Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 15 (2013), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Corrigenda and Addendas

Original Paper

Partnership for Health-2, A Web-Based Versus Print Smoking Cessation Intervention for Childhood and Young Adult Cancer Survivors: Randomized Comparative Effectiveness Study

Karen M Emmons[1,2], PhD; Elaine Puleo[3], PhD; Kim Sprunck-Harrild[1], MPH, MSW; Jennifer Ford[4], PhD; Jamie S Ostroff[4], PhD; David Hodgson[5], MD; Mark Greenberg[6], MD; Lisa Diller[1,7], MD; Janet de Moor[8], PhD; Vida Tyc[9], PhD

[1]Dana-Farber Cancer Institute, Boston, MA, United States
[2]Harvard School of Public Health, Boston, MA, United States
[3]University of Massachusetts, Amherst, MA, United States
[4]Memorial Sloan Kettering Cancer Center, New York, NY, United States
[5]Princess Margaret Hospital, Toronto, ON, Canada
[6]Hospital for Sick Children, Toronto, ON, Canada
[7]Harvard Medical School, Boston, MA, United States
[8]Ohio State University College of Public Health, Columbus, OH, United States
[9]St. Judes Children's Research Hospital, Memphis, TN, United States

Corresponding Author:
Karen M Emmons, PhD
Dana-Farber Cancer Institute
450 Brookline Ave, LW601
Boston, MA, 02215
United States
Phone: 1 617 632 2188
Fax: 1 617 632 5690
Email: Karen_M_Emmons@dfci.harvard.edu

Abstract

Background: Smoking among cancer survivors increases the risk of late effects and second cancers. This article reports on Partnership for Health-2 (PFH-2)—an effort to develop an effective and scalable version of Partnership for Health (PFH), which was a previously tested peer-delivered telephone counseling program that doubled smoking cessation rates among childhood cancer survivors who smoke.

Objective: This paper presents results from a randomized controlled trial evaluating the effectiveness of PFH-2 in targeted and tailored Web-based versus print formats. The overall goal was to determine whether the intervention outcomes in these self-guided scalable formats approximate what was found in a more intensive telephone counseling program.

Methods: This study was a randomized controlled trial with a 15-month follow-up that included 374 smokers who were survivors of childhood or young adult cancers, recruited from five survivorship clinics. Participants were randomly assigned to a Web-based or print format of the PFH intervention; all had access to free pharmacotherapy. The website was designed to provide new content at each log-on, and a peer counselor moderated a forum/chat feature. The primary outcome was smoking status at 15 months post randomization.

Results: In total, 58.3% (77/132) of Web participants logged on at least once (mean visits 3.25). Using multiple imputation methods for missing data, there were similar rates of cessation in the two arms (print: 20/128, 15.6%; Web: 33/201, 6.4%), and no differences in quit attempts or readiness to quit. The quit rates were equivalent to those found in our previous telephone counseling intervention. There were high rates of satisfaction with both of the PFH-2 interventions.

Conclusions: The print and Web formats yielded equivalent levels of success to those found with our telephone-delivered intervention and are comparable to other Internet treatment studies. This study provides important options for survivorship programs that may not have resources for interpersonal forms of cessation counseling. Efforts to increase patient use of the interventions may result in higher cessation rates.

Trial Registration: Clinicaltrials.gov NCT00588107; http://clinicaltrials.gov/ct2/show/NCT00588107 (Archived by WebCite at http://www.webcitation.org/6K1gJtFEC).

(J Med Internet Res 2013;15(11):e218) doi:10.2196/jmir.2533

KEYWORDS

cancer survivors; smoking cessation; cancer prevention

Introduction

Remarkable improvements in the treatment and long-term survival of childhood and young adult cancer survivors have resulted in prevention of adverse late effects and second primary cancers being a key part of survivorship care [1-5]. Smoking rates among this population are substantial [6-8]. In the Childhood Cancer Survivor Study, the largest US cohort of childhood cancer survivors, 28% of survivors reported having smoked at least 100 cigarettes in their lifetime and 17% reported current cigarette smoking [6]. In the British Childhood Cancer Survivor Study, a population-based cohort of 17,981 survivors, 20.0% were current regular smokers and 29.8% were regular smokers but no longer are [8].

This team's previous work has demonstrated the efficacy of Partnership for Health (PFH), a survivor-focused peer-delivered telephone counseling intervention for smoking cessation. PFH led to a doubling in quit rates compared with usual care and the intervention effect was sustained over 2-5 years of follow-up [9,10]. The connection that the peer-delivered intervention provided between survivors was an important way to engage participants in the intervention. However, it is challenging to scale interventions that include ongoing counseling.

