diverse individuals who have diabetes: the IDEATel project. Age Ageing 2009 Mar;38(2):219-225 [FREE Full text] [doi: 10.1093/ageing/afn299] [Medline: 19171951]

125. Trief PM, Teresi JA, Izquierdo R, Morin PC, Goland R, Field L, et al. Psychosocial outcomes of telemedicine case management for elderly patients with diabetes: the randomized IDEATel trial. Diabetes Care 2007 May;30(5):1266-1268. [doi: 10.2337/dc06-2476] [Medline: 17325261]

126. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hills SL, et al. Outcomes of a home telehealth intervention for patients with diabetes and hypertension. Telemed J E Health 2012 Oct;18(8):575-579. [doi: 10.1089/tmj.2011.0237] [Medline: 22873700]

127. Wakefield BJ, Koopman RJ, Keplinger LE, Bomar M, Bernt B, Johanning JL, et al. Effect of home telemonitoring on glycemic and blood pressure control in primary care clinic patients with diabetes. Telemed J E Health 2014 Mar;20(3):199-205 [FREE Full text] [doi: 10.1089/tmj.2013.0151] [Medline: 24404819]

128. Weinstock RS, Brooks G, Palmas W, Morin PC, Teresi JA, Eimicke JP, et al. Lessened decline in physical activity and impairment of older adults with diabetes with telemedicine and pedometer use: results from the IDEATel study. Age Ageing 2011 Jan;40(1):98-105 [FREE Full text] [doi: 10.1093/ageing/afq147] [Medline: 21081539]

129. Weinstock RS, Teresi JA, Goland R, Izquierdo R, Palmas W, Eimicke JP, et al. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. Diabetes Care 2011 Feb;34(2):274-279 [FREE Full text] [doi: 10.2337/dc10-1346] [Medline: 21270184]

130. Luchsinger JA, Palmas W, Teresi JA, Silver S, Kong J, Eimicke JP, et al. Improved diabetes control in the elderly delays global cognitive decline. J Nutr Health Aging 2011 Jun;15(6):445-449 [FREE Full text] [Medline: 21623465]

131. Barrera M, Glasgow RE, McKay HG, Boles SM, Feil EG. Do Internet-based support interventions change perceptions of social support?: An experimental trial of approaches for supporting diabetes self-management. Am J Community Psychol 2002 Oct;30(5):637-654. [Medline: 12188054]

132. Blackberry ID, Furler JS, Best JD, Chondros P, Vale M, Walker C, et al. Effectiveness of general practice based, practice nurse led telephone coaching on glycaemic control of type 2 diabetes: the Patient Engagement and Coaching for Health (PEACH) pragmatic cluster randomised controlled trial. BMJ 2013;347:f5272 [FREE Full text] [Medline: 24048296]

133. Dyson PA, Beatty S, Matthews DR. An assessment of lifestyle video education for people newly diagnosed with type 2 diabetes. J Hum Nutr Diet 2010 Aug;23(4):353-359. [doi: 10.1111/j.1365-277X.2010.01077.x] [Medline: 20497292]

134. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz DL. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract 2008 Jun;14(3):465-469. [doi: 10.1111/j.1365-2753.2007.00881.x] [Medline: 18373577]

135. Gibson B, Marcus RL, Staggers N, Jones J, Samore M, Weir C. Efficacy of a computerized simulation in promoting walking in individuals with diabetes. J Med Internet Res 2012;14(3):e71 [FREE Full text] [doi: 10.2196/jmir.1965] [Medline: 22576226]

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

137. Jennings CA, Vandelanotte C, Caperchione CM, Mummery WK. Effectiveness of a web-based physical activity intervention for adults with Type 2 diabetes-a randomised controlled trial. Prev Med 2014 Mar;60:33-40. [doi: 10.1016/j.ypmed.2013.12.011] [Medline: 24345601]

138. McKay HG, Glasgow RE, Feil EG, Boles SM, Barrera MJ. Internet-based diabetes self-management and support: initial outcomes from the Diabetes Network project. Rehabil Psychol 2002;47(1):31-48. [doi: 10.1037//0090-5550.47.1.31]

139. Odegard PS, Christensen DB. MAP study: RCT of a medication adherence program for patients with type 2 diabetes. J Am Pharm Assoc (2003) 2012;52(6):753-762. [doi: 10.1331/JAPhA.2012.11001] [Medline: 23229961]

140. Tildesley H, Mazanderani A, Chan J, Ross S. Efficacy of A1C reduction using Internet intervention in patients with type 2 diabetes treated with insulin. Can J Diabetes 2011 Jan;35(3):250-253. [doi: 10.1016/S1499-2671(11)53007-2]

141. Vigersky RA, Fonda SJ, Chellappa M, Walker MS, Ehrhardt NM. Short- and long-term effects of real-time continuous glucose monitoring in patients with type 2 diabetes. Diabetes Care 2012 Jan;35(1):32-38 [FREE Full text] [doi: 10.2337/dc11-1438] [Medline: 22100963]

142. Williams ED, Bird D, Forbes AW, Russell A, Ash S, Friedman R, et al. Randomised controlled trial of an automated, interactive telephone intervention (TLC Diabetes) to improve type 2 diabetes management: baseline findings and six-month outcomes. BMC Public Health 2012;12:602 [FREE Full text] [doi: 10.1186/1471-2458-12-602] [Medline: 22857017]

143. Young RJ, Taylor J, Friede T, Hollis S, Mason JM, Lee P, et al. Pro-active call center treatment support (PACCTS) to improve glucose control in type 2 diabetes: a randomized controlled trial. Diabetes Care 2005 Feb;28(2):278-282. [Medline: 15677779]

144. Liebreich T, Plotnikoff RC, Courneya KS, Boulé N. Diabetes NetPLAY: a physical activity website and linked email counselling randomized intervention for individuals with type 2 diabetes. Int J Behav Nutr Phys Act 2009;6:18 [FREE Full text] [doi: 10.1186/1479-5868-6-18] [Medline: 19327141]

145. Cumming A. Language assessment in education: tests, curricula, and teaching. Ann Rev Appl Linguist 2009 Apr 24;29:90-100. [doi: 10.1017/S0267190509090084]

146. Hughson J, Woodward-Kron R, Parker A, Hajek J, Bresin A, Knoch U, et al. A review of approaches to improve participation of culturally and linguistically diverse populations in clinical trials. Trials 2016 May;17(1):263 [FREE Full text] [doi: 10.1186/s13063-016-1384-3] [Medline: 27229153]

147. Central Intelligence Agency. The World Factbook. Washington, DC: Central Intelligence Agency; 2013. URL: https://www.cia.gov/library/publications/the-world-factbook/fields/2075.html [accessed 2016-04-19] [WebCite Cache ID 6gtAe8gZI]

148. National Insitutes of Health. 2001. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research: amended, October, 2001 URL: http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm [WebCite Cache ID 6gsnkYZ6K]

149. Kanter MH, Abrams KM, Carrasco MR, Spiegel NH, Vogel RS, Coleman KJ. Patient-physician language concordance: a strategy for meeting the needs of spanish-speaking patients in primary care. Perm J 2009;13(4):79-84 [FREE Full text] [Medline: 20740108]

150. Segalowitz N, Kehayia E. Exploring the determinants of language barriers in health care (LBHC): toward a research agenda for the language sciences. Can Mod Lang Rev 2011 Nov;67(4):480-507. [doi: 10.3138/cmlr.67.4.480]

151. Greenhalgh T. Health literacy: towards system level solutions. BMJ 2015;350:h1026. [Medline: 25712067]

152. Bibi R, Redwood S, Taheri S. Raising the issue of overweight and obesity with the South Asian community. Br J Gen Pract 2014 Aug;64(625):417-419 [FREE Full text] [doi: 10.3399/bjgp14X681145] [Medline: 25071053]

153. Donovan JL, Paramasivan S, de SI, Toerien M. Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials. Trials 2014;15:5 [FREE Full text] [doi: 10.1186/1745-6215-15-5] [Medline: 24393291]

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

155. Trofimovich P, Isaacs T. Disentangling accent from comprehensibility. Bilingualism 2012 May 25;15(04):905-916. [doi: 10.1017/S1366728912000168]

156. Derwing T, Munro M. Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research. Amsterdam: John Benjamins Publishing Company; 2015.

157. Mills N, Blazeby JM, Hamdy FC, Neal DE, Campbell B, Wilson C, et al. Training recruiters to randomized trials to facilitate recruitment and informed consent by exploring patients' treatment preferences. Trials 2014;15:323 [FREE Full text] [doi: 10.1186/1745-6215-15-323] [Medline: 25115160]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HbA$_{1c}$: glycated hemoglobin
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROB: risk of bias

Original Paper

Electronic Quality of Life Assessment Using Computer-Adaptive Testing

Chris Gibbons[1,2,3*], PhD; Peter Bower[4*], PhD; Karina Lovell[5*], PhD; Jose Valderas[6*], MPH, PhD; Suzanne Skevington[2*], PhD

[1]Cambridge Centre for Health Services Research, University of Cambridge, Cambridge, United Kingdom

[2]Manchester Centre for Health Psychology, University of Manchester, Manchester, United Kingdom

[3]The Psychometrics Centre, University of Cambridge, Cambridge, United Kingdom

[4]Centre for Primary Care, University of Manchester, Manchester, United Kingdom

[5]School of Nursing, Midwifery and Social Work, University of Manchester, Manchester, United Kingdom

[6]Collaboration for Academic Primary Care, University of Exeter, Exeter, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Chris Gibbons, PhD
Cambridge Centre for Health Services Research
University of Cambridge
17 Mill Lane
Cambridge,
United Kingdom
Phone: +44 1223 765 203
Email: drcgibbons@gmail.com

Abstract

Background: Quality of life (QoL) questionnaires are desirable for clinical practice but can be time-consuming to administer and interpret, making their widespread adoption difficult.

Objective: Our aim was to assess the performance of the World Health Organization Quality of Life (WHOQOL)-100 questionnaire as four item banks to facilitate adaptive testing using simulated computer adaptive tests (CATs) for physical, psychological, social, and environmental QoL.

Methods: We used data from the UK WHOQOL-100 questionnaire (N=320) to calibrate item banks using item response theory, which included psychometric assessments of differential item functioning, local dependency, unidimensionality, and reliability. We simulated CATs to assess the number of items administered before prespecified levels of reliability was met.

Results: The item banks (40 items) all displayed good model fit (P>.01) and were unidimensional (fewer than 5% of t tests significant), reliable (Person Separation Index>.70), and free from differential item functioning (no significant analysis of variance interaction) or local dependency (residual correlations < +.20). When matched for reliability, the item banks were between 45% and 75% shorter than paper-based WHOQOL measures. Across the four domains, a high standard of reliability (alpha>.90) could be gained with a median of 9 items.

Conclusions: Using CAT, simulated assessments were as reliable as paper-based forms of the WHOQOL with a fraction of the number of items. These properties suggest that these item banks are suitable for computerized adaptive assessment. These item banks have the potential for international development using existing alternative language versions of the WHOQOL items.

Introduction

Improving patient-centered care (PCC) is a key strategic priority for health care systems worldwide due to the increasing burden of non-communicable chronic disease and ageing populations [1]. In the United States, the Institute of Medicine enshrines PCC as one of the six elements of high-quality care [2]. In the United Kingdom, the National Health Service (NHS) Outcomes Framework provides a new focus on patient outcomes, rather than processes of care—a vision grounded in PCC and shared decision making [3,4]. Improving quality of life (QoL) and satisfaction with care for patients with chronic conditions is central to the NHS Outcomes Framework's objectives [5].

Increasing priority placed on PCC reflects a longstanding movement towards patient-centered metrics and away from sole reliance on disease-centered measures of severity, impact, and burden [6]. Such patient-centered metrics include satisfaction [7], activation [8], and subjective QoL [9]. Subjective QoL is of special interest as it seeks to quantify "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" [9].

Patient-reported outcome measures (PROMs) are measurements of any aspect of a patient's health status that come directly from the patient, usually through a paper-based questionnaire scale [10,11]. Measures of subjective QoL provide a comprehensive assessment of the patient's life encompassing physical, psychological, social, and environmental factors, which are rated as useful by clinicians [12]. These measures may alert clinicians to a patient's concerns and prompt a discussion between the two parties about these issues [13,14].

Clinical trials assessing PROM feedback report improvements in identification of clinical issues, emotional well-being, patient-centered discussions, and symptom recognition in pediatric, oncology, and respiratory settings [15-19]. We are unaware of any published randomized controlled trials that have used the World Health Organization's Quality of Life (WHOQOL) instruments to evaluate the impact of assessment and feedback on patient outcomes (search strategy published in [20]). A recent pilot study demonstrated modest benefits of WHOQOL feedback on psychological QoL alongside positive comments on the perceived usefulness of sharing this information with doctors [21,22]. These results indicate some promise for this developing field.

However, the overall level of evidence for the impact of using PROs in clinical practice is mixed [11,14]. In mental health research, the effectiveness of PROM interventions appears to be mediated by the quality of the feedback [23]. A Cochrane review is planned to assess the evidence relating to the use of the PROMs in clinical practice [20].

Despite the potential benefits of PROM administration, their use is frequently dismissed in many medical settings including family practice [24], which may be partially attributed to the impracticality of administering paper-based questionnaires in a time-pressured environment. Recent research has highlighted the time- and resource-consuming issues of data storage, dealing with missing data, and analyzing results as potential barriers to the uptake of paper-based PROMs [25]. In the research setting, the length of questionnaires is negatively associated with response rate [26], indicating a general preference for shorter assessments, which may be adversely affecting implementation in tandem with other human, financial, and logistical barriers. A lack of clear instruction, training, and feedback that is linked to specific clinical action may also contribute to poor implementation of PROMs in clinical practice [11].

The increased availability of modern computing technologies (eg, smartphones and tablets) provides an opportunity to computerize PROM administration, which has previously been primarily paper-based. In addition to practical advantages of not having to score and interpret paper-based questionnaires,

administration using a computer adaptive testing (CAT) algorithm is a quicker and potentially more relevant and accurate method of assessing patient reported outcomes [27,28]. The development of open-source platforms for deploying cloud-based CATs both on the Internet and within apps [29,30] allow researchers and clinicians to create and deploy CATs and presents new opportunities to disrupt traditional forms of PRO assessment and feedback and to deliver them directly to patients at scale.

In the United States, the patient-reported outcomes assessment system (PROMIS) program has developed computer adaptive tests for fatigue, pain, depression, and health-related QoL among other things, which are rigorously validated but unfortunately are currently limited to PROMIS-validated questionnaires only, which do not include subjective QoL [31].

CAT relies on model parameters derived from psychometrically calibrated item banks. Item bank information, including item parameters, is obtained by fitting scale data to item response theory (IRT) models, of which the Rasch model is a special case. The Rasch model has been widely used to assist the development and assess the psychometric properties of QoL scales [32,33], item banks for computer adaptive tests [34,35], short-form questionnaires [36,37], and for developing clinically oriented content-based interpretation tools. The Rasch model is closely related to other IRT models, but it has stricter assumptions that can result in concise banks of items devoid of uninformative or unnecessary items and that may yield the highest standard of measurement, including specific objectivity [38-40]. IRT affords some advantages over "classical test" methodologies. Most significantly, IRT models allow PROMs to be accurately administered using any number or combination of items from the original scale. This allows CAT algorithms to select the most relevant and informative items for the test taker [41].

We hypothesize that the application of item response theory and computer adaptive testing algorithms to QoL scale data will improve precision (ie, reliability) and efficiency (ie, the number of items needed to be administered). These improvements will be driven by the removal of unnecessary items during the calibration of the item bank using IRT and the "intelligent" administration of items using CAT algorithms. The study used items from the WHOQOL's 100-item measure to create four item banks to measure QoL in physical, psychological, social, and environmental domains and to test the performance of these item banks using simulated CAT.

Methods

Population

We conducted the current analysis on data collected from 320 people living in the United Kingdom [9]. The population consisted of 162 females (51%), 260 "sick" people (balanced across International Classification of Diseases-10 categories I-XVIII), and a mean age of 44 years (SD 17). Detailed descriptions of the sample may be found elsewhere [9]. English is the development language of the WHOQOL measures, which are all designed to be internationally relevant, but there is some

evidence that differential item functioning (DIF) between different countries exists for many items within the WHOQOL-100 [32]. We, therefore, chose to create item banks that will be psychometrically accurate for use in the United Kingdom in the first instance.

Measures

WHOQOL-100

The WHOQOL-100 is a generic 100-item measure of subjective QoL designed for use across a spectrum of populations, including sick and well people. The original scale is scored as 25 facets representing six domains of quality of life (physical, psychological, emotional, social, independence, and spiritual) [42,43]. Other versions of the WHOQOL, including the WHOQOL-BREF, include the same facets to represent four domains of QoL (physical, psychological, environmental, social) [6,44]. Four items in the WHOQOL-100 represent general QoL and overall health. High scores in each domain (recoded for negatively worded items) indicate a better QoL than lower scores in the same domain. Respondents judge their quality of life over the previous 2 weeks. The international WHOQOL-100, as well as the UK national instrument, show excellent psychometric qualities of internal consistency, reliability, and construct validity [9,42,43]. In our research, the domains and facets of the WHOQOL-100 were arranged to mirror the four-dimension structure of the popular WHOQOL-BREF measure [6]. This structure has been empirically supported using structural equation modeling [42] and will facilitate comparisons between the new item banks and the large body of research that has employed the shorter four-dimensional measure.

Previous studies have applied data from the WHOQOL instruments to IRT resulting in both unidimensional and multidimensional solutions [32,45,46]. These studies uncovered issues relating to category threshold ordering, DIF between countries, and item redundancy (see Multimedia Appendix 1).

Analysis

Item Response Theory

We assessed the advanced psychometric criteria and estimated item bank parameters by fitting scale data to the Partial Credit Model (PCM), a polytomous extension of the Rasch model suitable for Likert-type data [47]. Scalability and monotonicity were assessed using Mokken analysis before the more rigorous tests of PCM assumptions. Both the Mokken and Rasch models can be seen as probabilistic extensions of the deterministic Guttman scaling model. The probabilistic version is better suited to psychological constructs and real-world data [48]. Mokken

analysis is done prior to Rasch analysis to ensure that the scale structure is consistent with the Rasch model (ie, item response probabilities increase monotonically in line with the level of the underlying trait). The combination of the two methodologies in this order is recommended and has been shown to be useful in previous research conducted by members of our group [37,49]. Where scale data did not fit either the Mokken or the PCM, an iterative process of scale improvement was undertaken by removing items that violated the assumptions of either model. The iterative process involved stepwise assessments of scalability (indicated by a Loevinger's Ho value >.3), category threshold ordering, item fit to the PCM (chi-square $P>.010$), fit residuals (fit residuals within ±2.5), local dependency (residual correlations <.10), and DIF (no significant analysis of variance interactions by demographic group). Items that violated any of the above assumptions were individually removed, and the remaining items were reanalyzed. Disordered thresholds were collapsed for adjacent categories, while ensuring anchor semantics remained logical (ie, "Agree" would not be collapsed into "Neither Agree nor Disagree"). This process was repeated until no items failed to meet the assumptions of the PCM: presented category disordering, misfit to the model, high fit residuals, local dependency, or DIF. Unidimensionality and overall model fit was assessed once issues with items breaching the above assumptions had been resolved. Further details of the IRT analyses are given in Multimedia Appendix 1.

Computer Adaptive Testing

CAT is a process whereby items from an item bank are automatically chosen and administered one-by-one, based on an algorithm that attempts to choose items that will maximize the information gained about the test taker. While CATs may be of any length, they are usually governed by a "stopping rule." Estimations of person location on the underlying continuum (their level of QoL, in this context) are recalculated depending on previous item responses, and the item that has the greatest information function (IF) at the reestimated level of theta is then administered. This estimation process continues until the stopping rule is met. Stopping rules may demand that a questionnaire is finished once a certain number of items have been administered, or the test has been going on for a predefined amount of time, or until a level of measurement precision has been achieved. Measurement precision is defined using the standard error (SE) of measurement. SE is inversely related to (and thus comparable with) marginal reliability such that reliability$=1 - SE^2$, where the standard deviation of the distribution is equal to 1.

Figure 1. Disordered and reordered thresholds for item F9.3 "How much do any difficulties in mobility bother you?" (F9.3 has been rescored from 1-2-3-4-5 to 1-2-3-3-4 to account for the disordered category thresholds 3 and 4).
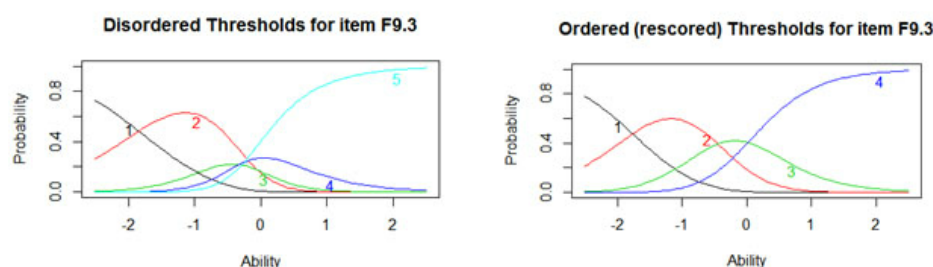
Simulation

We used the R-based CAT simulation engine Firestar [50] to simulate CATs in this study. The first item that the CAT administered for each domain was the item with the greatest information at the distribution mean. The IRT scaling constant was set to 1.7 [51]. We conducted 1000 iterations of the CAT with data simulated using distribution of person location (theta) values based on PCM estimations from the current dataset [47,52].

For this study, we defined three stopping rules for the CAT simulations based on standard errors equivalent to reliability values of .70, .80, and .90 (SE=.55, .45 and .32, respectively). These values were picked because they represent the minimum value for group measurement (.70) and the minimum value for individual-level measurement (.90), as well as a value in between [53]. Comparative analysis was facilitated by running a second set of simulations with stopping rules based on the published reliability values of the paper-based versions of the WHQOL-100 and the WHOQOL-BREF [6,54]. For example, where the published reliability for the Psychological QoL domain was .82, we set the stopping rule standard error to .42 (which is equivalent to alpha=.82) and compared the mean number of items administered before the stopping rule was met, with the number of items in the paper-based questionnaire.

Firestar uses a Bayesian expected a posteriori theta estimator and the maximum posterior weighted information (MPWI) item selection criterion. The MPWI selects items based on the IF weighted by the posterior distribution of trait/phenomena values [55]. This criterion has been shown to provide excellent measurement information for CAT using polytomous items.

Software

Analyses were conducted using Rasch Unidimensional Measurement Models 2030 [56] and the R Statistical Computing Environment [57] with the "mokken" and "ltm" packages installed [58-60]. Rasch analysis was conducted solely using RUMM, and "ltm" was used to draw Figure 1. Computer adaptive testing simulation was conducted using the FIRESTAR code generator for R [50].

Results

The 100 items of the WHOQOL-100 were arranged into four subscales, reflecting the factor structure of the shorter WHOQOL-BREF measure [42,54].

Domain One: Physical Quality of Life

Mokken analysis confirmed the initial scalability of all 28 items in the physical QoL subscale. All items returned a Ho value >.30 indicating acceptable scalability (see Multimedia Appendix 1 for the details of items removed from the scale).

Following Mokken analysis, the 28 items were fitted to the PCM to evaluate their advanced psychometric properties. Initial fit to the Rasch model was poor (see Table 1, Physical Initial). A number of items displayed disordered threshold and were rescored (see Figure 1 for an example of disordered and ordered thresholds). Misfit was apparently driven by 14 items that displayed high fit residuals ($\geq \pm 1.4$) and six locally dependent items, resulting in the removal of a total of 17 items. Details of the items removed from each domain may be found in Multimedia Appendix 2.

The final Physical QoL item bank consisted of 11 items, with excellent fit to the PCM and high reliability (see Table 1, analysis Physical Final). The scale was free from DIF and local dependency and was unidimensional (final item fit statistics, scoring, and threshold locations are shown in Multimedia Appendix 3). The item bank was well targeted, with fewer than 1.2% extreme scores.

Domain Two: Psychological Quality of Life

Mokken analysis indicated the removal of six items from Psychological facets on "Thinking" (1 item), "Self-esteem" (2 items), and "Spiritual" (3 items) domain, which did not scale appropriately (Ho <.30) with the rest of the items in the scale.

Following removal of the six items, the remaining 18 items did not fit the PCM (χ^2_{216}=474.6, P<.001; see Table 1, analysis Psychological Initial). A number of items required rescoring to account for disordered thresholds. Misfit was driven by three items with high positive fit residuals and three items displaying local dependency (see Multimedia Appendix 1). Following removal of these items, the final scale showed excellent fit to the Rasch model, including excellent reliability, targeting, unidimensionality, and an absence of DIF or local dependency (see Table 1, Psychological Final).

Domain Three: Social Quality of Life

Mokken analysis confirmed the scalability of the 12 items (Ho >.30; see Multimedia Appendix 1) but these 12 items did not fit the PCM (χ^2_{106}=143.49, P<.001; see Table 1, analysis Social Initial). A number of items were rescored to resolve issues caused by disordered thresholds. Misfit was driven by four

locally dependent items (14.2, 15.1, 14.3, and 15.2; see Multimedia Appendix 1). Following removal of these items, the final 8-item scale fit the Rasch model (χ^2_{72}=88.37, P=.09; see Table 1, analysis Social Final) including the absence of DIF or local dependency and excellent reliability, and unidimensionality. The Social QoL item bank was also exceptionally well targeted, with only 3 respondents falling outside the measurable range of the scale (0.94%).

Domain Four: Environmental Quality of Life

Mokken analysis indicated the removal of 16 items from the 32-item Environmental QoL scale (Loevinger's Ho <0.3; see Multimedia Appendix 1). The remaining 16 items did not fit the PCM (χ^2_{144}=191.23, P<.001). The iterative item removal and rescoring procedure led to a reduction of seven items that breached the assumption of local dependency (see Multimedia Appendix 1). The final scale has an excellent fit to the Rasch model (χ^2_{81}=65.11, P=.90) including good reliability, excellent scale targeting, and acceptable dimensionality (see Table 1, Environmental Final).

Table 1 displays overall summaries for the initial and final analyses performed to validate each item bank. None of the item banks showed immediate fit to the Rasch model without modification.

Table 1. Summary Rasch fit statistics and psychometric criteria for all subscales.

Analysis ID[a]		Items n	Item residual		Person residual		Chi square			Reliability	Extremes, %	% of t tests significant	t test 95% CI
			Mean	SD	Mean	SD	χ^2	df	P				
Physical	Initial	28	.53	2.7	-.28	1.71	388.5	171	<.01	.91	0	30	15.62-24.38
	Final	11	.32	1.2	-.37	1.42	109	108	0.46	.89	1.56	4.5	2.13-6.61
Psychological	Initial	24	.67	2.6	-.29	1.91	474.6	216	<.01	.91	0	22.85	18.21-27.49
	Final	14	.31	1.3	-.35	1.52	133.6	135	.52	.90	0	7.18	4.35-10.01
Social	Initial	12	.39	2	-.34	1.36	143.5	106	<.01	.87	0	17.58	13.36-21.8
	Final	8	-.03	1.5	-.38	1.05	88.37	72	.09	.81	.94	8.78	5.67-11.89
Environmental	Initial	16	.30	1.3	-.39	1.54	191.2	144	<.01	.88	0	8.86	5.73-11.99
	Final	9	.38	0.9	-.32	1.18	65.11	81	.90	.80	0	7.5	4.61-10.39
Ideal values			0	>1.4	0	>1.4			>0.01	>0.85	< 10%	<5%	<5%

[a]"Initial" refers to preanalysis values, "Final" to the final version.

Table 2. Summary of computer adaptive testing (CAT) simulation (1000 iterations).

Domain QoL	Stopping rule, SE(θ)	Number of items used		Range of items used	Mean SE	Reliability	Correlation between CAT θ and complete test θ
		Mean	SD				
Physical	<.32	10.01	1.22	8-11	0.32	0.9	1
	<.45	4.23	0.84	3-6	0.43	0.82	0.99
	<.55	2.46	0.5	2-3	0.52	0.73	0.98
Psychological	<.32	9.8	2	7-12	0.32	0.9	1
	<.45	4.5	0.94	3-6	0.42	0.82	0.98
	<.55	4.32	0.45	4-6	0.52	0.73	0.96
Social	<.32	7.3	1.06	5-8	0.36	0.87	1
	<.45	4.32	1.71	3-8	0.42	0.82	0.99
	<.55	2.44	0.7	2-4	0.5	0.75	0.97
Environmental	<.32	7.96	1.25	6-9	0.34	0.89	1
	<.45	3.61	1.39	2-7	0.43	0.82	0.98
	<.55	2.34	0.48	2-4	0.48	0.77	0.97

Table 3. Comparison of paper-based World Health Organization Quality of Life (WHOQOL) measures and the computer adaptive testing (CAT) simulations of the item banks.

Scale	Domain	Original scale information		Stopping rule	CAT simulation	
		Items, n	Reliability, alpha	Reliability-matched SE	Items administered, median	Actual SE
WHOQOL-BREF	Physical	7	0.82	0.42	4	0.42
	Psychological	6	0.81	0.44	4	0.42
	Social	3	0.68	0.55	2	0.5
	Environmental	8	0.8	0.45	3	0.43
WHOQOL-100[a]	Physical	16	0.86	0.37	7	0.36
	Psychological	20	0.82	0.42	4	0.42
	Social	12	0.73	0.52	2	0.5
	Environmental	32	0.85	0.39	5	0.38

[a]Independence and spirituality domains omitted.

Computer Adaptive Testing Simulations

The results of the initial computer adaptive testing (CAT) simulation are displayed in Table 2. Predefined stopping rules based on different SE values were used to assess the number of items that the CAT needed to administer to reach a given level of reliability. Despite the relatively small item banks, acceptable reliability was gained with a mean of four items across all administrations (alpha>.70) and a high standard of reliability (alpha>.90) could be gained with a mean of 9 items (alpha>.90).

The results of the CAT simulation for each item bank are presented in Table 3. Stopping rules based on SE values yielded tests of varying lengths. The reduced item versions correlated strongly with the full-length item banks.

A second reliability-matched simulation (where the stopping rule of the simulation was matched to the reliability of the published measures [6,54]) shows that the item banks can produce a measurement that is as reliable as the WHOQOL-BREF and the WHOQOL-100 using 43% and 75% fewer items, respectively.

Discussion

Principal Findings

We calibrated four item banks that measure physical, psychological, social, and environmental QoL. Simulated computer-adaptive administration of the item banks demonstrates their ability to create accurate measurements that are both significantly shorter and often more reliable than paper-based alternatives.

In this study, the decision was made to evaluate an item bank on a sample collected in the United Kingdom. This sample was chosen to avoid issues of differential item functioning that has had a significant impact on previous studies using multinational data [32], and with the aim of providing an item bank that could be used to inform a CAT suitable for use within clinical practice.

While this study plainly demonstrates the advantages of IRT and CATs in terms of their reliability and efficiency, these techniques can also improve the quality of clinical PROMs by other means. The ability of the algorithms to select the most relevant items based on a patient's previous responses may also provide utility in clinical measurement. Targeting in this manner not only makes assessments more relevant but also prevents patients from being asked to complete items that may be distressing or redundant. For example, a person at an early stage of a progressive disease may become distressed or concerned by completing items that assess symptoms they may experience in many later stages of the disease. A correctly calibrated item bank and CAT administration system could create accurate measurement for such patients without the need to present items that were not relevant to their level of functional impairment.

The results of this study are in line with findings from prior investigations of item bank performance, most notably and recently from the PROMIS group, insofar as CAT produced measurement estimates that were more precise, efficient, and flexible than paper-based tests for other constructs including fatigue and functional status (eg, [61-63]). To our knowledge, this study represents the first time that simulations of generic QoL item banks have been tested in this manner, though a recent study has developed an item bank suitable for assessing emotional functioning in cancer [64].

Previous studies that have applied WHOQOL scale data IRT models have employed different approaches. Studies using WHOQOL-100 data and the Rasch model have evaluated the suitability of an "index" solution, which assesses QoL as a single unidimensional, rather than multidimensional, construct. In these studies, the strict assumptions of the Rasch model led to the removal of a similar number of items, though they were not the same items that we removed from this study. Issues of DIF were also evident [32]. Other IRT analyses reported elsewhere have often presented caveats such as poor reliability or unclear dimensionality for one or more of the subscales, especially on analysis using the shorter WHOQOL-BREF (eg, [33,45]).

One notable advantage of the methods employed in the current analysis of the larger initial item banks (eg, using the WHOQOL-100 items arranged into the WHOQOL-BREF

format) led to acceptable measurement across all four domains of the WHOQOL measure and obtaining excellent measurement properties with each. It must be noted that the Social QoL domain displayed unidimensionality that was slightly above the recommended threshold (5.67%, rather than 5%) for strict unidimensionality.

Multi-item questionnaires measuring health outcomes are still widely used in clinical trials [65] and epidemiological research [66]. Due to the wide variance in the type and function of PROMs, it is no small task to develop recommendations for how often they should be recalibrated using contemporary data. Happily, the increased use of IRT and adaptive testing, rather than classical test theory, means that is possible to engage in a process of iterative calibration, and the addition of new items to an item bank while collecting data for other purposes. This practice of pretesting is common in educational testing, where items must be frequently changed to reduce cheating [67,68].

The cross-cultural development of the original WHOQOL instruments suggests good potential for the development of culturally sensitive item banks and CATs. Further analyses to the one presented here provided preliminary evidence on the use of the WHOQOL item banks for use in different cultures (eg, [69,70]).

From a technical perspective, there is clear potential to develop the IRT methods employed in this study further and to apply these data to a multidimensional item response theory (MIRT) model [71]. An MIRT solution using the bi-factor model could take account of the shared variance between items in the four domains to simultaneously produce a summary score for global QOL alongside the scores for the individual domains [49].

Sample size restrictions precluded such an analysis being conducted in our study. We must note that while such a bi-factor MIRT analysis would be cutting edge regarding its methodology, some work is yet to be done to demonstrate the clinical relevance and interpretability of MIRT questionnaires and adaptive tests, though multidimensional computer adaptive tests are beginning to emerge [72].

This study naturally provides the foundations for future work to develop and evaluate a CAT system than can deliver these item banks to clinical populations and to assess the performance of the item banks under "live" testing conditions, rather than use simulations. The recent development of free-to-use and open source CAT platforms, such as Concerto [29] opens the possibility for the widespread use of computer-assisted psychometrics in clinical research and practice. Additionally, the adoption of the WHOQOL questionnaire offers the availability of 15 different language-translated versions of the questionnaire items, increasing the feasibility of international assessment of QoL using CATs [29,70].

Conclusion

We have presented functional item banks that are capable of producing high-quality measurement across four domains of QoL using fewer items than equivalent paper-based measures. These item banks outperform the paper-based versions of the WHOQOL both in terms of reliability, length, and the flexibility in which they may be administered. The computer adaptive tests based on the WHOQOL would be suitable across a range of medical specialities, and hence particularly useful in primary care as an aid to understanding and quantifying the quality of life across diverse biopsychosocial domains.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional details of item response theory analysis procedure.

[PDF File (Adobe PDF File), 88KB - jmir_v18i9e240_app1.pdf]

Multimedia Appendix 2

Details of removed items.

[PDF File (Adobe PDF File), 130KB - jmir_v18i9e240_app2.pdf]

Multimedia Appendix 3

Summary item fit statistics.

[PDF File (Adobe PDF File), 308KB - jmir_v18i9e240_app3.pdf]

XSL•FO

RenderX

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 2015 Jan;385(9963):117-171. [doi: 10.1016/S0140-6736(14)61682-2]

2. Committee on Quality of Health Care in America Institute of Medicine. Crossing the Quality Chasm?: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.

3. Stewart M. Towards a global definition of patient centred care. BMJ 2001 Feb 24;322(7284):444-445 [FREE Full text] [Medline: 11222407]

4. Roter D. The enduring and evolving nature of the patient–physician relationship. Patient Education and Counseling 2000 Jan;39(1):5-15. [doi: 10.1016/S0738-3991(99)00086-5]

5. Department of Health. The NHS Outcomes Framework 2013/14. 2012. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213055/121109-NHS-Outcomes-Framework-2013-14.pdf [accessed 2016-08-29] [WebCite Cache ID 6k7zx030e]

6. Skevington S, Lotfy M, O'Connell K. The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A Report from the WHOQOL Group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00]

7. Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK--development and psychometric characteristics. BMC Fam Pract 2009;10:57 [FREE Full text] [doi: 10.1186/1471-2296-10-57] [Medline: 19698140]

8. 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]

9. Skevington SM. Measuring quality of life in Britain. Journal of Psychosomatic Research 1999 Nov;47(5):449-459. [doi: 10.1016/S0022-3999(99)00051-3]

10. US Food and Drug Administration. Health and Quality of Life Outcomes. 2006. Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims URL: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf [accessed 2016-08-29] [WebCite Cache ID 6k807HrrZ]

11. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Jan 4;17(2):179-193. [doi: 10.1007/s11136-007-9295-0]

12. Skevington SM, Day R, Chisholm A, Trueman P. How much do doctors use quality of life information in primary care? Testing the Trans-Theoretical Model of behaviour change. Qual Life Res 2005 May;14(4):911-922. [doi: 10.1007/s11136-004-3710-6]

13. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006 Oct;12(5):559-568. [doi: 10.1111/j.1365-2753.2006.00650.x]

14. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. J Eval Clin Pract 1999 Dec;5(4):401-416. [doi: 10.1046/j.1365-2753.1999.00209.x]

15. Velikova G, Keding A, Harley C, Cocks K, Booth L, Smith A, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. Eur J Cancer 2010 Sep;46(13):2381-2388. [doi: 10.1016/j.ejca.2010.04.030] [Medline: 20570138]

16. Aaronson N. Assessing quality of life in clinical practice in oncology. Eur J Cancer 1999 Sep;35:S222 [FREE Full text] [doi: 10.1016/S0959-8049(99)81295-7]

17. Wolfe J, Orellana L, Cook E, Ullrich C, Kang T, Geyer J. Improving the care of children with advanced cancer by using an electronic patient-reported feedback intervention: results from the PediQUEST randomized controlled trial. J Clin Oncol 2014:1119. [doi: 10.1200/JCO.2013.51.5981]

18. Velikova G, Booth L, Smith A, Brown P, Lynch P, Brown J, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]

19. DuBenske L, Gustafson D, Namkoong K, Hawkins R, Brown R, McTavish F. Effects of an interactive cancer communication system on lung cancer caregivers' quality of life and negative mood: a randomized clinical trial. 2010 Presented at: International Psycho-Oncology Society; May 2010; Quebec, QC.

20. Goncalves DCB, Gibbons C, Ricci-Cabello I, Bobrovitz N, Gibbons E, Kotzeva A, et al. Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2015(4):-. [doi: 10.1002/14651858.CD011589]

21. Llewellyn A, Skevington S. Using guided individualised feedback to review self-reported quality of life in health and its importance. Psychol Health 2015;30(3):301-317. [doi: 10.1080/08870446.2014.972396] [Medline: 25280244]

XSL•FO

RenderX

22.   Llewellyn AM, Skevington SM. Evaluating a new methodology for providing individualized feedback in healthcare on quality of life and its importance, using the WHOQOL-BREF in a community population. Qual Life Res 2016 Mar;25(3):605-614. [doi: 10.1007/s11136-015-1132-2] [Medline: 26370098]

23.   Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. Psychiatr Serv 2015 Mar 1;66(3):224-241. [doi: 10.1176/appi.ps.201400141] [Medline: 25727110]

24.   Jani BD, Purves D, Barry S, Cavanagh J, McLean G, Mair FS. Challenges and implications of routine depression screening for depression in chronic disease and multimorbidity: a cross sectional study. PLoS One 2013 Sep;8(9):e74610 [FREE Full text] [doi: 10.1371/journal.pone.0074610] [Medline: 24058602]

25.   Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ 2010 Jan 18;340:c186. [doi: 10.1136/bmj.c186]

26.   Nakash RA, Hutton JL, Jørstad-Stein EC, Gates S, Lamb SE. Maximising response to postal questionnaires--a systematic review of randomised trials in health research. BMC Med Res Methodol 2006;6:5 [FREE Full text] [doi: 10.1186/1471-2288-6-5] [Medline: 16504090]

27.   Jette AM, McDonough CM, Haley SM, Ni P, Olarsch S, Latham N, et al. A computer-adaptive disability instrument for lower extremity osteoarthritis research demonstrated promising breadth, precision, and reliability. J Clin Epidemiol 2009 Aug;62(8):807-815 [FREE Full text] [doi: 10.1016/j.jclinepi.2008.10.004] [Medline: 19216052]

28.   Petersen MA, Aaronson NK, Arraras JI, Chie W, Conroy T, Costantini A, EORTC Quality of Life Group. The EORTC computer-adaptive tests measuring physical functioning and fatigue exhibited high levels of measurement precision and efficiency. J Clin Epidemiol 2013 Mar;66(3):330-339. [doi: 10.1016/j.jclinepi.2012.09.010] [Medline: 23347855]

29.   Psychometrics Centre. Concerto Adaptive Testing Platform. Cambridge: University of Cambridge; 2013. URL: http://www.psychometrics.cam.ac.uk/newconcerto [accessed 2016-08-29] [WebCite Cache ID 6k82UkkZ4]

30.   Scalise K, Allen DD. Use of open-source software for adaptive measurement: Concerto as an R-based computer adaptive development and delivery platform. Br J Math Stat Psychol 2015 Nov;68(3):478-496. [doi: 10.1111/bmsp.12057] [Medline: 26061260]

31.   Amtmann D, Cook KF, Jensen MP, Chen W, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182 [FREE Full text] [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]

32.   Leplege A, Ecosse E, WHOQOL Rasch Project Scientific Committee. Methodological issues in using the Rasch model to select cross culturally equivalent items in order to develop a Quality of Life index: the analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, United Kingdom). J Appl Meas 2000;1(4):372-392. [Medline: 12077463]

33.   Pomeroy IM, Tennant A, Young CA. Rasch analysis of the WHOQOL-BREF in post polio syndrome. J Rehabil Med 2013 Sep;45(9):873-880 [FREE Full text] [doi: 10.2340/16501977-1186] [Medline: 23979562]

34.   Lai J, Cella D, Chang C, Bode R, Heinemann A. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res 2003:485-501. [doi: 10.1023/A:1025014509626]

35.   Forkmann T, Boecker M, Norra C, Eberle N, Kircher T, Schauerte P, et al. Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. Rehabil Psychol 2009 May;54(2):186-197. [doi: 10.1037/a0015612] [Medline: 19469609]

36.   Gibbons CJ, Mills RJ, Thornton EW, Ealing J, Mitchell JD, Shaw PJ, et al. Rasch analysis of the hospital anxiety and depression scale (HADS) for use in motor neurone disease. Health Qual Life Outcomes 2011;9:82 [FREE Full text] [doi: 10.1186/1477-7525-9-82] [Medline: 21955749]

37.   Bee P, Gibbons C, Callaghan P, Fraser C, Lovell K. Evaluating and Quantifying User and Carer Involvement in Mental Health Care Planning (EQUIP): Co-Development of a New Patient-Reported Outcome Measure. PLoS ONE 2016:e0149973. [doi: 10.1371/journal.pone.0149973]

38.   Perline R, Wright BD, Wainer H. The Rasch Model as Additive Conjoint Measurement. Applied Psychological Measurement 1979 Apr 01;3(2):237-255. [doi: 10.1177/014662167900300213]

39.   Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care 2004 Jan;42(1 Suppl):I7-16. [doi: 10.1097/01.mlr.0000103528.48582.7c] [Medline: 14707751]

40.   Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. J Appl Meas 2001;2(4):389-423. [Medline: 12011506]

41.   Smits N, Cuijpers P, van Straten A. Applying computerized adaptive testing to the CES-D scale: a simulation study. Psychiatry Res 2011:ePub. [doi: 10.1016/j.psychres.2010.12.001]

42.   The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [Medline: 9626712]

43.   The WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): Position paper from the World Health Organization. Social Science & Medicine 1995 Nov;41(10):1403-1409. [doi: 10.1016/0277-9536(95)00112-K]

44.   Skevington SM, McCrate FM. Expecting a good quality of life in health: assessing people with diverse diseases and conditions using the WHOQOL-BREF. Health Expect 2012 Mar;15(1):49-62. [doi: 10.1111/j.1369-7625.2010.00650.x] [Medline: 21281412]

45. Rocha NS, Power MJ, Bushnell DM, Fleck MP. Cross-cultural evaluation of the WHOQOL-BREF domains in primary care depressed patients using Rasch analysis. Med Decis Making 2012 Aug;32(1):41-55. [doi: 10.1177/0272989X11415112] [Medline: 21825270]

46. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur J Public Health 2006 Aug;16(4):420-428 [FREE Full text] [doi: 10.1093/eurpub/cki155] [Medline: 16141303]

47. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982 Jun;47(2):149-174. [doi: 10.1007/BF02296272]

48. Engelhard G. Historical Perspectives on Invariant Measurement: Guttman, Rasch, and Mokken. Measurement: Interdisciplinary Research and Perspectives 2008 Aug 20;6(3):155-189. [doi: 10.1080/15366360802197792]

49. Gibbons CJ, Kenning C, Coventry PA, Bee P, Bundy C, Fisher L, et al. Development of a multimorbidity illness perceptions scale (MULTIPleS). PLoS One 2013;8(12):e81852 [FREE Full text] [doi: 10.1371/journal.pone.0081852] [Medline: 24376504]

50. Choi S. Firestar: Computerized Adaptive Testing Simulation Program for Polytomous Item Response Theory Models. Applied Psychological Measurement 2009 Oct 22;33(8):644-645 [FREE Full text] [doi: 10.1177/0146621608329892]

51. Camilli G. Origin of the scaling constant d=1.7 in Item Response Theory. J Educ Behav Stat 1994;19(3):293-295.

52. Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Qual Life Res 2010 Feb;19(1):125-136 [FREE Full text] [doi: 10.1007/s11136-009-9560-5] [Medline: 19941077]

53. Reeve B, Hays R, Bjorner J, Cook K, Crane P, Teresi J, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007 May;45(5 Suppl 1):S22-S31. [doi: 10.1097/01.mlr.0000250483.85507.04] [Medline: 17443115]

54. The WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998 Jun;46(12):1569-1585. [Medline: 9672396]

55. Choi SW, Swartz RJ. Comparison of CAT Item Selection Criteria for Polytomous Items. Appl Psychol Meas 2009 Sep 1;33(6):419-440 [FREE Full text] [doi: 10.1177/0146621608327801] [Medline: 20011456]

56. Andrich D, Sheridan B, Luo G. Rasch models for measurement: RUMM2030. Perth, Western Australia: RUMM Pty Ltd; 2010.

57. R Development Team. R: A language and environment for statistical computing. URL: http://www.r-project.org/ [accessed 2016-08-25] [WebCite Cache ID 6k1U10E4V]

58. van der Ark LA. Mokken Scale Analysis in R. J Stat Soft 2007;20(11):1 [FREE Full text] [doi: 10.18637/jss.v020.i11]

59. van der Ark LA. New Developments in Mokken Scale Analysis in R. J Stat Soft 2007;20(11):- [FREE Full text] [doi: 10.18637/jss.v020.i11]

60. Rizopoulos D. ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses. J Stat Soft 2006;17(5):25. [doi: 10.18637/jss.v017.i05]

61. Cella D, Gershon R, Lai J, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Qual Life Res 2007 Mar;16 Suppl 1:133-141. [doi: 10.1007/s11136-007-9204-6] [Medline: 17401637]

62. Lai J, Cella D, Chang C, Bode R, Heinemann A. Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res 2003:501. [doi: 10.1023/A:1025014509626]

63. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. J Rheumatol 2009 Sep;36(9):2061-2066. [doi: 10.3899/jrheum.090358] [Medline: 19738214]

64. Petersen MA, Gamper E, Costantini A, Giesinger JM, Holzner B, Johnson C, EORTC Quality of Life Group. An emotional functioning item bank of 24 items for computerized adaptive testing (CAT) was established. J Clin Epidemiol 2016 Feb;70:90-100. [doi: 10.1016/j.jclinepi.2015.09.002] [Medline: 26363341]

65. Pi-Sunyer X, Astrup A, Fujioka K, Greenway F, Halpern A, Krempf M, et al. A randomized, controlled trial of 30 mg of liraglutide in weight management. New Engl J Med 2015;373:22. [doi: 10.1056/NEJMoa1411892]

66. Eekhout I, Enders CK, Twisk JWR, de Boer MR, de Vet HCW, Heymans MW. Including auxiliary item information in longitudinal data analyses improved handling missing questionnaire outcome data. J Clin Epidemiol 2015 Jun;68(6):637-645 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.01.012] [Medline: 25724894]

67. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Fort Worth, TX: Wadsworth Pub Co; 1986.