A recent evaluation of existing infrastructure for delivering smoking cessation services in the context of survivorship programs revealed relatively few resources; only 3% of programs assessed smoking status at every visit, as recommended by Public Health Service (PHS) guidelines, and only 25% offered cessation services [11]. Further, cancer survivors are quite geographically dispersed and thus, effective interventions are needed that can be scaled easily and delivered remotely regardless of survivors' location. The present study focused on the adaptation of the PFH peer-delivered intervention for a Web- and print-based format as a way to increase the intervention's dissemination potential and sustainability. We selected Web- and print-based interventions because of the relatively high penetration of Internet access in the target age group [12] and because these formats could be integrated easily into standard practice at survivorship programs across the country, compared with telephone-based interventions.

This paper presents results from a randomized control trial evaluating the effectiveness of Partnership for Health-2 (PFH-2) in targeted and tailored Web-based versus print formats. The overall goal was to determine the outcomes for these self-guided scalable formats and whether they approximate what was found in a more intensive telephone counseling program. Our hypothesis was that the Web format would yield higher quit rates than print and would be similar in effectiveness to that found in the original PFH counseling intervention.

Methods

Setting

PFH-2 was conducted in collaboration with five cancer centers in the United States and Canada, with Institutional Review Board (IRB) approval at all sites. The study was also advertised on childhood and young adult survivorship websites. Eligibility included: diagnosed with cancer before age 35, currently between ages 18-55, completed cancer treatment for ≥2 years, mentally able to provide informed consent, reachable by telephone, able to speak English, and a current smoker (defined as smoking within the previous 30 days). Participants were informed that the study was examining different ways to deliver health information, including information about tobacco use, to survivors. They were not required to be interested in smoking cessation in order to participate. Baseline data collection began on December 2005 and follow-up data collection ended in October 2009.

A preliminary screen for eligiblity was performed at each site via medical record review or brief telephone screening. After consent for sharing contact information was obtained, contact information was forwarded to the survey team to verify eligibility, obtain informed consent for study participation, and administer the baseline telephone survey.

PFH-2 Study Design

PFH-2 was a stratified randomized controlled trial with cancer center as strata. The goal was to test two scalable intervention formats for smoking cessation among childhood and young adult survivors, developed from an evidence-based intervention, and to determine whether the Web intervention, with the advantages that an interactive website has to offer, would outperform tailored and targeted print materials. Participants were randomized to one of two intervention conditions within strata, in a 5:3 randomization scheme: (1) PFH-2 Print Materials Intervention, or (2) PFH-2 Web Intervention. The random allocation sequence was generated by the study biostatistician. Randomization was done by the survey team and supervised by the biostatistician, following completion of the baseline survey. Study design is provided in detail elsewhere [13]. Intervention goals for both conditions included: (1) assess and enhance motivation to change, (2) address ambivalence about behavior change, (3) provide social support, (4) assess and build self-efficacy, (5) increase awareness of risks, (6) identify and address barriers to change, and (7) address nicotine dependence. Both conditions included: (1) a letter encouraging smoking cessation from the site oncologist, developed based on the principles of the National Cancer Institute's "5 A's" smoking counseling guidelines, (2) free pharmacotherapy for participants and spouses/significant others who want to quit, and (3) tailored

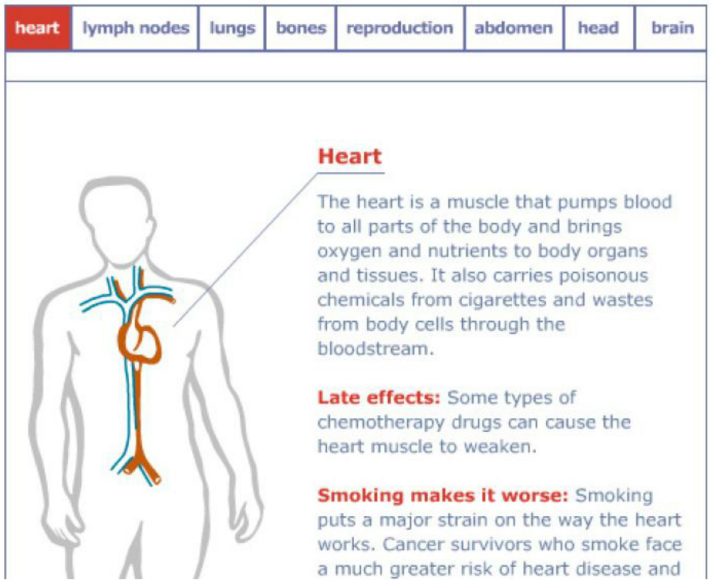
and targeted self-help content (Web or print) addressing participant-specific barriers to change and other survivor-related topics of interest. The intervention period was 6 months and a follow-up survey was completed by telephone at 15 months after randomization.

PFH-2 Print Materials Intervention

The Print Materials arm received tailored and targeted self-help materials that were developed for the peer counselor condition in PFH-1. The materials were organized into a series of manuals, based on readiness to change [14], that addressed participant-specific barriers to change and other survivor-related topics of interest (eg, addressing depression, handling stress, managing weight). The manuals were designed to be as interactive as possible, with worksheets and opportunities for personalizing the content included throughout. Testimonials and stories of other survivors' experiences were used to provide the survivor-to-survivor connection. Both PFH-2 conditions included other key features of the original PFH peer-delivered intervention, including a letter encouraging smoking cessation from the site oncologist and free pharmacotherapy (nicotine patch or Zyban) for themselves and any smoking partner/spouse who wished to quit. Those who were interested in pharmacotherapy contacted study staff and provided permission to contact their health care provider for approval using a fax-back form. We also sent the provider a copy of the pamphlet "Helping Smokers Quit: A guide for primary care clinicians" (US Department of Health and Human Service/Agency for Health Care Policy and Research) and a basic fact sheet about adult survivors of cancer. The pharmacotherapy was also sent to the provider and distributed to the patient by his or her office.

A baseline feedback report (BFR) was generated for each individual and sent between 5 and 10 business days following completion of the baseline survey. The BFR reflected information that patients gave on the baseline survey about their readiness to quit, perceived risk, nicotine dependence, and, based on these factors, which intervention manual to start with, drawing on approaches we have used previously [11,15]. The BFR also gave basic facts about cancer treatment and illustrated how cancer, cancer treatment, and smoking affects many of the same organs (see Figure 1). The Print Materials condition was designed to reflect as many of the realities as possible of how the intervention might be utilized when scaled to existing cancer survivorship clinics.

Figure 1. Illustration of the interaction of smoking and cancer treatment/late effects.



PFH-2 Web Intervention

Within 5-10 business days following completion of the baseline survey, those assigned to the Web intervention were sent log-on information. The Web intervention consisted of seven discrete tailored sessions designed to parallel the counseling sessions of the original PFH study and mirror the basic content of the PFH-2 print materials. The content was dynamically tailored, matching the participants' stage of readiness. Upon first log-on, participants saw their BFR, described above, on their home page. The home page also highlighted active components of the intervention that participants had not yet navigated. Patients saw refreshed content on the home page as they progressed from session to session. A peer counselor moderated the website's discussion forum and served as a resource for questions.

Participants who never logged in or stopped using the site received additional email prompts that highlighted content that survivors might find particularly important or engaging, along with biweekly emailed newsletters. Those who had not accessed the website within 11 weeks were sent a final letter along with the print materials to increase the likelihood of some exposure to the intervention content and to approximate likely approaches in a clinic setting. Participants had access to the website for 6

months regardless of their log-in status. For quality assurance purposes, several "test participants" were created and followed throughout the implementation period to identify glitches or issues with the Web system. These "participants" were not included in the tracking data and did not impact on intervention delivery in any way.

Measures

Sociodemographics and Medical History

Age, gender, race, ethnicity, marital status, education, height, weight, employment status, type of cancer, and cancer treatment were assessed.

Internet Access and Utilization

Internet access and utilization were measured with questions about whether participants owned a computer, whether they had access to the Internet at home and/or work, and how frequently they used a computer, email, and the Internet [16].

Smoking Behavior

"Smoking status" was assessed by self-reported assessment of smoking, even a puff, in the past 30 days. The bogus pipeline procedure was used to increase the accuracy of self-report, following standard implementation methods [17,18]. Specifically, participants were informed that they may be asked to provide a saliva sample to confirm their smoking status. This procedure has been found to improve the quality of self-reported smoking behavior. "Nicotine dependence" was assessed based on number of minutes after waking that participants smoked their first cigarette [19] (<30 minutes, nicotine dependent; ≥30 minutes, not nicotine dependent). "Quit attempts" were assessed by the number of quits in the previous 12 months with at least 24 hours abstinence. "Use of pharmacotherapy" was assessed with two questions about whether participants had ever used Zyban or nicotine replacement therapy to quit smoking.

Motivational Variables

"The Stages of Change Scale" assessed motivation to quit smoking [14], according to four categories: (1) precontemplation: not seriously thinking about quitting in the next 6 months; (2) contemplation: seriously thinking about quitting in the next 6 months; (3) preparation: intending to quit in the next month and have tried to quit in the past year; and (4) action: not currently smoking and quit within the past 6 months. "Self-efficacy" was assessed related to participants' level of confidence that they could quit smoking in the next 1 and 6 months [7].

Household and Workplace Smoking Policy

Participants were asked about rules regarding smoking at home and work (smoking unrestricted inside the building/house, limited to certain rooms, or forbidden inside the building/house).

Psychological Variables

"Cancer-related distress" was assessed with the Intrusive Thoughts subscale of the Impact of Events Scale (IES), which measures frequency with which thoughts regarding cancer recurrence intrude into consciousness [20]. "Perceived control" was assessed with a 3-item scale that measured the degree to which participants felt they could control physical side effects, future health, and chance of a cancer recurrence [21]. "Perceived vulnerability" was assessed with a question about the likelihood of experiencing serious health problems in the future [11,22]. "Depression" was measured with the 2-item Prime MD scale [23].