68. Parshall C. Item Development and Pretesting in a CBT Environment. In: Building the foundations for future assessments. UK: Routledge; 2002.

69. Tennant A, Penta M, Tesio L, Grimby G, Thonnard J, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004 Jan;42(1 Suppl):I37-I48. [doi: 10.1097/01.mlr.0000103529.63132.77] [Medline: 14707754]

70. Gibbons C. Assessing the performance of an item bank derived from the World Health Organisation Quality of Life-100 measure for computer adaptive testing across diverse cultures. Qual Life Res 2015;24:59-60.

71. Chalmers R. mirt: A Multidimensional Item Response Theory Package for the Environment. J Stat Soft 2012;48(6):- [FREE Full text] [doi: 10.18637/jss.v048.i06]

72. Hackshaw M. Association of patient-reported outcomes with progression-free survival in malignant pleural mesothelioma. Diss Abstr Int Sect B Sci Eng 2010;71:3001 [FREE Full text]

Abbreviations

CAT: computer adaptive testing
DIF: differential item functioning
IRT: item response theory
MPWI: maximum posterior weighted information
PCC: patient-centered care
PROMIS: patient-reported outcomes assessment system
PROMS: patient-reported outcome measures
QoL: quality of life
SE: standard error
WHO: World Health Organization
WHOQOL-100: World Health Organization Quality of Life 100 questionnaire
WHOQOL BREF: World Health Organization Brief questionnaire
MIRT: multidimensional item response theory

XSL•FO
RenderX

Original Paper

A Web-Based Telehealth Training Platform Incorporating Automated Nonverbal Behavior Feedback for Teaching Communication Skills to Medical Students: A Randomized Crossover Study

Chunfeng Liu[1*], BSc (Hons); Renee L Lim[1,2*], GCertEd; Kathryn L McCabe[1,3*], PhD; Silas Taylor[4*], MEd; Rafael A Calvo[1*], PhD

[1]School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia

[2]Sydney Medical School, The University of Sydney, Sydney, Australia

[3]Psychiatry and Behavioral Sciences, University of California, Davis, Davis, CA, United States

[4]UNSW Medicine, University of New South Wales, Sydney, Australia

[*]all authors contributed equally

Corresponding Author:
Rafael A Calvo, PhD
School of Electrical and Information Engineering
The University of Sydney
Bldg J03, School of Electrical and Information Engineering
Maze Crescent
Sydney, 2006
Australia
Phone: 61 293518171
Fax: 61 293513847
Email: rafael.calvo@sydney.edu.au

Abstract

Background: In the interests of patient health outcomes, it is important for medical students to develop clinical communication skills. We previously proposed a telehealth communication skills training platform (EQClinic) with automated nonverbal behavior feedback for medical students, and it was able to improve medical students' awareness of their nonverbal communication.

Objective: This study aimed to evaluate the effectiveness of EQClinic to improve clinical communication skills of medical students.

Methods: We conducted a 2-group randomized crossover trial between February and June 2016. Participants were second-year medical students enrolled in a clinical communication skills course at an Australian university. Students were randomly allocated to complete online EQClinic training during weeks 1–5 (group A) or to complete EQClinic training during weeks 8–11 (group B). EQClinic delivered an automated visual presentation of students' nonverbal behavior coupled with human feedback from a standardized patient (SP). All students were offered two opportunities to complete face-to-face consultations with SPs. The two face-to-face consultations were conducted in weeks 6–7 and 12–13 for both groups, and were rated by tutors who were blinded to group allocation. Student-Patient Observed Communication Assessment (SOCA) was collected by blinded assessors (n=28) at 2 time points and also by an SP (n=83). Tutor-rated clinical communications skill in face-to-face consultations was the primary outcome and was assessed with the SOCA. We used t tests to examine the students' performance during face-to-face consultations pre- and postexposure to EQClinic.

Results: We randomly allocated 268 medical students to the 2 groups (group A: n=133; group B: n=135). SOCA communication skills measures (score range 4–16) from the first face-to-face consultation were significantly higher for students in group A who had completed EQClinic training and reviewed the nonverbal behavior feedback, compared with group B, who had completed only the course curriculum components (*P*=.04). Furthermore, at the second face-to-face assessment, the group that completed a teleconsultation between the two face-to-face consultations (group B) showed improved communication skills (*P*=.005), and the one that had teleconsultations before the first face-to-face consultation (group A) did not show improvement.

Conclusions: The EQClinic is a useful tool for medical students' clinical communication skills training that can be applied to university settings to improve students clinical communication skills development.

Introduction

There is good evidence that effective patient-clinician communication can positively influence patient health outcomes [1-3]. For instance, a clinician's supportive expressions can help the patient to develop greater feelings of trust toward their clinician. These feelings of trust lead to greater patient self-efficacy, where the patient is more likely to follow recommended therapies, resulting in a better treatment outcome [4]. This evidence has meant that more training programs are being offered to students to help them learn clinical communication skills. Since students become competent through practice and feedback [5], medical students need practice with real or standardized patients (SPs) and feedback from patients and tutors. An SP normally refers to someone who has been trained to act as a patient in a medical situation. However, despite the importance of communication skills, the time allocated to such training within medical curricula is often limited. This is influenced, in part, by the logistics of providing large groups of medical students with access to SPs with whom they can practice and formulate their communication techniques.

The traditional method for clinical communication skills training is to provide students with feedback on video-recorded face-to-face consultations [6]. Students benefit from reviewing these videotapes of their clinical consultations with real patients or SPs [7], even more so when observers provide feedback about the verbal or nonverbal behaviors [8]. However, organizing large-scale face-to-face practice sessions and setting up the recording environment are a challenge for medical schools. Teleconferencing has been proposed as a solution for dealing with this challenge [9]. For example, the WebEncounter teleconference platform, developed to enable medical interns to communicate with SPs, showed that practicing on WebEncounter enhanced the communication skills of the interns when giving bad news [10]. Another recent study that related to medical students also suggested that involving telehealth consulting between medical practices and patients enhanced students' learning [11].

However, like WebEncounter, most clinical communication skills training systems tend to limit training to verbal communication skills and overlook important nonverbal communication behaviors. This is problematic, given that nonverbal communication is the major communication channel between individuals [12,13]. Manually annotating students' nonverbal behaviors from face-to-face consultations and providing this feedback to students are common ways to improve the learning of nonverbal communication skills [14]. However, the practicalities of providing this type of feedback means that

it is too time consuming to be widely adopted in medical education curricula.

We have previously described a platform called EQClinic [15,16]. Briefly, EQClinic is an e-learning platform that allows medical students to have recorded teleconsultations with SPs. The platform uses computer vision and audio processing techniques to automatically recognize, quantify, and visualize selected nonverbal behaviors (as well as human feedback) for student learning and reflection. Initial pilot application of EQClinic has shown that medical students' awareness of their nonverbal communication improved using EQClinic. However, the platform has not been applied within a typical university medical school curriculum.

Therefore, the goal of this study was to conduct a randomized crossover trial of the EQClinic incorporated into a university medical school curriculum (Multimedia Appendix 1 [17]). The EQClinic platform is designed to provide clinical communication skills training that integrates nonverbal behavior assessment for medical students. We used a randomized crossover design to initially test the effectiveness of the EQClinic, whereby we allocated medical students enrolled in a clinical communication skills course to 1 of 2 groups and asked them to complete a teleconsultation using EQClinic at different times during the semester. Interleaved with exposure to the EQClinic were face-to-face clinical consultation skills assessments. By staging exposure of students to the EQClinic, we evaluated the potential impact of the platform by comparing group performance on face-to-face clinical consultations before and after EQClinic exposure. To our knowledge, this is the first application of automated nonverbal behavior detection techniques for improved medical students' communication skills. We hypothesized that the use of EQClinic would improve medical students' learning about communication skills.

Methods

Participants

Participants were second-year undergraduate medical students from an Australian medical school. All students were enrolled in a communication skills training course provided by the medical school. Prior to this study, they were not offered any training about teleconsultations in the medical school. This study was approved by the University of New South Wales Research Ethics Committee (HC Reference Number: HC16048). Students were asked to sign an online consent form when they first accessed EQClinic. No content or methodological modifications were made after study commencement.

Questionnaires

The same 5 surveys previously reported by our group were used in the present study [15]. The pre- and postinterview questionnaires ascertained students' understanding of communication skills. The Post Interview Nonverbal Behavior Reflection Questionnaire asked students to estimate how often they engaged in certain nonverbal behaviors during the interview. The Reflection Questionnaire prompted students to reflect on the consultation. The primary outcomes measure was the Student-Patient Observed Communication Assessment (SOCA) form, which is an adapted version of the Calgary Cambridge Guides [18]. The SPs and tutors used the SOCA to rate students' communication skills. The form contained four aspects: providing structure, gathering information, building rapport, and understanding the patient's needs.

EQClinic

EQClinic comprises five components: an online training component, a personal calendar, a real-time interaction component, a nonverbal behavior detector, and a feedback generator. In the following sections, we briefly describe each of these.

Training Component and Personal Calendar

EQClinic provides training videos and documents for students and SPs to familiarize themselves with the platform. EQClinic also provides students and SPs with an automated personal calendar. SPs can offer their availability on the calendar for students to make a booking. All appointments are confirmed using the automated messaging system without need for human resources.

Real-Time Interaction Component

Once the appointment has been confirmed, videoconferencing enables a student and an SP to have a teleconsultation. The application works on most Web browsers of a personal computer or Android tablet. During the recorded consultation, the SP can record positive and negative moments using a "thumbs" tool and comment box.

To facilitate learning through reflection, online assessments were included for students. The SPs evaluated student performance immediately after the teleconsultation, during which time the students conduct a self-assessment using the same form. Students could immediately review the SP's rating.

Nonverbal Behavior Detector

Using audio processing and computer vision techniques, EQClinic automatically analyses the video recordings and detects the following nonverbal behaviors: head movements (nodding, head shaking, and head tilting), facial expressions (smiling and frowning), body movements (body leaning, hand gestures, and overall body movements), voice properties (volume and pitch), and speech patterns (turn taking and speaking ratio changes).

Feedback Generator

Feedback information includes computer-generated nonverbal behavior feedback (NVBF) and comment feedback from the SP. EQClinic visualizes students' nonverbal behavior using two types of feedback reports: single-feature and combined-feature reports. The single-feature feedback report illustrates each form of nonverbal behavior separately. The combined-feature feedback report displays multiple kinds of nonverbal behavior on one page. The comment feedback provides students a report that contains all the comments from the SP and tutor.

Study Design and Procedure

The administrator of this course randomly allocated a cohort of 268 students to group A (n=133) or group B (n=135) (see Figure 1) using a computer-generated random number sequence. One student was moved from group A to group B for administrative reasons. Following random allocation to a group, each participant was provided three opportunities to complete simulated clinical consultations with SPs: a teleconsultation using EQClinic, and two face-to-face consultations. In this study, all consultations focused on history-taking skills, to ensure a structured and consistent interaction. The allocation of the three consultations was varied between the 2 groups. The study was conducted over 13 weeks, and it included 4 periods (see Figure 1). (1) During weeks 1–5, group A completed a teleconsultation using EQClinic and group B was blocked from the platform. (2) During weeks 6 and 7, both groups completed a face-to-face consultation. In this period, group A was still able to access the platform for reviewing feedback only. (3) During weeks 8–11, group B completed an EQClinic consultation and group A was blocked from the platform. (4) During weeks 12 and 13, both groups were asked to complete another face-to-face consultation. In this period, group B was able to access the platform for reviewing feedback only. Due to the limited resources of setting up face-to-face consultations, not all enrolled students completed two face-to-face consultations. However, having a teleconsultation using EQClinic was mandatory for every student.

Figure 1. Flowchart of student participation in the EQClinic medical communication training program.
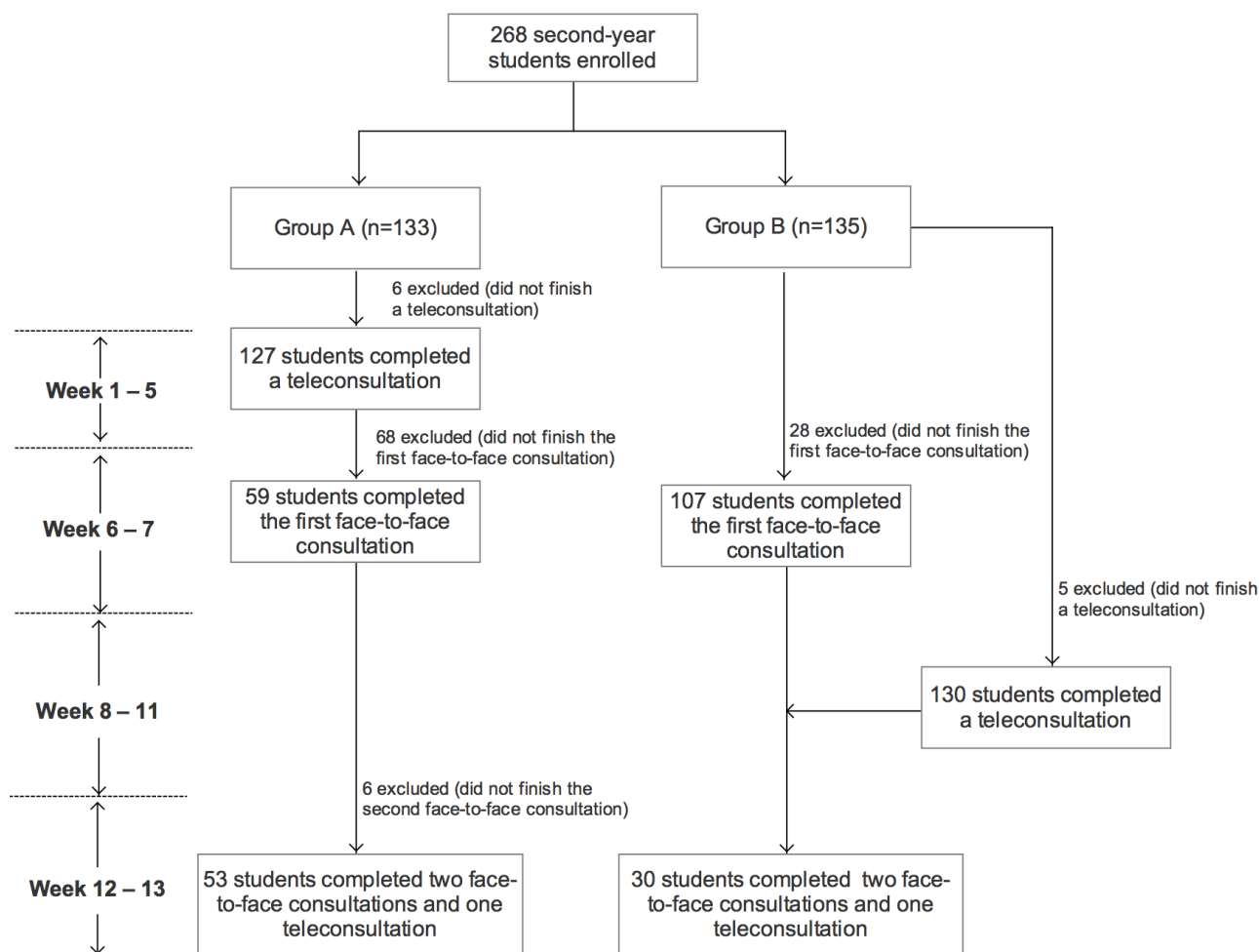
Teleconsultations Procedure

All participating SPs and students completed training via the training component of EQClinic. In the SPs' online training component, training videos demonstrated how to book appointments, conduct consultations with students, provide comments, and evaluate the student's performance. The patient scenario was also included in training and detailed the main symptoms of the SP and other historical information. All SPs were required to complete this online training. Following training, the SPs listed their availability for consultations on their EQClinic calendar.

Students were requested, by email, to complete one teleconsultation with an SP through EQClinic. The email described the details of the study and asked them to log in to EQClinic to complete the training module. It also informed them that, once they finished the training, they could request a consultation time from the slots available on their personal calendar. The SPs and students were allowed to have the teleconsultation anywhere as long as there was (1) a Web browser on a personal computer or an Android tablet with an external or built-in camera and microphone, (2) a good Internet connection, and (3) good lighting.

EQClinic teleconsultation comprised three sections: interviewing, assessing, and reviewing (see Figure 2) [15]. Interview and assessment components took approximately 40 minutes for a student and 25–30 minutes for an SP to complete. In the interviewing section, the student completed the preinterview questionnaire, and then the student and the SP conducted a 15-minute consultation via the teleconference component. The student and the SP then completed the online assessments. After each interview, the SP assessed the performance of the student using the SOCA form. Meanwhile, the student estimated their nonverbal behavior using the Post Interview Nonverbal Behavior Reflection Questionnaire, completed a personal SOCA form, and then reviewed the SOCA form completed by the SP and reflected on the interview using the Reflection Questionnaire.

Students were emailed to ask them to return to the system 24 hours after the consultation to review different kinds of feedback, which included the video recording, comments from the SP, and automated NVBF. Students also completed the postinterview questionnaire.

Figure 2. Workflow for the EQClinic consultation. SOCA: Student-Patient Observed Communication Assessment; SP: standardized patient.
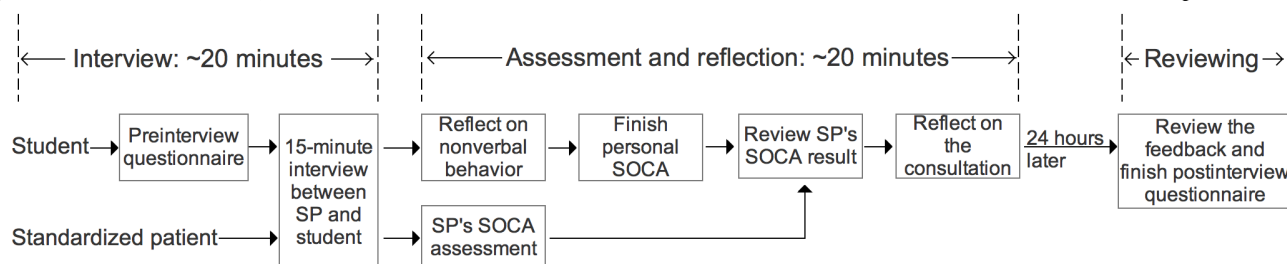
Face-to-Face Consultations Procedure

Face-to-face consultations were conducted in consultation rooms of a university-based clinical skills center. A trained tutor was present in the room to observe and assess the performance of the student during the consultation with the SP. The tutors were blinded to condition allocation (group A or group B). The tutor completed a SOCA form to assess the student, and the SP did not provide any evaluation and feedback for the student on this occasion. The students were asked to review the tutor's assessment and complete the Reflection Questionnaire. The scenario design and length of face-to-face consultations were the same as those for the teleconsultations.

Results

Figure 1 shows participants' flow through the trial. In the period of week 1 to week 5, 127 (46 male, 81 female) of 133 (95.5%) group A students completed the teleconsultation on EQClinic. In the second period (weeks 6–7), 166 of the 268 students from both group A (59/133, 44.4%) and group B (107/135, 79.3%) completed a face-to-face consultation. During weeks 8–11, a total of 130 (62 male, 68 female) of 135 (96.3%) group B students completed the teleconsultation using EQClinic. Lastly,

144 students from both group A (n=109) and group B (n=35) completed a face-to-face consultation during the period of weeks 12 and 13. In total, 11 students (6 from group A, 5 from group B) did not complete the teleconsultation in this study. At the second face-to-face consultation (weeks 12–13), 53 of 133 (39.9%) students from group A and 30 of 135 (22.2%) from group B had completed one teleconsultation and one face-to-face consultation before completing the second face-to-face consultation.

Table 1 and Table 2 describe the mean subgroup assessment results in the various study periods. Mean total SOCA scores from the first face-to-face consultation for group A (mean 13.02) and group B (mean 12.58) did not differ significantly between the groups (P=.08). To examine the influence of the NVBF component, we compared the group A mean SOCA scores (group A + NVBF: mean 13.21; 33/59, 55.9%) of those who had reviewed the NVBF component of the EQClinic before having their face-to-face consultations with the scores of group B students (mean 12.58). Mean SOCA scores were significantly higher on face-to-face SOCA total score in group A + NVBF ($t_{58.25}$=2.13, P=.04) than in group B. However, they did not differ statistically from group A students who did not review their nonverbal behavior component of the EQClinic (mean 12.77; 26/59, 44.1%, P>.05).

Table 1. Mean group medical communication skills (measured by Student-Patient Observed Communication Assessment score) assessment results (part 1: weeks 1–7).

Component	Weeks 1–5 (TC[a])		Weeks 6–7 (F2FC[b])					
	Group A (n=127)		Group A (n=59)		Group A (NVBF[c]) (n=33)		Group B (n=107)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total score	11.59	2.67	13.02	1.49	13.21	1.45	12.58	1.61
Providing structure	2.88	0.72	3.17	0.50	3.27	0.45	3.12	0.53
Gathering information	2.92	0.72	3.25	0.58	3.15	0.51	3.07	0.56
Building rapport	2.95	0.73	3.34	0.60	3.39	0.66	3.24	0.56
Understanding patient's needs	2.83	0.80	3.25	0.58	3.39	0.56	3.14	0.61

[a]TC: teleconsultation.

[b]F2FC: face-to-face consultation.

[c]NVBF: students who had a face-to-face consultation and reviewed the nonverbal behavior feedback.

Table 2. Mean group medical communication skills (measured by Student-Patient Observed Communication Assessment score) assessment results (part 2: weeks 8–13).

Component	Weeks 8–11 (TC[a])		Weeks 12–13 (F2FC[b])										
	Group B (n=130)		Group A (n=109)		Group A (ConA[c]) (n=53)		Group B (n=35)		Group B (ConB[d]) (n=30)		Group B (NVBF[e]) (n=13)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total score	13.13	2.31	13.28	1.54	13.28	1.46	13.43	1.63	13.53	1.52	13.62	1.64
Providing structure	3.31	0.67	3.25	0.49	3.23	0.46	3.31	0.52	3.37	0.48	3.38	0.49
Gathering information	3.34	0.72	3.26	0.55	3.30	0.53	3.51	0.55	3.53	0.50	3.46	0.50
Building rapport	3.16	0.73	3.46	0.58	3.47	0.54	3.34	0.58	3.37	0.55	3.46	0.63
Understanding patient's needs	3.32	0.68	3.31	0.57	3.28	0.59	3.26	0.44	3.27	0.44	3.31	0.46

[a]TC: teleconsultation.

[b]F2FC: face-to-face consultation.

[c]ConA: group A students who participated in two consultations (one face-to-face consultation, one teleconsultation) before week 12.

[d]ConB: group B students who participated in two consultations (one face-to-face consultation, one teleconsultation) before week 12.

[e]NVBF: students who had a face-to-face consultation and reviewed the nonverbal behavior feedback.

Following group B exposure to the EQClinic, the mean total SOCA scores from the second face-to-face consultation did not differ between the groups (group A: mean 13.28; group B: mean 13.53, $P>.05$). Mean SOCA scores of group B students (group B + NVBF: mean 13.62; 13/30, 43.3%) who reviewed the NVBF component of the EQClinic before their second face-to-face consultation did not differ from those in group A or group B who did not complete the nonverbal review (mean 13.47; 17/30, $P>.05$).

We used paired-samples t tests to compare the SOCA assessment scores for those students who completed EQClinic on both their two face-to-face consultations. Group B alone showed significant improvement in their mean SOCA score (mean preexposure score 12.58 vs postexposure score 13.53; $t_{48}=-2.96$; $P=.005$). Group A showed no significant increase in SOCA scores (mean preexposure score 13.02 vs postexposure score 13.28; $P>.05$). Comparison of the mean SOCA teleconsultation scores rated by SPs showed that group B's score (mean 13.13) was significantly higher than group A's score (mean 11.59; $t_{246.61}=-4.83$, $P<.001$).

Discussion

We incorporated EQClinic into a medical communication skills teaching curriculum to provide students with additional practice opportunities with SPs. Importantly, students could review their nonverbal communication behaviors. We examined the effects of EQClinic on medical students' learning of communication skills evaluated via the students' assessment (SOCA) scores. Results showed that students who completed a teleconsultation using EQClinic and reviewed the NVBF achieved higher SOCA scores in the first face-to-face consultation. In addition, students accomplished higher SOCA scores in their second face-to-face consultation if they completed a teleconsultation between the two face-to-face consultations.

Overall, adherence to the program was somewhat less than anticipated, with only 30% of student completing all components of the study. Dropout increased as the semester progressed. However, given the requirements of the undergraduate course, and the tendency for increased workload as the semester progresses, this result is unsurprising.

The results of the first face-to-face consultation show that the students who completed a teleconsultation and reviewed the NVBF component scored significantly higher in their face-to-face consultation than did students who did not interact with SPs on EQClinic. These results are promising. The difference in performance between the 2 groups seems to indicate that having EQClinic practice coupled with reviewing feedback improved medical communication skills in group A. As noted above, group B students achieved lower mean overall SOCA score in the first face-to-face consultation. However, overall, group B students showed significant improvement from their first to second face-to-face consultation. However, whether students reviewed their NVBF did not influence results for this group.

These findings are interesting because they suggest improvement in communication skills assessment after reviewing nonverbal feedback. While the need for medical communication skills training is widely accepted within the medical teaching community [1-3], there is less consensus on the need for specific teaching on the nonverbal aspects of communication. This is related to the lack of adequate resources, knowledge, and expertise in this aspect of communication [19]. To our knowledge, this is the first study to systematically incorporate nonverbal learning feedback into medical communication skills training.

Furthermore, that we showed no significant difference between group scores in the second face-to-face assessment seems to indicate that the timing of exposure to EQClinic within a teaching curriculum did not influence students' learning results.

In our study, group A was exposed to EQClinic at the beginning of the course; whereas group B was exposed in the middle of their course. We showed that at the commencement of the curriculum, when students did not have significant knowledge of clinical communication skills, exposure to EQClinic yielded a measurable bump in their clinical communication skills. For medical educators this seems to indicate that EQClinic could be incorporated at any period during the teaching curriculum.

We also showed that group B performed significantly better than group A on the SP-rated EQClinic teleconsultations. This difference could be explained in several ways. The first way relates to timing of EQClinic exposure, with group B completing the teleconsultation later in the semester than group A. Second, completing face-to-face consultations before being exposed to EQClinic, experience, and feedback garnered from the face-to-face consultation may have improved student performance in the teleconsultation. The third possibility is that the SPs who assessed students via the EQClinic increased their ratings across the semester. However, SP ratings neither contributed to student assessment nor were a central feature of the EQClinic.

Telehealth studies involving medical students and interns in urban, rural, and remote areas indicated that this medium was a useful learning tool [10,11]. In EQClinic, we enhance existing telehealth systems by providing students with multiple kinds of feedback. We contend that the primary functions of EQClinic are 2-fold: to facilitate student access to SPs to practice and refine their medical communication skills. The importance of SPs to facilitate the application of clinical communication theory, especially early one-on-one interactions, has been described previously [20]. The second function of the EQClinic is to facilitate reflective practice by providing human and computer-generated feedback, in particular in regard to nonverbal behaviors, in medical communication skills training.

However, based on our findings, it remains unclear which of the learning components were most useful to enhancing students' learning. Moreover, although a single exposure to the EQClinic led to a measurable improvement in students' medical communication skills scores, future studies will benefit from an examination of the appropriate "dose" of EQClinic. This will help determine the necessary exposure needed to provide sustained improvement and generalizable communication skills training. Finally, the growth of collected student data by EQClinic will aid the refinement of rules and models using machine learning algorithms to indicate to students what nonverbal behavior is associated with positive or negative responses and feedback from SPs in their clinical teleconsultations.

Study Limitations

There are several limitations to our study that should be considered when interpreting these findings. First, the absence of baseline measures limited our ability to observe change over time. Second, all the consultations conducted in this study were limited to a history-taking scenario. In reality, clinicians encounter many different scenarios. For example, when breaking bad news to patients, the clinician has to handle difficulties related to emotions. In addition, all the students in this study were second-year medical students who had limited knowledge about communication skills. Future studies may explore whether EQClinic is also useful for senior medical students and professionals. A third limitation is the relatively low proportion of students (30%) who completed all components of the study. While the sample was still appropriate for the statistical tests conducted, future investigations will benefit from exploring in greater detail the reason for student nonparticipation.

Conclusions

This study provided evidence that furnishing medical students with opportunities to conduct teleconsultations with SPs improved medical communication skills. In particular, offering enhanced and quantified feedback information facilitates their reflection and enhances their learning of clinical communication skills. Importantly, this study demonstrated that EQClinic was a useful and practical communication skills learning tool that is well suited to medical students within university settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH Checklist V1.6 [20].

[PDF File (Adobe PDF File), 1005KB - jmir_v18i9e246_app1.pdf]

References

1. Jackson JL. Communication about symptoms in primary care: impact on patient outcomes. J Altern Complement Med 2005;11(Suppl 1):S51-S56. [doi: 10.1089/acm.2005.11.s-51] [Medline: 16332187]
2. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. Soc Sci Med 2003 Sep;57(5):791-806. [Medline: 12850107]

XSL•FO

RenderX

3.   Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]

4.   Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]

5.   Mann KV. Theoretical perspectives in medical education: past experience and future possibilities. Med Educ 2011 Jan;45(1):60-68. [doi: 10.1111/j.1365-2923.2010.03757.x] [Medline: 21155869]

6.   Howley L, Martindale J. The efficacy of standardized patient feedback in clinical teaching: a mixed methods analysis. Med Educ Online 2004 Jun 01;9(18):[no page].

7.   Paul S, Dawson KP, Lanphear JH, Cheema MY. Video recording feedback: a feasible and effective approach to teaching history-taking and physical examination skills in undergraduate paediatric medicine. Med Educ 1998 May;32(3):332-336. [Medline: 9743791]

8.   Fukkink R, Trienekens N, Kramer L. Video feedback in education and training: putting learning in the picture. Educ Psychol Rev 2010 Oct 16;23(1):45-63. [doi: 10.1007/s10648-010-9144-5]

9.   Novack DH, Cohen D, Peitzman SJ, Beadenkopf S, Gracely E, Morris J. A pilot test of WebOSCE: a system for assessing trainees' clinical skills via teleconference. Med Teach 2002 Sep;24(5):483-487. [Medline: 12474813]

10. Daetwyler CJ, Cohen DG, Gracely E, Novack DH. eLearning to enhance physician patient communication: a pilot test of "doc.com" and "WebEncounter" in teaching bad news delivery. Med Teach 2010;32(9):e381-e390. [doi: 10.3109/0142159X.2010.495759] [Medline: 20795797]

11. Knight P, Bonney A, Teuss G, Guppy M, Lafferre D, Mullan J, et al. Positive clinical outcomes are synergistic with positive educational outcomes when using telehealth consulting in general practice: a mixed-methods study. J Med Internet Res 2016;18(2):e31 [FREE Full text] [doi: 10.2196/jmir.4510] [Medline: 26858152]

12. Ekman P, Friesen W, O'Sullivan M, Scherer K. Relative importance of face, body, and speech in judgments of personality and affect. J Pers Soc Psychol 1980;38(2):270.

13. O'Sullivan M, Ekman P, Friesen W, Scherer K. What you say and how you say it: the contribution of speech content and voice quality to judgments of others. J Pers Soc Psychol 1985 Jan;48(1):54-62. [Medline: 3981393]

14. Ishikawa H, Hashimoto H, Kinoshita M, Yano E. Can nonverbal communication skills be taught? Med Teach 2010;32(10):860-863. [doi: 10.3109/01421591003728211] [Medline: 20854164]

15. Liu C, Calvo RA, Lim R. Improving medical students' awareness of their nonverbal communication through automated nonverbal behavior feedback. Frontiers ICT 2016 Jun 20;3:[no page]. [doi: 10.3389/fict.2016.00011]

16. Liu C, Scott KM, Lim RL, Taylor S, Calvo RA. EQClinic: a platform for learning communication skills in clinical consultations. Med Educ Online 2016;21:31801 [FREE Full text] [Medline: 27476537]

17. 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]

18. Simmenroth-Nayda A, Heinemann S, Nolte C, Fischer T, Himmel W. Psychometric properties of the Calgary Cambridge guides to assess communication skills of undergraduate medical students. Int J Med Educ 2014;5:212-218 [FREE Full text] [doi: 10.5116/ijme.5454.c665] [Medline: 25480988]

19. Nestel D, Tabak D, Tierney T, Layat-Burn C, Robb A, Clark S, et al. Key challenges in simulated patient programs: an international comparative case study. BMC Med Educ 2011;11:69 [FREE Full text] [doi: 10.1186/1472-6920-11-69] [Medline: 21943295]

20. Joekes K, Noble LM, Kubacki AM, Potts HW, Lloyd M. Does the inclusion of 'professional development' teaching improve medical students' communication skills? BMC Med Educ 2011;11:41 [FREE Full text] [doi: 10.1186/1472-6920-11-41] [Medline: 21708000]

Abbreviations

NVBF: nonverbal behavior feedback
SOCA: Student-Patient Observed Communication Assessment
SP: standardized patient

XSL·FO
RenderX

Original Paper

Using Foreign Virtual Patients With Medical Students in Germany: Are Cultural Differences Evident and Do They Impede Learning?

Jens Walldorf[1], Dr. med; Tina Jähnert[1]; Norman B Berman[2], MD; Martin R Fischer[3], Prof Dr Med

[1]Universitätsklinik und Poliklinik für Innere Medizin I, Universitätsklinikum Halle, Martin-Luther-Universität Halle-Wittenberg, Halle, Germany

[2]Department of Pediatrics, Geisel School of Medicine at Dartmouth, Hanover, NH, United States

[3]Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany

Corresponding Author:
Jens Walldorf, Dr. med
Universitätsklinik und Poliklinik für Innere Medizin I
Universitätsklinikum Halle
Martin-Luther-Universität Halle-Wittenberg
Ernst-Grube-Strasse 40
06120 Halle
Germany
Phone: 49 345 557 2665
Fax: 49 345 557 2253
Email: jens.walldorf@uk-halle.de

Abstract

Background: Learning with virtual patients (VPs) is considered useful in medical education for fostering clinical reasoning. As the authoring of VPs is highly demanding, an international exchange of cases might be desirable. However, cultural differences in foreign VPs might hamper learning success.

Objective: We investigated the need for support for using VPs from the United States at a German university, with respect to language and cultural differences. Our goal was to better understand potential implementation barriers of a intercultural VP exchange.

Methods: Two VPs were presented to 30 German medical students featuring a cultural background different from German standards with respect to diagnostic and therapeutic procedures, ethical aspects, role models, and language (as identified by a cultural adaptation framework). Participants were assigned to two groups: 14 students were advised to complete the cases without further instructions (basic group), and 16 students received written explanatory supplemental information specifically with regard to cultural differences (supplement group). Using a 6-point scale (6=strongly agree), we analyzed the results of an integrated assessment of learning success as well as an evaluation of cases by the students on usefulness for learning and potential issues regarding the language and cultural background.

Results: The German students found it motivating to work with cases written in English (6-point scale, 4.5 points). The clinical relevance of the VPs was clearly recognized (6 points), and the foreign language was considered a minor problem in this context (3 points). The results of the integrated learning assessment were similar in both groups (basic 53% [SD 4] vs supplement 52% [SD 4] correct answers, P=.32). However, students using the supplemental material more readily realized culturally different diagnostic and therapeutic strategies (basic 4 vs supplement 5 points, P=.39) and were less affirmative when asked about the transferability of cases to a German context (basic 5 vs supplement 3 points, P=.048).

Conclusions: German students found English VPs to be highly clinically relevant, and they rated language problems much lower than they rated motivation to work on cases in English. This should encourage the intercultural exchange of VPs. The provision of supplemental explanatory material facilitates the recognition of cultural differences and might help prevent unexpected learning effects.

XSL·FO
RenderX

Introduction

Training with virtual patients (VPs) is a useful e-learning approach in medical education, complementing traditional curricular training strategies [1,2]. Using case-based e-learning software, practical knowledge and competence can be gained in a realistic setting [3]. Studies demonstrate that training with VPs improves aspects of knowledge and clinical reasoning in medical education [4-6]. Still, uncertainties exist regarding the optimal design and implementation of VPs [7,8], and the curricular implementation of this e-learning feature is rather limited [9].

A reason for the still limited use of VPs might be that creating high-quality medical e-learning material is a highly complex procedure and requires considerable time and knowledge resources [10,11]. Thus, it may be worthwhile to share existing high-quality VP cases among several medical institutions. In fact, an international exchange of VP cases might be attractive [12]. While English is widely accepted as the universal scientific language in medicine, the use of English e-learning material in general and VP cases in particular is not widespread in Germany. Therefore, the international exchange of VP cases might be quite challenging—obviously with regard to language barriers, but possibly also regarding cultural or national peculiarities in diagnostic or therapeutic procedures and guidelines as well as ethical aspects. Little is known about the effect of these specifics on the intercultural exchange of VPs. It is possible that cultural differences have an adverse effect on motivation or—even more problematic—on knowledge acquisition directly. German students using VPs from the United States for medical education might not be aware of cultural differences and thus acquire knowledge that might be, or seem, incorrect in their German medical context. On the other hand, it is possible that the recognition of these differences improves the motivation to deal with these cases and learn from contrasting effects.

We hypothesize that providing a supplemental description of the linguistic and cultural differences in VPs helps students identify and accept the differences and possibly use them as a stimulus for learning.

In this prospective study among German medical students, we investigated students' perception of virtual cases written in English and featuring a different cultural background. We also examined the effect of offering supplementary explanatory material on learning success, motivation, and recognition of differences.

Methods

To study the effect of cultural differences in working with VPs, we used the CASUS case-based learning environment [13]. The VPs were developed in a US medical context [10,14] and delivered in English. The CASUS platform uses a linear navigation concept, including varied task types, expert comments for feedback, and an integrated learning assessment of students' performance. In these intermittent assessments, the students are asked several questions in different formats (eg, multiple choice and free text) regarding diagnostic or therapeutic procedures. The learning environment allows recording the results of the learning assessments as well as the time on task to complete each step of a case.

One case from family medicine (*fmCase*) and one from internal medicine (*imCase*) were selected because they contain substantial cultural differences to German standards or procedures and were appropriately challenging for fifth-year German medical students. *fmCase* consisted of a 39-year-old male with epigastric pain, Mr. Rodriguez, and *imCase* was Mr. Ramirez, a 78-year-old man with fever, lethargy, and anorexia (see Multimedia Appendices 1 and 2 for details on the learning objectives each case).

Specifically, the *fmCase* presents a 39-year-old Latino immigrant with epigastric pain. In this patient, a *Helicobacter pylori* associated gastric ulcer is diagnosed and repeatedly treated with antibiotics according to the results of serum and stool tests. The resistance pattern and prevalence of *H. pylori* infection vary from region to region, resulting in different (national) guidelines regarding diagnosis (ie, timing of endoscopy) and treatment (eg, choice of antibiotics) [15,16].

The *imCase* involves a 78-year-old male who suffers from urosepsis complicated by mesenteric ischemia and ultimately dies. Here, the clinical management of sepsis and gastrointestinal bleeding and especially aspects of palliative care are of interest, which are clearly based on different cultural backgrounds. National recommendations, legislation, and regulations show international differences [17,18].

To facilitate the recognition and classification of cultural differences, a simple categorization of distinct features of VPs (cultural adaptation framework) was outlined. Accordingly, both VPs were screened for cultural differences using the following categories: (1) diagnostic procedures, (2) therapeutic procedures, (3) the professional role of the medical student or the physician (eg, interaction with the patient), (4) ethical principles, and (5) language (highly specific medical terms and abbreviations, units).

Based on the differences, we developed an explanatory worksheet (the supplement) for each case. In addition to a glossary of specific medical English terms, this supplement included a critical discussion of cultural differences in the cases, as well as information on the corresponding procedures within the German medico-cultural background.

We aimed to keep these instructions as short as possible (resulting in 736 words and 1207 words per instruction, respectively; see Multimedia Appendices 3 and 4 for the full text of instructions). In an initial test, 5 voluntary students commented on the supplemental material and the evaluation forms, which were revised accordingly.

We then instructed 14 voluntary students to complete the cases including the integrated learning assessments, allowing the use of textbooks, Web-based dictionaries, and unit converters (the basic group). Subsequently, 16 students were instructed to complete the same cases under the same conditions plus the supplementary material in German (the supplement group).

An evaluation form (using the software EvaSys V6.1, Electric Paper Evaluationssysteme GmbH) was developed and revised after initial testing. In this evaluation, we specifically addressed the recognition of cultural differences, in particular considering language, medical and ethical aspects, and the interaction with patients. Also, all students were asked to assess their own motivation to work with VPs in general and specifically in English. They used a 6-point Likert-like scale (with 1="strongly disagree" and 6="strongly agree"). The students in the supplement group were also asked to evaluate the usefulness and efficacy of the supplementary material. The students were also asked to comment on the cases using free text.

We processed the results using Microsoft Excel (2003) for statistical analysis. To describe the results in the following section, the median is given together with the mean and standard deviation, where appropriate. To test for statistical differences between the groups, the Mann-Whitney U test (for comparing ordinal responses of two groups) or the chi-square test (for comparing frequencies) were applied. After Bonferroni correction for multiple testing, statistical significance was defined as $P < .006$.

The ethics committee of the Medical Faculty of the Martin-Luther-University Halle-Wittenberg, Germany approved the project.

Results

Applying the suggested cultural adaptation framework on the two VPs, we identified 67 cultural differences. We identifed 58 differences (87%) as medical language issues, ten differences as diagnostic proceedings (eg, the relevance of endoscopy in diagnosing epigastric pain or lower gastrointestinal bleeding), five differences as therapeutic proceedings (eg, 81 mg vs 100 mg acetylsalicylic acid in coronary heart disease), three differences regarding the professional role of the physician or the medical student (eg, when obtaining informed consent), and one difference regarding decision making in a palliative situation (ethical/legal aspects).

In total, 33 students volunteered to participate in this study. Three students did not complete the cases and thus were not included in this analysis. The participants were fifth-year medical students (10 male and 20 female students, mean age 24 years [SD 2]; see Table 1), and all of them spoke German as their first language. The majority (29/33, 96%) of students reported learning English at school and many (22/33, 73%) of them had studied English for more than 7 years (Table 1). One student reported that he took English lessons at university. Six students (20%) spent more than 6 months in a country where English is the first language. During the last 6 months, the majority of participants (24/33, 83%) read at least one medical paper in English, and 5 students (16%) read at least one English medical textbook. There were no significant differences between the two groups regarding age, medical education (according to the year of medical education), or English language skills (as determined by classes at school or University).

When working with the cases, 23 students used Web-based English-German dictionaries (all students 77%, basic 73%, supplement 81%), 2 students used a paper-based dictionary, 13 students also used other online resources, and 7 students did not use any resources in addition to the material provided with the cases or supplements.

The results of the integrated learning assessments did not differ between the two groups of students: the basic and supplement groups answered 53% and 52% of the question items built into the VPs correctly (Table 1).

Table 1. Characteristics and performance of participants (learning assessment and required time).

	Basic	Supplement	All	*P* value
Students, n	14	16	30	
Age in years, mean (SD), median	24 (1), 24	25 (2), 24	24 (2), 24	.29[a]
Semester, mean (SD), median	8.9 (1.7), 9	9.9 (1.3), 10	9.4 (2.6), 10	.07[a]
Learned English >7 yrs in school, %	60%	81%	73%	.29[b]
Used e-dictionary, %	73%	81%	77%	.52[b]
Learning assessment, % correct answers, mean (SD), median	53 (4), 57	52 (4), 52	52 (4), 55	.32[a]
Time required (log files from both cases) in minutes, mean (SD), median	121 (52), 119	134 (38), 133.5	128 (44), 124	.45[a]

[a]Mann-Whitney U test

[b]Chi-square test.

The results of the evaluation are summarized in Table 2. Statements 1-4 refer to problems and motivation issues due to language problems. The students were clearly able to understand the instructions to the cases (Statement 3) and only reported minor language issues that potentially detracted from the actual cases (Statement 1). The students agreed very much that the clinical relevance of the cases was very understandable (Statement 2) and they felt motivated by working with English VPs (Statement 4). The provision of supplemental material did not influence language issues or motivation in this setting.

Statements 5 and 6 address differences in diagnostic and therapeutic procedures and the student-patient interaction . The transferability of the cases to a German cultural background

was considered possible by the basic students (Table 2, Statements 5-7), while supplement students were clearly more skeptical (Table 2, Statement 7).

Statements 8 and 9 (Table 2) were exclusively addressed by the supplement group. These students agreed that the additional supplement was helpful to identify cultural differences between the proceedings in the patient cases and the knowledge, which is imparted within a German cultural background. To a lesser extent, the supplemental material was considered helpful in working with the VPs in general.

The free text comments of the students were reviewed with respect to the identification of cultural differences. Three students specifically noted differences in diagnostic procedures, and 4 students noted differences in therapeutic procedures. One student commented that the role of the medical student in the case was different to the students' role in Germany. Three students pointed out differences in ethical aspects (insurance and accounting matters in medical care for immigrants, aspects of palliative care). Eight students specifically complained that they had to struggle with the English language or specific medical terms or abbreviations, and 2 students had problems with technical aspects of the Internet (eg, specific link not working, unpleasant layout).

Table 2. Results of the evaluation after completion of the two virtual patient cases. The students rated statements on a scale ranging from 6 (strongly agree) to 1 (strongly disagree).

	Basic	Supplement	All	P value[a]
1. I was busier with language problems than with the analysis of the case.	2.5	3	3	.29
2. The clinical relevance of the cases was understandable.	6	6	6	.42
3. For me it was difficult to understand the technical instructions in English.	1.5	2	2	.35
4. For me it was motivating to work with a case written in English.	4.5	4.5	4.5	.66
5. Diagnostic and therapeutic strategies are different to the strategies and procedures I am aware of.	4	5	4.5	.39
6. Medical students interacted with the patients in a way that was different from the interaction I am aware of.	2	3.5	2	.13
7. The described medical situation can be readily transferred to practices in Germany.	5	3	4	.048
8. I realized differences more readily using the additional information in the supplement.	n/a	5	n/a	
9. The supplements were helpful in working with the cases.	n/a	4	n/a	

[a]P values were calculated using the Mann-Whitney U Test.

Discussion

Principal Findings

Knowledge of cultural differences in medicine and medical education is very important, especially when interacting and communicating with patients or colleagues with a different cultural background. Typically, the term "cultural differences" is used to describe aspects of the doctor-patient or doctor-doctor relationship and refers to different explanatory models of health and illness, different cultural values, cultural differences in patient preferences for doctor-patient relationships, racism/perceptual biases, and linguistic barriers [19]. In this context, the development of "cultural competence" in medical education is considered to be highly important in medical education [20,21].

The cultural adaptation framework based on the experience with the 2 VPs in our study demonstrates that the term "cultural differences" should refer not only to the doctor-patient-relationship, but also to differences in diagnostic and therapeutic procedures, which may be caused by economic, epidemiologic, and historical differences. As no comparable framework for cultural adaptation exists so far, the proposed categorization of cultural differences in this context can be useful in future research as well as in adapting actual VPs for cross-cultural use.

One important result of this study is that VPs written in English are well accepted by German medical students: the students reported that working with the VPs was motivating, and despite minor language-related issues, the clinical relevance was very well understood (Table 2, Statements 1-4). Because intrinsic motivation is an important driver of learning activities [22], these results are clearly encouraging when discussing the interinstitutional and international exchange and availability of VPs in medical education [12].

Interestingly, although the supplement students received a concise list with translations of medical terms and abbreviations, they reported language-related difficulties similar to the students in the basic group (Table 2, Statement 1). Students more readily relied on their familiar (Web-based) translation resources than on looking up terms in the prepared list: 23 of the 30 students used online dictionaries, with no significant difference between the basic and supplement groups. Considering that the majority of cultural differences in our VPs consisted of linguistic peculiarities, this might indicate that the simultaneous use of supplements and Web-based dictionaries results in redundancy and possibly cognitive overload [23]. In order to reduce the cognitive load, it seems advisable to reduce the help translating

English medical terms, especially since the students are well able to help themselves through language problems.

Our results underline the importance of providing additional information on cultural aspects of VPs. Similar observations have been made when using TV dramas in medical education [24] . Williams et al described "unexpected learning outcomes" in this context, which are clearly related to cultural differences, and strongly recommend a "reflection component" when using TV dramas for medical education. In our study, this reflection component consisted of written supplemental information. Discussions in small groups might be more effective, but they are more demanding as well.

Limitations

The students' evaluation has to be interpreted with some caution, as the recruitment of the students (self-selective, not randomized sample: voluntary participation) might have resulted in a substantial selection bias. It cannot be excluded that above-average motivated students took part in this study.

Conclusion

The study results support our hypothesis that providing a supplemental description of the linguistic and cultural differences in VPs might help students identify and accept those differences, especially with regard to cultural differences in general. The supplemental material regarding linguistic differences was considered less valuable. If students are allowed to work in their normal context (ie, working online with access to Web-based dictionaries), linguistic information should be restricted to avoid redundant information and cognitive overload. However, it should not be left to the students themselves to identify important cultural differences in this context. The unsupervised use of such cases in extracurricular or private study generally might be beneficial but problematic if certain aspects of cultural differences are not noticed by students. When recommending the use of foreign VPs, it would be advisable to ensure that the medical procedures, ethical principles, or role models taught in the cases fit to the knowledge and competences that are intended. Specific cultural differences should be identified and discussed. This might result in an intensified work-up of the cases and might possibly lead to a better understanding of the learning objectives.

Conflicts of Interest

Multimedia Appendix 1

Learning objectives fmCASE.

[PDF File (Adobe PDF File), 152KB - jmir_v18i9e260_app1.pdf]

Multimedia Appendix 2

Learning objectives imCase.

[PDF File (Adobe PDF File), 139KB - jmir_v18i9e260_app2.pdf]

Multimedia Appendix 3

Supplemental material fmCase.

[PDF File (Adobe PDF File), 135KB - jmir_v18i9e260_app3.pdf]

Multimedia Appendix 4

Supplemental material imCase.

[PDF File (Adobe PDF File), 156KB - jmir_v18i9e260_app4.pdf]

References

1. Sandars J, Homer M, Pell G, Croker T. Web 2.0 and social software: the medical student way of e-learning. Med Teach 2008;30(3):308-312. [doi: 10.1080/01421590701798729] [Medline: 18608950]
2. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
3. Huwendiek S, Dern P, Hahn EG, Pediaditakis D, Tönshoff B, Nikendei C. [Training requirements, expertise and basic conditions of committed medical educators in Germany]. Z Evid Fortbild Qual Gesundhwes 2008;102(10):613-617. [Medline: 19402347]

XSL•FO

RenderX

4.   Funke K, Bonrath E, Mardin WA, Becker JC, 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]

5.   Salminen H, Zary N, Björklund K, Toth-Pal E, Leanderson C. Virtual patients in primary care: developing a reusable model that fosters reflective practice and clinical reasoning. J Med Internet Res 2014;16(1):e3 [FREE Full text] [doi: 10.2196/jmir.2616] [Medline: 24394603]

6.   Poulton T, Ellaway RH, Round J, Jivram T, Kavia S, Hilton S. Exploring the efficacy of replacing linear paper-based patient cases in problem-based learning with dynamic Web-based virtual patients: randomized controlled trial. J Med Internet Res 2014;16(11):e240 [FREE Full text] [doi: 10.2196/jmir.3748] [Medline: 25373314]

7.   Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]

8.   Kleinert R, Wahba R, Chang D, Plum P, Hölscher AH, Stippel DL. 3D immersive patient simulators and their impact on learning success: a thematic review. J Med Internet Res 2015;17(4):e91 [FREE Full text] [doi: 10.2196/jmir.3492] [Medline: 25858862]

9.   Waldmann U, Vollmar H, Stracke S, Fassnacht U, Gensichen J, Sönnichsen A, et al. Überblick über Patientensimulationsprogramme - Hintergründe, Möglichkeiten und Einsatz in der Lehre. Z Allg Med 2006 Dec;82(12):536-542. [doi: 10.1055/s-2006-942313]

10. Berman NB, Fall LH, Chessman AW, Dell MR, Lang VJ, Leong SL, et al. A collaborative model for developing and maintaining virtual patients for medical education. Med Teach 2011;33(4):319-324. [doi: 10.3109/0142159X.2011.540268] [Medline: 21456990]

11. Georg C, Zary N. Web-based virtual patients in nursing education: development and validation of theory-anchored design and activity models. J Med Internet Res 2014;16(4):e105 [FREE Full text] [doi: 10.2196/jmir.2556] [Medline: 24727709]

12. Berman NB, Durning SJ, Fischer MR, Huwendiek S, Triola MM. The Role for Virtual Patients in the Future of Medical Education. Acad Med 2016 Sep;91(9):1217-1222. [doi: 10.1097/ACM.0000000000001146] [Medline: 26959224]

13. Simonsohn AB, Fischer MR. [Evaluation of a case-based computerized learning program (CASUS) for medical students during their clinical years]. Dtsch Med Wochenschr 2004 Mar 12;129(11):552-556. [doi: 10.1055/s-2004-820543] [Medline: 14997405]

14. Leong SL. fmCASES: Collaborative Development of Online Cases to Address Educational Needs. Ann Fam Med 2009;7(4):374-375 [FREE Full text] [doi: 10.1370/afm.1028] [Medline: 19597176]

15. Lee S. Current progress toward eradicating Helicobacter pylori in East Asian countries: differences in the 2013 revised guidelines between China, Japan, and South Korea. World J Gastroenterol 2014 Feb 14;20(6):1493-1502 [FREE Full text] [doi: 10.3748/wjg.v20.i6.1493] [Medline: 24587624]

16. Malfertheiner P, Venerito M, Selgrad M. Helicobacter pylori infection: selected aspects in clinical management. Curr Opin Gastroenterol 2013 Nov;29(6):669-675. [doi: 10.1097/MOG.0b013e328365d443] [Medline: 24100726]

17. Van Beek K, Woitha K, Ahmed N, Menten J, Jaspers B, Engels Y, et al. Comparison of legislation, regulations and national health strategies for palliative care in seven European countries (Results from the Europall Research Group): a descriptive study. BMC Health Serv Res 2013 Jul;13:275 [FREE Full text] [doi: 10.1186/1472-6963-13-275] [Medline: 23866928]

18. Koh M, Hwee PC. End-of-life care in the intensive care unit: how Asia differs from the West. JAMA Intern Med 2015 Mar;175(3):371-372. [doi: 10.1001/jamainternmed.2014.7397] [Medline: 25581079]

19. Schouten BC, Meeuwesen L. Cultural differences in medical communication: a review of the literature. Patient Educ Couns 2006 Dec;64(1-3):21-34. [doi: 10.1016/j.pec.2005.11.014] [Medline: 16427760]

20. Betancourt JR. Cultural competence and medical education: many names, many perspectives, one goal. Acad Med 2006 Jun;81(6):499-501. [doi: 10.1097/01.ACM.0000225211.77088.cb] [Medline: 16728795]

21. Seeleman C, Suurmond J, Stronks K. Cultural competence: a conceptual framework for teaching and learning. Med Educ 2009 Mar;43(3):229-237. [doi: 10.1111/j.1365-2923.2008.03269.x] [Medline: 19250349]

22. Kusurkar RA, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: a review of the literature. Med Teach 2011;33(5):e242-e262. [doi: 10.3109/0142159X.2011.558539] [Medline: 21517676]

23. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist 2003 Mar;38(1):43-52. [doi: 10.1207/S15326985EP3801_6]

24. Williams R, Evans L, Alshareef NT. Using TV Dramas in Medical Education. Educ Prim Care 2015 Jan;26(1):48-49. [Medline: 25654807]

Abbreviations

VP: virtual patient

XSL•FO
RenderX

Impact of Game-Inspired Infographics on User Engagement and Information Processing in an eHealth Program

Maria Leonora G Comello[1], PhD; Xiaokun Qian[1], MA; Allison M Deal[2], MS; Kurt M Ribisl[2,3], PhD; Laura A Linnan[2,3], ScD; Deborah F Tate[2,3], PhD

[1]School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Maria Leonora G Comello, PhD
School of Media and Journalism
University of North Carolina at Chapel Hill
Carroll Hall, CB 3365
Chapel Hill, NC, 27599
United States
Phone: 1 919 928 2440
Fax: 1 919 962 0620
Email: comello@email.unc.edu

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2017/1/e12/

Abstract

Background: Online interventions providing individual health behavior assessment should deliver feedback in a way that is both understandable and engaging. This study focused on the potential for infographics inspired by the aesthetics of game design to contribute to these goals.

Objective: We conducted formative research to test game-inspired infographics against more traditional displays (eg, text-only, column chart) for conveying a behavioral goal and an individual's behavior relative to the goal. We explored the extent to which the display type would influence levels of engagement and information processing.

Methods: Between-participants experiments compared game-inspired infographics with traditional formats in terms of outcomes related to information processing (eg, comprehension, cognitive load) and engagement (eg, attitudes toward the information, emotional tone). We randomly assigned participants (N=1162) to an experiment in 1 of 6 modules (tobacco use, alcohol use, vegetable consumption, fruit consumption, physical activity, and weight management).

Results: In the tobacco module, a game-inspired format (scorecard) was compared with text-only; there were no differences in attitudes and emotional tone, but the scorecard outperformed text-only on comprehension (P=.004) and decreased cognitive load (P=.006). For the other behaviors, we tested 2 game-inspired formats (scorecard, progress bar) and a traditional column chart; there were no differences in comprehension, but the progress bar outperformed the other formats on attitudes and emotional tone (P<.001 for all contrasts).

Conclusions: Across modules, a game-inspired infographic showed potential to outperform a traditional format for some study outcomes while not underperforming on other outcomes. Overall, findings support the use of game-inspired infographics in behavioral assessment feedback to enhance comprehension and engagement, which may lead to greater behavior change.

(J Med Internet Res 2016;18(9):e237)   doi:10.2196/jmir.5976

KEYWORDS

infographics; game design; eHealth; personalized feedback; visuals

Introduction

Online health behavior assessments have become important tools for monitoring and motivating individual health behavior change [1-3]. Behavior change is possible when assessments provide a combination of information and personalized feedback [4-6]. However, to be effective, the feedback must be clear and engaging. Users who find the information confusing, uninteresting, or even discouraging may be deterred from further interaction. Thus, in addition to being comprehensible, content should also be engaging to retain as many users as possible who may benefit from continued participation [7].

This study focused on the potential of infographics to address these communication challenges when used to deliver feedback following online health behavior assessments. Infographics have been defined as a visualization of data or ideas to convey information in a way that is easily understood [8]. Although infographics are frequently used to deliver health information, there is a lack of research on the conditions under which different formats may be most effective [9]. We addressed this gap by examining multiple infographic formats as part of feedback for 6 health behaviors. The study included formats inspired by gamification, which is the use of game design elements in nongame contexts [10]. The key issue was whether game-inspired infographics (ie, infographics inspired by aesthetic and other elements of games) would be more effective at facilitating comprehension, positive attitudes toward the information, positive emotional responses, and other outcomes than the more traditional formats that have been used to represent behavioral feedback.

We explored this issue through pretesting visual elements for use in the Carolina Health Assessment and Resource Tool (CHART). CHART is an online tool developed at the University of North Carolina at Chapel Hill, USA, that offers evidence-based assessment for health behaviors that have been linked to the leading causes of death due to chronic disease in the United States [11]. The 6 behavioral modules addressed in this study were tobacco use, alcohol use, physical activity, vegetable consumption, fruit consumption, and weight management. CHART provides feedback that includes a statement of the individual's level of behavior (based on questions answered in the assessment) relative to the recommended level for that behavior. Although CHART output includes other personalized feedback components, this study focused only on the display of information regarding behavior relative to the recommended level.

We conducted formative research to guide selection of simple visuals that would accompany this feedback. Formative research refers to the stage prior to implementation of a health communication effort in which messages and strategies are evaluated for the likelihood of achieving intended effects [12]. Strategies can include experimental assessment of stimuli (sometimes called pretesting) to identify promising formats [13]. In particular, the evaluation of visual elements in the formative research stage is advocated by the US National Cancer Institute [14] for effective health communication; however, the process and results of such pretests are not often reported. Thus, this study contributes to knowledge of best practices in the development and systematic assessment of health infographics for use in online environments.

Visuals and Infographics in Health Communication

Infographics are an increasingly common feature in the online media landscape because of their reputation for presenting information in a visually compelling and easy to process manner [8]. The most recent reviews of the use of infographics and other visuals in health contexts [15-17] suggest that visuals add to the impact of health information in terms of attention, comprehension, recall, intention, and behavior change. For people with low health literacy, the beneficial effect of visuals as an aid to comprehension is even more pronounced [17]. Further, certain kinds of visuals such as avatars in risk infographics may help people better comprehend how statistical risk information can apply to an individual [18]. It should be noted that many of these studies involved the delivery of quantitative risk information (eg, number of people affected by a type of cancer in a population) and not information on levels of individual behavior relative to a recommended level (*personalized* behavioral feedback), which was the focus of our study. Given the importance of tailoring in eHealth interventions [19], our study makes an independent contribution by examining infographic effects in the context of promoting individual risk factor change.

At the most basic level, studies have shown that adding visuals to text results in desirable outcomes from a health promotion perspective. For example, patients receiving discharge information as text with visuals had better outcomes (compared with those receiving text-only) in terms of attention to information, recall of information, and adherence to instructions [20]. An explanation for the superiority of formats that combine visuals with text over text-only is based on dual coding theory [21], which proposes that there are 2 cognitive systems, one specialized for the processing of language and the other for nonlanguage-based information. Although the systems are distinct, activity in one system can initiate activity in the other. Thus, information combining text with visuals is more likely to be encoded and comprehended than is information presented in text only [22].

Graphical formats may also influence perceptions of risks in several ways. For example, some formats can make some aspects of risk information more salient than others. A systematic review [15] showed that graphs that visualize part-to-whole relationships may draw attention to the relationship between the number of people affected by a hazard and the entire population at risk, whereas graphs that show only the number of people at risk may increase perceived risk and increase the likelihood of protective behavior. Thus, visuals have a role in framing risk information by providing cues to interpretation that may not otherwise be present. Also, the visual complexity of displays can influence perceptions of risk. Displays of quantitative risk information that contained animation resulted in worse performance (compared with a static display) on indicators of comprehension [23]. Thus, although more complex displays might seem to be preferable given a higher potential

of being encoded (as would be suggested by dual coding theory), in some cases, simpler displays may be optimal.

Although assessments of graphs tend to focus on accuracy of perceptions and other cognitive outcomes, it is also important to look at the appeal of visual elements. In a real-world setting, people simply may not accept or attend to graphics they do not like [15]. This may be the case in self-guided assessments where people can exit at any time. Ancker and colleagues [15] found that participants preferred simple graphics and human figures over nonhuman figures. For example, in a qualitative study evaluating the responses of women to visual formats used to convey breast cancer risk, human figures were preferred over bar graphs because the human figures were perceived as relatable and as conveying a meaningful message [24].

A theoretical model of persuasion that provides a framework for thinking about both cognitive and affective outcomes as part of a larger process of persuasion is the hierarchy of effects model [25], which conceptualizes comprehension as one step among many in the behavior change process. Given the roots of the model in consumer behavior research, a key construct within the model is attitude toward the advertisement, an affective construct representing feelings of favorability that can be used in the assessment of advertising and other materials designed to have persuasive effects (eg, health messages). Studies have shown that attitude toward the advertisement mediates the effects of exposure on brand perceptions and purchase intentions [26-29] and on actual behavior [29]. In a health framework, positive evaluations of advertisements have an impact on perceptions of health-relevant product categories [30]. Although infographics are not commonly associated with advertising or promotion, the personalized feedback accompanying health behavior assessments is intended to promote progress toward healthy behavioral outcomes. Therefore, infographics used in such feedback should be evaluated for their impact on attitudes and emotions so that their role in supporting behavior change can be better understood.

Game-Inspired Design

The infographics employed in this study were inspired by concepts underlying gamification, which has been defined as the use of game features in nongame contexts to motivate and engage users [10]. Organizations have used gamification to encourage targets of influence to complete tasks that may be viewed as boring, unpleasant, or unnecessary by making the tasks seem more rewarding and fun. For example, gamification has been used for facilitating workplace training [31] and cultivating consumer engagement [32]. "Serious games" have been receiving attention particularly in the domain of health [33,34], where the games have been used for a variety of interventions, including increasing physical activity [35,36], slowing cognitive decline [37], improving driving skills [38], and educating on self-management of health conditions [39].

Common game design features include a mechanism for providing feedback on progress toward goals, as well as rewards for achieving desired behaviors. Feedback can be in the form of scorecards that provide information on user performance, much like a performance summary page that one might encounter in a video game. It is also common for websites to

encourage completion of tasks (eg, constructing user profiles) by adding a progress bar that depicts how close a user is to task completion. In a consumer behavior study, Cheema and Bagchi [40] found that the easy-to-visualize format of a progress bar may increase motivation relative to more difficult-to-visualize formats for people who are close to the goal. In addition to features that provide performance feedback and encourage task completion, another hallmark of game design is the use of playful elements that aim to elicit the experience of fun [41] and can include visuals that are amusing, interesting, provocative, or distinctive in ways that would be viewed as a pleasant surprise. Although some scholars have drawn a distinction between gamefulness and playfulness [10], in practice, users may not distinguish between them.

Because gamification encompasses multiple components in an integrated system [42], it should be noted that the focus of this pretest was not on testing an integrated gamified system designed to change the behaviors monitored by CHART. Rather, given our formative research goal, the study focused on pretesting visual elements (including some with progress bars and visually interesting elements) as potential improvements to the personalized feedback provided by CHART. In the gamification community, the term game-inspired design describes the general use of aesthetic features, narrative tones, and other elements borrowed from games in the design of other objects [43]. Therefore, we describe the infographics with these features as being game inspired rather than as part of a gamified system for this pretest.

There are compelling reasons to consider game-inspired design in the development of visuals for health behavior assessments. Game-inspired infographics can include playful elements that may elicit positive attitudes and affect. Based on the hierarchy of effects model [25], these attitudes may serve as antecedents to the desired behavior. Furthermore, game-inspired infographics may cue interpretation in ways that are helpful to health goals. The progress bar, for instance, implies the presence of a desirable goal, which is a concept that is not built into more conventional displays of quantity such as the standard column chart. In emphasizing the presence of a desirable goal, a progress bar may serve as a way to frame the advocated behavior in terms of gains from meeting the goal rather than losses from not meeting the goal. From a health perspective, the literature on gain-versus-loss framing shows that gain framing is especially well suited to prevention contexts [44], in line with the behavior change aims of the CHART intervention program. Moreover, because health behaviors are often associated with multiple and sometimes competing attitudes [45], infographics could increase the accessibility of associations that support health goals, which would in turn have the potential to strengthen the link between attitude and behavior [46].

In summary, visuals in health contexts can affect both information processing and attitudinal outcomes, and it is important to examine effects in both domains for the purposes of pretesting. Based on the studies reviewed, our formative test was guided by the following expectations and questions. First, we proposed that infographics would facilitate information processing (including comprehension, recall, and perceptions of processing ease) over text-based formats. Second, we

expected that game-inspired infographics would increase engagement (attitudes toward the information, emotional tone, and perceived effectiveness) more effectively than traditional infographic formats within the context of CHART personalized health feedback. Third, to gain further insight, we also explored whether any effects of format might depend on whether an individual has already successfully met the goal for the behavior. Individuals who have not met behavioral goals are of particular importance because they represent those who might benefit the most from persuasive efforts.

Methods

Participants and Procedure

To test the effects of infographic format, we conducted posttest-only, between-participants experiments administered online by Qualtrics (Provo, Utah). We recruited participants (N = 1162) on Amazon.com's Mechanical Turk (Amazon.com, Inc, Seattle, WA, USA), a crowdsourcing Internet marketplace for recruiting workers to complete tasks. Research has demonstrated the value of Mechanical Turk as a recruiting tool in experimental [47] and public health research [48]. The study was advertised as a short survey that offered US $1.25, which was a typical incentive for surveys of this duration at the time of data collection. Individuals were eligible if they were 18–65 years old, able to communicate in English, and living in the United States.

There were 6 concurrent experiments, with 1 for each behavior (ie, tobacco use, alcohol use, vegetable consumption, fruit consumption, physical activity, and weight management). We first randomly assigned participants to behavioral modules. Then, within modules, we randomly assigned participants to condition (traditional or game-inspired format, as described below under the Stimuli subheading). Participants answered questions assessing information processing and engagement. These were followed by items that were not related to the study and served as distractors to enable measurement of recall at the end. Finally, participants viewed items measuring potential moderators, demographics, and recall. Participants received a unique code to enter in Mechanical Turk to confirm completion. The study was approved by the University of North Carolina at Chapel Hill Institutional Review Board, where the research was conducted.

Stimuli

Overall Approach

Stimuli were developed by the same professional graphic designer to ensure equivalence in execution quality and style. Participants were exposed to sample results of a health behavior assessment that a hypothetical individual would receive. Within each module, all of the formats presented the same level of individual behavior and the same evidence-based goal. In all modules, the hypothetical user's behavior is short of the goal but would likely not be considered obviously unhealthy. To increase the likelihood that participants would attend to hypothetical feedback, we asked participants to imagine that the results belonged to a friend. By having the results attached to a specific person, we aimed to make the information more concrete, which is perceived as more useful and relevant to decision making than abstract information [49]. Figure 1 displays all formats used by condition and module.

Traditional Formats

In all modules, there was a condition that represented the existing traditional formats that were in use by CHART at the time of the study. In the tobacco module, this format was a text-only display. In the other modules, the format was a traditional column chart, with the vertical bar displaying the individual's level of behavior and a line indicating the recommended level. The traditional column charts also included text identifying the goal and the individual's level of behavior.

Game-Inspired Formats

To serve as comparisons with traditional formats, we created 2 game-inspired formats. In the *scorecard* format, the game-inspired element was the visual appearance of a game scorecard. As adapted for this study, the scorecard included a side-by-side display of "yes" and "no" checkboxes to indicate whether a person has met the goal. This visual was accompanied by text that stated the goal and the individual's level of behavior. We used the scorecard format in all modules in this executional style. In the *progress bar* format, a game-inspired element was the gamification concept of a horizontal bar displaying the level of achievement toward task completion. As with the scorecard, this format also included text that stated the goal and the individual's level. We did not use the progress bar format in the tobacco module, which provides only dichotomous yes/no feedback of whether one has met the goal of being tobacco free. However, we did use the progress bar in all other modules, which provide feedback on an ordinal numerical scale.

In addition to the visual appearance of a progress bar, other game-inspired features in this format included playful elements and cues to action that were consistent with the advocated behaviors. For example, the progress bar for eating fruits included pieces of fruit and a fork, and the progress bar for physical activity was represented by a figure running toward a finish line. Although the general concept of a progress bar entails an increase from left to right toward a fixed goal, we had to modify the concept for the modules of weight and alcohol because, in these cases, behavioral goals do not involve a straightforward increase. Rather, alcohol goals require ensuring that use does not exceed a particular level, while weight goals require keeping body mass index within a healthy range. Thus, modifications of the progress bar were developed for greater congruence with these goals.

Game-inspired conditions in each module were represented by single exemplars, which we chose as follows. The graphic designer developed several mock-ups as candidates to represent each condition. For the progress bar executions, the designer was instructed to include playful elements and behavioral cues where appropriate. The research team reviewed all options and selected the one that was the clearest and most reasonable example of each condition, given the behavior and the overall context of the website. The research team suggested modifications to the designs and reviewed final versions prior to use.

Figure 1. Stimuli by module and condition.[a]Conditions are traditional (text-only for tobacco; column for nontobacco), progress (game‐inspired bar), and scorecard (game‐inspired side‐by‐side comparison).[b]Progress bar format was not applicable (N/A) to tobacco module, which provides dichotomous feedback.

Dependent Measures

Information Processing

Comprehension was measured with multiple-choice questions that appeared along with the infographic. The purpose was to assess whether participants could interpret the graphic correctly. Questions asked what the health goal is, whether the goal was met, and (if appropriate for the module) whether meeting the goal required an increase or decrease in current behavior and by how much. The number of correct answers was summed. Because there was a large proportion of all-correct responses (reported in results), we dichotomized the variable as all-correct or not.

Recall of the goal was measured at the end of the survey with a single item that was not accompanied by the feedback display viewed earlier. The item appeared after many distractor items, and the purpose was to assess whether information could be retrieved. We coded responses as correct or not.

Time to answer comprehension questions was collected as an unobtrusive measure of information processing ease, with shorter times indicating greater ease. Collection of this data was made possible by selecting the option in Qualtrics that records response times to questions. Values were adjusted for the number of questions in the module. Also, because response latencies are typically skewed, we log transformed the data before analysis, per standard practice [50].

Perceived cognitive load measures self-reported difficulty in processing the information [51]. The following items appeared immediately after the comprehension questions: (1) "I had difficulty understanding the infographic," (2) "I felt 'lost' when interpreting the infographic," and (3) "It was clear how the information fit together" (reverse coded). Response options ranged from 1 (not at all) to 5 (a lot). Items were averaged (Cronbach alpha = .82).

Engagement

Attitude toward the infographic was measured with 7 bipolar adjectives adapted from the Attitude Toward the Ad scale [30]. Participants were asked to what extent (range 1–5) they thought the infographic looked not cool/cool, boring/interesting, unpleasant/pleasant, unappealing/appealing, not likable/likable, unexciting/exciting, and unattractive/attractive. Items were averaged (Cronbach alpha = .94).

Positive emotional tone included 3 items adapted from Pechmann and Reibling [52]. Participants were asked to what extent looking at the infographic made them feel amused, happy, and upbeat. Response options ranged from 1 (not at all) to 5 (a lot). Items were averaged (Cronbach alpha = .88).

Perceived effectiveness was measured with 3 items that asked the extent to which participants thought the infographic would be an effective way to provide information, would be valuable to the recipient, and would be motivating. The items are similar to those used in assessments of persuasive messages [53]. Response options ranged from 1 (not at all) to 5 (a lot). Items were averaged (Cronbach alpha = .88).

Moderator Variable

Participants' own behavior was assessed in the module to which they were assigned with the same items used in CHART to assess that behavior. We adapted these items from validated sources and describe them here briefly. For tobacco, we asked participants whether they smoked cigarettes every day, some days, or not at all; current smoking was defined as using every day or some days [54]. For alcohol, we asked participants about consumption in the past 30 days [54]. For physical activity, participants provided the number of minutes of physical activity they get in a typical week [55]. For the modules of vegetable and fruit consumption, participants reported how many cups they eat in a typical day [56]. For weight management, participants reported height and weight, which we used to compute their body-mass index. For all modules, we coded responses as meeting the goal or not. The goals for each behavior appear on the infographics in Figure 1. Percentages of participants not meeting goals are reported below.

Analysis Plan

General Approach

Although the 6 between-participants experiments conducted in this pretest would essentially allow analysis of multiple replication studies, such an approach would be lengthy to report and would not answer the question of which format overall would be best suited to the website and whether there were boundary conditions imposed by module on condition effects. Thus, we proceeded as follows. We analyzed the data in the tobacco and nontobacco modules separately because they differed in number of conditions compared (2 in tobacco, 3 in nontobacco). For data in nontobacco modules, we first examined the interaction of module and condition to see whether the effect of condition depended on module. There was no significant interaction, indicating the consistency of effects across the experiments in the nontobacco modules; therefore, we combined the data. A fixed-effects approach was used given the small number of nonrandom messages employed (see [57], for arguments supporting fixed-effects over random-effects approaches in similar contexts). We included module as a covariate to account for possible differences in outcomes based solely on module. There were no differences in distributions of demographic variables across the study groups (*P* values ranging from .15 to .80). As a result, we did not use demographics and other attributes as covariates.

Main Effects of Condition

For categorical outcome variables of comprehension and recall, we used logistic regression to model the effect of infographic condition on each outcome. For the continuous information processing outcomes (cognitive load and time to complete), we first ran multivariate analysis of variance to evaluate the overall effect of condition. If the Wilks' lambda test was significant, we ran univariate analysis of variance to evaluate the effect on individual outcomes. Significant univariate results were followed by pairwise comparisons when there were more than 2 groups being compared (ie, for nontobacco data). We used the same process for the continuous outcomes related to engagement (attitude, positive emotional tone, and perceived effectiveness).

XSL•FO
RenderX

Moderation of Condition Effects

We evaluated the potential moderator of whether one has met the goal by using an interaction term. If we found a significant interaction, we report results for the effect of infographic separately for those who did and did not meet the goal.

Results

Sample Characteristics and Assignment to Module/Condition

Table 1 summarizes sample characteristics. Table 2 provides the percentages of participants who did not meet the behavioral goals for each module. The percentages range from 25% for tobacco (ie, 25% used tobacco) to 86% for vegetable consumption (ie, 86% did not eat the recommended amount). Table 2 also shows the numbers of participants who were randomly assigned to each module and to each condition within the module.

Main Effects of Condition

Tobacco

We compared 2 conditions: game-inspired (scorecard) and traditional (text-only). For information processing outcomes, participants who viewed the scorecard (vs text-only) were more likely to get all comprehension questions correct (98.1% vs 71.2%, $P=.004$) and to get the recall question correct (94.2% vs 82.7%), although the difference in percentages for recall was only marginally significant ($P=.08$). Multivariate analysis showed a significant association of infographic type on the outcomes of perceived cognitive load and time to answer questions ($P=.02$). Univariate analyses revealed that participants who viewed the scorecard reported lower cognitive load than did those who viewed text-only (estimate $-.54$, $P=.006$). However, in terms of time to answer, the difference was not significant ($P=.70$). Figure 2 shows comparisons of means and percentages for information processing outcomes by condition.

For engagement-related outcomes, multivariate analysis with the 3 dependent variables (attitude toward the infographic, positive emotional tone, and perceived effectiveness) did not reveal an effect of condition ($P=.42$), so we did not probe them further (Figure 3).

Table 1. Characteristics of participants (N=1162) in experiments assessing infographics.

Characteristic	Mean (SD)	No.	%
Age in years	32.5 (11.1)		
Sex			
Male		624	53.70
Female		538	46.30
Race			
White		939	80.81
Black		89	7.66
American Indian		5	0.43
Asian		84	7.23
Native American		7	0.60
Hawaiian/Pacific Islander		6	0.52
Multiple		32	2.75
Ethnicity			
Hispanic		76	6.54
Non-Hispanic		1085	93.37
Education			
Less than high school		6	0.52
High school/GED[a]		115	9.90
Some college		341	29.35
2-year college		127	10.93
4-year college		452	38.90
Master's degree		92	7.92
PhD, JD, or MD degree		29	2.50

[a]General equivalency diploma.

XSL•FO

RenderX

Table 2. Cell sizes by module and condition, and percentages of participants not meeting behavioral goals by module.

Module	Condition[a]			No. (%)[b]	No. (%) not meeting goal
	Traditional	Progress	Scorecard		
Tobacco use	52	N/A[c]	52	104 (8.95)	26 (25.0)
Alcohol use	58	59	58	175 (15.06)	88 (50.3)
Physical activity	56	60	55	171 (14.72)	93 (54.4)
Vegetable consumption	51	52	53	156 (13.43)	134 (85.9)
Fruit consumption[d]	96	88	93	277 (23.84)	195 (70.4)
Weight management[d]	95	91	93	279 (24.01)	154 (55.2)

[a]Conditions are traditional (text-only for tobacco; column for nontobacco), progress (game‐inspired bar), and scorecard (game‐inspired side‐by‐side comparison).

[b]Total participants for all modules: N=1162.

[c]Progress bar format was not applicable (N/A) to tobacco module, which provides dichotomous feedback.

[d]More participants were recruited for modules of fruits and weight (relative to other modules) to allow for analysis of variables not related to this study.

Figure 2. Percentages and means for information processing outcomes by condition and module. Bars extend to upper and lower bounds of the 95% CIs. For comprehension and recall, percentages reflect those who answered correctly. For perceived cognitive load, means are on a 1–5 scale; lower means denote lower perceived load (ie, easier processing). Time to answer comprehension questions was measured in seconds and was adjusted for number of questions in the module; lower means denote less time taken to answer (ie, easier processing). Traditional condition was text-only for the tobacco module, and traditional column chart for nontobacco modules.
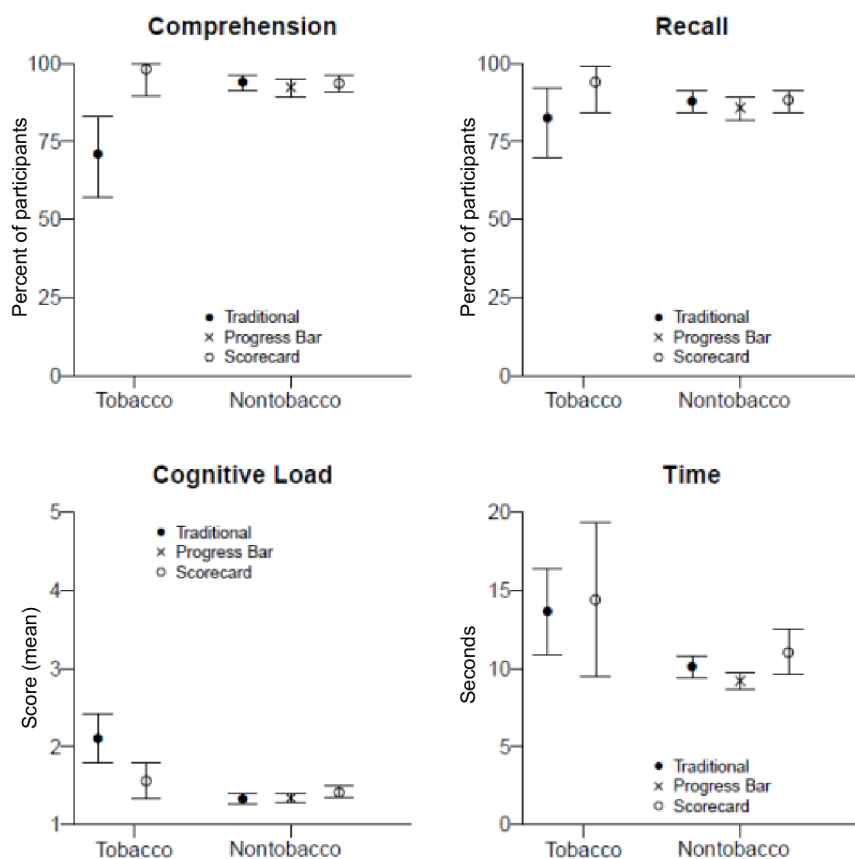
Figure 3. Means for engagement outcomes by condition and module. Bars extend to upper and lower bounds of the 95% CIs. Higher means denote more positive emotional tone, greater perceived effectiveness, and more positive attitude toward the infographic. Traditional condition was text-only for the tobacco module and traditional column chart for nontobacco modules.
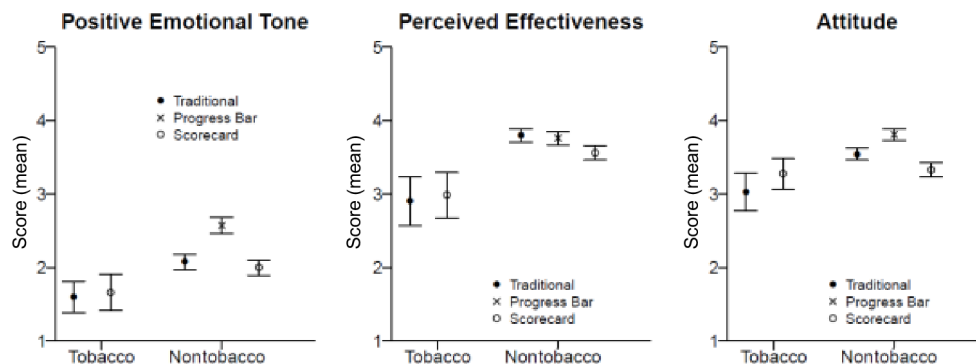
Health Behaviors Other Than Tobacco

We compared 3 conditions: 2 game-inspired (scorecard and progress bar) and 1 traditional (column chart). For information processing outcomes, both comprehension and recall (≥85%) were high across conditions, and pairwise contrasts among the 3 conditions did not show differences (all $P \geq .20$ for both outcomes). Multivariate analysis of perceived cognitive load and time to answer questions was not significant ($P=.09$) (Figure 2).