Contact With the Health Care System

Participants were asked if they had a regular health care provider and if they had been seen by their primary care physician or their oncologist in the past year.

Intervention Use

Use of the Web-based intervention was assessed using Web analytics. Use of the print intervention was assessed on the follow-up survey, with questions about the percentage of materials read and frequency of use.

Data Analysis

Original sample size calculations were adjusted due to the discovery during implementation that the participating survivorship programs' estimates of smokers were an over-estimate. Thus, we used a 5:3 allocation scheme to ensure an adequate sample size to evaluate the use of the Web-based intervention. With this approach, we had 71% power to detect a difference of 9% in quit rate between the Web and print groups. Baseline comparisons of patient characteristics between intervention arms were assessed. Depression level was the only variable of significance in this comparison and was, in addition to site, controlled for in all future analyses. All outcome analyses were conducted using multiple imputation methods based on the assumption of arbitrary missing patterns and thus used Markov chain Monte Carlo methods. This assumes multivariate normality to impute missing values (11.7%, 27/230 of the Web group and 12.5%, 18/144 of the print group were missing at follow-up). All analyses followed procedures described in the SAS OnlineDoc Version 8 for multiple imputation [24]. Logistic regression models were created to assess the impact of a priori determined predictor variables on the primary outcome—smoking status at follow-up. Variables that were significant in these models were considered potential variables to be entered into a multivariable logistic model predicting smoking status at follow-up. Any a priori determined variable (such as study site and depression level) were included in the final model selection regardless of bivariate significance. A parsimonious model that made clinical sense was developed by a process of forward and backward stepwise regression based on the individual variable significance in the model and the effect its removal had on other variable coefficients in the model. This was followed by assessing potential interaction effects, effect modifiers, and confounders. For the secondary outcome variables—quit attempts and readiness to quit smoking—similar model development occurred using polytomous logistic regression models with categorical outcome. All analyses were conducted in SAS Version 9.3.

Results

Participant Characteristics

We assessed eligibility among 4399 survivors; 4025 people were excluded (35.88%, 1444/4025 were not reached; the majority of the remainder were ineligible due to smoking status); 46.7% (374/801) of eligible survivors were enrolled in the study. In total, 88.0% (329/374) of enrollees completed the 15-month follow-up (see Figure 2). In Figure 2, "lost to follow-up" were subjects for which no contact information was available at the 15-month time period whereas "missing" were those subjects who did not respond to the survey after multiple contact attempts.

Mean age at enrollment was 32 years (SD 7.94) and at cancer diagnosis was 12 years (SD 8.06). The sample was evenly split by gender (51.3%, 192/374 male) and was predominantly white (86.4%, 323/374). In total, 36.1% (135/374) had a high school education or less and 29.7% (111/374) had at least a college degree. About half (176/374) were married or living with a partner and the majority (79.4%, 297/374) were employed in the last year. There were no differences between the two arms at baseline on any demographic or cancer-related variables (see Table 1).

A total of 59.9% (224/374) of the sample smoked less than one pack of cigarettes per day (range <1 to 60; mean 10/day). About half of the sample was nicotine dependent. The majority (63.1%, 236/374) were in preparation to quit smoking and 55.9% (209/374) screened positive for possible depression.

There were no differences in any baseline demographic variables among completers and drop-outs except employment status; drop-outs were more likely to be employed.

Primary Outcomes—Smoking Cessation

There were no significant differences between the two interventions arms in terms of smoking status at follow-up. At the final assessment, 16.5% of Web participants (22/132) and 15.5% of print participants (20/127) reported being abstinent for the previous 30 days (see Table 2).

Several factors were associated with smoking abstinence in multivariate analyses, including gender ($P \leq .04$, males less likely to have quit at follow-up) and cancer treatment factors (higher cessation rates were associated with a cancer diagnosis of leukemia [$P=.03$] or non-Hodgkin's lymphoma [$P=.01$], not having received radiation [$P=.01$], and having received surgery [$P=.01$]). Higher levels of perceived vulnerability for serious illness in the future was marginally associated with lower rates of cessation ($P=.08$ for more likely vulnerable) and less restrictive rules about smoking in the home were associated with significantly lower rates of cessation ($P=.05$).

An exploratory comparison of quit rates in the original PFH peer-delivered telephone intervention with the current Web and print condition suggests that the PFH-2 interventions attained equivalent levels of cessation. The original telephone-delivered intervention resulted in quit rates of 15% at 12-month follow-up. Although this was not a randomized comparison, these results do suggest that both Web- and print-based intervention methods developed specifically for cancer survivors do have a similar level of intervention impact to that found with a more intensive peer-counseling telephone intervention.

Secondary Outcomes—Quit Attempts and Readiness to Change

There were no significant differences between the two arms in terms of quit attempts (see Table 3). On average, 35.1% (91/259) of participants made no quit attempts, 37.8% (98/259) made limited (1-3) attempts, and 27.0% (70/259) made extensive (4+) quit attempts over the 15-month follow-up period.