For engagement-related outcomes, multivariate analysis revealed a significant association ($P<.001$), and univariate analyses revealed significant differences for each of the 3 outcomes (attitude toward the infographic, positive emotional tone, and perceived effectiveness) (Figure 3).

For attitude toward the infographic, participants who viewed the progress bar had the most positive attitudes, followed by those who saw the traditional bar, and then the scorecard (mean scores 3.81, 3.54, and 3.33, respectively). Pairwise contrasts showed that all means were significantly different from each other ($P<.001$ for all).

For positive emotional tone, participants who viewed the progress bar gave the highest ratings (ie, extent to which the infographic made them feel amused, happy, and upbeat), followed by those who saw the traditional bar, and then the scorecard (mean scores 2.57, 2.08, and 2.00, respectively). The significant contrasts were progress versus scorecard, and progress versus traditional bar ($P<.001$ for both contrasts).

For perceived effectiveness, there were no differences between participants who viewed the progress bar and the traditional bar (in terms of perceptions of providing information, motivating behavior, and being valuable). The mean scores were 3.76 and 3.80 ($P=.59$). The scorecard had the lowest mean on this outcome (mean 3.56), and it was significantly lower than the means for progress and traditional bar ($P=.004$ and $P<.001$, respectively).

Moderation of Condition Effects

Finally, we asked whether any effects of infographic condition would be moderated by whether the individual has met the behavioral goal. No effect was seen for having met the goal in the tobacco module, but for the nontobacco modules, having met the goal influenced the relationship between infographic condition and positive emotional tone (interaction $P=.003$). Because the interaction indicates that the effect of condition changes as a function of whether the goal has been met, we probed the interaction to understand the nature of the difference.

Results showed that, regardless of whether or not one had met the goal, the progress bar performed best; however, differences between the progress bar and the other 2 formats were more pronounced for those who did *not* meet the goal. The mean scores for the progress bar, traditional bar, and scorecard were 2.69, 1.98, and 1.96, respectively, for those who did not meet the goal, and 2.38, 2.24, and 2.08, respectively, for those who did meet the goal. The progress bar was rated significantly higher than both the traditional bar and the scorecard among those who did not meet the goal (both $P<.001$), and significantly higher than the scorecard among those who did meet the goal ($P=.01$).

Discussion

Health behavior assessments that provide personalized feedback face the challenge of conveying to individuals how their behaviors stack up to evidence-based recommendations and motivating behavior change. We conducted this study as formative research to examine whether game-inspired infographics would be more or less effective than traditional formats in their ability to convey information in ways that are easy to process and engaging. We report these results to document the process of development and to potentially benefit other eHealth programs that seek to convey similar information in similar contexts.

We found that, across multiple health behaviors, a game-inspired infographic was superior to a traditional format for some outcomes, and for other outcomes there was no observable difference. Specifically, in the tobacco module, where the scorecard was the only game-inspired design that we tested, the scorecard outperformed the traditional text-only format on comprehension and perceived cognitive load, and there were no differences in performance on the other information processing and engagement outcomes. In the modules for behaviors other than tobacco, there were no differences between the progress bar and the traditional bar for the information processing outcomes, but the progress bar outperformed the

traditional bar on outcomes related to engagement (positive emotional tone and attitude). Thus, based on the potential for a game-inspired design to outperform in the cases described above (and not to underperform otherwise), game-inspired designs show promise as a component of behavioral assessment output, relative to more traditional formats.

Moreover, in the nontobacco modules, the appeal of the progress bar over the traditional was pronounced among people who had not yet met the national recommendation for that health behavior. Individuals in this group may represent a priority audience because they would benefit the most from persuasive efforts. Prior work suggests that intervention efforts aimed at people who see no need to change health behavior should emphasize positive consequences of change more than negative consequences of not changing [58]. Game-inspired infographics that explicitly frame feedback in a positive light (eg, progress toward a desirable state) may therefore be more effective with this group than other formats, consistent with our results.

Comparing the findings for tobacco and nontobacco modules, a question that arises is why there were effects on information processing but not engagement outcomes in the tobacco module, and on engagement but not information processing outcomes in the nontobacco modules. The asymmetry may be due to the comparisons that were available in the tobacco versus nontobacco modules. In terms of information processing, the comparisons in the nontobacco modules were all among equally information-rich conditions, in the sense that all conditions provided 3 informational components (goal statement, whether the goal was met, and a visual representation of that information). In contrast, the comparison available in the tobacco module was between 2 conditions that differed in information richness, in the sense that the text-only traditional condition (which provided only 2 components: the goal statement and whether the goal was met) was compared with text plus visual (3 components) in the game-inspired scorecard condition. Thus, it makes sense that we observed a difference in information processing in the tobacco module but not in the other modules, as would be expected based on dual coding theory [21].

Similarly, in terms of engagement outcomes, both conditions in the tobacco module lacked design elements that may have contributed to emotional appeal, whereas the progress bar conditions in the nontobacco modules were designed to contain such elements (eg, playful design and sense of movement toward a desirable goal). Therefore, it makes sense that we observed differences on this dimension in nontobacco modules but not in the tobacco module. The dichotomous feedback in the tobacco module ruled out a progress bar as a potential format in this study, so we used only the scorecard. However, future research could examine the effectiveness of other visual representations of tobacco cessation goal attainment that may convey greater emotional appeal. In the nontobacco modules, it is interesting that the scorecard fared worse than the traditional bar on 2 of the 3 engagement outcomes. It may be that the dichotomization of behavior in the scorecard is less appealing than formats that

present feedback within a numerical range, as both other formats did. However, it is clear that the progress bar outperformed the scorecard on all 3 engagement outcomes and is therefore the more promising of the 2 game-inspired formats in the context of this study. The differences in performance between these 2 game-inspired formats further highlight the value of pretesting different executions during the formative research process.

Limitations

It is important to address the limitations introduced by the exploratory nature of the study. We used Mechanical Turk as a recruitment tool for this initial test because Mechanical Turk made it possible to obtain large samples at a reasonable cost; however, the resulting sample was not as diverse as would have been ideal in terms of age, race, ethnicity, and education level. Future research should employ methods that would yield a more diverse sample. The design of the study was posttest-only, which does not allow observation of changes before and after exposure. Although a posttest-only design helps avoid some potential threats to validity that might be introduced by preexposure measures (such as sensitizing individuals to treatment or later measurement), future studies should take a longitudinal approach to enable the observation of changes in key variables over time.

In terms of stimuli, we used only one execution to represent each format in each module, and although we selected the game-inspired exemplars as the best representatives of the formats among many alternatives reviewed by the research team, it is possible that other executions would have been more effective. We combined modules in the analysis after ensuring that the effect of condition did not differ across modules; as a consequence, though, it is not possible to pinpoint which specific features of the game-inspired infographics were responsible for observed effects. Subsequent research should systematically test multiple versions in which elements are present or absent to identify which had the most impact. Participants viewed sample feedback that was standardized within a module for experimental control and not tailored to their own behavior because it was not feasible to produce all possible variations across modules, formats, and behavior levels. However, we are planning research to evaluate the infographics on the actual CHART platform, which would provide personalized assessments and would also enable us to study the effects of infographics on intentions and outcomes related to behavior change.

Conclusions

The personalized feedback provided by health behavior assessments must tell people who do not meet a behavioral goal that they are short of the mark without leading to demotivation and disengagement. This is a challenging task. The principles of game design informed the development of infographics that were better than or did not differ from traditional formats in terms of effects on engagement and information processing outcomes. The results of our pretest have the potential to help online health assessment tools provide personalized feedback in a way that may facilitate progress toward health goals.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Forjuoh S, Ory M, Wang S, des Bordes JK, Hong Y. Using the iPod Touch for patient health behavior assessment and health promotion in primary care. JMIR Mhealth Uhealth 2014;2(1):e14 [FREE Full text] [doi: 10.2196/mhealth.2927] [Medline: 25100308]

2. Douglas S, Caldwell B. Design and validation of an individual health report (IHR). Int J Ind Ergonomics 2011 Jul;41(4):352-359. [doi: 10.1016/j.ergon.2011.02.004]

3. Bardus M, Smith J, Samaha L, Abraham C. Mobile phone and Web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]

4. Soler R, Leeks K, Razi S, Hopkins DP, Griffith M, Aten A, et al. A systematic review of selected interventions for worksite health promotion. The assessment of health risks with feedback. Am J Prev Med 2010 Feb;38(2 Suppl):S237-S262. [doi: 10.1016/j.amepre.2009.10.030] [Medline: 20117610]

5. Willey S, Walsh JK. Outcomes of a mobile health coaching platform: 12-week results of a single-arm longitudinal study. JMIR Mhealth Uhealth 2016;4(1):e3 [FREE Full text] [doi: 10.2196/mhealth.4933] [Medline: 26747611]

6. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]

7. Schubart J, Stuckey H, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21372645]

8. Smiciklas M. The Power of Infographics: Using Pictures to Communicate and Connect With Your Audiences. Indianapolis, IN: Que Biz-Tech; 2012.

9. King AJ, Jensen JD, Davis LA, Carcioppolo N. Perceived visual informativeness (PVI): construct and scale development to assess visual information in printed materials. J Health Commun 2014;19(10):1099-1115. [doi: 10.1080/10810730.2013.878004] [Medline: 24742307]

10. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. New York, NY: ACM; 2011 Presented at: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]

11. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA 2004 Mar 10;291(10):1238-1245. [doi: 10.1001/jama.291.10.1238] [Medline: 15010446]

12. Atkin CK, Freimuth V. Guidelines for formative evaluation research in campaign design. In: Rice RE, Atkin CK, editors. Public Communication Campaigns. 4th edition. Thousand Oaks, CA: Sage Publications, Inc; Mar 06, 2012:53-68.

13. Pechmann C, Andrews C. Copy test methods to pretest advertisements. In: Sheth JN, Malhotra NK, editors. Wiley International Encyclopedia of Marketing. Hoboken, NJ: John Wiley & Sons, Ltd; 2010.

14. International Cancer Screening Network. Designing Print Materials: A Communications Guide for Breast Cancer Screening. Bethesda, MD: National Cancer Institute; 2007.

15. Ancker J, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc 2006;13(6):608-618 [FREE Full text] [doi: 10.1197/jamia.M2115] [Medline: 16929039]

16. Garcia-Retamero R, Cokely E. Communicating health risks with visual aids. Curr Dir Psychol Sci 2013 Sep 25;22(5):392-399. [doi: 10.1177/0963721413491570]

17. Houts P, Doak C, Doak L, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]

18. Witteman H, Fuhrel-Forbis A, Wijeysundera H, Exe N, Dickson M, Holtzman L, et al. Animated randomness, avatars, movement, and personalization in risk graphics. J Med Internet Res 2014;16(3):e80 [FREE Full text] [doi: 10.2196/jmir.2895] [Medline: 24642037]

19. Noar S, Harrington N. eHealth Applications: Promising Strategies for Health Behavior Change. New York, NY: Routledge; 2012.

20. Delp C, Jones J. Communicating information to patients: the use of cartoon illustrations to improve comprehension of instructions. Acad Emerg Med 1996 Mar;3(3):264-270 [FREE Full text] [Medline: 8673784]

21. Paivio A. Mental Representations: A Dual Coding Approach. New York, NY: Oxford University Press; 1990.

22. David P. News concreteness and visual-verbal association: Do news pictures narrow the recall gap between concrete and abstract news? Human Comm Res 1998 Dec;25(2):180-201. [doi: 10.1111/j.1468-2958.1998.tb00442.x]

23. Zikmund-Fisher B, Witteman H, Fuhrel-Forbis A, Exe N, Kahn V, Dickson M. Animated graphics for comparing two risks: a cautionary tale. J Med Internet Res 2012;14(4):e106 [FREE Full text] [doi: 10.2196/jmir.2030] [Medline: 22832208]

24. Schapira M, Nattinger A, McHorney C. Frequency or probability? A qualitative study of risk communication formats used in health care. Med Decis Making 2001;21(6):459-467. [Medline: 11760103]

25. McGuire WJ. Public communication as a strategy for inducing health-promoting behavioral change. Prev Med 1984 May;13(3):299-319. [Medline: 6387698]

26. MacKenzie S, Lutz R, Belch G. The role of attitude toward the ad as a mediator of advertising effectiveness: a test of competing explanations. J Marketing Res 1986 May;23(2):130-143. [doi: 10.2307/3151660]

27. Mitchell A, Olson J. Are product attribute beliefs the only mediator of advertising effects on brand attitude? J Marketing Res 1981 Aug;18(3):318-332. [doi: 10.2307/3150973]

28. Shimp T. Attitude toward the ad as a mediator of consumer brand choice. J Advertising 1981 Jun;10(2):9-48. [doi: 10.1080/00913367.1981.10672756]

29. Olney T, Holbrook M, Batra R. Consumer responses to advertising: the effect of ad content, emotions, and attitude toward the ad on viewing time. J Consumer Res 1991 Mar;17(4):440-453. [doi: 10.1086/208569]

30. Kelly K, Slater M, Karan D. Image advertisements' influence on adolescents' perceptions of the desirability of beer and cigarettes. J Public Policy Marketing 2002 Sep;21(2):295-304. [doi: 10.1509/jppm.21.2.295.17585]

31. Mollick E, Rothbard N. Mandatory fun: gamification and the impact of games at work. SSRN J 2013:1. [doi: 10.2139/ssrn.2277103]

32. Lounis S, Neratzouli X, Pramatari K. Can gamification increase consumer engagement? A qualitative approach on a green case. In: Douligeris C, Polemi N, Karantjias A, Lamersdorf W, editors. IFIP Advances in Information and Communication Technology. Berlin: Springer; 2013:200-212.

33. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]

34. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]

35. Ahola R, Pyky R, Jämsä T, Mäntysaari M, Koskimäki H, Ikäheimo TM, et al. Gamified physical activation of young men: multidisciplinary population-based randomized controlled trial (MOPO study). BMC Public Health 2013 Jan 14;13(1):a. [doi: 10.1186/1471-2458-13-32] [Medline: 23311678]

36. Consolvo S, Klasnja P, McDonald D, Avrahami D, Froehlich J, LeGrand L, et al. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. New York, NY: ACM; 2008 Presented at: Proceedings of the 10th International Conference on Ubiquitous Computing; 2008; Seoul, South Korea p. 54-63. [doi: 10.1145/1409635.1409644]

37. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform 2012;177:85-96. [Medline: 22942036]

38. Sue D, Ray P, Talaei-Khoei A, Jonnagaddala J, Vichitvanichphong S. Assessing video games to improve driving skills: a literature review and observational study. JMIR Serious Games 2014;2(2):e5 [FREE Full text] [doi: 10.2196/games.3274] [Medline: 25654355]

39. Cafazzo J, Casselman M, Hamming N, Katzman D, Palmert M. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]

40. Cheema A, Bagchi R. The effect of goal visualization on goal pursuit: implications for consumers and managers. J Marketing 2011 Mar;75(2):109-123. [doi: 10.1509/jmkg.75.2.109]

41. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business. Philadelphia, PA: Wharton Digital Press; 2012.

42. Schell J. The Art of Game Design: A Book of Lenses. Boca Raton, FL: CRC Press; 2008.

43. Marczewski A. Gamasutra. 2013 Oct 24. Game Thinking - Breaking Down Gamification and Games URL: http://www.gamasutra.com/blogs/AndrzejMarczewski/20131024/203054/Game_Thinking__Breaking_down_gamification_and_games.php [accessed 2016-08-23] [WebCite Cache ID 6jyvusCwG]

44. Gallagher K, Updegraff J. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]

45. Lipkus IM, Pollak KI, McBride CM, Schwartz-Bloom R, Lyna P, Bloom PN. Assessing attitudinal ambivalence towards smoking and its association with desire to quit among teen smokers. Psychol Health 2005 Jun;20(3):373-387. [doi: 10.1080/08870440512331333988]

46. Fazio R, Powell M, Williams C. The role of attitude accessibility in the attitude-to-behavior process. J Consumer Res 1989 Dec;16(3):280-288. [doi: 10.1086/209214]

XSL•FO

RenderX

47.   Buhrmester M, Kwang T, Gosling S. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]

48.   Carter R, DiFeo A, Bogie K, Zhang GQ, Sun J. Crowdsourcing awareness: exploration of the ovarian cancer knowledge gap through Amazon Mechanical Turk. PLoS One 2014;9(1):e85508 [FREE Full text] [doi: 10.1371/journal.pone.0085508] [Medline: 24465580]

49.   Borgida E, Nisbett R. The differential impact of abstract vs. concrete information on decisions. J Appl Social Pyschol 1977 Sep;7(3):258-271. [doi: 10.1111/j.1559-1816.1977.tb00750.x]

50.   Fazio RH. A practical guide to the use of response latency in social psychological research. In: Hendrick C, Clark MB, editors. Research Methods in Personality and Social Psychology. Newbury Park, CA: Sage Publications; 1990:74.

51.   Eveland WP, Dunwoody S. User control and structural isomorphism or disorientation and cognitive load? Commun Res 2001 Feb 01;28(1):48-78. [doi: 10.1177/009365001028001002]

52.   Pechmann C, Reibling E. Antismoking advertisements for youths: an independent evaluation of health, counter-industry, and industry approaches. Am J Public Health 2006 May;96(5):906-913. [doi: 10.2105/AJPH.2004.057273] [Medline: 16571709]

53.   Dillard J, Shen L, Vail R. Does perceived message effectiveness cause persuasion or vice versa? 17 consistent answers. Human Commun Res 2007 Oct;33(4):467-488. [doi: 10.1111/j.1468-2958.2007.00308.x]

54.   Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Questionnaire 2011. 2011 Jan 27. URL: http://www.cdc.gov/brfss/questionnaires/pdf-ques/2011brfss.pdf [accessed 2016-08-23] [WebCite Cache ID 6jyxM3OOp]

55.   Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Questionnaire 2008. 2007 Dec 31. URL: http://www.cdc.gov/brfss/questionnaires/pdf-ques/2008brfss.pdf [accessed 2016-08-23] [WebCite Cache ID 6jyxV29LC]

56.   Centers for Disease Control and Prevention. 2009-2010 National Health and Nutrition Examination Survey. 2009. URL: https://www.cdc.gov/nchs/nhanes/nhanes2009-2010/questionnaires09_10.htm [accessed 2016-08-23] [WebCite Cache ID 6jyxWCeuV]

57.   Slater M, Peter J, Valkenberg PM. Message variability and heterogeneity: a core challenge for communication research. Commun Yearbook 2015;39:3-31 [FREE Full text] [Medline: 26681816]

58.   Hall KL, Rossi JS. Meta-analytic examination of the strong and weak principles across 48 health behaviors. Prev Med 2008 Mar;46(3):266-274. [doi: 10.1016/j.ypmed.2007.11.006] [Medline: 18242667]

Abbreviations

CHART: Carolina Health Assessment and Resource Tool

Original Paper

Evaluation of a Serious Self-Regulation Game Intervention for Overweight-Related Behaviors ("Balance It"): A Pilot Study

Jorinde Spook[1,2,3], PhD; Theo Paulussen[1], PhD; Gerjo Kok[2], PhD; Pepijn van Empelen[1], PhD

[1]TNO, Life Style, Leiden, Netherlands

[2]Maastricht University, Department of Work and Social Psychology, Maastricht University, Maastricht, Netherlands

[3]Wageningen University and Research, Department of Communication, Philosophy and Technology: Centre for Integrative Development, Wageningen, Netherlands

Corresponding Author:
Jorinde Spook, PhD
Wageningen University and Research
Department of Communication, Philosophy and Technology: Centre for Integrative Development
PO Box 8130
6700 EW
Wageningen,
Netherlands
Phone: 31 0317481446
Fax: 31 0317481446
Email: jorinde.spook@wur.nl

Abstract

Background: Serious games have the potential to promote health behavior. Because overweight is still a major issue among secondary vocational education students in the Netherlands, this study piloted the effects of "Balance It," a serious self-regulation game intervention targeting students' overweight-related behaviors: dietary intake and physical activity (PA).

Objective: We aimed to pilot the effects of Balance It on secondary vocational education students' dietary intake and PA.

Methods: In total, 501 secondary vocational education students participated at baseline (intervention: n=250; control: n=251) in this pre-post cluster randomized trial. After 4 weeks, at immediate posttest, 231 students filled in the posttest questionnaire (intervention: n=105; control: n=126). The sample had a mean age of 17.28 (SD 1.26, range 15-21) years, 62.8% (145/231) were female, and 26.8% (62/231) had a non-Dutch background. Body mass index (BMI kg/m^2) ranged from 14.4 to 31.1 (mean 21.1, SD 3.3). The intervention and control groups were compared on the primary (behavioral) outcomes of dietary intake (fruit and vegetable consumption, snack consumption, and soft drink consumption) and PA (moderate and vigorous). Additionally, we explored (1) differences between the intervention and control groups in determinants of dietary intake and PA, including attitude, self-efficacy, intention, barrier identification, action planning, and action control, and (2) differences between active (intervention) users and the control group in dietary intake, PA, and associated determinants.

Results: After corrections for multiple testing, we did not find significant differences between the intervention group and control group in terms of dietary intake, PA, and determinants of dietary intake and PA. Exploratory research indicated that only 27.6% (29/105) of the intervention group reported actual intervention use (ie, active users). For exploratory reasons, we compared the active users (n=29) with the control group (n=124) and corrected for multiple testing. Results showed that active users' snack consumption decreased more strongly (active users: mean change=–0.20; control group: mean change=–0.08; beta=–0.36, P=.01, R^2 change=.05), and their use of active transport had a stronger increase (active users: mean change=0.92; control group=–0.12; beta=1.58, P=.02, R^2 change=.03) than the control group. Results also revealed significant differences in action planning (active users: mean change=0.42; control group: mean change=0.07; beta=0.91, P=.01, R^2 change=.04) and action control (active users: mean change=0.63; control group: mean change=–0.05; beta=1.25, P=.001, R^2 change=.08) in terms of unhealthy eating.

Conclusions: The Balance It intervention did not show favorable effects on dietary intake and PA compared to the control condition. However, only a small number of people in the intervention condition actually used Balance It (27.6%). Exploratory analyses did suggest that, if used as planned, Balance It could contribute to changing dietary intake and PA behaviors, albeit it remains debatable whether this would be sufficient to prevent overweight.

XSL•FO
RenderX

Introduction

Overweight and obesity are related to various chronic health problems, including type 2 diabetes mellitus, cardiovascular disease, cancer, and also psychosocial problems [1-6]. In the Netherlands, approximately 20% of youth (aged 16-20 years) from low socioeconomic status (SES) families are overweight or obese [7], and prevalence is even higher among youth with Turkish or Moroccan descent [8-10]. Treatment of overweight remains a challenge; hence, it is important to target overweight-related behaviors (eg, dietary intake and physical activity [PA]) in intervention studies designed to prevent overweight in low SES youth.

Recent advances in technology enable researchers to tailor dietary intake and PA interventions to the needs of the target population. Moreover, it is possible to design a program that is cost effective, that has a wide reach, and that can function as a standalone program [11,12]. Reviews highlight the potential of computer-tailored interventions in terms of effectively changing and promoting health-related behaviors [13-15], yet targeting young people; immigrant groups; people with a low, primary, or basic vocational education; and people with weak health motivation can still be challenging [16]. To overcome hurdles such as low reach and limited adoption of computer-tailored interventions, several strategies have been recommended, including increasing the interactivity and visual attractiveness of the program [17-19]. Serious gaming is a promising method that can be used to stimulate intervention use because such games are designed to be highly enjoyable, attention grabbing, and intrinsically motivating [20-23]. In previous research [24], serious gaming interventions (eg, "Diab" and "Nano") appeared to increase fruit and vegetable intake. However, playing these games did not increase water consumption, PA, or body composition. Thompson et al [25] indicated that action intentions may be an important component of successful interventions to stimulate youth fruit and vegetable intake [25], which in combination with coping plans may also account for PA [26]. As such, these studies showed that serious games have great potential to change health-related behaviors. However, according to DeSmet et al [19], serious games generally fall short in applying effective behavior change methods to change health-related behaviors. DeSmet et al [19] advocate the use of dual theoretical frameworks, stressing the importance of a theoretical foundation in both behavioral prediction and game theories. To this end, we combined effective behavior change techniques (as applied in computer-tailored interventions) with serious gaming strategies to encourage intervention use and target health behavior change simultaneously. As such, we developed a serious self-regulation game intervention called "Balance It."

Balance It combines behavior change techniques derived from self-regulation theory [27,28] with serious game elements. It is a serious self-regulation game designed to target dietary intake and PA among secondary vocational education students. Balance It was systematically developed by means of Intervention Mapping, a protocol that enables the systematic planning of theory- and evidence-based interventions [29]. Further elaboration on the design rationale of Balance It can be found elsewhere [30]. Our key research objectives in this pilot study were to (1) identify the effectiveness of Balance It on changes in (determinants of) secondary vocational education students' dietary intake and PA, and (2) evaluate the uptake and usage of the game and the game elements.

Methods

Study Design

A cluster randomized trial was conducted in 2014/2015 with measurements taken at baseline, immediately posttest (after 4 weeks of game play), and at a 4-week follow-up. Fifteen vocational education schools in the Netherlands were approached to participate in this study. In total, 4 schools agreed to participate and were randomly assigned to the intervention or waiting list control group. To counteract contamination effects between participant groups and to increase participants' compliance with the study, random allocation to conditions took place at the level of schools [31]. All procedures were approved by The Research Ethics Board of the School of Psychology and Neuroscience (Maastricht University).

Participants

The power calculation (alpha=.05, beta=.80) was based on a mean effect size of 31 for dietary intake and PA intervention [32]. This required a minimum of 130 participants for both the intervention and the control group. Students who did not have a mobile phone operating on iOS or Android were exempted from participation, as were students younger than 16 years or older than 21 years. Students younger than 16 years were exempted from participation because, within the Netherlands, individuals are allowed to provide informed consent from the age of 16 years. Participants older than 21 years were also exempted from participation because they had outgrown puberty and were not targeted by the intervention. All other students between the ages of 16 and 21 years were eligible for participation.

Procedures

One week before the study, participants received passive consent forms addressed to their parents or caregivers. At baseline, a research assistant went to the schools to introduce the study and to collect the survey data. In total, 238 students gave their consent to participate in this study. At baseline, participants filled out an online baseline questionnaire regarding their mean dietary intake and PA, social cognitive factors (ie, attitude, self-efficacy, and intention), perceived barriers, self-regulation

skills, action planning, and action control. After they finished the questionnaire, participants received a link to the Balance It website and further instructions about downloading the Balance It app from the research assistant. All students received a posttest questionnaire 4 weeks after the baseline measure was taken. Participatory incentives of €20 vouchers were randomly distributed among participants. The chance of winning a voucher increased with the number of measures completed (one measure 1:8, two measures 2:8).

Balance It

Balance It was designed as a tailored, interactive multimedia game in which each game could be played either individually or competitively with others, at any time and place desired. It was designed as an educational, strategic game that could be played on a daily basis for 4 continuing weeks or on a weekly basis for 6 continuing weeks. Within each game, players set their own graded tasks (eg, to eat two pieces of fruit per day), which were selected from a multiple-choice list (Figure 1; [30]). They monitored and evaluated these goals on a daily or weekly

basis, depending on the type of game they chose to play. Each day or week, players were prompted with their goals and reminded to return to the game. Visual feedback on self-reported goal attainment was provided for each goal, and players were prompted to reflect on their condition and on the perceived barriers or facilitators of goal accomplishment. Finally, participants were encouraged to formulate implementation intentions. In turn, these implementation intentions, or strategies, could be set as reminder prompts at any specific time point the player preferred. Information about formulating implementation intentions was provided on the Balance It website. The website also provided a general overview of the participant's progress and a peer-support system (ie, the Balance It forum). Reinforcement was given in the form of obtainable "Tetris-shaped" building blocks and the allocation of "super powers" after goal accomplishment and self-evaluation of the targeted behavior. With these building blocks, players were encouraged to build a tower and to keep the tower in balance (see Figure 1; for a full description of the game design and content see [30]).

Figure 1. Screenshots of task initiation in the Balance It app.
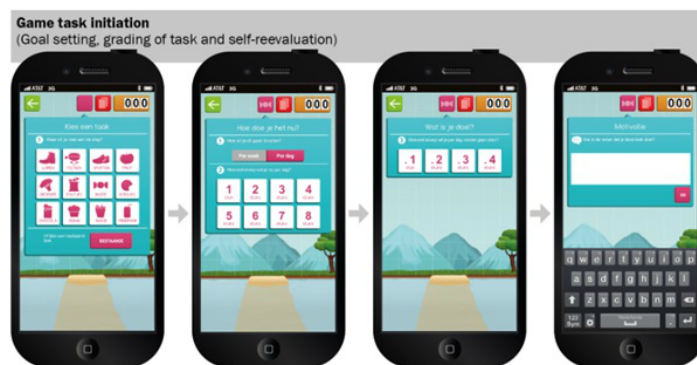
Waiting List Control Group

At baseline, the control group was instructed to fill in the baseline questionnaire and informed that the researcher would return in 4 weeks for a posttest measure. Between measures, no interventions were offered by the researchers. Immediately following the posttest, students were provided with information about Balance It and given the opportunity to play.

Behavioral Outcome Measures

Dietary Intake

The assessment of dietary intake was derived from a validated food frequency questionnaire [33]. Questions were related to the participant's mean daily fruit and vegetable intake, snack consumption, and soft drink consumption. Answers were given on an 8-point scale on which participants could record the number of days they consumed specific foods, ranging from 0 (never or almost never) to 7 (every day). In addition, the quantity of their dietary intake was assessed. Response categories ranged from 1 (half portion or piece a day) to 7 (three or more portions or pieces a day). Based on these scores, the mean intake per day was calculated.

Physical Activity

The PA measures were derived from the Injuries and Physical Activity in the Netherlands ("Ongevallen en Bewegen in Nederland") questionnaire (validated; [34]). Questions were related to the participant's mean moderate PA (walking and cycling) and vigorous PA (exercise). For example, moderate PA was operationalized as "During the last week, how many days did you carry out 30 minutes of moderate PA?" Answers were given on an 8-point scale on which participants could rate the number of days they were moderately or vigorously active, ranging from 0=never or almost never to 7=every day.

Determinants of Dietary Intake

In addition to the behavioral outcomes, social cognitive factors were measured for healthy dietary intake (fruit and vegetable intake) and unhealthy dietary intake (snacks, sweets, and soft drink consumption). All measures of determinants were preceded by a stem, followed by the behavioral outcome measures as subcategories.

Attitude

Attitudes toward dietary intake were assessed by three items using semantic differential response scales, such as "I think that eating two pieces of fruit a day is..." (1=very bad to 5=very good; 1=very unpleasant to 5=very pleasant; 1=very unhealthy

to 5=very healthy) derived from [35,36]. Cronbach alpha for the healthy dietary intake attitude items was .87 and Cronbach alpha for the unhealthy dietary intake attitude items was .87.

Self-Efficacy

Self-efficacy toward dietary intake was assessed by one item preceded by a question stem: "If I want to, I am capable of..." Items were derived from van der Horst et al [35] and from Van Genugten et al [36]. Response options ranged from 1=definitely not to 5=definitely. Cronbach alpha for the healthy dietary intake self-efficacy items was .82. Cronbach alpha for the unhealthy dietary intake self-efficacy items was .88.

Intention

Dietary intake intention was assessed with one item preceded by the stem: "I planned to..." derived from [35,36]. Response options ranged from 1=definitely not to 5=definitely. Cronbach alpha for the healthy dietary intake intention items was .82. Cronbach alpha for the unhealthy dietary intake intention items was .93.

Barrier Identification

Barriers to healthy dietary intake were assessed separately from barriers to unhealthy dietary intake because different barriers influence fruit and vegetable intake and unhealthy dietary intake. Healthy dietary intake was assessed with five items using 5-point Likert scales: "I am capable of eating sufficient fruit and vegetables, also when I am..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to when I am alone, during the weekend, when I am in a hurry, when I experience difficulties preparing fruits and vegetables, and when there is a lack of choice. Items were derived from previous measures [37,38]. Cronbach alpha for the healthy dietary intake barrier identification items was .86.

Barriers to unhealthy dietary intake were assessed with 13 items using 5-point Likert scales: "I am capable of eating a limited amount of unhealthy snacks, also when I am..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to physical settings (eg, when I am at home), sedentary activities (eg, when I am watching TV), social settings (eg, when I am at a party), and mood (eg, when I am sad). Items were derived from previous measures [37,38]. Cronbach alpha of the unhealthy dietary intake barrier identification items was .96.

Action Planning

Action planning in terms of dietary intake was assessed by four items, such as "I have a clear plan for when I..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to when, where, how, and how often participants planned to eat more healthy or less unhealthy foods (derived from [35,36]). Cronbach alpha for the healthy dietary intake action planning items was .97. Cronbach alpha for the unhealthy dietary intake action planning items was .96.

Action Control

Action control in terms of dietary intake was measured with four items using 5-point Likert scales, such as "During the last month, I have constantly monitored my..." Response options

ranged from 1=definitely not to 5=definitely. Subcategories referred to self-monitoring of fruit and vegetable consumption, awareness of fruit and vegetable standards, self-regulatory effort to eat more healthy and less unhealthy foods, and self-regulatory effort to conform to norm behavior (eg, eat two pieces of fruits a day) (derived from [39]). Cronbach alpha for the healthy dietary intake action control items was .94. Cronbach alpha for the unhealthy dietary intake action control items was .94.

Determinants of Physical Activity

Social cognitive factors were also measured for moderate PA (eg, walking and cycling) and vigorous PA (eg, exercising). All measures of PA determinants were preceded by a stem followed by the behavioral outcome measures subcategories.

Attitude

Attitudes toward PA was assessed by three items using semantic differential response scales, such as "I think that exercising is..." (1=very bad to 5=very good; 1=very unpleasant to 5=very pleasant; 1=very unhealthy to 5=very healthy) (derived from [35,36]). Cronbach alpha for the PA attitude items was .79.

Self-Efficacy

Self-efficacy toward PA was assessed by one item preceded by a question stem: "If I want to, I am capable of..." Items were derived from Van der Horst et al [35] and Van Genugten et al [36]. Response options ranged from 1=definitely not to 5=definitely. Cronbach alpha for the PA self-efficacy item was .75.

Intention

Intention was assessed with one item preceded by the stem: "I planned to..." (derived from [35,36]). Response options ranged from 1=definitely not to 5=definitely. Cronbach alpha for the PA intention item was .73.

Barrier Identification

Barriers to PA were assessed with seven items using 5-point Likert scales, such as "I am capable of being more physically active, also when I am..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to when I am busy, when I am stressed, if I failed last time, when I am tired, when it is raining, if I do not have the time, and if I do not get social support (derived from [37,38]). Cronbach alpha for the PA barrier identification items was .93.

Action Planning

Action planning in terms of PA was assessed by four items using 5-point Likert scales, such as "I have a clear plan for when I..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to when, where, how, and how often participants planned to be more physically active (derived from [35,36]). Cronbach alpha for the PA action planning items was .96.

Action Control

Action control in terms of PA was measured with four items using 5-point Likert scales, such as "During the last month, I have constantly monitored my..." Response options ranged from 1=definitely not to 5=definitely. Subcategories referred to

self-monitoring of PA, awareness of PA standards, self-regulatory effort to be more physically active, and self-regulatory effort to conform to norm behavior (eg, to be moderately active) (derived from [39]). Cronbach alpha for the PA action control items was .94.

Demographics

Items regarding gender, age, BMI, educational level, cultural background, accommodation, and living situation were included at the beginning of the baseline measure. Ethnicity was defined according to the procedures of Statistics Netherlands; individuals were considered to have a Dutch background if both parents were born in the Netherlands. If one of the parents was born outside the Netherlands, the student was considered to have a non-Dutch background [40].

Self-Reported Intervention Evaluation

To evaluate subjective experience of using the Balance It app or website, 19 items regarding the Balance It app in general were preceded by the stem: "What did you think of..." Response options ranged from 1 (very bad) to 5 (very good) (compare attitude measures [35,36]). Cronbach alpha for attitude toward the Balance It app in general was .98. The 13 items were preceded by the same stem and referred to the specific game elements included (eg, "What did you think of the theme of Balance It?). Response options ranged from 1 (very stupid) to 5 (very funny) (compare attitude measures [35,36]). Cronbach alpha for the attitude toward game elements was .98.

Statistical Analyses

Descriptive statistics were used to characterize both study groups at baseline (ie, gender, age, educational level, ethnicity, and and body mass index [BMI]). Chi-square tests and *t* tests were conducted to evaluate whether participant characteristics were related to drop out during the study. Because there was no significant differentiation between school levels, linear regression analyses were performed to study dietary intake and PA change over time, differences between the intervention and control groups, and differences between active users and the control group. In these analyses, primary outcomes were analyzed and differences between groups on determinants of the primary outcomes were explored (controlling for condition and baseline differences). After doing the linear regression analyses, multiple testing adjustment procedures were taken into account according to the Benjamini-Hochberg procedures (ie, we calculated the false discovery rates [FDR] for all primary outcomes and exploratory determinants of these outcome measures). A *P* value of .05 or lower was considered to be statistically significant. All analyses were conducted with IBM SPSS version 20.0 (IBM Corporation).

Results

Participants and Dropout Analysis

In total, 501 students were invited to participate in this study (intervention: n=250; control: n=251; Figure 2). Of all students invited, 488 participated (97.4%). We excluded 6 students because they were younger than 15 years and 29 students because they were older than 21 years. After exclusion, 228 participants in the intervention group and 225 in the control group remained at baseline. After 4 weeks, 117 participants dropped out from the intervention group, and 92 participants dropped out from the control group. Logistic regression analyses revealed that participants who dropped out were significantly older (mean 17.69, SD 1.53 years) than nondropouts (mean 17.28, SD 1.26 years; OR 0.81, 95% CI 0.71-0.93). Tests also showed that students with a non-Dutch background were more likely to drop out (119/209, 56.9%) than students with a Dutch background (86/209, 41.1%; OR 1.95, 95% CI 1.31-2.92; unknown background: 4/209, 1.9%). We did not find any significant differences in terms of gender, level of education, year of education, or BMI. The BMI distribution of the sample was comparable with previous research on secondary vocational education students' health and weight [41]. At final count, 105 participants were included in the intervention group and 126 participants were included in the control group.

Active Users

Of all participants who remained in the intervention group at posttest (n=105), 27.6% (29/105) reported actual intervention use. Compared to the control group (n=200), self-reported active users were less likely to follow vocational education related to care and well-being (*P*<.001) and more likely to follow vocational education in economics (*P*<.001). Active users were also more likely to be in their first year (n=100) as compared to the control group (ie, n=124; *P*=.01). The two groups did not significantly differ in age, gender, ethnicity, BMI, social vocational education sector, technique vocational education sector, or level of education.

Baseline Between-Group Differences

Table 1 presents the demographic background of Balance It participants at baseline (N=231). Compared to the control group (mean 17.52, SD 1.36 years), participants in the intervention group were younger (mean 16.96, SD 1.10 years; *P*=.05). They were also more likely to participate in the economics vocational education sector and less likely to participate in care and well-being, social work, and economy vocational education sectors. Finally, they were more likely to be in the first year of secondary vocational education. Therefore, we included these variables as covariates in all further analyses. We also controlled for baseline differences between the intervention group and the control group in case they differed in behavioral outcome and determinant measures. As such, the intervention group at baseline was more likely to use active transport as compared to the control group (*P*=.04). Therefore, we controlled for the use of active transport at baseline in further analyses concerning active transport.

Figure 2. Flow diagram of the enrollment and selection of study participants.
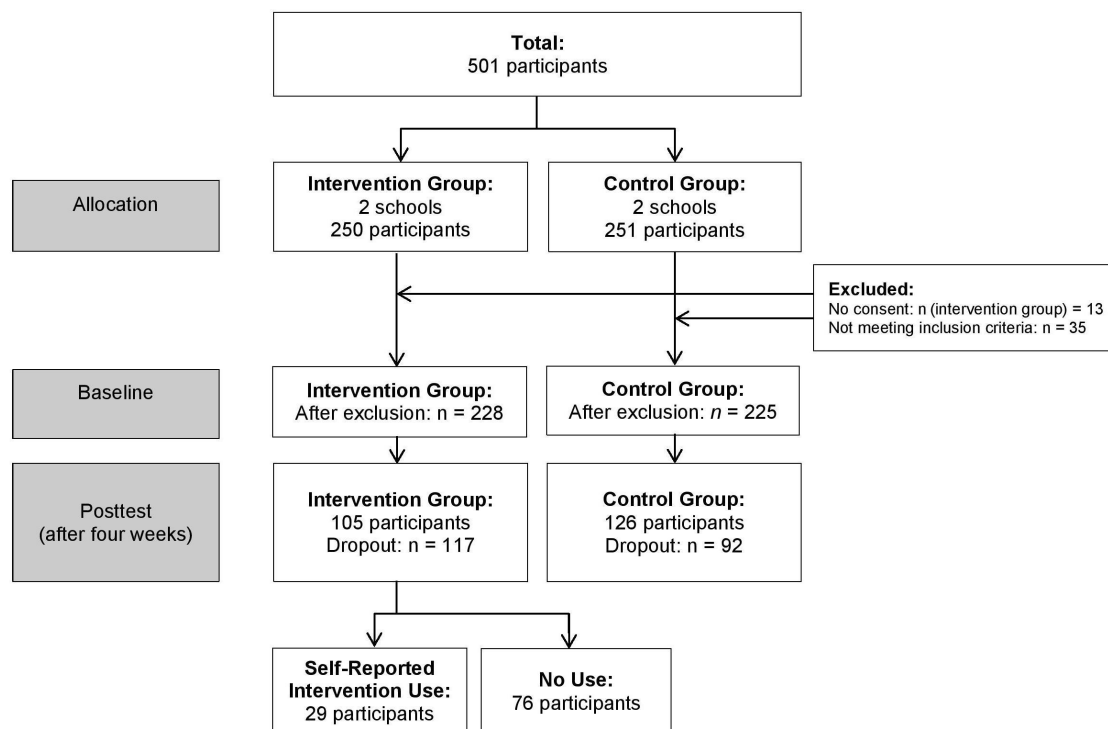
Table 1. Demographic background of Balance It participants at baseline (N=231).