Demographic factors associated with quit attempts in multivariable analyses included male gender ($P.001$ and $P=.06$ for limited and extensive quit attempts, respectively), having less education ($P=.01$ for limited quit attempts), using pharmacotherapy ($P<.001$), being a more frequent computer user ($P=.03$-$.06$), perceiving a moderate ($P=.03$) or high level of vulnerability ($P=.07$) to serious future illness (for limited quit attempts), and not being nicotine dependent ($P=.002$ for extensive attempts).

There were also no significant differences between the two arms in terms of impact on readiness to quit smoking. Only two variables, using pharmacotherapy and nicotine dependence, were associated with readiness to quit in multivariable analyses: using pharmacotherapy was associated with being in the action (OR 5.33, 95% CI 1.20-23.64) ($P=.03$) and preparation stage (OR 6.05, 95% CI 1.81-20.17) ($P=.01$) (vs precontemplator). Those who were not nicotine dependent were more likely to be in later stages at the follow-up (OR 4.27, 95% CI 1.21-12.95) ($P \leq .01$) and preparation versus precontemplation (OR 2.03, 95% CI 1.02-4.02) ($P=.006$).

Table 1. Baseline variables by treatment condition (n=329).

		Treatment condition				*P* value[a]
		Print (n=128)		Web (n=201)		
		n	%	n	%	
Demographics						
Gender (Female)		70	55.6	93	45.8	.19
Education						.85
	≤ high school or GED	43	34.1	75	36.9	
	Some college or training after college	50	39.7	63	31.0	
	College graduation	35	27.8	63	31.0	
Age (LS means)		33.59		32.50		.23
Employed, past year		97	77.0	159	78.3	.46
Race						.47
	White	111	88.1	173	85.2	
	Non-white	17	13.5	28	13.8	
Marital status						.71
	Married/living with partner	61	48.4	100	49.3	
	Never married and not living with partner	55	43.7	83	40.9	
	Divorced/no longer living with partner	12	9.5	18	8.9	
Cancer diagnosis						.79
	Leukemia	34	27.0	45	22.2	
	Hodgkins disease	21	16.7	40	19.7	
	CNS malignancy	15	11.9	17	8.4	
	Non-Hodgkins Lymphoma	6	4.8	14	6.9	
	Bone cancer	10	7.9	15	7.4	
	Other	42	33.3	70	34.5	
	Received radiation	81	64.3	122	60.1	.71
	Received chemotherapy	96	76.2	153	75.4	.91
	Received surgery	93	73.8	141	69.5	.53
Psychological variables						
Spent at least part of the day worried about getting cancer again in past week						.13
	Yes	22	17.5	24	11.8	
Impact of Events scale, Intrusive Thoughts subscale - categorized						.22
	None	47	37.3	78	38.4	
	Little	22	17.5	47	23.2	
	Some	29	23.0	35	17.2	
	More	30	23.8	41	20.2	
Screened positive for possible depression		85	67.5	100	49.3	.003
Perceived personal control (LS means)		10.88		11.76		.07
Perceived risk for serious health problems in the future						.61
	No chance/very unlikely/unlikely	29	23.0	41	20.2	
	Moderate chance	37	29.4	60	29.6	
	Likely	36	28.6	53	26.1	

			Treatment condition				*P* value[a]
			Print (n=128)		Web (n=201)		
			n	%	n	%	
		Very likely/certain to happen	25	19.8	44	21.7	
Health							
	General health						.67
		Excellent/very good	43	34.1	66	32.5	
		Good	53	42.1	77	37.9	
		Fair/poor	32	25.4	58	28.6	
	Stage of change						.86
		Precontemplation	15	11.9	33	16.3	
		Contemplation	34	27.0	39	19.2	
		Preparation	77	61.1	127	62.6	
Computer Use							
	Computer use at baseline						.79
		Daily	81	64.3	124	61.1	
		Less than daily, but monthly or more	20	15.9	43	21.2	
		Less than monthly or never	27	21.4	34	16.7	

[a]Controlling for site.

Figure 2. Recruitment and retention rates (Consolidated Standard of Reporting Trials Statement - CONSORT).

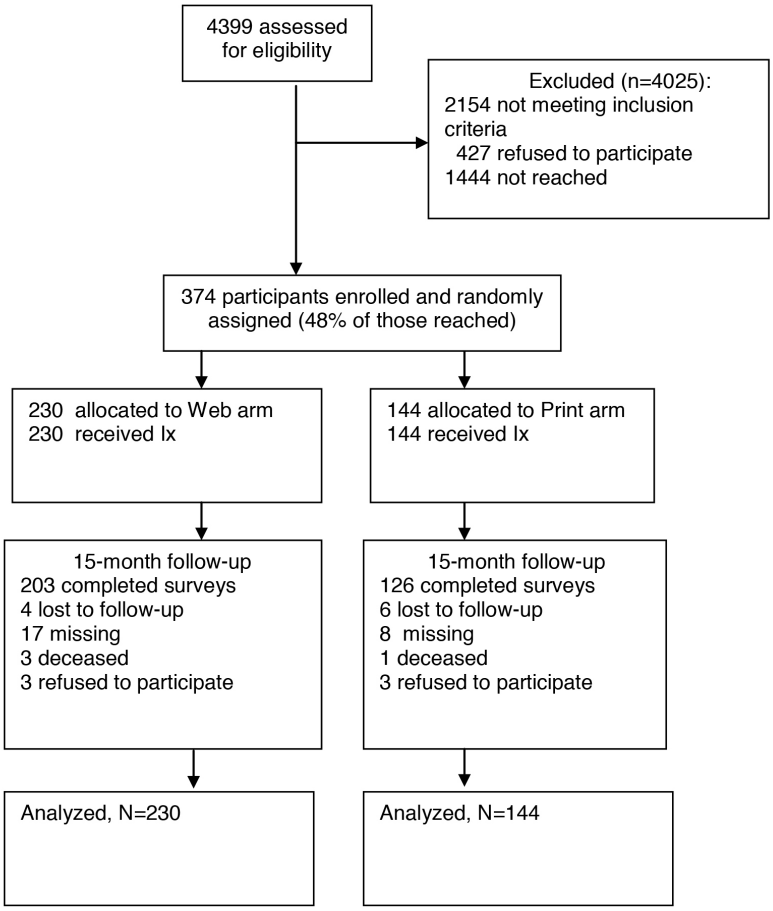


Table 2. Multivariable models predicting smoking at 15-month follow-up (n=374).

		Abstainer vs smoker at follow-up (adjusted for site)	
		P value	Adjusted OR (95% CI)
Treatment condition			
	Print	REF[a]	1.00
	Web	.87	1.07 (0.50-2.26)
Gender			
	Male	.04	0.49 (0.25-0.97)
	Female	REF	1.00
Race			
	White	.57	1.37 (0.46-4.06)
	Non-white	REF	1.00
Screened positive for possible depression			
	Yes	.22	0.62 (0.29-1.34)
	No	REF	1.00
Diagnosis			
	Leukemia	.03	2.64 (1.07-6.52)
	Hodgkins Disease	1.00	1.00 (0.31-3.20)
	CNS malignancy	.63	0.65 (0.12-3.68)
	Non-Hodgkins Lymphoma	.01	6.35 (1.48-27.17)
	Bone cancer	.86	1.13 (0.27-4.68)
	Other	REF	1.00
Received radiation for CA			
	Yes	.01	0.39 (0.19-0.81)
	No	REF	1.00
Received surgery for CA			
	Yes	.01	3.39 (1.29-8.88)
	No	REF	1.00
Perceived vulnerability for serious illness in future			
	No chance/very unlikely/unlikely	REF	1.00
	Moderate chance	.77	0.88 (0.37-2.11)
	Likely	.08	0.40 (0.14-1.12)
	Very likely/certain to happen	.12	0.45 (0.16-1.23)
Rules about smoking in home at follow-up			
	There are no rules	.05	0.38 (0.14-0.98)
	People can only smoke in certain rooms	.16	0.35 (0.08-1.53)
	People cannot smoke inside	REF	1.00

[a]REF: reference point.

Table 3. Multivariable models predicting quit attempts among smokers at 15-month follow-up (n=374).

		4 or more times vs none (adjusted for site)		1-3 times vs none (adjusted for site)	
		P value	Adjusted OR (95% CI)	*P* value	Adjusted OR (95% CI)
Treatment condition					
	Web	.90	1.05 (0.51-2.18)	.96	1.02 (0.54-1.90)
	Print	REF[a]	1.00	REF	1.00
Gender					
	Male	.06	2.06 (0.98-4.33)	.01	2.27 (1.18-4.40)
	Female	REF	1.00	REF	1.00
Race					
	White	.09	0.38 (0.12-1.18)	.01	0.28 (0.10-0.74)
	Non-white	REF	1.00	REF	1.00
Screened positive for possible depression					
	Yes	.30	0.67 (0.32-1.43)	.60	0.83 (0.43-1.63)
	No	REF	1.00	REF	1.00
Education					
	Did not complete HS / GED	.78	1.34 (0.16-11.36)	.01	5.24 (1.50-18.33)
	Completed HS / GGED	.23	0.53 (0.19-1.49)	.58	0.77 (0.31-1.93)
	Some college or training after HS	.79	1.13 (0.45-2.85)	.64	1.23 (0.51-2.97)
	College graduate	REF	1.00	REF	1.00
Use medication to help quit smoking at follow-up					
	Yes	<.001	9.45 (2.90-30.74)	<.001	9.27 (3.04-28.28)
	No	REF	1.00	REF	1.00
Computer use					
	Daily	.03	3.33 (1.12-9.90)	.06	2.20 (0.97-4.96)
	< than daily, but monthly or more	.03	4.01 (1.12-14.40)	.04	2.89 (1.05-7.95)
	< than monthly/never	REF	1.00	REF	1.00
Experience any serious illness in future					
	No chance/very unlikely/unlikely	REF	1.00	REF	1.00
	Moderate chance	.24	1.93 (0.64-5.80)	.03	3.33 (1.16-9.56)
	Likely	.36	1.67 (0.55-5.10)	.07	2.52 (0.91-6.97)
	Very likely/certain to happen	.52	0.69 (0.22-2.17)	.55	1.39 (0.47-4.09)
Smoke within 30 minutes of waking					
	Yes	REF	1.00	REF	1.00
	No	.002	3.70 (1.66-8.25)	.15	1.65 (0.83-3.26)