Demographic variables	Intervention group (n=105)	Control group (n=126)	χ^2 (df)	t_{227}	P value
Age (years), mean (SD)	16.96 (1.10)	17.52 (1.36)		−3.34	.003
Gender (male), n (%)	39 (37.1)	47 (37.3)	0.0 (1)		.93
Ethnicity (Dutch), n (%)	77 (73.3)	92 (73.0)	0.1 (1)		.77
Vocational education sector track, n (%)					
Care and well-being	25 (23.8)	115 (91.3)	108.4 (1)		.001
Economics	71 (67.6)	0 (0.0)	128.9 (1)		.001
Technique	1 (1.0)	2 (1.6)	0.2 (1)		.69
Social work	0 (0.0)	1 (1.0)	0.8 (1)		.55
Educational level, n (%)			0.5 (1)		.49
Level 3	5 (4.8)	9 (7.1)			
Level 4	95 (90.5)	115 (91.3)			
Year of education, n (%)			39.3 (1)		.001
Year 1	100 (100.0)	84 (67.7)			
Year 2	0 (0.0)	40 (34)			
Living situation			5.3 (4)		.26
Both parents	79 (76.7)	100 (79.4)			
One parent	19 (18.4)	16 (12.7)			
Alone	2 (1.9))	2 (1.6)			
Other	3 (2.9)	8 (6.3)			
BMI categories, n (%)			4.7 (3)		.66
Underweight (BMI <18.5)	8 (12.7)	10 (11.2)			
Normal weight (BMI 18.5-25)	47 (74.6)	58 (65.1)			
Overweight (BMI 25-30)	8 (12.7)	18 (20.2)			
Obese (BMI >30)	0 (0.0)	3 (3.4)			

Table 2. Effects of Balance It on behavioral outcomes and determinants.[a]

Outcome variable	T0, mean (SD)		T1, mean (SD)		T0-T1, Δ mean		Difference test	
	Intervention (n=103)[b]	Control (n=125)[c]	Intervention (n=103)[d]	Control (n=125)[e]	Intervention	Control	B (95% CI)	R^2 change
Behavioral outcomes								
Fruit intake (mean portion/day)	0.81 (0.68)	0.80 (0.68)	1.05 (0.75)	0.81 (0.62)	0.14	0.01	0.21 (–0.07 to 0.49)	.01
Vegetable intake (mean portion/day)	1.26 (0.33)	1.32 (0.38)	1.21 (0.41)	1.28 (0.36)	–0.05	–0.04	–0.03 (–0.15, 0.10)	.00
Snack consumption (mean portion/day)	0.91 (0.50)	0.98 (0.51)	0.86 (0.51)	0.90 (0.48)	–0.05	–0.08	0.01 (–0.17 to 0.19)	.00
Soft drink consumption (mean portion/day)	1.07 (0.53)	1.11 (0.59)	0.92 (0.57)	1.07 (0.57)	–0.15	–0.04	–0.25 (–0.45 to –0.05)	.03
Moderate PA (days) [f]	4.30 (2.41)	3.82 (2.67)	3.91 (2.54)	3.31 (2.51)	–0.39	–0.51	0.20 (–0.87 to 1.27)	.00
Vigorous PA (days)	5.21 (2.26)	5.25 (2.07)	4.74 (2.47)	4.78 (2.27)	–0.47	–0.47	0.10 (–0.12 to 1.33)	.00
Active transport (days)	2.55 (1.99)	2.50 (2.37)	3.20 (2.51)	2.38 (2.13)	0.65	–0.12	0.94 (0.06 to 1.81)	.02
Determinants: fruit and vegetable intake (5-point scale)								
Attitude	4.01 (0.53)	3.98 (0.60)	3.93 (0.80)	4.00 (0.64)	–0.08	0.02	–0.26 (–0.51 to –0.02)	.02
Self-efficacy	4.33 (0.80)	4.29 (0.81)	4.04 (1.02)	4.13 (0.92)	–0.29	–0.16	–0.44 (–0.85 to 0.03)	.02
Intention	3.86 (1.00)	3.73 (1.05)	3.66 (1.09)	3.67 (1.11)	–0.20	–0.06	–0.32 (–0.74 to 0.09)	.01
Perceived barriers	3.45 (0.89)	3.52 (0.96)	3.53 (1.03)	3.49 (0.91)	0.08	–0.03	0.13 (–0.24 to 0.51)	.00
Action planning	3.04 (1.13)	3.05 (1.14)	3.37 (1.05)	2.86 (1.20)	0.33	–0.19	0.36 (–0.10 to 0.82)	.01
Action control	2.89 (1.26)	2.80 (1.17)	3.40 (1.14)	2.86 (1.20)	0.51	0.06	0.53 (0.04 to 1.02)	.02
Determinants: snack and soft drink consumption (5-point scale)								
Attitude	3.63 (0.66)	3.42 (0.61)	3.72 (0.87)	3.63 (0.67)	0.09	0.19	–0.23 (–0.53 to 0.06)	.01
Self-efficacy	4.28 (0.77)	4.10 (0.83)	3.92 (1.04)	3.91 (0.94)	–0.36	–0.19	–0.37 (–0.84 to 0.09)	.01
Intention	3.75 (1.03)	3.40 (1.03)	3.60 (1.05)	3.35 (1.08)	–0.15	–0.05	–0.14 (–0.55 to 0.28)	.00
Perceived barriers	3.52 (0.98)	3.35 (0.92)	3.48 (1.01)	3.37 (0.89)	–0.04	0.02	0.21 (–0.18 to 0.59)	.01
Action planning	3.11 (1.13)	2.99 (1.11)	3.37 (1.06)	3.06 (1.07)	0.27	0.07	0.33 (–0.12 to 0.77)	.01
Action control	2.96 (1.27)	2.78 (1.16)	3.28 (1.15)	2.73 (1.20)	0.32	–0.05	0.48 (–0.01 to 0.97)	.02

Outcome variable	T0, mean (SD)		T1, mean (SD)		T0-T1, Δ mean		Difference test	
	Intervention (n=103)[b]	Control (n=125)[c]	Intervention (n=103)[d]	Control (n=125)[e]	Intervention	Control	B (95% CI)	R^2 change
Determinants: PA (5-point scale)								
Attitude	4.21 (0.53)	4.21 (0.58)	4.02 (0.85)	4.13 (0.60)	–0.20	–0.09	–0.18 (–0.43 to 0.07)	.01
Self-efficacy	4.42 (0.66)	4.30 (0.82)	3.98 (1.04)	4.08 (0.89)	–0.44	–0.22	–0.39 (–0.80 to 0.02)	.02
Intention	4.01 (0.96)	3.87 (1.03)	3.70 (1.07)	3.68 (1.04)	–0.31	–0.19	–0.44 (–0.87 to –0.01)	.02
Perceived barriers	3.47 (1.02)	3.16 (1.05)	3.39 (1.09)	3.16 (1.04)	–0.08	0.00	–0.03 (–0.40 to 0.35)	.00
Action planning	3.26 (1.17)	3.21 (1.08)	3.43 (1.06)	3.16 (1.06)	0.17	–0.05	0.27 (–0.19 to 0.74)	.01
Action control	3.05 (1.26)	2.93 (1.16)	3.38 (1.09)	2.82 (1.16)	0.33	–0.11	0.60 (0.15 to 1.04)	.03

[a] Differences between the intervention group and control group at posttest measurement are derived via linear regression analyses for linear variables (B and 95% CI are reported), correcting for the baseline score of Y, and demographic variables for which differences were found between groups at baseline (age, vocational education sector, year of education, and the use of active transport); corrected for multiple testing (based on false discovery rate).

[b] Except for action planning and action control for fruit and vegetable intake, snack and soft drink consumption, and PA (n=99).

[c] Except for action planning and action control for fruit and vegetable intake, snack and soft drink consumption, and PA (n=124).

[d] Except for fruit and vegetable intake and snack and soft drink consumption (n=126), moderate PA (n=124), and active transport (n=123) for behavioral outcomes, and as follows for fruit and vegetable intake, snack and soft drink consumption, and PA: attitude (n=99), self-efficacy and intention (n=96), perceived barriers (n=95), and action planning and action control (n=92).

[e] Except for fruit and vegetable intake and soft drink consumption (n=104), moderate PA (n=101), vigorous PA (n=98), and active transport (n=99) for behavioral outcomes, and action planning and action control for fruit and vegetable intake, snack and soft drink consumption, and PA (n=124).

[f] Physical Activity.

Group Comparison: Intervention Versus Control Group

Change scores for the intervention group were compared with change scores for the control group for both behavioral (primary) outcome measures and determinants (secondary outcome measures). All findings of the linear regressions are presented in Table 2.

Exploratory Analysis: Primary Outcomes and Determinants of Primary Outcomes

After correcting for multiple testing, we did not find significant differences in change scores between the intervention group and the control group for dietary intake and PA (see Table 2). There were no significant differences in change scores between the two groups on determinants of dietary intake and PA.

Exploratory Analyses: Active Users Versus the Control Group

The same regression analyses performed to compare the intervention and control groups were also used to compare the groups "active users" in the intervention group (29/103, 28.2%) and the control group (n=124). Allocation to these groups was based on self-reported intervention use. There were no significant baseline differences between active and nonactive users in the intervention group.

Baseline Differences Between Active Users and the Control Group

Compared to the control group, active users were more likely to participate in the economics vocational education sector (χ^2_1=90.4, P<.001) and active users were more likely to be in the first year of secondary vocational education (χ^2_1=12.3, P<.001). Therefore, we controlled for these differences in the following analyses. We also controlled for baseline differences between the active users and the control group in case they differed significantly on primary behavioral outcomes and exploratory determinant measures. We found that the active users at baseline were more likely to use active transport as compared to the control group (P=.04). Therefore, we controlled for the use of active transport at baseline in all subsequent analyses concerning active transport.

Exploratory Analysis: Active Users Versus the Control Group

After correcting for multiple testing, we found that active users reported marginally stronger increases in fruit intake (active users: mean change=0.51; control group: mean change=0.01; beta=0.34, P=.06, R^2 change=.02), stronger decreases in snack consumption (active users: mean change=–0.20; control group: mean change=–0.08; beta=–0.36, P=.01, R^2 change=.05), and

stronger increased use of active transport (active users: mean change=0.92; control group: mean change=–0.12; beta=1.58, P=.02, R^2 change=.03). In terms of unhealthy eating, results also revealed significant differences in action planning (active users: mean change=0.42; control group: mean change=0.07; beta=0.91, P=.01, R^2 change=.04), and PA (active users: mean change=0.44; control group: mean change=–0.05; beta=0.83, P=.03, R^2 change=.03), and action control (active users: mean change=0.63; control group: mean change=–0.05; beta=1.25, P=.001, R^2 change=.08).

Process Evaluation of Self-Reported and Registered Game Play

User data (an objective measure) showed that Balance It was played 771 times in total. These games primarily consisted of daily tasks (671/771, 87.0%) and individual game play (632/771, 82.0%). Of all the goals set (ie, type of tasks), players chose to improve their fruit intake in 15.0% of all cases (116/771 of which 44.0%, 51/116 of the goals were accomplished), 3.0% opted to increase their vegetable intake (23/771 of which 39%, 9/23 of the goals were accomplished), 29.1% opted to decrease their snack consumption (224/771 of which 70.1%, 157/224 of the goals were accomplished), 8.9% opted to decrease their soft drink consumption (69/771 of which 63%, 44/69 of soft drink-related goals were accomplished), 31.0% opted to increase their moderate PA (239/771 of which 54.8%, 131/239 of moderate PA goals were accomplished), and 13.0% opted to increase their vigorous PA (100/771 of which 39%, 39/100 of vigorous PA goals were accomplished). Goal accomplishment was more likely when participants were motivated (OR 2.6, 95% CI 1.9-3.5), and less likely when they did not have the time (OR 0.6, 95% CI 0.4-0.9) or when they experienced the location as a barrier (OR 0.6, 95% CI 0.4-0.9).

At posttest, 50% (15/29) of the participants who used the intervention reported that they played Balance It because they wished to have a healthier lifestyle, and 50% (14/29) played the game because they were asked to for the purpose of our study (29/103). Of the participants who did not play Balance It, 24% (18/74) reported that they did not have the time to play. The participants who used the intervention were, on average, neutral to positive about the Balance It app. When asked whether they were planning to recommend Balance It to others, participants gave a mean score of 3.14 (SD 1.03) on a scale ranging from 1 (very bad) to 5 (very good); likewise, the mean rating for the tutorial (using the same scale) was 3.72 (SD 0.75). Also, the specific game elements were evaluated neutrally to positively, on average, ranging from 3.43 (SD 1.00) on a scale ranging from 1 (very stupid) to 5 (very nice) for the construction worker, to 3.62 (SD 0.90) for the option of using special powers on the tower of an opponent. The mean overall rating given for the Balance It app (on a scale of 1 to 10, 1=the lowest grade, 10=the highest grade) was 6.71 (SD 1.96). The mean overall rating for the website (using the same scale) was 6.50 (SD 1.40).

Discussion

The aim of this study was to pilot the effects of Balance It, a serious game intervention targeting secondary vocational education students' dietary intake and PA.

Main Findings

No significant differences between the intervention and control groups in terms of dietary intake and PA (the primary outcomes) were observed. Additional exploratory analyses did not reveal significant differences in change scores between the intervention and control group in terms of psychological determinants of dietary intake and PA, as targeted by Balance It.

The study also revealed that the number of people that used the Balance It intervention was less than expected because only 27.6% used it as intended. For exploratory purposes, we examined the potential of Balance It among active users by comparing participants in the intervention group who reported that they had used the intervention with the control group. We did find that active users increased their fruit consumption marginally and active transport significantly, and showed stronger decreases in snack consumption compared to the control group. Although we should acknowledge that other factors could explain these differences (ie, self-selection), the findings could indicate Balance It may contribute to changes in PA and dietary intake if used as planned.

Taking into account that a difference of 100 kcal in daily caloric intake/expenditure can contribute to overweight prevention [42], the increase in active transportation by 0.92 days on average may contribute to the prevention of overweight. Snack consumption was only decreased by a mean 0.20 snack portions per day, which may not be sufficient to make a change in daily energy balance. Nevertheless, active users showed an improvement in action planning and action coping skills to decrease their snack consumption, which may be related to students' increase in fruit consumption as a healthy alternative to unhealthy snacks. Nevertheless, the changes observed may not be large enough to prevent overweight. In general, active users rated the intervention moderately positively and registered data showed that active users mainly opted to decrease their snack consumption (29.1%) or to increase their moderate PA (31.0%).

Consistent with previous serious gaming studies targeting dietary intake and PA, our results suggest that the use of a self-regulation game intervention could improve dietary intake and active transport among youth [24,43]. Despite the significant increase in use of active transport for the intervention users, they did not report a significant increase in moderate or vigorous PA. That we found no direct effect of the intervention on PA may be because posttest measures took place after 4 weeks of game play, whereas PA goals were partially set on a weekly basis, taking up to 6 weeks of game play, and therefore were not yet completed at the time of the posttest measure. The 4-week period may also be too short to expect change in PA [44].

Previous research shows that youth from low SES families are less engaged with health behaviors and not as successful in

XSL•FO

RenderX

terms of translating their health intentions into behavior [45]. Therefore, we incorporated gameplay, reminders, professional support, and social contracting into our serious game to stimulate intervention use. Moreover, we plan to embed the intervention within an existing student tracking system as suggested by Crutzen and other authors [46-48] (see [30] for a more detailed description of Intervention Mapping step 5: program implementation) to stimulate initial use. Despite these attempts to encourage implementation use, only 27.6% of the intervention group reported actual use of Balance It. Although these active users were moderately positive about Balance It, it may be that players were not sufficiently transported into the game by the narrative of Balance It because the narrative that was used to stimulate immersion is more in line with what Lu et al [49] describe as an "instruction," which does not adequately facilitate immersion. According to Lu et al, narratives should have attractive features (eg, a plot, a beginning, middle, and an end), and should allow players to experience a character's happiness on their journey toward adoption of a healthy behavior more directly and vividly than didactic instruction alone [49]. However, in practice, professional game designers often stress the importance of game simplicity to enhance motivation to play (see also [50]), limiting the space for extensive narratives to be included within a game. One way of resolving this issue would be to use so-called novellas, which can be defined as highly immersive stories designed to increase engagement with the game intervention provided prior to the game [51]. Placing novellas outside the Balance It app may also be beneficial in terms of exposure because students were stimulated to follow the link to the Balance It website by the research assistant during pretest, but the research assistant was not able to check if all students downloaded the app. As such, participants would be more likely to be exposed to the novella, increasing the likelihood of intervention (or Balance It app) use. These so-called "novellas" are still in their infancy, but because they seem to be a rather promising way of countering low levels of engagement, further research is recommended.

Limitations

Some limitations of this study should be acknowledged. First, the study is a cluster randomized trial, which was chosen over a randomized controlled trial because of practical considerations (ie, school coordinators who wanted their students to be in the same condition), to prevent contamination effects and to enhance participant compliance [31,52]. It should also be noted that participants who were at criteria for the outcomes (ie, students who already eat healthy and have sufficient PA) at baseline were included in the cluster randomized trial, which may have reduced the overall effect size. A second limitation is that only a pretest and posttest were included in this study, and follow-up measures to evaluate long-term effects were lacking.

Consequently, the results of this study are based on change scores collected over a 4-week period, whereas many PA-related goals that were set by game players took 6 weeks to accomplish. As such, we might not expect the full effects of Balance It to reveal themselves until participants had used it for 6 weeks. Participants were contacted via email and requested to fill in the follow-up questionnaire 4 weeks after the posttest measure, but with a response rate of only 4%, we did not analyze these data. Finally, email addresses were inaccurately reported (or not reported) in the Balance It app, which was unfortunate because we needed them to merge (objective) user data with self-reported survey data at baseline and at posttest. Consequently, user data could not be merged with self-reported data and conclusions about the objective use of the intervention in relation to the learning outcomes should be interpreted with care. We recommend that future research use a randomized controlled trial design and incorporate appropriate follow-up measures and checks regarding the collection of email addresses or the inclusion of other merge variables.

Finally, it should be noted that despite the potential of the peer-support component as included in the Balance It website to increase intervention effects on self-regulation skills [53], usage of the peer-support system was rather disappointing (ie, only 7% of the users reported actual use of the peer-support system). The lack of peer-support system use is a commonly reported problem in online interventions [54], although if used, peer-support systems could have facilitated behavior change and problem-solving processes [29]. Because the peer-support system was placed outside the Balance It app as a result from a trade-off that was made between developers and behavior change experts about the number of layers within the game, visiting the system online may have been a barrier. A second explanation for the lack of peer-support system use may also be derived from the limited number of schools (ie, n=4) included in this cluster randomized trial. In total, only two schools were allocated to the intervention group, indicating that most participants knew one another in daily life. If students preferred to receive or provide social support, this may have taken place in real life instead of through the online peer-support system of Balance It.

Conclusion

The Balance It intervention did not show favorable effects on dietary intake and PA compared to the control condition. However, only a small number of people in the intervention condition actually used Balance It (27.6%). Exploratory analyses did suggest that, if used as planned, Balance It could contribute to changing dietary intake and PA behaviors, albeit it remains debatable whether this would be sufficient to prevent overweight.

Conflicts of Interest

None declared.

References

1. Eisenberg ME, Neumark-Sztainer D, Story M. Associations of weight-based teasing and emotional well-being among adolescents. Arch Pediatr Adolesc Med 2003 Aug;157(8):733-738. [doi: 10.1001/archpedi.157.8.733] [Medline: 12912777]

2. Formiguera X, Cantón A. Obesity: epidemiology and clinical aspects. Best Pract Res Clin Gastroenterol 2004 Dec;18(6):1125-1146. [doi: 10.1016/j.bpg.2004.06.030] [Medline: 15561643]

3. Roberts RE, Deleger S, Strawbridge WJ, Kaplan GA. Prospective association between obesity and depression: evidence from the Alameda County Study. Int J Obes Relat Metab Disord 2003 Apr;27(4):514-521. [doi: 10.1038/sj.ijo.0802204] [Medline: 12664085]

4. Storch EA, Milsom VA, Debraganza N, Lewin AB, Geffken GR, Silverstein JH. Peer victimization, psychosocial adjustment, and physical activity in overweight and at-risk-for-overweight youth. J Pediatr Psychol 2007;32(1):80-89 [FREE Full text] [doi: 10.1093/jpepsy/jsj113] [Medline: 16601255]

5. Stunkard AJ, Faith MS, Allison KC. Depression and obesity. Biol Psychiatry 2003 Aug 1;54(3):330-337. [Medline: 12893108]

6. Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases. Diet, Nutrition and the Prevention of Chronic Diseases. Geneva: Word Health Organization; 2004. URL: http://apps.who.int/iris/bitstream/10665/42665/1/WHO_TRS_916.pdf [accessed 2016-08-05] [WebCite Cache ID 6jWyyOjeN]

7. Bernaards C, Van Buuren S. Rapportage veranderingen in het beweeggedrag van mbo studenten: TNO (No. R10185). Leiden, the Netherlands: TNO Behavioural and Societal Sciences; 2012. URL: http://www.stefvanbuuren.nl/publications/2012%20Veranderingen%20in%20het%20beweeggedrag%20van%20mbo%20studenten.pdf [accessed 2016-08-04] [WebCite Cache ID 6jVZ5MVsr]

8. Schönbeck Y, van Buuren S. Factsheet resultaten vijfde landelijke groeistudie. Leiden, the Netherlands: TNO; 2010. URL: https://www.tno.nl/media/1996/20100608-factsheet-resultaten-vijfde-landelijke-groeistudie1.pdf [accessed 2016-08-05] [WebCite Cache ID 6jWzD3v0P]

9. Wang Y, Monteiro C, Popkin BM. Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China, and Russia. Am J Clin Nutr 2002 Jun;75(6):971-977 [FREE Full text] [Medline: 12036801]

10. Wang Y, Liang H, Tussing L, Braunschweig C, Caballero B, Flay B. Obesity and related risk factors among low socio-economic status minority students in Chicago. Public Health Nutr 2007 Sep;10(9):927-938. [doi: 10.1017/S1368980007658005] [Medline: 17381954]

11. Tate D. Application of innovative technologies in the prevention and treatment of overweight in children and adolescents. In: Jelalian E, Stelle RG, editors. Handbook of Childhood and Adolescent Obesity. New York, NY: Springer Science and Business Media; 2008:378-404.

12. van Keulen HM, Mesters I, Ausems M, van Breukelen G, Campbell M, Resnicow K, et al. Tailored print communication and telephone motivational interviewing are equally successful in improving multiple lifestyle behaviors in a randomized controlled trial. Ann Behav Med 2011 Feb;41(1):104-118 [FREE Full text] [doi: 10.1007/s12160-010-9231-3] [Medline: 20878293]

13. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51:214-221. [doi: 10.1016/j.ypme.2010.06.004]

14. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006;31(3):205-223. [doi: 10.1207/s1532496abm3103_2]

15. Nguyen B, Kornman KP, Baur LA. A review of electronic interventions for prevention and treatment of overweight and obesity in young people. Obes Rev 2011 May;12(5):e298-e314. [doi: 10.1111/j.1467-789X.2010.00830.x] [Medline: 21348921]

16. 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 Sep;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]

17. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Internet-delivered interventions aimed at adolescents: a Delphi study on dissemination and exposure. Health Educ Res 2008 Jun;23(3):427-439 [FREE Full text] [doi: 10.1093/her/cym094] [Medline: 18209115]

18. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]

19. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]

20. Baranowski T, Baranowski J, Cullen KW, Marsh T, Islam N, Zakeri I, et al. Squire's Quest! Dietary outcome evaluation of a multimedia game. Am J Prev Med 2003 Jan;24(1):52-61. [Medline: 12554024]

21. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]

22. Graesser A, Chipman P, Leeming F, Biedenback S. Deep learning and emotion in serious games. In: Ritterfeld U, Cody M, Vorderer P, editors. Serious Games, Mechanisms and Effects. New York, NY: Routledge; 2009.

XSL•FO
RenderX

23.   Prensky M. Digital Game-Based Learning. St Paul, MN: Paragon House; 2007.

24.   Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]

25.   Thompson D, Bhatt R, Vazquez I, Cullen KW, Baranowski J, Baranowski T, et al. Creating action plans in a serious video game increases and maintains child fruit-vegetable intake: a randomized controlled trial. Int J Behav Nutr Phys Act 2015 Mar;12:39 [FREE Full text] [doi: 10.1186/s12966-015-0199-z] [Medline: 25890060]

26.   Araújo-Soares V, McIntyre T, MacLennan G, Sniehotta FF. Development and exploratory cluster-randomised opportunistic trial of a theory-based intervention to enhance physical activity among adolescents. Psychol Health 2009 Sep;24(7):805-822. [doi: 10.1080/08870440802040707] [Medline: 20205028]

27.   Maes S, Karoly P. Self-regulation assessment and intervention in physical health and illness: a review. App Psych 2005 Apr;54(2):267-299. [doi: 10.1111/j.1464-0597.2005.00210.x]

28.   van Genugten L, van Empelen P, Flink I, Oenema A. Systematic development of a self-regulation weight-management intervention for overweight adults. BMC Public Health 2010 Oct;10:649 [FREE Full text] [doi: 10.1186/1471-2458-10-649] [Medline: 20979603]

29.   Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernández ME. Planning Health Promotion Programs: An Intervention Mapping Approach. New York, NY: Jossey-Bass; 2011.

30.   Spook JE, Paulussen T, Paulissen R, Visschedijk G, Kok G, van Empelen P. Design rationale behind the serious self-regulation game intervention "Balance It": overweight prevention among secondary vocational education students in the Netherlands. Games Health J 2015 Oct;4(5):387-400. [doi: 10.1089/g4h.2014.0142] [Medline: 26181575]

31.   Donner A, Klar N. Design and analysis of cluster randomization trials in health research. Int J Epidemiol 2001;30(2):407-408. [doi: 10.1093/ije/30.2.407-a]

32.   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]

33.   van Assema P, Brug J, Ronda G, Steenhuis I. The relative validity of a short Dutch questionnaire as a means to categorize adults and adolescents to total and saturated fat intake. J Hum Nutr Diet 2001 Oct;14(5):377-390. [Medline: 11906579]

34.   Douwes M, Hildebrandt VH. Vragen naar de mate van lichamelijke activiteit; onderzoek naar de test-hertest betrouwbaarheid en congruente validiteit van een vragenlijst. Geneeskunde En Sport 2000;33(1):9-16.

35.   van der Horst K, Oenema A, van de Looij-Jansen P, Brug J. The ENDORSE study: research into environmental determinants of obesity related behaviors in Rotterdam schoolchildren. BMC Public Health 2008 Apr;8:142 [FREE Full text] [doi: 10.1186/1471-2458-8-142] [Medline: 18442370]

36.   van Genugten L, van Empelen EP, 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 Jul;3(3):e31 [FREE Full text] [doi: 10.2196/resprot.2599] [Medline: 25057122]

37.   Spook JE, Paulussen T, Kok G, Van EP. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]

38.   Spook J, Paulussen T, Kok G, Van Empelen P. Keeping it Real: Monitoring and Understanding Health Behavior in Daily Life [dissertation]. Meppel, The Netherlands: PrintSupport4U; 2016.

39.   Sniehotta F, Nagy G, Scholz U, Schwarzer R. The role of action control in implementing intentions during the first weeks of behaviour change. Brit J Soc Psychol 2006;45:87-106. [doi: 10.1348/014466605X624600]

40.   Statistics Netherlands. Hoe doet het CBS dat nou? Standaarddefinitie allochtonen [How Does Statistics Netherlands Do This? Standard Definition Immigrants].. Voorburg: Statistics Netherlands; 2009. URL: https://www.cbs.nl/ [accessed 2016-08-04] [WebCite Cache ID 6jVbGNT3Q]

41.   Rijpstra A, Bernaards C. De leefstijl van MBO studenten in Nederland 2009/2010 (No. TNO/LS 2011.014). Leiden, the Netherlands: TNO; 2011. URL: https://www.gezondeschool.nl/content/de-leefstijl-van-mbo-studenten-nederland-20092010 [accessed 2016-08-04] [WebCite Cache ID 6jVbLsDMF]

42.   Hill JO, Wyatt HR, Reed GW, Peters JC. Obesity and the environment: where do we go from here? Science 2003 Feb 7;299(5608):853-855. [doi: 10.1126/science.1079857] [Medline: 12574618]

43.   Baranowski T, Davis M, Resnicow K, Baranowski J, Doyle C, Lin LS, et al. Gimme 5 fruit, juice, and vegetables for fun and health: outcome evaluation. Health Educ Behav 2000 Feb;27(1):96-111. [Medline: 10709795]

44.   Prochaska JO. Transtheoretical model of behavior change. In: Gellman M, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer-Verlag; 2013:1997-2000.

45.   Conner M, McEachan R, Jackson C, McMillan B, Woolridge M, Lawton R. Moderating effect of socioeconomic status on the relationship between health cognitions and behaviors. Ann Behav Med 2013;46:19-30. [doi: 10.1007/a12160-013-9481-y]

46.   Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]

XSL•FO
RenderX

47. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]

48. Crutzen R, Kohl L, De Vries NK. Kennissynthese online preventie. Maastricht: Datawyse; 2013.

49. Lu AS, Baranowski T, Thompson D, Buday R. Story immersion of videogames for youth health promotion: a review of literature. Games Health J 2012 Jun;1(3):199-204 [FREE Full text] [doi: 10.1089/g4h.2011.0012] [Medline: 24416639]

50. Fogg BJ. A behavior model for persuasive design. 2009 Presented at: Persuasive '09: the 4th International Conference on Persuasive Technology; April 26-29, 2009; Clarement, CA.

51. Baranowski T, Baranowski J. Using fictional novellas as narrative companion for vegetable parenting game. 2015 Presented at: ISBNPA 2015: 14th meeting of the International Society for Behavioral Nutrition and Physical Activity; June 3-6, 2015; Edinburgh p. A.

52. McKenzie J, Ryan R, Di Tanna TG. Cochrane Consumers and Communication Review Group: cluster randomised controlled trials. 2014. URL: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/ClusterRCTs.pdf [accessed 2016-08-04] [WebCite Cache ID 6jVcB61eZ]

53. Heisler M. Different models to mobilize peer support to improve diabetes self-management and clinical outcomes: evidence, logistics, evaluation considerations and needs for future research. Fam Pract 2010 Jun;27 Suppl 1:i23-i32 [FREE Full text] [doi: 10.1093/fampra/cmp003] [Medline: 19293400]

54. 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]

Abbreviations

BMI: body mass index
OR: odds ratio
PA: physical activity
SES: socioeconomic status

XSL•FO
RenderX

Original Paper

Lessons Learned for Online Health Community Moderator Roles: A Mixed-Methods Study of Moderators Resigning From WebMD Communities

Jina Huh[1], PhD; Rebecca Marmor[1], MD; Xiaoqian Jiang[1], PhD

University of California San Diego, Department of Medicine, La Jolla, CA, United States

Corresponding Author:
Jina Huh, PhD
University of California San Diego
Department of Medicine
9500 Gilman Dr. MC 0881
La Jolla, CA, 92093-0881
United States
Phone: 1 8582462562
Fax: 1 858 534 8469
Email: jinahuh@ucsd.edu

Abstract

Background: Online health community (OHC) moderators help facilitate conversations and provide information to members. However, the necessity of the moderator in helping members achieve goals by providing the support they need remains unclear, with some prior research suggesting that moderation is unnecessary or even harmful for close-knit OHCs. Similarly, members' perceptions of moderator roles are underexplored. Starting January of 2013, WebMD moderators stopped working for WebMD communities. This event provided an opportunity for us to study the perceived role of moderators in OHCs.

Objective: We examine the OHC members' perception on OHC moderators by studying their reactions toward the departure of moderators in their communities. We also analyzed the relative posting activity on OHCs before and after the departure of moderators from the communities among all members and those who discussed moderators' departures.

Methods: We applied a mixed-methods approach to study the posts of all 55 moderated WebMD communities by querying the terms relating to discussions surrounding moderators' disappearance from the WebMD community. We performed open and axial coding and affinity diagramming to thematically analyze patients' reactions to the disappeared moderators. The number of posts and poster groups (members and moderators) were analyzed over time to understand posting patterns around moderators' departure.

Results: Of 821 posts retrieved under 95 threads, a total of 166 open codes were generated. The codes were then grouped into 2 main themes with 6 total subthemes. First, patients attempted to understand why moderators had left and what could be done to fill the void left by the missing moderators. During these discussions, the posts revealed that patients believed that moderators played critical roles in the communities by making the communities vibrant and healthy, finding solutions, and giving medical information. Some patients felt personally attached with moderators, expressing they would cease their community participation. On the other hand, patients also indicated that moderators were not useful or sometimes even harmful for peer interactions. The overall communities' posting activity, which was already in decline, showed no significant difference before and after the moderators' departure. In fact, the overall posting activities of the communities were declining well before the moderators' departure. These declining posting activities might be the reason why WebMD removed the moderators.

Conclusion: Compassionate moderators who provide medical expertise, control destructive member posts, and help answer questions can provide important support for patient engagement in OHCs. Moderators are in general received positively by community members and do not appear to interfere with peer interactions. Members are well aware of the possibility of misinformation spreading in OHCs. Further investigation into the attitudes of less vocal community members should be conducted.

XSL•FO

RenderX

Introduction

In-person patient support groups, often organized by hospitals and clinical moderators, are a well-established mechanism to encourage peer-patient interaction, help patients improve self-efficacy, and educate patients about self-care management [1-4]. As Web 2.0 and social media spread as one of the main Internet activities, OHCs have also proliferated, often without moderators [5]. Unmoderated communities can suffer from the negative consequences of misinformation and poor social dynamics (eg, trolling) if not well-maintained by community members [6-8], especially when the interest of the community is health. The addition of moderators or active commitment by the members can diminish such negative consequences of OHCs [6-9]. However, the cost of resources is high for hiring moderators, preferably those with clinical backgrounds. In addition, moderating thousands of posts [5], and motivating moderators to voluntarily participate in OHCs can be difficult [10].

To successfully administer OHCs, we need to understand the critical role that moderators have in OHCs. A study revealed effective moderation styles for various negative online behaviors (eg, trolling) [11]. Although the effectiveness of moderation styles (eg, rewarding vs punishing) has been studied, there is no consensus regarding the necessity and role of moderators on OHC retention and improving levels of social support, where prior research reveals conflicting results. A study showed that moderators may be important for both the vibrancy of forums and improving patient outcomes [7,12]. Moderators review postings, redirect conversations, and stimulate dialogue when forum activity lags. They also execute a "process function" and help establish and enforce community rules [13]. Moderators offer valuable help that clinicians cannot provide, including suggesting ways to communicate with health care providers and finding useful health information resources for self-management [14]. The necessity of such external governance in moderating troll conversations may be dependent on the specific community. For those OHCs where patients have already established strong rapport with one another and are self-policing community conversations, external governance can be unnecessary [15] or sometimes even disruptive [16]. Online health communities independently run by patients only can self-maintain high information quality [17], although a systematic review showed that the effectiveness of purely peer-patient–based OHCs in terms of clinical outcomes lacks RCT-based evidence [18].

Unlike in face-to-face support groups, where moderators are often clinicians known to participants, moderators in OHCs do not have an immediate connection and the trust of members. Moderators are also unable to provide medical consultation. Furthermore, one prior study found that moderators urging patients to talk to their health care providers instead of consulting the community was highly associated with decreased peer-patient interaction [16]. Health experts might not understand the needs of patients, generating potential

communication breakdown in OHC settings unless explicitly addressed [19]. Although young people who self-harm were willing to share their experiences with health professionals, the health professionals had low participation in an OHC [10]. The conflicting findings of prior research on patients' expected moderator roles and utility of using moderators demonstrate the need for further investigation into the necessity of moderators in OHCs.

WebMD [20], an online health information portal, has run over 60 OHCs, in which staff moderators (STMs) and health professional moderators (HPMs) participated in peer-patient conversations as moderators. Staff moderators in WebMD were those without clinical backgrounds, who help facilitate and moderate discussions. For instance, they would ask questions such as, "What is your plan for this week? Please share with us" to facilitate conversations among the members and build rapport. As for responding to others' threads, the moderators in WebMD were given guidelines as delineated in their policy statement [21], which says that the moderators are not supposed to provide medical consultations. Thus, STMs shared general information resources, such as pointers to accredited websites (eg, ADA.org).

Health professional moderators at WebMD were those with clinical backgrounds, providing clinical expertise to patient members by answering members' questions and participating in conversations. Nevertheless, the HPMs would follow the same guidelines given to STMs, where HPMs would not be allowed to give personalized medical consultation. Health professional moderators could clarify clinical concepts and knowledge that would be useful for self-management, or direct ways to discuss with the members' own health care providers.

WebMD, however, made a decision to let go of the STMs starting January of 2013. Members were not notified of this change in staffing prior to its occurrence, but quickly picked up on the absence of moderators. The departure of STMs triggered patient members to talk about their experiences with moderators, including HPMs, as part of the communities. The sudden departure of moderators from WebMD provided critical information about how patients perceived the benefits and disadvantages of having moderators in their communities when faced with loss of moderation.

While prior research has clearly demonstrated the benefit of moderators in OHCs, there is a limited understanding of the community member's perceptions of moderators in their participation with OHCs. In this paper, we will investigate: (1) member's perceptions of moderators in OHCs, and (2) the summary statistics analysis on OHC retention and moderators' removal to understand both qualitative and quantitative aspects of moderators' roles and influences in OHCs.

Methods

We first walk through the data collection process, including community selection, followed by the descriptions on our mixed-methods analysis.

Data Collection

Our institutional review board (IRB) determined this study to be unregulated because our data were publicly available and did not involve human subjects. WebMD has a total of 55 moderated communities. In January of 2015, we wrote an automated crawling program, which downloaded all 55 communities and saved the data to a local database software, MySQL (Oracle) [22]. The crawling program accessed the URLs of the WebMD OHCs, which are publicly available, and downloaded the content of the page (title, post content, date of post, and poster ID) in which OHC members exchanged discussions. The data in MySQL were then queried and analyzed for summary statistics analysis and exported to Microsoft Excel for thematic coding. Overall posting activity level of all 55 communities (Multimedia Appendix 1) was analyzed, which were the total number of posts, members, HPMs, and STMs to use the information for selecting the communities to study.

To best select active communities that can lend us more data regarding members' discussions on moderators' roles at WebMD, the following process was adopted: (1) From the total of 55 WebMD moderated communities, the top 20 communities which had the most posts in total after excluding nonhuman health-related community (eg, pet health) were selected. (2) Posts made before November 2012 and those made by moderators were excluded. (3) A regular expression query with the keywords related to moderators was conducted: "moderator[s]," "mod[s]," "[moderator names of that community]." The query was conducted on both the title and the message body of each post; (4) All posts under the threads that contained the query result posts were then collected. (5) The communities that came with zero query results when (3) was performed were excluded. (6) To further filter whether the thread was about (*a*) *moderators leaving* or (*b*) *members' perceived moderators' roles*, 2 coders manually went through each post to determine the relevance (further delineated in the Analysis section) and (7) The communities to investigate using our qualitative and quantitative analysis procedures were finalized. The result of this process is delineated in Figure 1 in the results section.

Figure 1. From all moderated WebMD communities (n=55), the communities were ranked based on total posting number. Then, top 20 communities with the most posting activities after excluding communities about nonhuman health-related topic (eg, pet health) were selected. Keyword search to find posts related to moderators was performed. All communities with more than zero query results when keyword search was performed (n=14) were included.
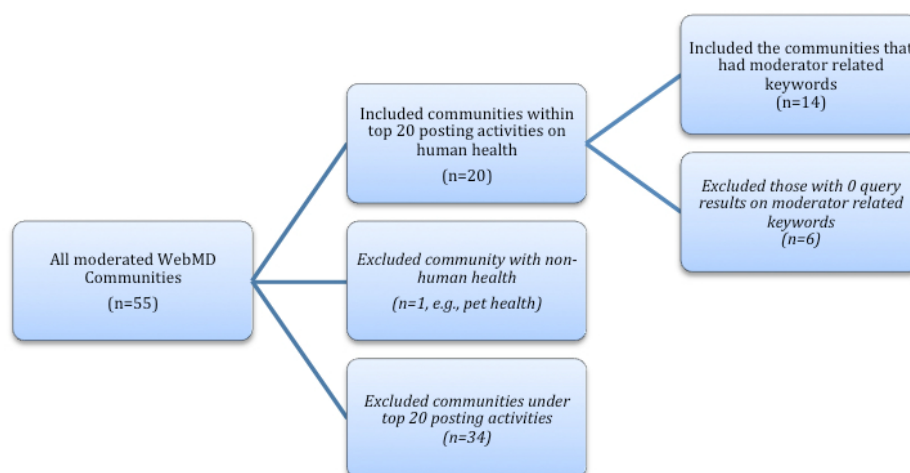
Figure 2. This figure describes further extracting related posts for this analysis from the 14 selected WebMD communities. From 14 selected communities, 11 communities contained related discussions of moderators. These posts were then tracked back to its original threads. All replies under those threads were retrieved, which came to be 821 total posts made by 199 unique posters.
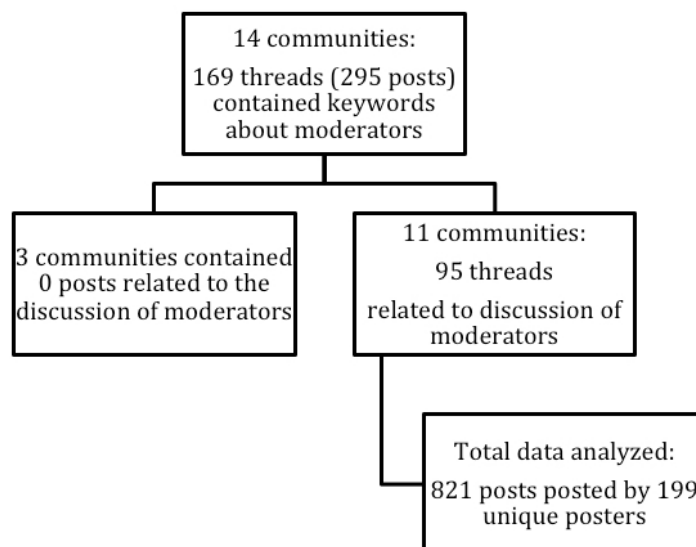
Analysis

We analyzed the total number of posts and unique posters grouped by STM, HPM, and members. Each poster group's posting activities were recorded by month throughout the data collection period. Another study [14] showed that moderators triggered conversations among members by initiating threads, whereas replies were used to ensure members who asked questions were responded with welcoming responses. To understand such moderators' member participation inducing effort together with members' posting pattern, we also analyzed general posting activities such as making distinctions between replies versus thread initiating posts. The posting activities were visualized to investigate the overall posting pattern over time.

From the retrieved posts in step (6) of the data collection, 2 coders worked together with the first 5 communities on determining the criteria for relevance, asking about the posts that were related to discussion about moderators. If the post contained any dialogue about WebMD moderators—either STMs or HPMs, the post was flagged as relevant. The resulting criteria for relevance were as follows: First, the post should be about moderators, not other synonym that did not mean community moderators (eg, mod podge). Second, even though the post does not explicitly mention moderators' removal from WebMD, if it contains any opinion about the moderators, it is included. The 2 coders divided the rest of the communities to code for relevance.

We analyzed the whole thread in which related posts were its part. One main reason for excluding the thread was that when the whole thread of a post ended without someone intervening to inform the poster that the moderator had left, in case a moderator left without the poster's knowledge.

For the posts determined as relevant, a mixed-methods approach was adopted to understand the phenomenon using thematic analysis in addition to the quantitative methods employed. The thematic analysis [23] was used to identify topics and themes that emerged around discussions about the moderators' disappearance or their perceived roles. These open codes were applied at the sentence level and were mutually inclusive.

With the resulting codes, an affinity diagram exercise was performed [24] (See Multimedia Appendix 2) to conduct axial coding. Affinity diagramming allowed us to find salient affinities among the codes. This process led to finding common and distinct themes across the codes, allowing main themes to emerge from the data. As a result, 2 main themes with three subthemes each were identified. These themes were applied at the post level, so each post would have multiple distinct themes, but a theme would not be repeated at the post level.

Results

Summary

In this section, we first report the community selection and the summary statistics of posting activities. We then further examine qualitative aspects of how members perceived having moderators, providing us with insights for how moderators should be implemented in OHCs. The follow-up posting activity analysis of the members who participated in the discussions was then reported.

Community Selection

A total of 14 communities were found to have included moderator-related keywords after having filtered out nonhuman health-related communities and relatively less active communities.

Filtering and Posting Activity Analysis

In this section, the results of the overall posting activities, identification of related posts for further qualitative analysis, and the posting activity analysis of the members who participated in the discussions about moderators leaving WebMD communities have been reported.

Overall Posting Activities of Staff Moderators, Health Professional Moderators, and Members

For the 14 resulting WebMD communities for us to analyze, 6-15 STMs and 1-7 HPMs have been moderating each community. It was found that these communities had, on average, 650 members per STM (ranging from 260 members per moderator on the Breast Cancer community to 1526 members per moderator on the Sex and Relationships community). On average, STMs started 255 threads and posted 1115 replies (from 98 to 747 thread starting posts and from 102 replies to 3209 replies, n=14). On average, HPMs started 36 threads and posted 690 replies (from 0 to 119 thread starting posts and from 0 to 3348 replies). Further detailed breakdown of the activities are delineated in the Multimeda Appendix 1.

Identifying Related Posts

Table 1 shows the query results and the filtering results on further identifying which posts were about moderators leaving the WebMD communities. From the 14 communities, a total of 293 posts under 169 threads were retrieved, which contained keywords about moderators. Of the 169 threads, 95 threads were related to the discussion about moderators' exit. During this process, 3 communities were excluded from the analysis because all their query results were not found to be about the moderators' exit. All replies from the 95 threads were then retrieved, which resulted in a total of 821 posts made by 199 unique posters under 95 threads by 11 communities.

Table 1. Eleven WebMD communities were identified as containing related posts on moderators leaving WebMD.

WebMD community	Number of posts identified as containing keywords (posts or number of threads the posts belong to)	Number of posts identified as related to moderators leaving (Number of total posts under the threads the related post belongs to)
Fibromyalgia	122 posts in 68 threads	78 posts from 33 threads (261 posts)
Breast_cancer	40 posts in 22 threads	31 posts from 14 threads (123 posts)
Diabetes	29 posts in 14 threads	23 posts from 9 threads (87 posts)
Back_pain	24 posts in 15 threads	20 posts from 11 threads (103 posts)
Pain_management	21 posts in 16 threads	12 posts from 9 threads (130 posts)
Sexual_conditions_and_sexually transmitted diseases	1 post in 1 thread	1 post from 1 thread (3 posts)
Sex_and_relationships	15 posts in 4 threads	15 posts from 4 threads (55 posts)
Bipolar_disorder	12 posts in 11 threads	9 posts from 8 threads (36 posts)
Lupus	12 posts in 3 threads	11 posts from 2 threads (13 posts)
Anxiety_and_panic_disorders	5 posts in 3 threads	1 post from 1 thread (2 posts total)
Depression	4 posts in 4 threads	3 posts from 3 threads (8 posts)
Diet	4 posts in 4 threads	0 threads (0 posts)
Pregnancy	3 posts in 3 threads	0 threads (0 posts)
Infertility_and_reproduction	1 post in 1 thread	0 threads (0 posts)
Total	293 posts in 169 threads	204 posts from 95 threads (821 posts)

Activities of the 11 Communities

As shown in Figure 3, the overall posting activity of the members in 11 communities started declining during early 2010. This is 2 years before the STMs stopped moderating WebMD communities. For those communities with declining participation pattern, it can be seen that the moderators' activities, both STM and HPM, have increased dramatically. However, members' participation did not increase thereafter, and STMs' activities have slowly decreased since late 2010 until STMs left WebMD communities. Staff moderators' proportion of thread initiating posts ranged from 9% to 49% (Mean=26%). For HPMs, it ranged from 0% to 14% (Mean=6%). Detailed breakdown of this analysis is included in the Multimedia Appendix 1.

The last date of the STMs' posts ranged between November 17 and December 20, 2012. Members' posts related to moderators'

departure started as early as November 17, 2012 (the last posting date of the Sex and Relationships community) and persisted till December 20, 2012. There was a one-time post, an identical one, made by one STM in March to a few of the communities, asking members to share their experiences around insurance to a WebMD email address. Other than this particular post, no other posts have been made by STMs after December 20, 2012. One STM came back to the Fibromyalgia community with a regular member profile, updating about her current situation of finding a new job and health issues she is dealing with.

In December 2012 (when STMs last posted) and January 2013, the discussion about moderator leaving WebMD peaked. Since then, members continued to talk about how moderators left, sharing opinions and solutions around STMs' departure.

Figure 3. Posting activities of the members, STMs, and HPMs over time in the context of when STMs stopped moderating WebMD communities.

Members' Qualitative Perceptions Toward Their Moderators

The thematic analysis of 821 posts resulted in generating 166 codes. From these codes, 6 subthemes grouped into 2 main themes from conversations about moderators during the period from December 2012 to February 2014 (see Table 2) were identified. All these coded posts were centered around moderators' disappearance from the communities, which led to further discussions about members' perceptions toward having moderators on the forum.

The process by which the members figured out the news that the STMs have left WebMD was consistent across the 11 communities and is as follows: members first noticed moderators were gone; next, the groups shared the news individuals received from WebMD that the moderators had been removed from the forums; and finally, members pondered why the moderators had been removed, shared members' initial reaction to the decision, and discussed how the community would continue to function in their absence (Figure 4).
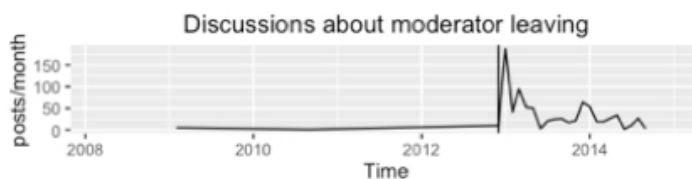
Figure 4. Members' discussion on moderator leaving peaked when moderators left and continued until the data collection ended. Some of the posts about moderator being gone were made as a reply to a post made a few years back (2009-2012). That is why earlier discussion points before December, 2012 are included in this graph.

Table 2. The table shows the number of coding instances and example quotes on WebMD users' discussions around moderators' departure.

Coding instances (frequency of code application) and example quotes	
Code name	Example quotes
Figuring out that moderators are gone	
Receiving the news	*I emailed about the absence of the moderators as well... I know many have been missing them and wondering where they went. This is the answer I got:*
	WebMD has decided to shift the focus of the communities from being WebMD-managed to now being more member-managed, allowing members to shape their conversations and their communities in the direction that suits their needs. We have seen this as a growing trend among other social networking sites or message boards and feel it will better facilitate interaction among our members. Even with this change, we will continue to invite experts to keep answering member questions as they have been in the past. Along with this, we have also recently made some significant updates to our Answers tool (answers.webmd.com) and we hope you will make use of this resource as well. [Diabetes]
	I think it is a shame that WebMD did not communicate better with the members here [Fibromyalgia]
Attempting to understand reasons for why moderators left	*I guess WebMD went through a change and decided that these people were no longer needed...I think though this is how companies do things like this these days. The less we know the better it goes is what they think, maybe. [Fibromyalgia]*
Discussing solutions to moderators' departure	*We need to stay strong for each other and the new ones that come here for our help. [Breast Cancer]*
	... As far as I know there aren't any live moderators on but if there is a problem with a post, such as spam, etc, we all use the "Report This" link to bring it to one of the offline mods' attention. [Depression]
Patient members' experience with moderators	
Moderators were not useful: Peer support is more important	*WebMD's policy has always been to simply delete any thread that anyone ever complains about, which is ridiculous... I really can't think of more than a handful of times over the years I've posted here that I saw a post including clearly inappropriate slang language. This is the one area in which the reduced moderator presence these boards is a feature, not a bug, IMHO. [Sex and Relationships] I have always maintained that the successful treatment of diabetes is 90% patient and only 10% doctor or caregiver [...] a lot of the ivory tower advice being given by those not affected by the disease may or may not apply to a particular patient. [Diabetes]*
Moderators provided pragmatic help	*Remember the good-ole-days when our Mods would assist and watch over us and delete such nonsense? [Breast Cancer]*
	Some of you may recall I left the WebMD communities for several months earlier this year. I decided to come back and try again. Sadly, I can no longer continue supporting any organization or website that condones rude posts filled with vulgar language and free unethical advertising. While this does not happen often in the Back Pain Community, it is happening in other communities that I visit. Posts are reported but are not deleted in a timely fashion as they were when there were moderators in each forum. The attacks are getting more serious and are staying around for days before being deleted. I have decided to contact WebMD and have this account deleted. I can still read, but will no longer participate. I pray you all find effective ways to manage your back pain issues. Click on my username or avatar picture to read my story. Blessings - [Poster name anonymized] [Back pain]
	Why is no one removing the spam that is saturating this board????? [Bipolar Disorder]
Established personal tie with moderators	*I just liked having her here and getting to know her and see her little ones pictures on here I will miss them also [Breast Cancer]*
	Miss them all [Bipolar disorder]

Figure 5. 11 community members' way to figure out moderators' departure from WebMD. This process helped us understand members' perceived role about moderators in online health communities.

Figuring Out That Moderator has Gone

In figuring out moderators' absence, members (1) noticed that moderators no longer participated, and (2) together they attempted to understand reasons for why moderators left. These discussions led to (3) discussions about solutions for not having moderators in their communities.

Noticing That Moderators were Gone: Receiving the News

WebMD community members' posts revealed that they had not received any public notification about STMs' departure from the communities. Instead, members slowly started realizing in mid-January 2013 that their moderators were not posting on their communities any more. This finding demonstrates that moderators' roles were salient enough that their absence was quickly detected by members. Members started noticing that

moderators were no longer posting, and they were seeking reasons for the absence of moderators' posts:

> *It seems that there are no more moderators on the community boards, and nobody is posting. Is anyone still here?* [Lupus]

Members who participated in multiple communities were able to publicize the idea that something suspicious was happening at the WebMD level, not just in one community:

> *I am active in four WebMD communities. No mods posting in any of them. Maybe they were given time off for the holidays?* [Diabetes]

To answer these questions, a number of members emailed the website directly and shared the replies from WebMD that WebMD had removed STMs from communities in December 2012 (see Table 1). Members who participated in multiple communities shared the news with other communities when suspicions from their members arose:

> *These communities no longer have moderators. The members are expected to inform others of the rules and policies. I am simply informing you of the facts.* [Pain management]

Members voiced their frustration about not being told about the change in the communities in advance:

> *NO GOODBYE? Just over. The top of the site still says moderator mediated. No one has said anything to us?* [Fibromyalgia]

As members realized moderators no longer participated in their communities, they attempted to figure out why moderators left.

Attempting to Understand Reasons for Why Moderators Left

Members reacted strongly against moderators being removed from the communities and attempted to understand the rationale behind the decision. Members from the community suspected that budget cuts were the primary reasons for removal of the moderators. Members shared concerns about the moderators' employment status, demonstrating the close relationships many had forged with the moderators. One member from a community shared evidence that WebMD had a budget cut, of which laying off moderators was a part. The following quote exemplifies the strong reaction to WebMD's decision:

> *Why??? Why would they deliberately sabotage a great site like Web.M.D.??? For what purpose??? Where are the BIG CHIEFS ??? What have they to say for themselves???* [Sex and Relationships]

> *this is part of the changes which take place every day....businesses finding ways to cut or reduce costs....I am sure the moderators were paid for being here as are the other professional people who were here as well.* [Fibromyalgia]

In one community, a member posted a press release on WebMD's budget cut, which they believed provided a reason for the layoffs:

> *For all those who have made inquiries about our missing leaders, apparently this is the answer:* Forbes news release of December, 11, 2012 at 9:29AM

> WebMD Cutting Staff 14%--250 Jobs. In Cost Cutting Program WebMD this morning disclosed plans to cut 250 jobs about 14% of the company's staff as part of a plan to reduce annual operating expenses by about $45 million. The company said the plan is designed to "streamline its operations, reduce costs, and better focus its resources on increasing user engagement, improving customer satisfaction, and driving innovation." *(Breast Cancer)*

As members accepted that moderators were no longer with them and attempted to understand why such a decision was made, the discussion moved on to how they, as WebMD community members, should proceed in maintaining their communities.

Discussing Solutions to Moderators' Departure

Members discussed ways to independently manage their communities—to "stand united" (Breast Cancer, Diabetes, Sex and Relationships, and Pain management) and make sure members regulate and facilitate conversations themselves. Some members predicted lack of moderators would produce problems, and advocated for strong mechanisms for self-regulation; others believed that communities could thrive without moderators, so long as they could provide a strong support system for each other.

A few members stated that they no longer wanted to participate in the communities, as they were upset by WebMD's decision to remove the moderators and bothered by increased spammers and trolls. One participant in the Pain Management community urged members to "ignore the trolls" and not "give [the trolls] any power to mess up" the community.

To sustain the communities, members needed to stay and support each other. For instance, members discussed "staying strong for each other and the new ones that come [to WebMD] for help" and urged that they "do [their] best to help [new comers] in any way [they] know how" (Breast Cancer). In response to some members stating they would leave, other members asked them to stay:

> *I'm sorry to see you go [David]. There are too few of us who have remained here over all the changes during the years (I've been around about 15 or more). Please reconsider. You help so many people. Anyway, hope you stay well and will contribute as you can to other groups.* [Diabetes]

> *I understand your needing a break, but you will be missed. Your insight and experience are a valuable asset to the other members of this community.* [Pain management]

Members also discussed that peer knowledge would not be enough and moderators were essential in communities requiring immediate answers to medical problems. Members discussed concrete actions they could take to fill the void left by the departure of the moderators, such as facilitating conversations by posting questions to elicit responses from "the wonderful people here" (Breast Cancer). After focused conversations about

XSL•FO
RenderX

solutions to moderators being gone, members continued to encourage community members to monitor spam and advertisements. A member also shared one tactic of controlling malicious posts: "to not reply." (Diabetes)

we can report it and see what happens and then FLAG IT AS SPAM by the FMILY here so no one bothers to read it! (Fibromyalgia)

Members began to practice self-moderation on their communities. They would tell other members not to share medical advice and provided alternative ways to get help at low cost (Pain Management), thus mirroring the prior activities of paid moderators. Members also put self-moderation into practice by reporting people; this task was not favorably received and caused one member to note that she or he "really miss(es) the moderators" (Pain Management). Furthermore, members shared websites that members should be careful of visiting. Members encouraged those who stated they would leave the community to stay, in an effort to sustain the community.

Members' Experiences on Having Moderators

While making sense of moderators being gone, patients shared their past experiences of having moderators as part of their communities. Throughout this process, even though HPMs still stayed on WebMD, members discussed their perceptions about health professionals as moderators. Contrasting opinions emerged about HPMs. Patients had conflicting views of whether moderators gave pragmatic help. Patients expressed that moderators were critical in ensuring the community did not include hostile messages, spams, and misleading information. Patients reported they felt personally attached with moderators. On the other hand, some expressed moderators were not useful and that peer support was more critical.

Moderators Provided Pragmatic Help

Of the members who commented on the utility of moderators, most mentioned moderators to be helpful. Although STMs were the ones who left the communities, members discussed experiences around HPMs as well, allowing us to understand members' experiences around not just STMs but also HPMs:

> *I notice very few of the boards have doctors posting to them anymore. They were one of the reasons why I joined WebMD.* [Sex and Relationships]

Many members did not draw a clear distinction between HPMs and STMs until some members clarified that there is a difference between STMs and HPMs. This perception was best illustrated by the sentiment that medical expertise was a unique resource that moderators could provide the communities. For instance, members in Sex and Relationships discussed how they missed doctors on the forum and expressed "personal experience doesn't always match having a medical opinion" and that although members "have tons of personal experience, [they] sure as heck don't have the required learning . " Similarly, a member stated, "While the support from people who are "suffering" from the same health problems is the most important thing on these blogs, expert advice [was] invaluable to [them] all. Members from the Diabetes community also felt that "the presence of mods experts in their fields to answer questions" was helpful. Members from Breast Cancer also noted the importance of moderators who are

"(medical) experts" on the topic: and that they "had always awaited a promised 'expert' for this board" *(Breast Cancer).*

Although the STMs did not have the medical expertise, members appreciated the unique help provided by these STMs. These moderators helped keep the communities interested by posting articles, news, and information about WebMD. Members felt confident that moderators would find solutions to problems that arose on the forum. Members believed moderators ensured communities kept their "focus and no bickering." Members did not "feel as safe posting here" (Diabetes, Back pain) as they did when the presence of moderators was obvious. Being monitored by moderators made some members feel safe to participate in the community. Members declared their intent to leave, saying that they "joined because it was not like all the other social networking sites" (Diabetes). They "did not feel comfortable here now" because "mods used to tend to these issues [monitoring spams] in a timely fashion and keep an eye out" (Breast Cancer). Apart from a policing role, moderators helped maintain communities by picking up conversation when it slowed down. Members felt people were not posting anymore because the moderators were gone.

Members from all communities saw an effect of not having moderators—spams lasting too long on the boards before being deleted, and that "New members are posting with rather radical ideas and trying to convince others that they are right" (Diabetes). The community being a "public, world-wide site with the communities open to anyone" (Diabetes), not having moderation made many members no longer feel safe posting.

Patients Felt Personally Attached With the Moderators

Members not only considered moderators to be helpful, but developed personal ties with them. They demonstrated these ties by voicing concerns about them losing their jobs. Several posters commented that moderators were like family members; for example, one wrote, "It is like losing a family member [...] To you all we will miss you" (Breast Cancer) and another similarly posted "we really loved having you on this board. We felt we were part of a large family" (Breast Cancer). In the Fibromyalgia community, members coordinated sending a card for one of the moderators. In the Depression community, a member started a thread to thank one of the moderators and wish her well. Some posters tried to petition WebMD to bring back the moderators, on the Sex and Relationships community, a thread called "BRING BACK OUR MODERATORS!!!" was posted with 17 replies in which the members vented on why moderators had to leave: "Complete madness to fire them."

Several members even threatened to leave the communities after learning about STMs leaving WebMD. One member wrote how she had been on WebMD for over 10 years and said:

It was and has been my life-‐line and support for info from very good ladies. But Excuse ME-‐-‐-‐-‐-‐You are NOT the only Medical Information site on the Internet. I do believe you have just Stepped in IT [Breast Cancer].

Some members wondered where they could find their former community members on other websites. One member suggested to start a discussion titled "Refugees from that other site" for the members to meet up and continue to support one another

(Breast Cancer). Another suggested Facebook site where members can find the other members. Several Breast Cancer community members discovered that former members had posted about a discussion thread called "Friends from WebMDs Support Board" at another breast cancer forum.

Moderators Were Not Useful: Peer Support Is More Important

While most patients who commented about moderators believed they were helpful and important members of their communities, some patients discussed how they did not find moderators' help to be useful. These members considered peer-support as the main benefit to participation in OHCs, and felt that moderators could deter the dynamic exchange of ideas among patients and peers.

For example, a member stated that HPMs did not weigh in with medical expertise when needed. Certain moderators' questions were thought to be unhelpful for the community, such as "What turns you on?" in the case of Sex and Relationships community, which seemed to minimize the seriousness of the medical focus of the community. Other examples of ineffective moderation included inappropriate language spreading out even in the presence of moderators.

Members who saw moderators' roles as providing medical expertise did not find HPMs to be helpful, as they cannot "treat, diagnose, or exam the people" (Sex and Relationships). In addition, members discussed how the "one size fit all" advice coming from the HPMs was not helpful; rather, peers' "one-on-one, spiritually rewarding" support would be more beneficial for the community (Diabetes).

A few members considered others' shared personal experience as more critical than moderators' help. In an exchange between 2 members on the Sex and Relationships community, they shared their belief that their community did not seem to need HPMs' involvement:

> Member 1: It's got to a point where we do not need a doctor, we're too good Member 2: I think you are right? We really are better than any doctor, any day, Right???

A similar sentiment was found in the Fibromyalgia community:

It's the people that make the community and the support we gain from each other that keep us coming back.

Some members believed that the absence of moderators might facilitate increased sharing among peers. Members believed that self-efficacy was essential and that moderators could not help with disease management. This thought was especially apparent on the Diabetes community. Patients felt that the communities did a good job of self-regulating and making sure that no one was claiming to share medical advice; rather, they were sharing anecdotal experiences. As one poster commented:

I think there are a lot of very smart people who post on the various boards and I have never seen anyone claim to be an expert. We post what we know through personal experiences or something that has worked for us [Sex and Relationships].

Although members believed that moderators may not improve the chemistry of communities, they did acknowledge that the website was losing long-term membership with the absence of moderators. Several members of the Diabetes community mentioned that it is up to the patients themselves to manage their disease, and that moderators cannot help. Similar sentiments were echoed on the Sex and Relationships community, noting that peer's expertise was important.

Discussion

Principal Findings

OHCs were primarily developed to connect patients to others with similar disease processes in an attempt to provide emotional and informational support [25-28]. While some communities, such as WebMD, have experimented with incorporating paid moderators to guide discussions, it is generally believed that the primary utility of OHCs lies in peer-patient interactions [29,30] . Many believe that patients gravitate to these communities because they are free from professional governance [15,31]. By introducing and then removing STMs from the communities, WebMD has provided a critical test case to demonstrate the impact of moderators on OHCs.

This analysis of the departure of STMs from WebMD communities offers several insights into the role that moderators play on OHCs. Longstanding community members quickly detected moderators' absence. The departure spurred conversation and debate on the communities about the utility of moderators. This rapid detection of the departure of moderators appeared to be indicative of the important role moderators played on many of the communities, such that the void was noted quickly. This void was validated by our quantitative analysis demonstrating that many of the moderators were active on communities. The close personal ties that many community members felt with moderators was demonstrated by the fact that many members worried about the employment status of the moderators, and some further petitioned WebMD to bring the moderators back.

Although many participants appeared to be especially appreciative of the pragmatic help that moderators had provided to their communities, such as monitoring for trolls and stimulating conversation when it had slowed, many of these roles were quickly assumed by community members. However, members did not enjoy these tasks, which often made them comment on how they missed the moderators. Community members also worked to create a sense of camaraderie and encourage their peers to continue to post, thus filling in the void left by the moderators. These tasks appeared to be less bothersome to posters. Such altruistic behavior is often observed in closely knit online communities, which is one of the strongest sources of maintaining online participation [32,33].

Several posters commented they did not feel as safe participating after moderators had left their communities, even as members tried to step in and report inappropriate posting activity. This finding suggests that even on vibrant communities, the role of the professional moderators may be unique and irreplaceable. This role of professional moderators, however, should be

distinguished from general moderators or informal leaders in online community literature [34,35]. The sense of credibility and trust toward information source, including the author of the information, has a unique role in OHCs compared with general communities. The altruistic volunteerism common in most closely knit online communities [35-37] will not fulfill the absence of professional moderators in OHCs.

Although the general sentiment of community members toward moderators was positive, this sentiment was not universal. The departure of the moderators spurred debate on some communities, which felt that the utility of the moderators might vary according to the disease process the community dealt with. Members were insightful in their discussions about the tension between having moderators versus not having moderators. Having moderators can potentially impede conversation but limit the exchange of erroneous information. The lack of moderators or self-moderated communities can have strong community dynamics but requires strong commitment by the members to self-moderate bothersome, unmoderated exchanges by visitors and new comers with either ill intent or lack of knowledge around the community norms and common ground. Furthermore, not having moderators and committed members would lead to reduced access to health professional knowledge or validated information.

A few members threatened to, and eventually left their communities, after removal of the moderators. Some spoke about attempting to find people who had left for other websites. The remaining members believed that their communities were less active. However, this analysis has revealed that the overall posting activity of the WebMD health communities was declining throughout the study period, so it is difficult to infer that the decline in posting activity was a result of the moderators' departure. Rather, the decline of member participation was followed by STMs' and HPMs' increased posting activity. Moderators' increased activities showed WebMD's effort to increase participation, which was evidenced by a study that reported qualitative analysis on the post content of WebMD's moderators [14]. The researchers of this study found that moderators' thread initiating posts were facilitating conversations through asking questions for members to share their experiences. Moderators' replies were used to respond to threads to ensure members feel their posts were being read [38,39]. We saw that STMs generated more initiating of conversations compared with HPMs, and HPMs mainly replied to members' posts, considering HPMs' roles in WebMD was to provide clinical expertise [14].

The participation decline observed in WebMD is a canonical retention problem in social media. Thus, user migration has been widely studied to understand how to retain and attract users in social media platforms [40-42]. Main strategies in retaining users include providing useful content and improving socialization [42]. WebMD similarly made an attempt to spark conversations by adding moderators and their conversation facilitating messages together with HPMs' contribution of clinical expertise. Even with increased moderators' activities, however, members' participation did not increase, which could have been a possible reason for WebMD's decision to remove the moderators.

Another reason for decline in member participation might be due to the members not seeing health benefits as Su et al have discussed in their literature review of eHealth Systems [43]. However, the link between such individual evaluation of the community and the aggregate decline in participation is still unclear.

We also found the lack of distinction community members drew between HPMs and STMs. Although WebMD only dismissed STMs, members did not initially pick up on the fact that HPMs were still active members of their communities. Similarly, even on communities where there were HPMs participating (eg, Breast Cancer), some members were frustrated because they did not realize that there were HPMs (and such moderators were desired). Online health information consumers are keen on finding indicators for the source of information [44-46]. Thus, such WebMD members' behavior to confuse moderators' background expertise was a unique challenge worth noting. Our findings contest any perception that patients gravitate to OHCs because they are free from health professional involvement and that OHCs spread misinformation due to patients' critical consumption of shared information online [47,48]. Many participants were concerned about the validity of information that would be shared on their communities without the expertise of paid moderators to help curate the information available to them. At the same time, assessing posters' credibility and source of information was a challenging task, considering members' observed confusion between STMs and HPMs.

Our study has implications for future OHC development. By exploring the impact of the removal of STMs from several well-established communities and the discussions ensued among participants, an insight was gained into what aspects of moderators' presence and posting activity are most appreciated by members and helpful:

OHCs need explicit, systematic policing activities put into place to help members feel safe and agree on continued participation. This was a universal sentiment expressed on the communities; members appreciated this role that moderators had previously played and did not enjoy assuming this responsibility. Given the disproportionate number of verbal members who volunteer to take on the moderating role, strong dependence on these members will threaten the sustainability of the communities.

The importance of HPMs may vary according to health topics of the communities. It is difficult to characterize which communities benefited most from moderators from our study, given the more or less similar participation pattern across all communities. However, members' conversations revealed the needs are different among different communities, where communities such as diabetes or heart disease might need immediate help from health professionals. Further work should clarify and identify what characteristics of illness experiences drive members to seek health professionals' knowledge online rather than peers' social support.

The background of moderators (eg, clinical vs nonclinical) should be explicit. Some WebMD community members demonstrated a poor understanding of the distinction between HPMs and STMs, where some thought HPMs have left the communities. Even in communities with an HPM, posters noted

that they had been promised such an expert on the community and yet had still not encountered him. Explicit signatures should be available for every post and profile, clarifying if the moderator has undergone any professional health training, or is an STM. If an STM, whether the STM has any professional knowledge on health should be presented explicitly.

Major community changes, such as removing moderators, should be communicated to members with advanced notice. Much of the discussion about moderators centered around communities attempting to understand what had happened to them after they noticed their absence. Members' frustration about the decision to remove moderators, and the lack of notification, was apparent in their posts. Members felt betrayed, and as many of them had developed close bonds with moderators, were upset about not being able to say their goodbyes to these members of their communities.

Limitations

Our data source was limited to observed, publicly available posts. Thus, our analysis is limited to those who are verbal and have expressed opinions. Other rich metadata sources can be used to further analyze deeper insights about OHC activities, such as user logs, page views, and click streams [43]. The results reported in this study only account for a small percentage of the activities that can be observed in OHCs. However, access to such rich data on webpage activity of online health support group will require multiple layers of privacy protection methods, partnering with the support group provider, and close

communication with the users who are willing to share such private data with the public for research purposes. Su et al [43] has also found that none of the eHealth systems studies on OHCs they reviewed examined private data, such as page views and other use logs. Studying such nonverbal behavior online will have continued challenges in obtaining the data and analyzing them.

Conclusions

The analysis of STMs' departure from WebMD communities provided a critical insight into the utility of moderators in OHCs. By analyzing the reactions of community members in 11 different illness communities to moderators' departure, it was found that members perceive that moderators play an important role in OHCs, stimulating discussion and making them feel safe. However, this quantitative analysis of the posting activities shows that moderators' efforts were not enough to increase the already decreasing member participation. Our work disputes a widely held belief about patients' blindly accepting other patients' shared information in OHCs and preferring nongovernance. Although moderators' efforts did not play a large-scale influence on member participation, their presence was effective and favored by the core members. The discussion on member attrition and attraction in OHCs contributes to the call for analyzing attrition as a core research in eHealth systems [49]. This study contributes to the medical informatics community in understanding the utility of adding moderators and optimizing their roles in sustaining member participation in OHCs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table of posting activity of members, STMs, and HPMs in the WebMD communities.

[PDF File (Adobe PDF File), 34KB - jmir_v18i9e247_app1.pdf]

Multimedia Appendix 2

Affinity diagram example of the codes generated.

[PDF File (Adobe PDF File), 395KB - jmir_v18i9e247_app2.pdf]

References

1. Borkman T. Self-help groups at the turning pointmerging egalitarian alliances with the formal health care system? American Journal of Community Psychology 1990 Apr 18;18(2):321-332 [FREE Full text] [doi: 10.1007/BF00931307]
2. Lavoie F, Borkman T, Gidron B. Self-help and mutual aid groups: international and multicultural perspectives. New York: Haworth Press; 1994.
3. Norcross JC, Campbell LF, Grohol JM, Santrock JW, Selagea F, Sommer R. Self-help that worksresources to improve emotional health and strengthen relationships. Oxford, UK: Oxford University Press; 2013.
4. Powell TJ. Professional help and informal help: Competing or complementary systems. In: Working with self-help. Silver Spring, MD: National Association of Social Workers; 1990:31-49.

5.   Huh J, Patel R, Pratt W. Tackling Dilemmas in Supporting "The Whole Person" in Online Patient Communities. Proc SIGCHI Conf Hum Factor Comput Syst 2012;2012:923-926 [FREE Full text] [doi: 10.1145/2207676.2208535] [Medline: 24634893]

6.   Kraut RE, Resnick P, Kiesler S, Burke M, Chen Y, Kittur N, et al. Building Successful Online Communities: Evidence-Based Social Design. Cambridge, MA: The MIT Press; Mar 2012.

7.   O'Grady L, Bender J, Urowitz S, Wiljer D, Jadad A. Promoting and participating in online health forums: A guide to facilitation and evaluation for health professionals. Journal of Communication in Healthcare 2013 Jul 18;3(3-4):246-257. [doi: 10.1179/175380710X12870623776478]

8.   Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]

9.   Tiwana A. Do bridging ties complement strong ties? An empirical examination of alliance ambidexterity. Strat Mgmt J 2008 Mar;29(3):251-272 [FREE Full text] [doi: 10.1002/smj.666]

10. Owens C, Sharkey S, Smithson J, Hewis E, Emmens T, Ford T, et al. Building an online community to promote communication and collaborative learning between health professionals and young people who self-harm: an exploratory study. Health Expect 2015 Feb;18(1):81-94. [doi: 10.1111/hex.12011] [Medline: 23075133]

11. Matzat U, Rooks G. Styles of moderation in online health and support communities: An experimental comparison of their acceptance and effectiveness. Computers in Human Behavior 2014 Jul;36:65-75 [FREE Full text] [doi: 10.1016/j.chb.2014.03.043]

12. Resnick PJ, Janney AW, Buis LR, Richardson CR. Adding an online community to an internet-mediated walking program. Part 2: strategies for encouraging community participation. J Med Internet Res 2010;12(4):e72 [FREE Full text] [doi: 10.2196/jmir.1339] [Medline: 21169161]

13. Edwards AR. The Moderator As an Emerging Democratic Intermediary: The Role of the Moderator in Internet Discussions About Public Issues. Information Policy. 2002 Jan. (1) p. 3-20 URL: http://dl.acm.org/citation.cfm?id=1412444.1412446 [WebCite Cache ID 6kKBDymoJ]

14. Huh J, McDonald DW, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]

15. Maloney-Krichmar D, Preece J. A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM Trans Comput-Hum Interact 2005 Jun 01;12(2):201-232. [doi: 10.1145/1067860.1067864]

16. Huh J. Clinical Questions in Online Health Communities: The Case of "See your doctor" Threads. CSCW 2015;2015:1488-1499 [FREE Full text] [doi: 10.1145/2675133.2675259] [Medline: 26146665]

17. Cole J, Watkins C, Kleine D. Health Advice from Internet Discussion Forums: How Bad Is Dangerous? J Med Internet Res 2016;18(1):e4 [FREE Full text] [doi: 10.2196/jmir.5051] [Medline: 26740148]

18. 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]

19. Ekberg J, Timpka T, Angbratt M, Frank L, Noren A, Hedin L, et al. Design of an online health-promoting community: negotiating user community needs with public health goals and service capabilities. BMC Health Services Research 2013 Jul 4;13(1):258 [FREE Full text] [doi: 10.1186/1472-6963-13-258]

20. WebMD. URL: http://www.webmd.com [WebCite Cache ID 6ipdEqkyn]

21. WebMD. Additional Information URL: http://www.webmd.com/about-webmd-policies/additional-info?ss=ftr [accessed 2016-07-06] [WebCite Cache ID 6ipd5YOCt]

22. Liu F, Yu C, Meng W, Chowdhury A. Effective Keyword Search in Relational Databases. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. New York, NY: ACM; 2006:563-574.

23. Corbin JM, Strauss AL. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications, Inc; Nov 19, 2007.

24. Beyer H, Holtzblatt K. Contextual design: defining customer-centered systems. San Francisco, CA: Morgan Kaufmann Publishers; 1998.

25. Ahmad S, Fergus K, McCarthy M. Psychosocial issues experienced by young women with breast cancer: the minority group with the majority of need. Curr Opin Support Palliat Care 2015 Sep;9(3):271-278. [doi: 10.1097/SPC.0000000000000162] [Medline: 26147915]

26. Tixier M, Gaglio G, Lewkowicz M. Translating Social Support Practices into Online Services for Family Caregivers. Proceedings of the ACM 2009 International Conference on Supporting Group Work 2009:71-80 [FREE Full text] [doi: 10.1145/1531674.1531685]

27. Vilhauer RP. Perceived benefits of online support groups for women with metastatic breast cancer. Women Health 2009;49(5):381-404. [doi: 10.1080/03630240903238719] [Medline: 19851944]

28. Yoo W, Namkoong K, Choi M, Shah DV, Tsang S, Hong Y, et al. Giving and Receiving Emotional Support Online: Communication Competence as a Moderator of Psychosocial Benefits for Women with Breast Cancer. Comput Human Behav 2014 Jan;30:13-22 [FREE Full text] [doi: 10.1016/j.chb.2013.07.024] [Medline: 24058261]

29. Hoey LM, Ieropoli SC, White VM, Jefford M. Systematic review of peer-support programs for people with cancer. Patient Educ Couns 2008 Mar;70(3):315-337. [doi: 10.1016/j.pec.2007.11.016] [Medline: 18191527]

30. Setoyama Y, Nakayama K, Yamazaki Y. Peer support from online community on the internet among patients with breast cancer in Japan. Stud Health Technol Inform 2009;146:886. [Medline: 19593035]

31. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas In Online Health Communities. J Biomed Inform 2016 Aug 25:- Epub ahead of print. [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]

32. Yuqing R, Robert K, Sara K, Paul R. Encouraging commitment in online communities. In: Kraut RE, Resnick P, Kiesler S, Burke M, Chen Y, Kittur N, et al, editors. Building successful online communities: Evidence-based social design. Cambridge, MA: The MIT Press; Mar 2012:77-124.

33. Kraut RE, Resnick P, Kiesler S, Burke M, Chen Y, Kittur N, et al. Building Successful Online Communities: Evidence-Based Social Design. Cambridge, MA: The MIT Press; 2012.

34. Liao Q, Pan Y, Zhou MX, Ma F. Chinese Online Communities: Balancing Managementcontrol and Individual Autonomy. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA p. 2193-2202 URL: http://doi.org/10.1145/1753326.1753658 [doi: 10.1145/1753326.1753658]

35. O'mahony S, Ferraro F. The emergence of governance in an open source community. Academy of Management Journal 2007 Oct 01;50(5):1079-1106 [FREE Full text] [doi: 10.5465/AMJ.2007.27169153]

36. Butler B, Sproull L, Kiesler S, Kraut R. Community Effort in Online Groups: Who Does the Work and Why? In: Weisband SP, editor. Leadership at a distance: research in technologically-supported work. New York: Psychology Press; Jul 22, 2007.

37. Farzan R, Kraut R, Pal A, Konstan J. Socializing Volunteers in an Online Community: A Field Experiment. 2012 Presented at: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work; February 11-15, 2012; Seattle, Washington p. 325-334. [doi: 10.1145/2145204.2145256]

38. Kim AJ. Community building on the Web. Berkeley, CA: Peachpit Press; 2000.

39. Ludford PJ, Cosley D, Frankowski D, Terveen L. Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. 2004 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2004; Vienna, Austria p. 631-638.

40. Komito L. Social media and migration: Virtual community 2.0. J AM Soc Inf Sci 2011 Mar 22;62(6):1075-1086. [doi: 10.1002/asi.21517]

41. Kumar S, Zafarani R, Liu H. Understanding User Migration Patterns in Social Media. : AAAI; 2011 Presented at: AAAI; August 7-11, 2011; San Francisco, CA.

42. Xu Y, Yang Y, Cheng Z, Lim J. Retaining and attracting users in social networking services: An empirical investigation of cyber migration. The Journal of Strategic Information Systems 2014 Sep;23(3):239-253 [FREE Full text] [doi: 10.1016/j.jsis.2014.03.002]

43. Su WC, Chih MY. Is More eHealth System Use Better for Cancer Patients and Family Caregivers? A Literature Review. 2016 Presented at: American Medical Informatics Association Poster; November 12-16, 2016; Chicago, IL URL: http://doi.org/10.13140/RG.2.1.4077.7200 [doi: 10.13140/RG.2.1.4077.7200]

44. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]

45. Rains SA, Karmikel CD. Health information-seeking and perceptions of website credibility: Examining Web-use orientation, message characteristics, and structural features of websites. Computers in Human Behavior 2009 Mar;25(2):544-553. [doi: 10.1016/j.chb.2008.11.005]

46. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. Health Commun 2008 Jul;23(4):358-368. [doi: 10.1080/10410230802229738] [Medline: 18702000]

47. Merckelbach H, Jelicic M, Pieters M. Misinformation increases symptom reporting: a test–retest study. JRSM short reports 2011;2(10):75-78. [doi: 10.1258/shorts.2011.011062]

48. Travers RL. Information consumerism on the World Wide Web: implications for dermatologists and patients. Semin Cutan Med Surg 2002 Sep;21(3):223-231. [doi: 10.1053/sder.2002.34949] [Medline: 12322997]

49. 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]

Abbreviations

STM: staff moderator
HPM: health professional moderator
OHCs: online health communities
RCT: randomized controlled trial

XSL•FO

RenderX

Original Paper

Does Digital Video Advertising Increase Population-Level Reach of Multimedia Campaigns? Evidence From the 2013 Tips From Former Smokers Campaign

Kevin C Davis[1], MA; Paul R Shafer[1,2], MA; Robert Rodes[3], MS; Annice Kim[1], MPH, PhD; Heather Hansen[1], MPP; Deesha Patel[3], MPH; Caryn Coln[4], MCIS, PMP; Diane Beistle[3], BA

[1]Center for Health Policy Science and Tobacco Research, RTI International, Research Triangle Park, NC, United States

[2]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, GA, United States

[4]Northrop Grumman Information Systems, Atlanta, GA, United States

Corresponding Author:
Kevin C Davis, MA
Center for Health Policy Science and Tobacco Research
RTI International
3040 East Cornwallis Road
Research Triangle Park, NC, 27709
United States
Phone: 1 919 541 5801
Fax: 1 919 541 6683
Email: kcdavis@rti.org

Abstract

Background: Federal and state public health agencies in the United States are increasingly using digital advertising and social media to promote messages from broader multimedia campaigns. However, little evidence exists on population-level campaign awareness and relative cost efficiencies of digital advertising in the context of a comprehensive public health education campaign.

Objective: Our objective was to compare the impact of increased doses of digital video and television advertising from the 2013 Tips From Former Smokers (Tips) campaign on overall campaign awareness at the population level. We also compared the relative cost efficiencies across these media platforms.

Methods: We used data from a large national online survey of approximately 15,000 US smokers conducted in 2013 immediately after the conclusion of the 2013 Tips campaign. These data were used to compare the effects of variation in media dose of digital video and television advertising on population-level awareness of the Tips campaign. We implemented higher doses of digital video among selected media markets and randomly selected other markets to receive similar higher doses of television ads. Multivariate logistic regressions estimated the odds of overall campaign awareness via digital or television format as a function of higher-dose media in each market area. All statistical tests used the .05 threshold for statistical significance and the .10 level for marginal nonsignificance. We used adjusted advertising costs for the additional doses of digital and television advertising to compare the cost efficiencies of digital and television advertising on the basis of costs per percentage point of population awareness generated.

Results: Higher-dose digital video advertising was associated with 94% increased odds of awareness of any ad online relative to standard-dose markets (*P*<.001). Higher-dose digital advertising was associated with a marginally nonsignificant increase (46%) in overall campaign awareness regardless of media format (*P*=.09). Higher-dose television advertising was associated with 81% increased odds of overall ad awareness regardless of media format (*P*<.001). Increased doses of television advertising were also associated with significantly higher odds of awareness of any ad on television (*P*<.001) and online (*P*=.04). The adjusted cost of each additional percentage point of population-level reach generated by higher doses of advertising was approximately US $440,000 for digital advertising and US $1 million for television advertising.

Conclusions: Television advertising generated relatively higher levels of overall campaign awareness. However, digital video was relatively more cost efficient for generating awareness. These results suggest that digital video may be used as a cost-efficient

XSL•FO
RenderX

complement to traditional advertising modes (eg, television), but digital video should not replace television given the relatively smaller audience size of digital video viewers.

KEYWORDS

social marketing; smoking; health campaigns; digital advertising; television advertising

Introduction

In 2012, the US Centers for Disease Control and Prevention (CDC) launched the first federally funded national tobacco education campaign, Tips From Former Smokers (Tips). The campaign aired nationwide on cable television networks in addition to radio, online, print, and out-of-home (eg, billboard) outlets. Tips consisted of evidence-based, graphic, and emotional messages that portrayed the devastating consequences of several smoking- and secondhand smoke-related diseases and conditions, including Buerger's disease, tracheotomy, and heart attack. All video, radio, online, print, and out-of-home ads from the Tips campaign can be viewed at the Tips website [1]. The campaign was associated with approximately 1.6 million additional quit attempts among US smokers and an estimated 100,000 sustained quits of at least 6 months [2]. In addition, the campaign was found to be highly cost effective based on several accepted thresholds for costs per life year saved [3].