[a]REF: reference point.

Process Outcomes

Intervention Use

Of the Web participants, 58.3% (77/132) logged on at least once (mean visits 3.25). Among visitors to the site, 13% (10/77) logged on once, 20% (15/77) twice, 13% (10/77) three times, and 54% (42/77) logged on four or more times (range 6-98 times). Those who reported using the website more frequently (3+ times) had higher quit attempts and lower smoking rates at follow-up, compared with less frequent users (see Table 4). We

explored abstinence as an outcome related to different levels of use between groups; however, due to the distribution of quitters across intervention groups and use patterns, attempts to model the relationships were unstable.

Among the print condition participants, 58.3% (74/127) reported reading most or all of the print materials. About half reported using the materials on multiple occasions. Those who used more of the print materials (most/all) had higher quit attempts and lower smoking rates at follow-up, compared with those who used the materials less.

In total, 13.9% (36/259) of participants requested pharmacotherapy, with no differences between arms. Among those who requested this assistance, it was provided in 87% (31/36) of the cases; the primary reason for not providing it was that the participant did not have a regular physician who could confirm that there were no contraindications.

Satisfaction With Intervention

Of the Web participants who logged in, 87.9% (116/132) reported being satisfied or very satisfied with the site. Further, 75.8% (100/132) reported that the site provided new information about smoking, 83.3% (110/132) felt it provided new information about survivorship, and the majority felt that it was updated often enough; 81.1% (107/132) would recommend the site to other survivors.

In the print condition, 92.1% (117/127) reported being satisfied or very satisfied with the materials. Further, 93.7% (119/127) reported that the materials provided new information about smoking and 89.0% (113/127) felt they provided new information about survivorship; 88.2% (112/127) would recommend the materials to other survivors.

Access to the Web Intervention

In total, 79.9% (299/374) of the sample owned a computer and had access to the Internet at home and/or work; the majority (77.0%, 288/374) used the Internet at least once per week. Those who did not have regular Internet access had less education (P=.001), were more likely to be unemployed (P=.001), divorced (P=.001), and to be heavier smokers (P=.003). We offered MSN TV to those who did not have regular Internet; roughly two-thirds declined it.

Table 4. Quit attempts and smoking rate at follow-up among high/low intervention users.

	Quit Attempts		Smoking Rate (cigs/day)	
	Print	Web	Print	Web
Low Use	2.0	3.42	13.7	9.8
High Use	4.45	6.47	9.84	3.42

Discussion

Principal Findings

This paper assesses the efficacy of a print versus Web version of the Partnership for Health intervention. Both interventions were developed to address scalability of smoking cessation interventions among childhood and young adult cancer survivors. PFH-2 was designed to translate core components of PFH [9] to more scalable formats, and to determine whether equivalent levels of cessation could be achieved via a website tailored to cancer survivorship. Although we did indeed find that the PFH-2 intervention achieved similar levels of cessation to the peer counseling intervention, contrary to our hypothesis, there were no differences in cessation rates in the print versus Web arms and no differences in quit rates or changes in readiness to quit. Both interventions were viewed as substantive and appealing and were relatively comparable in terms of intervention "dose" based on participant report of use. These findings suggest that either the print or Web-format intervention could be recommended for survivors who smoke, as these cessation interventions yield equivalent levels of success to those found in our previous telephone-delivered intervention. The outcomes are also comparable to other Internet treatment studies [25] and PHS clinical guidelines for cost-effective interventions.

A Web-based approach was selected because of the presumed computer affinity of this younger population, as well as the potential for dissemination. We did offer access to the Internet via MSN TV to the 20% that did not have Internet access. Surprisingly, the majority of these participants declined the offer. This may reflect an active choice among these individuals not to engage with this technology and suggests that, at least in the population of childhood and young adult cancer survivors, efforts to increase access may not be helpful. Among those who received the print materials, engagement was good, which may be preferable in some settings and to some participants.