As consumers increasingly use digital devices to view new information and advertising content, public education campaigns have followed suit in using digital media as an advertising platform. In early 2013, CDC launched a second wave of the Tips campaign, which involved similar creative content to that in the 2012 campaign. The 2013 campaign was also supported by a robust digital advertising effort, including online video ads (with the same video content as in the television ads), display ads, mobile ads, and paid search to drive awareness of and traffic to the Tips campaign website. Recently published work has shown that advertising doses on television and digital platforms had significant effects on traffic to the Tips campaign website during the 2013 campaign [4]. Other recent data suggest that exposure to digital advertising during the 2012 Tips campaign was associated with increases in confirmed visits to the Tips campaign website and other cessation-oriented websites for several weeks after exposure [5]. However, gaps remain in understanding digital advertising's impact on real-world campaign exposure at the population level, practical use as a driver of an overall public health campaign message, and relative cost efficiency. This study extended this work and attempted to address these gaps by directly comparing the effects (and related cost efficiencies) of increased doses of digital video and television advertising on campaign awareness at the population level.

Several studies have examined the use of digital advertising for recruitment to interventions via online ads, social media, and text messaging. These studies found digital advertising to be useful and cost efficient for targeting smokers generally and subpopulations of smokers, such as Latinos, in particular. A 2008 study examined reach and cost effectiveness of a digital campaign to recruit smokers in New Jersey to a cessation treatment program and found that digital advertising yielded costs per enrollee that were competitive with traditional media [6]. In addition, a 2012 study found that digital advertising was cost effective for reaching Spanish-speaking Latino smokers and recruiting them to participate in an online cessation intervention [7].

To date, there is little evidence on the extent to which digital advertising can drive awareness of public education campaigns at the population level. In addition, to our knowledge, no major studies have explored the cost efficiency of digital advertising, relative to traditional broadcast platforms, as part of a comprehensive public health education campaign. However, several studies have demonstrated the potential for digital media to reach tobacco users, as there are already high rates of daily Internet use (82%) and online information seeking about health (80%) among the general adult population in the United States [8]. Google AdWords and search engines were shown to be useful tools for reaching smokeless tobacco users as early as 2005 [9]. In addition, search engine referrals were found to contribute over 70% of traffic to a CDC website on chronic fatigue syndrome over an 18-month period in 2006 and 2007 [10]. As noted above, digital advertising has been shown more recently to be effective at recruiting participants for tobacco cessation interventions [6,11], including among more difficult-to-reach minority subpopulations [7]. Campaign planners must carefully consider the mix of not only the message content, but also the platforms on which those messages will be delivered. In light of constrained budgets for advertising, the relative cost efficiency of advertising platforms is a key element of campaign decision making.

In this study, we compared the impacts of higher doses of digital video and television advertising from the 2013 Tips campaign on overall campaign reach at the population level. This is the first study, to our knowledge, that used a dual-mode design that included higher dosing of digital video and higher dosing of television advertising to identify the independent contributions of each media format to overall campaign awareness. In addition, we used data on adjusted advertising costs for the additional digital video and television advertising doses to compare the cost efficiencies of each format on the basis of costs per percentage point of population awareness generated. These comparisons provide new insights into the role of digital video advertising in driving overall audience exposure in the context of a broad multimedia campaign.

Methods

Television Advertising

The 2013 Tips campaign included purchasing advertising on cable television networks, aimed to deliver on average

approximately 800 ad gross rating points (GRPs) nationwide. Television ads aired during daytime and primetime hours on television shows frequently watched by the campaign's target audience of adult smokers. GRPs measure the relative "dose" of advertising delivered to a target audience in a given media market and time period. They are defined as the product of the proportion of an audience that is exposed (ie, audience reach) and the frequency of that exposure (ie, number of times an ad was seen). For example, if a television ad reaches 50% of an audience twice in 1 week, the GRP for this ad during that week is 100 (50 × 2) [12]. In addition to this base national ad buy, we randomly selected 67 designated market areas (DMAs) to receive an additional planned 1600 television GRPs during the 2013 Tips campaign to facilitate a range of analyses on the dose-response impact of additional television advertising.

Figure 1 summarizes the assignment of DMAs to each condition of higher media dosing. We randomly assigned higher-dose television advertising across 190 of the available 210 US DMAs. We excluded the 20 largest US DMAs from this randomization due to the high costs of additional local advertising in these markets. We stratified the remaining 190 DMAs by several characteristics that are associated with smoking, including race/ethnicity, income, and education, and then randomly assigned the DMAs within these strata. The probability of assignment to higher-dose television advertising was set at 35% based on the available budget for local television ad buys, resulting in 67 DMAs assigned to higher-dose television and 123 DMAs assigned to standard-dose television. The television campaign and methods of market-level randomization of the television media dose are discussed in more detail in a recent study of the 2013 Tips television campaign's impact on cessation-related outcomes among smokers [13].

Figure 1. Flow diagram of assignment of designated market areas (DMAs) to higher-dose and standard-dose television and digital video advertising for Tips From Former Smokers 2013 campaign. GRPs: gross rating points.
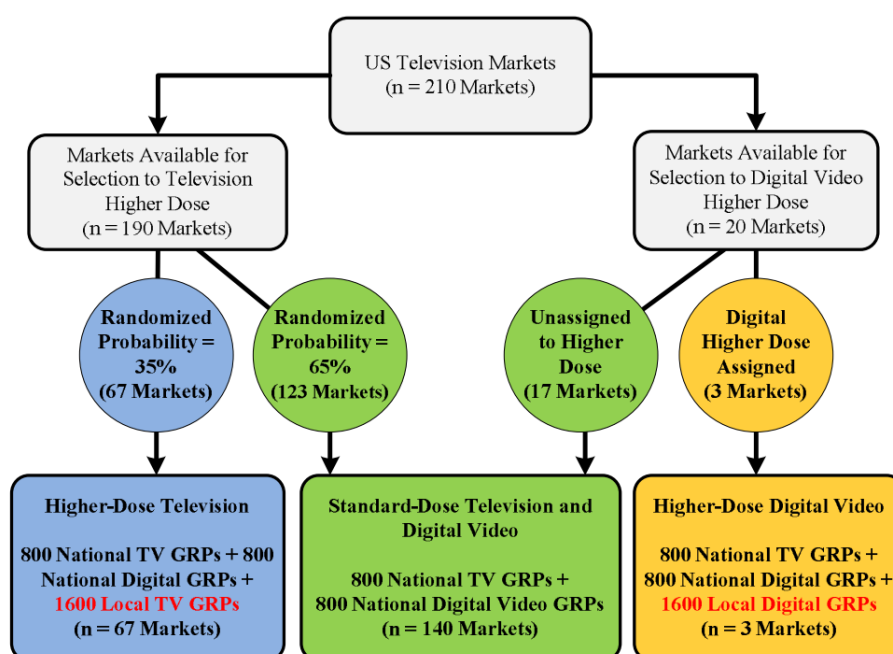
Digital Video Advertising

To enable comparisons of costs and ad awareness between television and digital video, we implemented a standard-dose and higher-dose digital video ad buy to mimic the standard-dose and higher-dose television advertising described above. We applied higher-dose digital video advertising within a subset of the remaining 20 largest DMAs that were excluded from the random assignment of higher-dose television described above. The digital video campaign included digital video ads (featuring the same 30-second ads used for the national television campaign) placed on a variety of online advertising networks such as Adotube (Exponential Interactive, Inc., Emeryville, CA, USA) and Tremor (Tremor Video, Inc., New York, NY, USA), as well as several online media networks, including YouTube (YouTube, LLC, San Bruno, CA, USA), Turner networks (Turner Broadcasting System, Inc., Atlanta, GA, USA), and

Discovery (Discovery Communications, Inc., Silver Spring, MD, USA). These networks are frequently used by the campaign's target audience of adult smokers. All digital video ads were clickable and directed viewers to the campaign's website.

We chose 3 DMAs—Tampa, Florida; Cleveland, Ohio; and Sacramento, California—to receive a higher dose of digital video advertising of roughly 1600 digital video-equivalent GRPs, mimicking the scale of the higher-dose television ad buy. We geographically targeted digital video ads based on the Internet protocol (IP) address of the users of websites where the ads were placed. In cases where the IP address of the user could not be attributed to a specific media market, that user was not included in the calculation of ad impressions and thus was not counted as a contribution to the digital GRPs. We chose these markets based on the affordability of local digital media

buys and to provide some geographic diversity in the group of digital higher-dose markets. Furthermore, the markets selected were very similar in terms of overall target population size, allowing for a nearly equal delivery of digital video GRPs across these markets. The remaining 17 large markets that we excluded

from higher-dose television and digital advertising made up the remainder of the standard-dose condition (140 DMAs in total). Figure 2 illustrates the allocation of the 3 media assignment conditions across DMAs in the United States.

Figure 2. Assignment of media exposure conditions to designated market areas in the United States for the Tips From Former Smokers 2013 campaign.
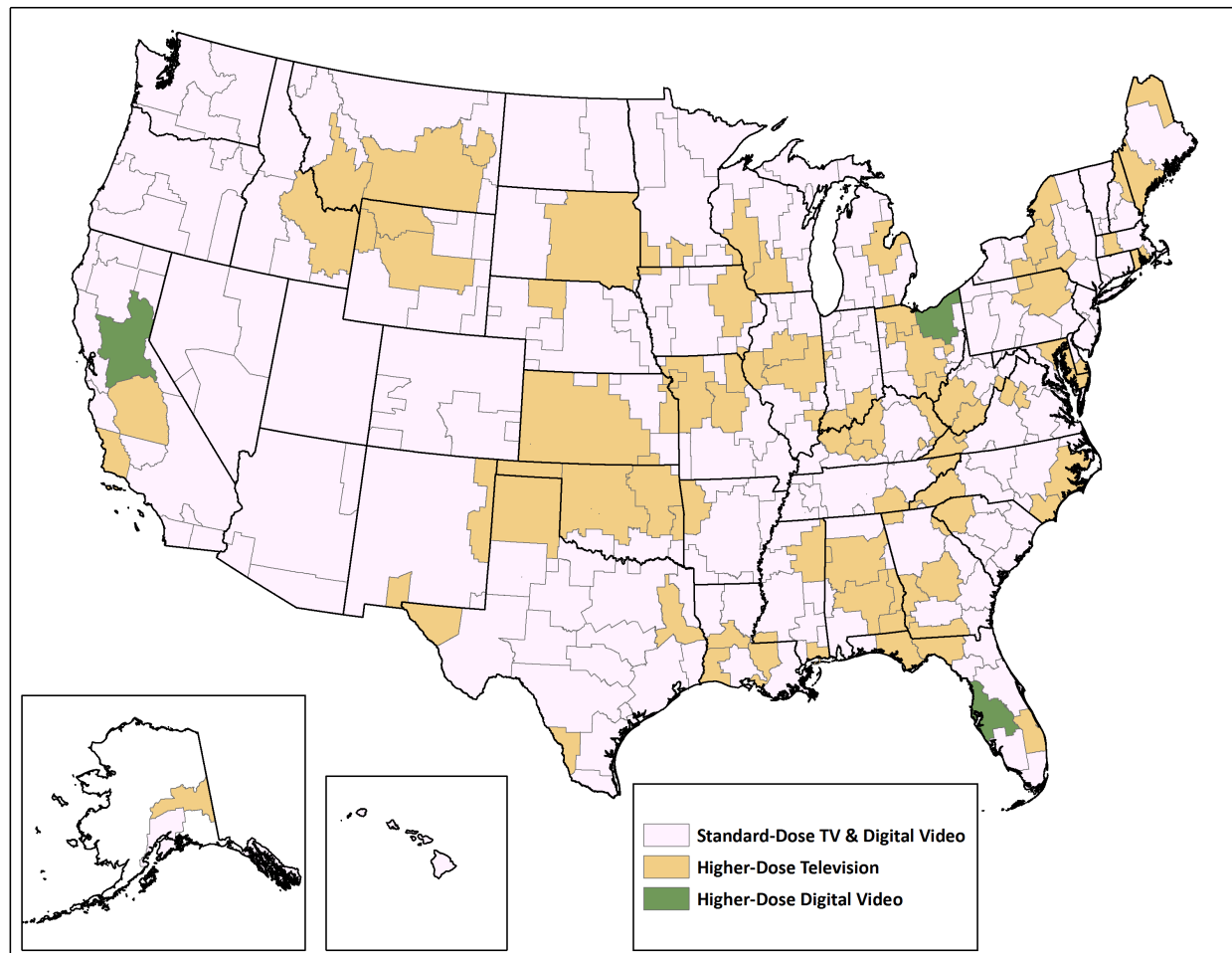
Survey Data

To measure population-level ad exposure rates, we used survey data from a national online survey of cigarette smokers conducted immediately after the conclusion of the 2013 Tips campaign. The survey sample was recruited from the online GfK KnowledgePanel (GfK Custom Research, LLC, Nuremberg, Germany) and included all previously available and newly recruited smokers in this panel. KnowledgePanel is recruited using address-based probability sampling, covering more than 95% of all US households. All panelists have a known probability of selection and cannot volunteer to be a part of the panel. The KnowledgePanel recruitment procedures are described in greater detail elsewhere [2,14,15]. Smokers were defined as adults aged ≥18 years who had smoked at least 100 cigarettes in their lifetime and reported currently smoking either every day or some days at the time of the 2013 Tips launch.

The final analytic dataset included a total of 15,400 current cigarette smokers aged ≥18 years in the United States. By market areas, there were 578 smokers in the 3 higher-dose digital

markets and 4632 smokers in the 67 higher-dose television markets. There were 10,190 smokers in total across the remaining 140 markets that received standard-dose television and digital video advertising. We weighted the survey data to be representative of the US adult cigarette smoker population. The weighting procedure used is similar to the weighting methods of the Behavioral Risk Factor Surveillance System, which uses demographic benchmarks from the US Census to yield a weighted survey sample that matches the US Census distributions for age, sex, race/ethnicity, and education [16,17].

Outcome Variables

The outcome variables in this study include a range of measures of self-reported exposure to the Tips campaign. We measured self-reported exposure to campaign ads using a standard ad recognition protocol [18]. Respondents first viewed 7 of a possible 11 Tips television and digital video ads via video stream within the survey to prompt recall and then immediately completed a battery of questions assessing their exposure to the ad in the past 3 months since the 2013 Tips campaign launch. Those who indicated any awareness of the viewed ad were then

asked to report the media format on which they recalled seeing the ad: (1) computer desktop or laptop, (2) mobile device, or (3) television. Respondents could indicate multiple media sources, since all Tips ads were available on television and digital formats. We repeated this process of displaying ads and assessing awareness for each ad, and randomized the display order of the ads. Respondents who were unable to view the ads via the within-survey video stream were shown a storyboard of screenshots from the ad along with text of the ad script. Among all respondents, 86.2% (13,275/15,400) were able to view the video ad streams, while the remaining participants viewed them as screenshots. There was no statistically significant difference in the rate of self-reported ad awareness across these 2 modes of ad viewing.

Using this ad recognition protocol, we created 3 dichotomous indicators of ad awareness within the past 3 months: (1) awareness of any ad via digital formats (computer desktop, laptop, or mobile device), (2) awareness of any ad via television, and (3) awareness of any ad via digital *or* television format. The third measure gauged overall campaign reach. We also created 2 additional dichotomous outcome indicators for awareness of Tips ads exclusively via digital formats and exclusively via television to examine the extent of simultaneous exposure through both formats. Finally, we created an index that measured overall frequency of ad exposure via either digital or television format among individuals who indicated that they had seen at least one Tips ad in the past 3 months. This index was defined as the sum of recall frequency (1=rarely saw ad, 4=saw ad very often) across each of the 7 ads that respondents viewed within the survey. Respondents who saw none of the 7 ads shown received a value of 0, whereas respondents who saw all 7 ads "very often" received a value of 28 for frequency of exposure (total range from 0 to 28).

Independent Variables

The primary independent variables in our analysis were dichotomous indicators of each media dose area: (1) higher-dose television markets, (2) higher-dose digital markets, and (3) standard-dose television and digital markets. In addition to these variables, we measured a wide range of potential confounders at the individual, state, and media market levels that may have been associated with ad recall. We examined whether these factors varied significantly across the exposure conditions to identify relevant control variables to include in our statistical analysis. These variables included demographic covariates for age, sex, race/ethnicity, education, annual household income, television viewing hours per day, household presence of children aged ≤17 years, household presence of a cigarette smoker, and having a chronic or mental health condition. In addition, we merged external state- and market-level variables with the survey data, including cumulative state per capita tobacco control program funding (1985–2012), state cigarette excise tax (2012), market-level population size, median income (in tens of thousands of dollars), and percentage of the population with a bachelor's degree. We also measured market-level cigarette smoking prevalence by aggregating recently published county-level data on smoking prevalence [19] to the market level, weighted by county population.

Advertising Costs

Cost information for the higher doses of television and digital video advertising were provided by the Tips campaign contractor, PlowShare Group (Stamford, CT, USA). These included the total costs for purchasing the higher doses of television and digital video advertising in each geographic area. Because the higher-dose digital markets have a larger audience than higher-dose television markets, average media costs for any platform are higher in the higher-dose digital markets. Therefore, we adjusted costs for the higher dose of digital video advertising for differences in total audience sizes. This was done to create per-market cost estimates that were comparable between the television and digital higher-dose markets. This adjustment was based on the estimated audience size for digital media (ie, total households with broadband Internet access in 2013) across the 67 higher-dose television markets. This enabled the Tips campaign's advertising agency to calculate the per-market costs for the increased dose of digital advertising as if it had been applied to the 67 higher-dose television markets.

Statistical Analysis

To assess the impact of additional digital and television advertising on audience exposure across each group of markets, we used logistic regression models to estimate each of the dichotomous awareness variables as a function of higher-dose digital and higher-dose television advertising markets, with standard-dose digital and television markets as the reference category. We used similar linear regression models to estimate the cumulative index of frequency of ad exposure as a function of the media dosing markets. All models included covariates for any of the aforementioned individual, market, or state-level variables that were significantly different across each media dose condition in bivariate analysis. Descriptive analyses of these variables showed that income, presence of a mental health condition, media market population, and media market median income varied significantly across the exposure conditions. Hence, our models included covariates for each of these variables.

Standard errors for each model were clustered at the media market level. We calculated predicted values for the outcome of awareness of any ad in digital or television format to estimate the percentage increases in overall awareness associated with digital and television higher-dose advertising. We used 1-tailed tests as our primary test of significance in the regression models, since our study was a real-world dosing test where all past research, including the 2012 Tips evaluation [2], suggested that there are no reasonable expectations that increased dosing of media would have negative effects on ad awareness [20-25]. All statistical tests used the .05 threshold for statistical significance. Given the smaller sample size of the higher-dose digital exposure condition, we also report marginally nonsignificant results below the .10 level, as these differences may be qualitatively meaningful but have more limited statistical power.

We calculated ad buy costs for each additional percentage point of predicted ad awareness attributable to the higher-dose digital and higher-dose television advertising. We compared costs per additional point of overall ad awareness and frequency of

exposure between the increased digital and television advertising doses to assess the marginal benefits of each advertising channel in increasing overall campaign reach at the population level. All analyses were conducted using Stata statistical software, Release 13 (StataCorp LP).

Results

Survey Sample Demographics

The unweighted sample was 77.0% (11,864/15,400) non-Hispanic white, 9.1% (1405/15,400) Hispanic, 8.1% (1241/15,400) non-Hispanic black, and 5.8% (890/15,400) non-Hispanic other races/ethnicities. Sample weighting appropriately increases representation of subgroups that are underrepresented in the unweighted data. For example, the weighted race/ethnicity distribution was 67.4% non-Hispanic white, 13.7% Hispanic, 12.1% non-Hispanic black, and 6.8% non-Hispanic other races/ethnicities. Table 1 presents the unweighted and weighted distributions of participants by age, sex, race/ethnicity, and education.

Table 1. Sample demographics of smokers, Tips From Former Smokers 2013 evaluation survey (N=15,400).

Characteristics	Unweighted data		Weighted data	
	n (%)	95% CI	%	95% CI
Age range (years)				
18–24	1311 (8.5)	8.1–9.0	13.3	12.4–14.3
25–34	2872 (18.7)	18.0–19.3	17.5	16.7–18.4
35–54	6256 (40.6)	39.9–41.4	36.8	35.7–37.9
55+	4961 (32.2)	31.5–33.0	32.4	31.4–33.5
Sex				
Male	6441 (41.9)	41.1–42.7	48.7	47.5–49.9
Female	8941 (58.1)	57.3–58.9	51.3	50.1–52.5
Race/ethnicity				
Non-Hispanic white	11,864 (77.0)	76.4–77.7	67.4	66.2–68.7
Non-Hispanic black	1241 (8.1)	7.6–8.5	12.1	11.3–13.0
Hispanic	1405 (9.1)	8.7–9.6	13.7	12.7–14.7
Non-Hispanic other	890 (5.8)	5.4–6.2	6.8	6.2–7.4
Education				
Less than high school	948 (6.2)	5.8–6.6	15.0	13.8–16.3
High school	3802 (24.7)	24.0–5.4	29.0	27.9–30.0
Some college	7008 (45.5)	44.7–46.3	30.	29.8–31.6
College graduate	3642 (23.7)	23.0–24.3	25.4%	24.4–26.3

Impact of Higher-Dose Digital Video and Television Advertising on Campaign Reach

Table 2 summarizes logistic and linear regression results for the relationships between higher-dose advertising and ad awareness. Higher-dose digital advertising was associated with 94% higher odds of awareness of any ad online relative to standard-dose markets (odds ratio, OR 1.94, *P*<.001). Higher-dose digital advertising was not significantly associated with increased awareness of ads via television (OR 1.33, *P*=.12). Higher-dose digital advertising was associated with a marginally nonsignificant increase in overall campaign awareness regardless of media format (OR 1.46, *P*=.09). Increased doses of television advertising were associated with significantly higher odds of ad awareness via both digital (OR 1.12, *P*=.04) and television formats (OR 1.87, *P*<.001). Higher-dose television advertising

was also associated with significantly higher odds of overall ad awareness regardless of media format (OR 1.81, *P*<.001).

Higher-dose digital advertising was associated with significantly higher odds of ad awareness only through online formats (OR 1.93, *P*<.001), while higher-dose television advertising was associated with a higher odds of ad awareness exclusively via television (OR 1.26, *P*<.001). Higher-dose digital advertising was associated with increased frequency of overall campaign exposure via any media channel (b=1.84, *P*=.01), as was higher-dose television advertising (b=2.79, *P*<.001). Based on predicted values from these models, we estimated that higher-dose digital advertising generated an approximate 6.6 percentage point increase in overall campaign awareness, while higher-dose television generated an estimated 8.6 percentage point increase.

Table 2. Regression model results for association between 2013 Tips From Former Smokers higher-dose digital and television advertising and ad awareness outcomes.

Media dose indicator (reference: standard-dose digital and television)	Logistic regression adjusted odds ratios					Linear regression coefficients
	Aware of any ad online	Aware of any ad on television	Aware of any ad on television or online	Aware of any ad on television only	Aware of any ad online only	Frequency of exposure on television or online
Higher-dose digital market	1.94	1.33	1.46	0.64	1.93	1.84
95% CI	1.39–2.70	0.82–2.17	0.84–2.52	0.58–0.71	1.40–2.67	0.25–3.42
P value	<.001	.12	.09	<.001	<.001	.01
Higher-dose television market	1.12	1.87	1.81	1.26	0.38	2.79
95% CI	0.99–1.28	1.59–2.19	1.55–2.12	1.12–1.41	0.19–0.75	2.09–3.49
P value	.04	<.001	<.001	<.001	.005	<.001

Digital Video and Television Cost Comparisons

The estimated cost of the additional television ad buy in the higher-dose markets was approximately US $9 million, while the comparable estimated cost of the additional digital ad buy was US $2.9 million (Table 3). Based on these cost estimates, the total estimated cost per additional percentage point of overall campaign reach (awareness of any ad on television or online) generated by the higher dose of digital video advertising was approximately US $440,000. By comparison, the cost of each additional percentage point of population reach generated by the higher dose of television advertising was approximately US $1 million.

Table 3. 2013 Tips From Former Smokers campaign costs for higher-dose digital and television advertising.

Media channel for higher dose of advertising	Cost (millions of US$)	Increase in awareness (%)	Estimated cost per percentage point of increased awareness (millions of US$)
Digital	2.9[a]	6.6	0.44
Television	9.0[b]	8.6	1.0

[a]Adjusted higher-dose digital costs.

[b]Actual higher-dose television costs.

Discussion

Principal Findings

Our findings suggest that, although digital video advertising may complement the overall reach of a broader multimedia campaign, television remains the strongest driver of overall campaign reach. The additional dosing of digital video advertising generated a significant increase in Tips ad awareness via online channels and also resulted in a marginally nonsignificant increase in overall ad awareness. However, the boost that the additional digital video advertising provided to the overall reach of the campaign was smaller (6.6 additional percentage points of awareness) than the impact of additional television advertising (8.6 additional percentage points of awareness). Awareness of the Tips campaign ads, and the resulting increased awareness of the specific health consequences of smoking, is an important precursor to action in reducing the prevalence and burden of tobacco use. For example, previous evaluation research on the 2012 Tips campaign showed that the campaign was associated with 100,000 new quit attempts among smokers [2], resulting in sizable downstream benefits, including more than 17,000 premature deaths averted and nearly 180,000 quality-adjusted life years saved [3].

In addition, we found that, although higher-dose digital video advertising generated larger increases in awareness via online channels, the higher-dose television advertising was also associated with a significant increase in online awareness. This likely reflects the fact that ad content was identical across digital video and television formats. Hence, primary awareness via television may have resulted in many individuals subsequently searching for the ads online or through social channels, such as Facebook or YouTube, and gaining additional online exposure through those digital activities. For campaign implementation practice, these results suggest that television advertising remains the most effective approach for driving overall campaign exposure at the population level, even when complementary digital video advertising is present. However, digital video advertising may be used to augment the exposure of a comprehensive campaign by reaching digital audiences.

In postestimation analyses, we found that only 2.3% of the sample in higher-dose digital markets reported seeing Tips ads exclusively online. This suggests that almost all persons exposed to Tips ads via digital channels were also exposed to television ads. Given the similarity in content and viewing experience between digital video and television formats, this raises the possibility that accurately differentiating between media platforms may be difficult. Emerging technologies and formats on which video content can be viewed (eg, smart televisions,

tablets, and smartphones) may necessitate measures of ad exposure that are not reliant on self-reports due to possible screen or device indifference among viewers. Campaign practitioners should consider using alternative data sources that provide passive measures of exposure to digital advertising as part of their evaluation efforts for these components of the campaign.

Our analysis suggests that the gains generated by higher doses of digital advertising were somewhat more cost efficient than those generated by higher doses of television advertising. Therefore, digital video advertising may be a cost-efficient way to boost the overall reach of a broad multimedia campaign. However, these findings do not imply that digital video is recommended as a predominant mode of advertising. The overall population of those watching digital video on the Internet in the United States is estimated to be 146 million compared with an estimated 285 million who watch traditional television [26]. Hence, the Internet digital video audience is smaller than the traditional television audience and may be younger or otherwise different from the general population, which has implications for the messaging and targeting of public health education campaigns. In addition, digital video advertising is delivered with more precision through targeted buys based on consumer online behaviors and demographic attributes, whereas television advertising is purchased for broader, less-defined audiences. This creates inherent limits to the maximum population reach that is achievable via digital video advertising channels, but it also facilitates more efficient reach among well-defined target audiences.

Limitations

This study has a few limitations. First, the assignment of higher-dose digital advertising was not randomized. Although we controlled for market-level characteristics that were statistically significant across media dose conditions, there may have been other unobserved and unmeasured factors that were also related to ad exposure. Second, our assessment of the impact of higher-dose digital advertising on recall of campaign ads was based on a relatively smaller sample of smokers in the 3 higher-dose digital markets. Hence, estimates of the effect of higher-dose digital advertising on campaign awareness may be less precise than estimates of the effect of higher-dose television advertising. Third, self-reported exposure to television ads may have been underreported because of our use of a Web-based survey, which potentially underrepresents populations that are not Internet enabled. Awareness of television ads may be underestimated if these populations rely more on television for media exposure. However, as described elsewhere [2,14,15], the KnowledgePanel sample we used provided coverage for non-Internet households by either providing a free laptop and Internet service or other means such as additional honoraria for completing the survey at a location with public Internet access. Fourth, it is possible that there was some imprecision in DMA-based media assignments based on TV media "leakage"

between adjacent DMAs or imprecision in the IP-based geographic assignments of digital media [27]. Fifth, the overlap of ad exposures via television and digital formats highlights potential difficulties in measuring the true contributions of digital video advertising as part of a broader multimedia campaign. In the case of the 2013 Tips campaign, the television and digital video ad content were identical. Given the low rates of online-only ad exposure, many individuals who reported seeing Tips ads online may have seen them on television first and subsequently searched for the ads online or discovered the ads after visiting the Tips website (which was also promoted by the television ads) or social media. We do not know the proportion of the smoker population that was exposed to a placed digital ad, as opposed to secondary exposure through searches after initial exposure via television. Additional measures on how digital ads are "discovered" when awareness has been claimed are important to understand the extent to which digital video advertising drives message exposure at the population level.

This is the first study, to our knowledge, to draw direct efficiency comparisons between digital video and television advertising in increasing the overall reach of a multimedia campaign at the population level. Previous studies have primarily focused on the use of digital media for purposes of recruiting smokers to cessation interventions [6,7,9] but have not systematically analyzed the use of digital media for promoting overall campaign reach. This study helps fill this gap with an assessment of digital video advertising from the 2013 Tips campaign and compared its efficiency with that of traditional television advertising.

Conclusions

In summary, this study provides new evidence on the role of digital video advertising in boosting the overall population-level reach of a multimedia campaign. Our results suggest that the use of digital advertising enhances overall exposure to a large television-based campaign. We also found that digital advertising was more cost efficient than television in driving ad awareness. However, the smaller size of the digital video viewer audience and relatively smaller effect of digital ad exposure relative to television suggests that digital video is best used as a complement to a main television campaign and not as a replacement for television.

In addition, further research is important to better understand the utility of digital video advertising in light of the high concomitancy of digital and television ad exposure. Specifically, new measures are critical to isolate population-level reach attributable solely to digital video advertising. Additional measurement of format-exclusive ads (ie, ads available in only one format or the other) and exogenous measures of ad exposure are also important to assess the extent to which digital video advertising alone can yield significant population-level rates of exposure.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Tips From Former Smokers. Atlanta, GA: CDC; 2016 Jan 26. URL: http://www.cdc.gov/tobacco/campaign/tips/ [accessed 2016-02-17] [WebCite Cache ID 6fMhOz4pb]
2. McAfee T, Davis KC, Alexander RL, Pechacek TF, Bunnell R. Effect of the first federally funded US antismoking national media campaign. Lancet 2013 Dec 14;382(9909):2003-2011. [doi: 10.1016/S0140-6736(13)61686-4] [Medline: 24029166]
3. Xu X, Alexander RL, Simpson SA, Goates S, Nonnemaker JM, Davis KC, et al. A cost-effectiveness analysis of the first federally funded antismoking campaign. Am J Prev Med 2015 Mar;48(3):318-325. [doi: 10.1016/j.amepre.2014.10.011] [Medline: 25498550]
4. Shafer PR, Davis KC, Patel D, Rodes R, Beistle D. Association between media dose, ad tagging, and changes in web traffic for a national tobacco education campaign: a market-level longitudinal study. J Med Internet Res 2016;18(2):e39 [FREE Full text] [doi: 10.2196/jmir.5343] [Medline: 26887959]
5. Kim A, Hansen H, Duke J, Davis K, Alexander R, Rowland A, et al. Does digital ad exposure influence information-seeking behavior online? Evidence from the 2012 Tips From Former Smokers National Tobacco Prevention Campaign. J Med Internet Res 2016;18(3):e64 [FREE Full text] [doi: 10.2196/jmir.4299] [Medline: 26983849]
6. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]
7. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. J Med Internet Res 2012;14(4):e116 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]
8. Emery S, Aly EH, Vera L, Alexander RL. Tobacco control in a changing media landscape: how tobacco control programs use the internet. Am J Prev Med 2014 Mar;46(3):293-296 [FREE Full text] [doi: 10.1016/j.amepre.2013.11.005] [Medline: 24512869]
9. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1:S35-S41. [Medline: 17491169]
10. Tian H, Brimmer DJ, Lin JS, Tumpey AJ, Reeves WC. Web usage data as a means of evaluating public health messaging and outreach. J Med Internet Res 2009;11(4):e52 [FREE Full text] [doi: 10.2196/jmir.1278] [Medline: 20026451]
11. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the utility of Web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
12. Farrelly MC, Davis KC, Nonnemaker JM, Kamyab K, Jackson C. Promoting calls to a quitline: quantifying the influence of message theme, strong negative emotions and graphic images in television advertisements. Tob Control 2011 Jul;20(4):279-284. [doi: 10.1136/tc.2010.042234] [Medline: 21289028]
13. McAfee T, Davis KC, Shafer P, Patel D, Alexander R, Bunnell R. Increasing the dose of television advertising in a national antismoking media campaign: results from a randomised field trial. Tob Control 2015 Dec 16:052517 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052517] [Medline: 26678518]
14. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet. Public Opin Q 2009 Dec 01;73(4):641-678. [doi: 10.1093/poq/nfp075]
15. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. Public Opin Q 2011 Oct 05;75(4):709-747. [doi: 10.1093/poq/nfr020]
16. Behavioral Risk Factor Surveillance System: Overview: BRFSS 2013. Atlanta, GA: Centers for Disease Control and Prevention; 2014 Aug 15. URL: http://www.cdc.gov/brfss/annual_data/2013/pdf/overview_2013.pdf [accessed 2016-08-26] [WebCite Cache ID 6k3IIuyjk]

17.   Behavioral Risk Factor Surveillance System: Weighting BRFSS Data: BRFSS 2013. Atlanta, GA: Centers for Disease Control and Prevention URL: http://www.cdc.gov/brfss/annual_data/2013/pdf/weighting_data.pdf [accessed 2016-08-26] [WebCite Cache ID 6k3IVFzJl]

18.   Southwell BG, Barmada CH, Hornik RC, Maklan DM. Can we measure encoded exposure? Validation evidence from a national campaign. J Health Commun 2002;7(5):445-453 [FREE Full text] [doi: 10.1080/10810730290001800] [Medline: 12455763]

19.   Dwyer-Lindgren L, Mokdad AH, Srebotnjak T, Flaxman AD, Hansen GM, Murray CJ. Cigarette smoking prevalence in US counties: 1996-2012. Popul Health Metr 2014;12(1):5 [FREE Full text] [doi: 10.1186/1478-7954-12-5] [Medline: 24661401]

20.   Ludbrook J. Should we use one-sided or two-sided P values in tests of significance? Clin Exp Pharmacol Physiol 2013 Jun;40(6):357-361. [doi: 10.1111/1440-1681.12086] [Medline: 23551169]

21.   Cho H, Abe S. Is two-tailed testing for directional research hypotheses tests legitimate? J Bus Res 2013 Sep;66(9):1261-1266. [doi: 10.1016/j.jbusres.2012.02.023]

22.   Wakefield MA, Spittal MJ, Yong H, Durkin SJ, Borland R. Effects of mass media campaign exposure intensity and durability on quit attempts in a population-based cohort study. Health Educ Res 2011 Dec;26(6):988-997 [FREE Full text] [doi: 10.1093/her/cyr054] [Medline: 21730252]

23.   Vallone DM, Duke JC, Cullen J, McCausland KL, Allen JA. Evaluation of EX: a national mass media smoking cessation campaign. Am J Public Health 2011 Feb;101(2):302-309. [doi: 10.2105/AJPH.2009.190454] [Medline: 21164094]

24.   Sims M, Salway R, Langley T, Lewis S, McNeill A, Szatkowski L, et al. Effectiveness of tobacco control television advertising in changing tobacco use in England: a population-based cross-sectional study. Addiction 2014 Jun;109(6):986-994 [FREE Full text] [doi: 10.1111/add.12501] [Medline: 24467285]

25.   Farrelly MC, Duke JC, Davis KC, Nonnemaker JM, Kamyab K, Willett JG, et al. Promotion of smoking cessation with emotional and/or graphic antismoking advertising. Am J Prev Med 2012 Nov;43(5):475-482. [doi: 10.1016/j.amepre.2012.07.023] [Medline: 23079169]

26.   Connecting With the Cosmos: The Total Audience Media Uuniverse. New York, NY: The Nielsen Company; 2015 Mar 19. URL: http://www.nielsen.com/us/en/insights/news/2015/connecting-with-the-cosmos-the-total-audience-media-universe.html [accessed 2016-08-23] [WebCite Cache ID 6jyyaBclr]

27.   Jones RB, Goldsmith L, Williams CJ, Kamel Boulos MN. Accuracy of geographically targeted internet advertisements on Google AdWords for recruitment in a randomized trial. J Med Internet Res 2012;14(3):e84 [FREE Full text] [doi: 10.2196/jmir.1991] [Medline: 22718043]

Abbreviations

CDC: Centers for Disease Control and Prevention
DMA: designated market area
GRP: gross rating point
IP: Internet protocol
OR: odds ratio
Tips: Tips From Former Smokers

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Do Health Care Providers Use Online Patient Ratings to Improve the Quality of Care? Results From an Online-Based Cross-Sectional Study

Martin Emmert[1*], MHMM, PhD; Nina Meszmer[2*], MSc; Uwe Sander[3], MD

[1]Health Services Management, Institute of Management, School of Business and Economics, Friedrich-Alexander-University Erlangen-Nuremberg, Nuremberg, Germany

[2]Health Care Management, Institute of Management, School of Business and Economics, Friedrich-Alexander-University Erlangen-Nuremberg, Nuremberg, Germany

[3]Media, Information and Design, Department of Information and Communication, University of Applied Sciences and Arts, Hannover, Hannover, Germany

[*]these authors contributed equally

Corresponding Author:
Martin Emmert, MHMM, PhD
Health Services Management, Institute of Management
School of Business and Economics
Friedrich-Alexander-University Erlangen-Nuremberg
Lange Gasse 20
Nuremberg, 90403
Germany
Phone: 49 9115302 ext 253
Fax: 49 9115302285
Email: Martin.Emmert@fau.de

Abstract

Background: Physician-rating websites have become a popular tool to create more transparency about the quality of health care providers. So far, it remains unknown whether online-based rating websites have the potential to contribute to a better standard of care.

Objective: Our goal was to examine which health care providers use online rating websites and for what purposes, and whether health care providers use online patient ratings to improve patient care.

Methods: We conducted an online-based cross-sectional study by surveying 2360 physicians and other health care providers (September 2015). In addition to descriptive statistics, we performed multilevel logistic regression models to ascertain the effects of providers' demographics as well as report card-related variables on the likelihood that providers implement measures to improve patient care.

Results: Overall, more than half of the responding providers surveyed (54.66%, 1290/2360) used online ratings to derive measures to improve patient care (implemented measures: mean 3.06, SD 2.29). Ophthalmologists (68%, 40/59) and gynecologists (65.4%, 123/188) were most likely to implement any measures. The most widely implemented quality measures were related to communication with patients (28.77%, 679/2360), the appointment scheduling process (23.60%, 557/2360), and office workflow (21.23%, 501/2360). Scaled-survey results had a greater impact on deriving measures than narrative comments. Multilevel logistic regression models revealed medical specialty, the frequency of report card use, and the appraisal of the trustworthiness of scaled-survey ratings to be significantly associated predictors for implementing measures to improve patient care because of online ratings.

Conclusions: Our results suggest that online ratings displayed on physician-rating websites have an impact on patient care. Despite the limitations of our study and unintended consequences of physician-rating websites, they still may have the potential to improve patient care.

Introduction

Over the last years, physician-rating websites (PRWs) have become a popular tool to create more transparency surrounding the quality of health care providers in the United States and other industrialized countries [1-3]. They usually provide structural information about a physician's office as well as online-derived patient satisfaction results [3]. Regarding the latter, the rating systems usually contain both scaled-rating systems (eg, stars, grades) and narrative comments. Although scaled-rating systems with standardized questions present a more structured way to receive answers about different aspects of care [4], free-text commentaries allow patients to express themselves in their own words [5]. The comments are intended to provide a more complete picture of the patient experience with that provider, incorporating emotional reactions and the meaning that patients ascribe to their experiences [4,6]. When comparing those two features, narrative comments are meant to be more powerful [7] because users are drawing more attention to words than to numbers [8]. Furthermore, narrative comments have been suggested as one possibility to provide performance metrics that are more easily understood, raising the willingness of users to provide a substantial feedback and creating a more personal feedback than other rating formats [4,9,10].

In general, PRWs and other established public reporting instruments (eg, the Wisconsin Collaborative for Healthcare Quality [11], the New York State Coronary Artery Bypass Surgery Report [12], or Nursing Home Compare [13]) intend to improve the quality of health care via two different pathways. According to the first pathway ("selection"), health care services are shifted to caregivers who provide better quality of care. However, there seems to be little or no impact on the selection of health care providers by patients and families or their representatives so far [14]. Nevertheless, a previously published cross-sectional study has shown that 65% of PRW users consulted a particular physician based on the online ratings and 52% avoided consulting a particular physician because of the ratings [15]. In addition, the second pathway ("changes") describes a mechanism by which providers are motivated to improve their quality of care for patients [16]. Regarding results from the latest systematic reviews, it seems that publicly releasing performance data stimulates quality improvement activity by means of offering new services, changing policies, change in personnel, and an increase in quality improvement activities [14,17].

So far, most of the evidence on whether public reporting instruments improve the quality of care comes from the inpatient sector [12,18,19]; little evidence is available in the outpatient sector [8,11,20]. Even further, no evidence is available on whether PRWs have an impact on the quality of care by motivating physicians to improve their quality of care [21]. In addition, little is known about the characteristics of physicians who use online rating websites or for what purposes [22]. Therefore, it is important to gain a scientific understanding of

whether online-based rating websites have the potential to contribute to a better standard of care. In this paper, we present the results of an online-based cross-sectional study that examines (1) which health care providers use online rating websites (2) for what purposes and (3) assesses whether health care providers use online patient ratings to improve patient care.

Methods

Design and Data Source

We conducted an online-based cross-sectional study by surveying outpatient physicians and other outpatient health care providers (eg, midwives, speech and language therapists) who have registered on the German PRW jameda (September 2015). These providers either subscribed to a monthly newsletter that contained an overview of all individual ratings posted on jameda over the previous four weeks or booked a jameda service product. On jameda, providers can register for a free-of-charge basic account that permits them to modify their personal or structural office data and ensures that they receive notification of online ratings on a monthly basis, as well as including the possibility of commenting on the patients' ratings on the website. Three products are offered on jameda that contain different tools, such as special presentation formats of the practice or the uploading of additional data. To date, jameda is likely to play the most important role in the German PRW movement [23]. For our purposes, the Web service provider of jameda sent out a special issue newsletter via email to registered health care providers (N=25,000) that contained an invitation and an online link to participate in the survey. The newsletter also contained some information about the study and its purpose. As an incentive, we held drawings for four Amazon vouchers with a value of €150 each.

We designed the survey by using QuestBack's Internet-based Enterprise Feedback Suite survey software. The questionnaire contained 26 questions about PRWs in general and consisted of three parts. After collecting sociodemographic information, the second part asked questions about the knowledge and usage of PRWs. In the third part of the survey, we aimed to assess whether and how health care providers react to online ratings and whether they had implemented any of 20 listed measures in order to improve patient care. Those measures were derived from a systematic review regarding the impact of public reporting on the quality of patient care (described subsequently). We distinguished between scaled-survey questions and narrative comments to investigate which are of greater importance to the providers (see Multimedia Appendix 1 for the developed survey instrument.)

Before conducting the study, the questionnaire was piloted by 25 individuals to ensure the proper working of the survey, the randomization process of some questions, and the comprehensibility of the wording. The pretest resulted in minor adaptations to the wording of some questions. We conducted a systematic search procedure in the databases Medline (via PubMed) and the Cochrane Library (May 2015) to identify

studies that have researched the impact of public reporting on the provider behavior with respect to the implementation of quality measures. Our search yielded 12 studies that were related to the outpatient (n=3) or inpatient sector (n=9). However, no study had investigated the usage and impact of online rating websites on provider behavior.

Data Analysis

Results are presented as both mean and standard deviation for continuous variables and as numbers and percentages for categorical variables. We performed comparisons between more than two groups by using a chi-square test (two-sided) for categorical variables and applying the Kruskal-Wallis test for continuous nonparametric variables. The Shapiro-Wilk test was applied to examine the normality of the data distribution. Multilevel multivariate logistic regression models were used to assess how much of the providers' reaction to implement measures could be explained by demographic and online report card-related variables. The dependent variable indicated whether a health care provider had implemented any measure(s) to improve patient care (yes/no). Regarding the independent variables, we used the following sequence of models: (1) adjusted for demographics (age, gender, marital status, Internet use, medical specialty), (2) adding information regarding any booked jameda service products (yes/no), and (3) adding online report card-related issues (ie, frequency of use, appraisal of the trustworthiness of the online ratings). All statistical analyses were conducted using SPSS version 22.0 (IBM Corp, Armonk, NY, USA). Observed differences were considered statistically significant if $P<.05$.

Results

In total, our final study sample consisted of 2360 respondents who completed the survey (response rate=9.44%, 2360/25000; completion rate=67.29%, 2360/3507). We excluded 49 participants from subsequent analysis because of extremely short answer times and/or inconsistent answer patterns. The completed online surveys took a mean 9.63 (SD 9.03) minutes. The overall mean age was 49.63 (SD 8.69) years, 66.67% (1560/2340) respondents were male, and the mean duration of practice in the doctors' office was 12.99 (SD 9.10) years (Table 1).

Regarding the medical specialty, 17.50% (413/2360) of the surveyed sample were general practitioners, 69.36% (1637/2360) were specialists, and 13.14% (310/2360) were other health care-related professions (eg, midwives). The three groups did not differ significantly in terms of age and Internet use, but did in terms of gender, marital status, and years of practice in the office.

Awareness and Usage of Online Rating Websites

Most providers became aware of the websites through the Internet (71.91%, 1697/2360), contact with the providers of PRWs (20.08%, 474/2360), or advertisements (16.74%, 395/2360) (Table 2). Differences regarding the sources of awareness between the three provider groups (ie, general practitioners, specialists, others) could all be shown to be statistically significant ($P<.05$ each), except for recommendations by friends and relatives as well as others. Furthermore, specialists used PRWs more frequently than general practitioners and other providers did ($P<.001$). Most providers (87.08%, 2055/2360) used online rating websites to read comments for their individual practice. Almost half of all providers read comments for other providers (48.69%, 1149/2360) and slightly more than one in three providers (35.97%, 849/2360) to know which measures might be implemented to increase patient satisfaction. In addition, 12.08% (285/2360) of respondents stated that they used online ratings for referring patients to other providers. Therefore, numbers for general practitioners (12.1%, 50/413) and specialists (10.38%, 170/1637) were significantly lower than those for other providers (21.0%, 65/310, $P<.001$). In addition, specialists evaluated their ratings significantly more often than general practitioners and other providers did ($P<.001$). Most providers evaluated the ratings themselves (84.87%, 2003/2360).

Table 1. Characteristics of respondents according to their medical discipline.

Characteristics	Overall (N=2360)	General practitioners (n=413)	Specialists (n=1637)	Others (n=310)	P^a value
Age (years)					
Mean (SD)	49.63 (8.69)	50.61 (8.65)	49.41 (8.76)	49.45 (8.33)	.08
Ranges, n (%)					.06
<35	109 (5.05)	9 (2.4)	87 (5.83)	13 (4.5)	
36-45	598 (27.71)	110 (29.0)	411 (27.55)	77 (26.8)	
46-55	890 (41.24)	142 (37.5)	619 (41.49)	129 (45.0)	
56-65	494 (22.89)	104 (27.4)	330 (22.12)	60 (20.9)	
≥66	67 (3.10)	14 (3.7)	8 (2.8)		
		45 (3.02)			
Gender, n (%)					<.001
Male	1560 (66.67)	278 (67.8)	1175 (72.31)	107 (35.1)	
Female	780 (33.33)	132 (32.2)	450 (27.69)	198 (64.9)	
Marital status, n (%)					<.001
Married	1706 (77.86)	316 (81.9)	1218 (79.87)	172 (61.4)	
Widowed	20 (0.91)	10 (2.6)	6 (0.39)	4 (1.4)	
Single	294 (13.42)	31 (8.0)	206 (13.51)	57 (20.6)	
Divorced	171 (7.80)	29 (7.5)	95 (6.23)	47 (16.8)	
Length of practice in the doctor's office (years)					
Mean (SD)	12.99 (9.10)	14.35 (9.67)	12.93 (8.95)	11.48 (8.83)	<.001
Ranges, n (%)					<.001
<5	616 (27.76)	97 (24.8)	427 (27.80)	92 (31.5)	
6-10	479 (21.59)	78 (20.0)	323 (21.03)	78 (26.7)	
11-15	333 (15.01)	56 (14.3)	238 (15.49)	39 (13.4)	
16-20	262 (11.81)	33 (8.4)	191 (12.43)	38 (13.0)	
21-25	303 (13.65)	72 (18.4)	211 (13.74)	20 (6.9)	
≥26	226 (10.18)	55 (14.1)	146 (9.51)	25 (8.6)	
Internet use, n (%)					.41
Several times a day	2124 (90.00)	366 (88.6)	1485 (90.71)	273 (88.1)	
Once a day	149 (6.31)	27 (6.5)	98 (5.99)	24 (7.7)	
Less than once a day	87 (3.69)	20 (4.8)	54 (3.30)		
			13 (4.2)		
Jameda service product, n (%)					<.001
Basic product	1601 (67.84)	365 (88.4)	1031 (62.98)	205 (66.1)	
Any service product (eg, gold, silver, platinum)	759 (32.16)	48 (11.6)	606 (37.02)	105 (33.9)	

^{a}*P* value was calculated using Kruskal-Wallis [1] and chi-square test [2].

Table 2. Awareness and use of physician-rating websites.

Characteristics	Overall, n (%) (N=2360)	General practitioners, n (%) (n=413)	Specialists, n (%) (n=1637)	Others, n (%) (n=310)	P^a value
How did you become aware of physician-rating websites?					
Contact with physician-rating website provider	474 (20.08)	78 (18.9)	352 (21.50)	44 (14.2)	.01
Internet	1697 (71.91)	295 (71.4)	1198 (73.18)	204 (65.8)	.03
Contact with patients	292 (12.37)	58 (14.0)	210 (12.83)	24 (7.7)	.02
Recommendations by peers	281 (11.91)	15 (3.6)	210 (12.83)	56 (18.1)	<.001
Advertisement	395 (16.74)	66 (16.0)	300 (18.33)	29 (9.4)	<.001
Newspapers or magazines	141 (5.97)	29 (7.0)	112 (6.84)	0 (0)	<.001
Recommendations by friends or relatives	89 (3.77)	12 (2.9)	61 (3.73)	16 (5.2)	.29
Others	111 (4.70)	23 (5.6)	79 (4.83)	9 (2.9)	.23
How often do you use physician-rating websites?					<.001
At least once per day	155 (6.57)	16 (3.9)	130 (7.94)	9 (2.9)	
Several times a week	273 (11.57)	29 (7.0)	209 (12.77)	35 (11.3)	
Once per week	500 (21.19)	92 (22.3)	346 (21.14)	62 (20.0)	
Once per month	655 (27.75)	121 (29.3)	445 (27.18)	89 (28.7)	
Less frequently	640 (27.12)	129 (31.2)	417 (25.47)	94 (30.3)	
Never	137 (5.81)	26 (6.3)	90 (5.50)	21 (6.8)	
For what purpose(s) do you use physician-rating websites?					
Reading own ratings	2055 (87.08)	363 (87.9)	1448 (88.45)	244 (78.7)	<.001
Commenting on own ratings	656 (27.80)	94 (22.8)	506 (30.91)	56 (18.1)	<.001
Reading ratings of other physicians because of interest	1149 (48.69)	181 (43.8)	816 (49.85)	152 (49.013)	.09
Readings ratings of other physicians for patient referral	285 (12.08)	50 (12.1)	170 (10.38)	65 (21.0)	<.001
Own practice marketing	785 (33.26)	78 (18.9)	579 (35.37)	128 (41.3)	<.001
I use physician-rating websites for other purposes	68 (2.88)	12 (2.9)	46 (2.81)	10 (3.23)	.92
How often do you evaluate your ratings on physician-rating websites?					<.001
At least once per week	443 (18.77)	44 (10.7)	348 (21.26)	51 (16.5)	
Several times a month	236 (10.00)	29 (7.0)	174 (10.63)	33 (10.7)	
Once per month	765 (32.42)	142 (34.4)	535 (32.68)	88 (28.4)	
Less frequently	734 (31.10)	161 (39.0)	477 (29.14)	96 (31.0)	
Never	182 (7.71)	37 (9.0)	103 (6.29)	42 (13.6)	
Who is responsible for evaluating the online ratings for your practice?					
I evaluate the online ratings for my practice myself	2003 (84.87)	354 (85.7)	1385 (84.61)	264 (85.2)	.84
Medical assistant(s)	115 (4.87)	24 (5.8)	89 (5.44)	2 (0.7)	<.001
Practice manager	180 (7.63)	17 (4.1)	159 (9.71)	4 (1.3)	<.001
Others	47 (1.99)	6 (1.5)	41 (2.50)	0 (0)	.01

[a]*P* value was calculated using chi-square test for all variables.

XSL•FO
RenderX

Implemented Measures to Improve Patient Care Because of Online Ratings

Overall, 54.66% (1290/2360) of respondents stated that they had used online ratings to implement some measure(s) to improve patient care (Tables 3 and 4). Numbers for specialists (57.79%, 946/1637) were significantly higher than those for general practitioners (50.1%, 207/413) and other providers (44.2%, 137/310, $P<.001$). The most widely implemented measures were related to communication with patients (28.77%, 679/2360) or the appointment scheduling process (23.60%, 557/2360). Staff-related measures also played an important role; 10.38% (245/2360) of respondents stated that they had invested

in further training of staff and 5.89% (139/2360) had recruited additional staff. Scaled-survey results seemed to have a greater impact on developing measures to improve patient care. The results showed that half of respondents (50.08%, 1182/2360) developed measures from scaled-survey ratings, whereas only 43.90% (1036/2360) stated that they had used information from narrative comments. In line with these results, specialists made significantly more use of the information in both scaled-survey ratings and narrative comments (53.21%, 871/1637 and 47.16%, 772/1637, respectively) than general practitioners (44.6%, 184/413 and 41.7%, 172/413, respectively) and other providers (41.0%, 127/310 and 29.7%, 92/310, respectively; $P<.001$).

Table 3. Measures that were implemented to increase patient satisfaction because of online ratings by results type and discipline (N=2360).

Measures	Overall by results type, n (%)			Overall by discipline, n (%)			
	Overall	Scaled-survey results	Narrative comments results	General practitioners (n=413)	Specialists (n=1637)	Others (n=310)	P^a value
Improvement of the communication with patients	679 (28.77)	527 (22.33)	444 (18.81)	107 (25.9)	517 (31.58)	55 (17.7)	<.001
Improve appointment scheduling process	557 (23.60)	456 (19.32)	317 (13.43)	89 (21.6)	423 (25.84)	45 (14.5)	<.001
Change in office workflow	501 (21.23)	379 (16.06)	310 (13.14)	88 (21.3)	378 (23.09)	35 (11.3)	<.001
Improvement of the waiting room equipment	266 (11.27)	225 (9.53)	135 (5.72)	42 (10.2)	202 (12.34)	22 (7.1)	.02
Training of the staff	245 (10.38)	173 (7.33)	152 (6.44)	39 (9.4)	196 (11.97)	10 (3.2)	<.001
Reassigning staff responsibilities	231 (9.79)	152 (6.44)	144 (6.10)	38 (9.2)	185 (11.30)	8 (2.6)	<.001
Investments in new technologies/ equipment (IT[b] equipment)	200 (8.47)	162 (6.86)	104 (4.41)	31 (7.5)	150 (9.16)	19 (6.1)	.16
Expand office hours	189 (8.01)	155 (6.57´)	104 (4.41)	26 (6.3)	132 (8.06)	31 (10.0)	.19
Introduction of patient reminders (eg, email reminders)	184 (7.80)	145 (6.14)	91 (3.86)	23 (5.6)	141 (8.61)	20 (6.5)	.08
Further educational training myself	157 (6.65)	113 (4.79)	90 (3.81)	14 (3.4)	102 (6.23)	41 (13.2)	<.001
Recruitment of additional staff	139 (5.89)	115 (4.87)	75 (3.18)	22 (5.3)	109 (6.66)	8 (2.6)	.02
Improvement of the communication with other providers	108 (4.58)	77 (3.26)	55 (2.33)	14 (3.4)	81 (4.95)	13 (4.2)	.38
Introduction of guidelines	100 (4.24)	76 (3.22)	51 (2.16)	2 (0.5)	83 (5.07)	15 (4.8)	<.001
Dismissing staff	78 (3.31)	56 (2.37)	46 (1.95)	12 (2.9)	61 (3.73)	5 (1.6)	.14
Higher usage of guidelines	77 (3.26)	57 (2.42)	39 (1.65)	10 (2.4)	58 (3.54)	9 (2.9)	.48
Planning of follow-up tests	65 (2.75)	46 (1.95)	35 (1.48)	15 (3.6)	44 (2.69)	6 (1.9)	.37
Hygiene improvement measures	64 (2.71)	48 (2.03)	33 (1.40)	9 (2.2)	46 (2.81)	9 (2.9)	.76
Others	140 (5.93)	86 (3.64)	108 (4.58)	22 (5.3)	97 (5.93)	21 (6.8)	.72
I have not implemented any measures	1070 (45.34)	1178 (49.92)	1324 (56.10)	206 (49.9)	691 (42.21)	173 (55.8)	<.001

[a]P value was calculated using chi-square test.

[b]IT: Information Technology.

Table 4. Measures that were implemented to increase patient satisfaction because of online ratings by scaled-survey and negative comments results (N=2360).

Measures	Scaled-survey results, n (%)				Narrative comments results, n (%)			
	General practitioners (n=413)	Specialists (n=1637)	Others (n=310)	P [a]	General practitioners (n=413)	Specialists (n=1637)	Others (n=310)	P value
Improvement of the communication with patients	83 (20.1)	399 (24.37)	45 (14.5)	<.001	73 (17.7)	341 (20.83)	30 (9.7)	<.001
Improve appointment scheduling process	65 (15.7)	355 (21.69)	36 (11.6)	<.001	54 (13.1)	239 (14.60)	24 (7.7)	.005
Change in office workflow	68 (16.5)	281 (17.17)	30 (9.7)	.004	55 (13.3)	238 (14.54)	17 (5.5)	<.001
Improvement of the waiting room equipment	37 (9.0)	169 (10.32)	19 (6.1)	.06	23 (5.6)	103 (6.29)	9 (2.9)	.06
Training of the staff	32 (7.8)	136 (8.31)	5 (1.6)	<.001	25 (6.1)	122 (7.45)	5 (1.6)	<.001
Reassigning staff responsibilities	24 (5.8)	122 (7.45)	6 (1.9)	<.001	26 (6.3)	115 (7.03)	3 (1.0)	<.001
Investments in new technologies/ equipment (IT[b] equipment)	22 (5.3)	124 (7.57)	16 (5.2)	.12	19 (4.6)	80 (4.89)	5 (1.6)	.04
Expand office hours	22 (5.3)	109 (6.66)	24 (7.7)	.42	12 (2.9)	74 (4.52)	18 (5.8)	.16
Introduction of patient reminders (eg, email reminders)	21 (5.1)	109 (6.66)	15 (4.8)	.29	14 (3.4)	69 (4.22)	8 (2.6)	.34
Further educational training myself	9 (2.2)	71 (4.34)	33 (10.7)	<.001	9 (2.2)	63 (3.85)	18 (5.8)	.04
Recruitment of additional staff	18 (4.4)	91 (5.56)	6 (1.9)	.02	10 (2.4)	60 (3.67)	5 (1.6)	.11
Improvement of the communication with other providers	9 (2.2)	58 (3.54)	10 (3.2)	.38	9 (2.2)	42 (2.57)	4 (1.3)	.38
Introduction of guidelines	2 (0.5)	63 (3.85)	11 (3.6)	.002	1 (0.2)	41 (2.50)	9 (2.9)	.01
Dismissing staff	10 (2.4)	43 (2.63)	3 (1.0)	.21	6 (1.5)	37 (2.26)	3 (1.0)	.23
Higher usage of guidelines	7 (1.7)	42 (2.57)	8 (2.6)	.58	6 (1.5)	31 (1.89)	2 (0.7)	.27
Planning of follow-up tests	8 (1.9)	32 (1.95)	6 (1.9)	.99	9 (2.2)	24 (1.47)	2 (0.7)	.24
Hygiene improvement measures	8 (1.9)	33 (2.02)	7 (2.3)	.96	5 (1.2)	25 (1.53)	3 (1.0)	.70
Others	13 (3.2)	57 (3.48)	16 (5.2)	.30	17 (4.1)	76 (4.64)	15 (4.8)	.88
I have not implemented any measures	229 (55.5)	766 (46.79)	183 (59.0)	<.001	241 (58.4)	865 (52.84)	218 (70.3)	<.001

[a]P value was calculated using chi-square test.

[b]IT: Information Technology.

Table 5 presents the results for implementing measures to improve patient care according to the medical specialty (here shown for medical disciplines with at least 20 providers and the seven most frequently implemented measures). As displayed, ophthalmologists (68%, 40/59) and gynecologists (65.4%, 123/188) were most likely to implement any measure to improve patient care. In contrast, the lowest percentages were calculated for psychiatrists (38%, 29/77) and pediatrics (40%, 16/40). Thereby, the mean number of implemented measures (mean 3.06, SD 2.29) was calculated to range between 1.81 (SD 1.05) for pediatrics and 4.29 (SD 3.05) for urologists, respectively. The association between the two variables could be shown to be marginally significant (Spearman P=.47 and P=.07, respectively).

Table 5. An overview of the seven most relevant measures that were implemented to improve patient care because of online ratings according to the medical specialty (medical disciplines with n>20; N=2360) (see Multimedia Appendix 2 for a complete overview).

Medical discipline	Any measure implemented, n (%)	Mean measures, mean (SD)	Measure, n (%)[a]						
			M1	M2	M3	M4	M5	M6	M7
Ophthalmology	40 (67.80)	3.48 (2.61)	20 (33.90)	23 (38.98)	17 (28.81)	13 (22.03)	11 (18.64)	13 (22.03)	6 (10.17)
Gynecology/obstetrics	123 (65.43)	3.29 (2.28)	58 (30.85)	59 (31.38)	57 (30.32)	26 (13.83)	29 (15.43)	26 (13.83)	18 (9.57)
Physical and rehabilitative medicine	15 (65.22)	2.43 (1.83)	7 (30.43)	3 (13.04)	6 (26.09)	2 (8.70)	1 (4.35)	2 (8.70)	1 (4.35)
Otorhinolaryngology (ENT[b])	59 (62.11)	3.44 (2.42)	38 (40.00)	25 (26.32)	28 (29.47)	10 (10.53)	11 (11.58)	13 (13.68)	16 (16.84)
Neurosurgery	17 (60.71)	3.07 (1.62)	12 (42.86)	10 (35.71)	6 (21.43)	1 (3.57)	4 (14.29)	2 (7.14)	2 (7.14)
Surgery/orthopedists	140 (60.61)	3.02 (2.27)	75 (32.47)	70 (30.30)	70 (30.30)	30 (12.99)	29 (12.55)	29 (12.55)	20 (8.66)
Oral and maxillofacial surgery	21 (60.00)	3.45 (1.21)	10 (28.57)	13 (37.14)	9 (25.71)	6 (17.14)	3 (8.57)	2 (5.71)	6 (17.14)
Urology	38 (58.46)	4.29 (3.05)	25 (38.46)	23 (35.38)	22 (33.85)	8 (12.31)	10 (15.38)	15 (23.08)	6 (9.23)
Dentistry	350 (57.47)	3.12 (2.45)	197 (32.35)	127 (20.85)	113 (18.56)	84 (13.79)	77 (12.64)	61 (10.02)	60 (9.85)
Dermatology and sexually transmitted diseases	41 (55.41)	2.79 (1.91)	24 (32.43)	21 (28.38)	14 (18.92)	4 (5.41)	8 (10.81)	8 (10.81)	3 (4.05)
Internal medicine (specialist)	45 (53.57)	2.93 (2.10)	24 (28.57)	26 (30.95)	20 (23.81)	7 (8.33)	6 (7.14)	9 (10.71)	9 (10.71)
General medicine	158 (52.32)	2.94 (2.08)	81 (26.82)	66 (21.85)	67 (22.19)	34 (11.26)	31 (10.26)	31 (10.26)	28 (9.27)
Psychosomatic medicine and psychotherapy	18 (50.00)	3.22 (2.44)	9 (25.00)	7 (19.44)	3 (8.33)	6 (16.67)	2 (5.56)	3 (8.33)	4 (11.11)
Internal medicine (GP[c])	40 (45.45)	3.10 (2.15)	21 (23.86)	22 (25.00)	22 (25.00)	5 (5.68)	10 (11.36)	8 (9.09)	4 (4.55)
Alternative practitioner	120 (42.25)	2.70 (2.14)	49 (17.25)	38 (13.38)	28 (9.86)	19 (6.69)	10 (3.52)	7 (2.46)	17 (5.99)
Pediatrics and adolescent medicine	16 (40.00)	1.81 (1.05)	10 (25.00)	6 (15.00)	3 (7.50)	4 (10.00)	1 (2.50)	1 (2.50)	0 (0.00)
Psychiatry and psychotherapy	29 (37.66)	2.61 (1.64)	16 (20.78)	13 (16.88)	10 (12.99)	6 (7.79)	5 (6.49)	0 (0.00)	3 (3.90)
Total	1290 (54.66)	3.06 (2.29)	679 (28.77)	557 (23.60)	501 (21.23)	266 (11.27)	245 (10.38)	231 (9.79)	200 (8.47)

[a] M1=Improvement of the communication with patients; M2=improve appointment scheduling process; M3=change office workflow; M4=improvement of the waiting room equipment; M5=training of the staff; M6=reassigning staff responsibilities; M7=investments in new technologies/equipment.

[b] ENT: ear, nose, and throat.

[c] GP: general practitioner.

XSL•FO

RenderX

Table 6. Multivariate regression analyses, including adjusted odds ratio (AOR), 95% confidence interval (CI), and *P* values, of the association between the implementation of measures to increase patient satisfaction because of online ratings (both scaled-survey ratings and narrative comments) and independent variables.

Characteristics	Model 1[a] AOR (95% CI)	*P* value	Model 2[b] AOR (95% CI)	*P* value	Model 3[c] AOR (95% CI)	*P* value
Age (years)		.57		.71		.85
<35 (ref)						
36-45	0.76 (0.49-1.19)	.23	0.77 (0.49-1.20)	.24	0.85 (0.53-1.36)	.50
46-55	0.77 (0.50-1.19)	.24	0.80 (0.52-1.25)	.33	0.91 (0.57-1.45)	.69
56-65	0.75 (0.47-1.19)	.22	0.79 (0.50-1.25)	.32	0.90 (0.55-1.47)	.67
≥66	0.57 (0.29-1.10)	.10	0.63 (0.32-1.23)	.18	0.70 (0.35-1.42)	.32
Gender						
Male (ref)						
Female	1.21 (0.99-1.48)	.07	1.16 (0.95-1.43)	.15	1.09 (0.88-1.36)	.43
Marital status		.38		.43		.55
Married (ref)						
Widowed	1.94 (0.60-5.46)	.21	1.99 (0.71-5.59)	.20	1.32 (0.46-3.80)	.61
Single	1.14 (0.86-1.50)	.37	1.13 (0.85-1.49)	.40	1.18 (0.88-1.59)	.27
Divorced	1.20 (0.85-1.69)	.31	1.15 (0.81-1.62)	.44	1.20 (0.83-1.74)	.34
Internet use		.70		.77		.21
Several times a day (ref)						
Once a day	1.06 (0.72-1.56)	.77	1.13 (0.76-1.66)	.55	1.39 (0.92-2.10)	.12
Less than once a day	0.82 (0.50-1.35)	.44	0.91 (0.55-1.50)	.71	1.29 (0.75-2.24)	.36
Medical specialty		<.001		<.001		<.001
General practitioner (ref)						
Specialist	1.31 (1.04-1.66)	.03	1.16 (0.91-1.48)	.23	1.17 (0.91-1.51)	.23
Others	0.77 (0.55-1.08)	.13	0.68 (0.48-0.96)	.03	0.59 (0.41-0.85)	.004
Jameda service product						
Basic product (ref)						
Any service product (eg, gold, silver, platinum)			1.61 (1.32-1.96)	<.001	1.13 (0.90-1.41)	.29
Use of physician-rating websites (frequency)						<.001
At least once per day (ref)						
Several times a week					1.39 (0.86-2.23)	.18
Once per week					0.98 (0.63-1.51)	.91
Once per month					0.59 (0.38-0.90)	.01
Less frequently					0.39 (0.25-0.61)	<.001
Never					0.18 (0.09-0.34)	<.001
Appraisal of the trustworthiness of scaled-rating results						.02
Not at all trustworthy (ref)						
Not trustworthy					1.68 (1.01-2.84)	.06
More or less trustworthy					1.97 (1.17-3.33)	.01
Somewhat trustworthy					1.93 (1.12-3.31)	.02
Very trustworthy					1.16 (0.59-2.26)	.67

Characteristics	Model 1[a]		Model 2[b]		Model 3[c]	
	AOR (95% CI)	P value	AOR (95% CI)	P value	AOR (95% CI)	P value
Appraisal of the trustworthiness of narrative comments						.09
Not at all trustworthy (ref)						
Not trustworthy					2.11 (1.21-3.69)	.009
More or less trustworthy					1.89 (1.08-3.28)	.03
Somewhat trustworthy					1.97 (1.13-3.44)	.02
Very trustworthy					2.26 (1.22-4.19)	.01

[a] Model 1: Adjusted for demographics (age, gender, marital status, Internet use, medical specialty) (χ^2_{12}=28,891, P=.004, Nagelkerke R^2=.019).

[b] Model 2: Adjusted for demographics, jameda service product (χ^2_{13}=50,980, P<.001, Nagelkerke R^2=.034).

[c] Model 3: Adjusted for demographics, jameda service product, use of physician-rating websites, appraisal of the trustworthiness of scaled-rating results/narrative comments (χ^2_{26}=251,463, P<.001, Nagelkerke R^2=.160).

Multilevel logistic regression models were performed to ascertain the effects of providers' demographics as well as report card-related variables on the likelihood that providers implemented measures to improve patient care (Table 6). Thereby, the dependent variable indicates whether a health care provider had implemented any measure(s) to improve patient care (yes/no).

All models revealed medical specialty to be a significantly associated predictor (P<.001 each). Thereby, the higher odds for specialists were proven to be statistically significant in our baseline model (AOR 1.31, 95% CI 1.04-1.66, P=.03). In contrast, the lower odds for other providers were proven to be statistically significant in our more comprehensive models and ranged between AOR 0.59 (95% CI 0.41-0.85, P<.001) and AOR 0.68 (95% CI 0.48-0.96, P=.03). In addition, the frequency of report card use (P<.001) and the appraisal of the trustworthiness of scaled-survey ratings (P=.02) were determined to be significantly associated predictors. Further regression analyses were run to determine whether the results differed when related only to implemented measures because of scaled-survey ratings or narrative comments (see Multimedia Appendices 3 and). As presented, the results could be shown to be very robust in terms of significantly associated predictors.

Discussion

The aim of this study was to examine which health care providers use online rating websites and for what purposes, as well as to investigate whether health care providers use online ratings to improve their quality of care. Therefore, we conducted an online-based cross-sectional study by surveying 2360 physicians and other health care providers. Our surveyed sample is slightly younger and contains a lower percentage of female respondents than providers who are registered on jameda or those who work in the German outpatient setting, respectively [24]. However, the distribution of our surveyed physicians across Germany as well as the medical specialties is similar to all physicians in the German outpatient setting (see Multimedia Appendix 5). Nevertheless, our sample may not be representative of all providers in the German outpatient setting despite the similarities in its characteristics. First, compared to the German outpatient setting, approximately 8% (25,000/304,818; 147,948 physicians, 69,370 dentists, and 87,500 other providers) of all providers were invited to participate in this survey, of whom slightly more than 9% completed the survey. Second, providers may differ in further characteristics among those who are registered on jameda and those who are not. For example, those less interested in online-related topics (eg, those without a practice website), or PRWs in general, are less likely to be represented by our results.

Our results show that more than half of the surveyed providers (55%) use online ratings to develop measures to increase the quality of patient care. This result is lower than the findings by Friedberg et al [25], who reported that 83% of interviewed physician group leaders used patient experience to improve the performance in the US outpatient setting. Furthermore, Smith and colleagues [20] showed an increase in the percentage of providers that implemented at least one of 22 possible diabetes improvement interventions from 75% to 94% after the onset of public reporting. The study approach here was quite similar because in both studies a literature-based list of possible interventions was presented. Thereby, the mean number of implemented measures to increase patient care in our study (mean 3.1, SD 2.3) was lower than in the study by Smith et al (mean 8.7, SD 4.5) [20].

The most widely implemented quality measures in our study are related to the communication with patients (28.77%, 679/2360), the appointment scheduling process (23.60%, 557/2360), and the office workflow (21.23%, 501/2360). These findings are in line with the results of Friedberg et al [25], who identified the change in the office workflow (eg, procedures for handling test results or incoming mail) as the most widely implemented initiative (70%), much more common even than in our study. The improvement of the appointment scheduling process was similarly reported by 27% of providers. Other common implemented quality measures in the Friedberg et al study (eg, training nonclinicians: 57%; reassigning staff responsibilities: 45%; or hiring or firing clinicians or staff: 36%) were higher in the US studies. Nevertheless, staff-related measures were implemented by approximately 16% of all respondents in our survey and thus do also play a significant

role in the German context. Although 10.38% (245/2360) of the respondents have invested in further training of staff, 5.89% (139/2360) have recruited additional staff, and 3.31% (78/2360) have dismissed staff as a consequence of the online ratings.

Another study showed the adoption of guidelines (87%) and the introduction of patient reminders (82%) to be the most frequently introduced quality measures [11]. Results from our study are far below those percentages and were shown to range between 4% (introduction of guidelines) and 8% (introduction of patient reminders). It is likely that differences in the surveyed samples (ie, individual physicians vs physician group leaders) may account for some of the difference. For example, physician groups in the study of Friedberg et al [25] represent all physicians who work at one or more practices with at least one group-level manager. This manager coordinates the correspondence with health plans and controls the performance of physician groups [25]. Furthermore, successful physician groups share a unified physician-led network, an infrastructure to support group performance, and incentives to engage individual physicians toward group goals [26]. Consequently, these physician groups might be more active in implementing quality measures due to an established infrastructure and developed competencies. In contrast, our sample comprises individual physicians who may not have the infrastructure or the competencies to introduce quality measures on such a large scale. Another study does not present detailed information about the frequency of implemented measures, but states several measures that have been widely implemented, such as the use of phone and mail patient reminders about existing upcoming appointments or to schedule follow-up tests, guideline-based reminders for providers at appointments about services that their patients should receive, or the adoption of care guidelines [20].

As shown, most providers use the websites for reading comments for their individual practice (87.08%, 2055/2360). Of those, 56.69% (1165/2055) stated that they had implemented quality measures because of the online ratings. This also means that 43.31% (890/2055) of those providers did not implement any quality measure(s), possibly because of a lack of time or trust in the online ratings. Others might find it challenging to develop quality measures at all because anonymous ratings sometimes do not provide enough information to learn about quality deficits [27]. Reading other physicians' ratings out of interest (49%) provides an opportunity for physicians to draw comparisons. This may also have a positive impact on the overall quality of care because some physicians are becoming engaged in implementing quality measures in order to perform better than their colleagues. According to our numbers, 12.08% (285/2360) of the respondents use online ratings for referring patients to other providers. This result exceeds those of other studies, which determined that physicians do not use publicly available quality information [28] or change their referral practices due to public reports [29,30]. The remaining providers might not use online ratings for referral decisions because subjective patient ratings do not accurately display the quality of health care providers [31,32].

We were able to show further that scaled-rating systems seem to have a greater impact on implementing quality measures than narrative comments in absolute terms. This might be because scaled ratings systems offer a more structural approach to rate health care providers in comparison to narrative comments [4]. Therefore, physician-rating systems provide standardized categories that more promptly enable the indication of quality deficits. However, even though narrative comments provide a more complete picture of the patients' experiences with the health care provider [6], they are more complicated and time-consuming to analyze. For example, it might be easier to determine deficits such as long waiting times for an appointment out of scaled-rating systems. This is because most scaled-rating systems contain this issue, but only 13% of narrative comments address this topic [33].

Finally, even though our results could demonstrate positive effects of public reporting on patient care, several unintended effects should be regarded for the special case of PRWs. For example, public reporting might reduce the access to care because health care providers tend to avoid high-risk patients and prefer low-risk patients ("cherry picking" or "cream skimming") [29,34]. This effect is likely to be even greater in the case of online rating websites compared to traditional reporting instruments (see previous) because no risk-adjustment of online ratings is carried out [3]. Furthermore, Lindenauer et al [35] mentioned the neglect of more important aspects of care. So far, all PRWs in Germany, and most in the United States, solely contain patient satisfaction as a performance measure. Results for guideline-based treatment (eg, for chronic conditions such as type 2 diabetes) are not presented on many report cards, meaning ratings for physicians are only based on patient satisfaction scores. From the perspective of a physician, this might be challenging because most (narrative) ratings concern the friendliness and general impression of a physician rather than the medical outcomes [33]. For instance, a physician might neglect the patients' desire for prescribed antibiotics in the case of a nonbacterial infection. Even though this behavior is correct from a medical point of view, patients might experience this differently and leave a low rating for a physician. Lindenauer et al [35] also stated the unintended risk of an increase in performance rates through coding and documentation ("gaming"). So far, this risk might be of lower importance on most reporting websites because no secondary data are used (which might be gamed). However, online rating websites are vulnerable to other gaming aspects [3,27]. For example, physicians may rate themselves, leave a low rating for their colleagues, or ask the more satisfied patients to leave a rating [27].

Our results demonstrate that health care providers in Germany use PRWs for reading comments for their individual practice. More than half of those providers stated that they had implemented quality measures in order to improve patient care. The most widely implemented quality measures are related to communication with patients, the appointment scheduling process, and office workflow. In addition, staff-related measures also play a significant role in the German context. Scaled-survey results seem to have a greater impact on deriving measures than narrative comments. This might be because the latter are more complicated and time-consuming to analyze. Thus, our results

confirm findings from the United States showing that online rating websites may have a positive impact on patient care.

There are some limitations that have to be considered when interpreting the results of this study. First, this study applied a cross-sectional design. Thus, we were able to identify an association between exposure and outcomes, but could not infer cause and effect. Next, even though our sample is representative of those physicians who are registered on the largest German PRW, representation of all practicing health care providers in the outpatient setting in Germany is not achievable (eg, those less interested in online-related topics, those who do not want to deal with this topic in general). In addition, some participants in our sample might be more familiar with Internet-related topics, which may account for some of the high awareness and usage levels of rating websites. Furthermore, the participants were recruited online by the provider of jameda. This means that we only surveyed providers who are active on jameda. Thus, our findings cannot be generalized for providers on other rating websites. For our regression analysis, we used whether providers have implemented any quality measure or not as the binary dependent variable. As shown previously, there were severe differences in the extent of the implementation between the individual measures. Therefore, results might be different if we used the implementation of the individual measures as the dependent variable.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 519KB - jmir_v18i9e254_app1.pdf]

Multimedia Appendix 2

Overview of all measures that were implemented to improve patient care because of online ratings according to the medical specialty (medical disciplines with N>20) (N=2360).

[PDF File (Adobe PDF File), 51KB - jmir_v18i9e254_app2.pdf]

Multimedia Appendix 3

Multivariate regression analyses; adjusted odds ratio (OR), 95% confidence interval (CI), and *P* value of the association between the implementation of measures to increase patient satisfaction because of scaled survey online ratings and independent variables.

[PDF File (Adobe PDF File), 30KB - jmir_v18i9e254_app3.pdf]

Multimedia Appendix 4

Multivariate regression analyses; adjusted odds ratio (OR), 95% confidence interval (CI), and *P* value of the association between the implementation of measures to increase patient satisfaction because of narrative comments and independent variables.

[PDF File (Adobe PDF File), 30KB - jmir_v18i9e254_app4.pdf]

Multimedia Appendix 5

Comparison of the study sample with the registered providers on jameda and providers in the outpatient sector in Germany.

[PDF File (Adobe PDF File), 21KB - jmir_v18i9e254_app5.pdf]

References

1. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
2. 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]
3. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
4. 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]

5. Schlesinger M, Grob R, Shaller D, Martino SC, Parker AM, Finucane ML, et al. Taking patients' narratives about clinicians from anecdote to science. N Engl J Med 2015 Aug 13;373(7):675-679. [doi: 10.1056/NEJMsb1502361] [Medline: 26267629]

6. Brennan PF. Patient satisfaction and normative decision theory. J Am Med Inform Assoc 1995;2(4):250-259 [FREE Full text] [Medline: 7583649]

7. Shaffer VA, Zikmund-Fisher BJ. All stories are not alike: a purpose-, content-, and valence-based taxonomy of patient narratives in decision aids. Med Decis Making 2013 Jan;33(1):4-13. [doi: 10.1177/0272989X12463266] [Medline: 23065418]

8. Ferstl K, Bruskiewicz K. Self-other agreement and cognitive reactions to multirater feedback. 2000 Aug 04 Presented at: 15th annual conference of the Society of Industrial and Organizational Feedback; Apr 14-16, 2000; New Orleans.

9. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]

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. Lamb GC, Smith MA, Weeks WB, Queram C. Publicly reported quality-of-care measures influenced Wisconsin physician groups to improve performance. Health Aff (Millwood) 2013 Mar;32(3):536-543 [FREE Full text] [doi: 10.1377/hlthaff.2012.1275] [Medline: 23459733]

12. Jha AK, Epstein AM. The predictive accuracy of the New York State coronary artery bypass surgery report-card system. Health Aff (Millwood) 2006;25(3):844-855 [FREE Full text] [doi: 10.1377/hlthaff.25.3.844] [Medline: 16684751]

13. Mukamel DB, Spector WD, Zinn J, Weimer DL, Ahn R. Changes in clinical and hotel expenditures following publication of the nursing home compare report card. Med Care 2010 Oct;48(10):869-874. [doi: 10.1097/MLR.0b013e3181eaf6e1] [Medline: 20733531]

14. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M. Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy). Evid Rep Technol Assess (Full Rep) 2012 Jul(208.5):1-645. [Medline: 24422977]

15. 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]

16. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I30-138. [Medline: 12544814]

17. Fung CH, Lim Y, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008 Jan 15;148(2):111-123. [Medline: 18195336]

18. Brutus S. Words versus numbers: a theoretical exploration of giving and receiving narrative comments in performance appraisal. Hum Resour Manage R 2010;20(2):144-157.

19. Werner RM, Konetzka RT, Kim MM. Quality improvement under nursing home compare: the association between changes in process and outcome measures. Med Care 2013 Jul;51(7):582-588 [FREE Full text] [doi: 10.1097/MLR.0b013e31828dbae4] [Medline: 23756645]

20. Smith MA, Wright A, Queram C, Lamb GC. Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. Health Aff (Millwood) 2012 Mar;31(3):570-577 [FREE Full text] [doi: 10.1377/hlthaff.2011.0853] [Medline: 22392668]

21. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]

22. Halling F, Halling C, Pohl-Dernick K. Arztbewertungsportale aus zahnaerztlicher Perspektive: physician rating websites from the perspective of dentists. Deutsche Zahnaerztliche Zeitschrift 2012;67(11):716-723.

23. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]

24. Bundesärztekammer. (Muster-)Weiterbildungsordnung 2003. 2013. URL: http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Weiterbildung/MWBO.pdf [accessed 2015-12-03] [WebCite Cache ID 6gsykf2IC]

25. Friedberg MW, SteelFisher GK, Karp M, Schneider EC. Physician groups' use of data from patient experience surveys. J Gen Intern Med 2011 May;26(5):498-504 [FREE Full text] [doi: 10.1007/s11606-010-1597-1] [Medline: 21161419]

26. Advisory Board. The high-performance medical group: from aggregations of employed practices to an integrated clinical enterprise URL: https://www.advisory.com/research/health-care-advisory-board/studies/2011/the-high-performance-medical-group [accessed 2016-04-19] [WebCite Cache ID 6gsyyBAI6]

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. Hermeling P, Geraedts M. Kennen und nutzen Ärzte den strukturierten Qualitätsbericht? Das Gesundheitswesen 2013;75(3):155-159.

29. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. N Engl J Med 1996 Jul 25;335(4):251-256. [doi: 10.1056/NEJM199607253350406] [Medline: 8657242]

XSL•FO
RenderX

30.   Hannan EL, Stone CC, Biddle TL, DeBuono BA. Public release of cardiac surgery outcomes data in New York: what do New York state cardiologists think of it? Am Heart J 1997 Dec;134(6):1120-1128. [Medline: 9424074]

31.   Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. BMC Health Serv Res 2015;15(1):414 [FREE Full text] [doi: 10.1186/s12913-015-1051-5] [Medline: 26404452]

32.   Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]

33.   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]

34.   Moscucci M, Eagle KA, Share D, Smith D, De Franco AC, O'Donnell M, et al. Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases. J Am Coll Cardiol 2005 Jun 7;45(11):1759-1765 [FREE Full text] [doi: 10.1016/j.jacc.2005.01.055] [Medline: 15936602]

35.   Lindenauer PK, Lagu T, Ross JS, Pekow PS, Shatz A, Hannon N, et al. Attitudes of hospital leaders toward publicly reported measures of health care quality. JAMA Intern Med 2014 Dec;174(12):1904-1911 [FREE Full text] [doi: 10.1001/jamainternmed.2014.5161] [Medline: 25286316]

Abbreviations

AOR: adjusted odds ratio
GP: general practitioner
PRW: physician-rating website

Corrigenda and Addenda

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

Iris M Kanera[1], MSc; Roy A Willems[1], MSc; Catherine AW 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 M Kanera, MSc
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
Chiba 3.21
Valkenburgerweg 177
Heerlen, Heerlen, 6401 DL
Netherlands
Phone: 31 455762448
Email: iris.kanera@ou.nl

Related Article:

Correction of: http://www.jmir.org/2016/8/e229

The authors of "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) have overlooked errors in the abstract during the proofreading process. The method section of the abstract was missing, the results section was displayed under 'methods', and the conclusion was displayed twice, under 'results' as well as under 'conclusions'. The methods section in the abstract need to be the following:

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.

This section needs to be added under 'methods', the results needs to be displayed under 'results', and the conclusion needs to displayed under 'conclusions'.

Furthermore, the degrees of the authors were corrected. As per AMA Style Guide, only the highest degrees should be presented, thus we deleted "BSc" from the authors Kanera and Gijsen. In addition, the degree of author Brigitte CM Gijsen (Master of Science in Public Health) was corrected from "MPH" to "MSc" in order to provide uniformity in degrees.

The affiliations of the authors have not been changed. This correction has been made in the online version of the paper on the JMIR website on September 13, 2016, together with publishing this corrigendum.

A correction notice has been sent to PubMed, and the publication was resubmitted to Pubmed Central and other full-text repositories.

XSL•FO

RenderX

XSL•FO

RenderX