A key issue facing any disseminated, patient-guided intervention is level of patient engagement. Low levels of intervention use is a common problem in both Web and print interventions [26-30] and rapid and consistent declines in intensity of use occur over time [31,32]. A dose-response between website use and abstinence outcomes has been documented [33-35]. Consistent with this literature, in PFH-2 use of the website was associated with cessation-related efforts. Chiu and Eysenbach [36] call for development of a research agenda targeting how to improve use of Web-based interventions, which is critical if they are to be helpful in taking more labor-intensive interventions to scale. The increasing penetration of social media and technology may present opportunities for increasing engagement, as noted by Graham et al [33,37]. Active participation in online communities is associated with higher rates of cessation, and integrating smokers into an online social network can increase support and may also increase utilization of cessation tools and nicotine replacement therapy [38]. Text messages and Twitter could be utilized as efficient ways for peer counselors to engage survivors. The increased access to video cameras on laptops and smartphones also provides opportunities to evaluate Skype and videoconferencing as

engagement strategies, although these strategies may also present issues when taken to scale. In the present study, we targeted all smokers—not just those who were interested in quitting smoking. The intervention was designed and presented with a focus on general survivorship issues, in addition to smoking, and was organized by stage of readiness to change. That said, those with less interest in cessation would be expected to have lower levels of engagement with cessation content, although integrating smoking-related information into other forms of interventions or materials for survivors may be a way to increase their reconsideration of smoking.

Use of pharmacotherapy was quite low, particularly given the high level of readiness to quit and the fact that it was available at no cost. This is somewhat puzzling, but may be a function of survivors' heavy use of medications as part of cancer treatment and possibly survivorship, which may make them more reticent to use pharmacologic treatments for non-cancer-related issues. In 2014, CMS (Center for Medicare and Medicaid Services) programs will cover all FDA-approved medications and cessation counseling [39]. There are very tangible and important health benefits associated with population-level access to evidence-based cessation treatments in the general population [40-42], and the benefits for survivors may be even greater, given the synergistic effects between smoking and their increased risk of late effects. Our findings suggest that while access is important, alone it may not be sufficient in this high-risk population to achieve large-scale increases in pharmacotherapy use. This is an area that warrants further research attention.

This study highlights the need to develop an effective infrastructure for delivery of smoking cessation services to childhood and young adult cancer survivors. The infrastructure for identifying smokers within long-term survivorship care programs is largely missing [11] and a more systematic approach to patient tracking and follow-up is needed. Survivorship programs should be strongly encouraged to follow the PHS guidelines for delivery of smoking cessation services in clinical care settings [43,44]. Given the current movement toward electronic health records [45], it is likely that there will be greater opportunity for developing this infrastructure within survivorship programs. The American Society of Preventive Oncology recently called for a paradigm change related to cancer prevention after cancer, highlighting a critical need to reduce preventable risk factors [46]. Tobacco use is not systematically assessed in US cancer programs—less than half of comprehensive cancer centers have a strategy in place for effective identification of tobacco use [47] and only 28% use any tobacco-related quality improvement measures. As a result, a large proportion of smokers with cancer do not receive formal assistance with quit attempts. An important question would be whether survivorship-focused interventions are needed or if survivors could be integrated into widely available Internet-based cessation programs, such as QuitNet or BecomeAnEx. Our previous qualitative work suggests that the survivorship identity may be quite important in terms of being willing to engage in smoking cessation [48]. However, additional experimental work to determine the added value of tailoring to survivorship would be extremely valuable.

Cancer survivors are a unique population that should be uniquely aware of risks for cancer. However, not all childhood and young adult survivors are well informed about their cancer, its treatment, and the associated late effects, in large part because of the young age at diagnosis, disease, and treatment complexity, and the fact the knowledge about late effects has been emerging over time [49,50]. The Childhood Cancer Survivor Study has demonstrated that emotional distress is related to health behaviors, including smoking in long-term survivors, and subgroups that are at risk for becoming smokers (eg, lower income survivors, those with multiple medical morbidities) have higher distress levels [51,52]. A study of adolescent and young adult survivors did not find statistically significant differences between survivors and controls in terms of psychological distress or health-related quality of life, but survivors had less positive health beliefs [53]. Of note, these survivors' beliefs reflected their perceived bad luck and uncertainty about their health, including concerns about future medical problems (ie, health perceptions) and beliefs (ie, cognitive competence) that they have cognitive challenges that could impact their function (eg, memory, attention, intellect). These types of beliefs could impact on positive coping, such as smoking cessation.

The lack of an effective implementation infrastructure posed a challenge in the design of PFH-2, as it does in many implementation research studies. We considered comparing the PFH-1 peer counseling condition with the website. However, if we had concluded that the peer counseling was superior, we would not have moved the field forward in terms of having an evidence-based intervention that could be sustained and scaled in real-world conditions. Given that we did not have a mechanism to deploy peer counselors that would be sustainable following completion of the study, we felt it was best to learn from the peer counselor model and adapt to a sustainable format. Thus, PFH-2 was designed with the real-world constraints of survivorship care in mind and compared best survivorship-focused print materials with the website, which provided some interactivity but was more sustainable and scalable than peer counseling. This approach is consistent with several of the broad principles outlined by Kottke et al [54], which are necessary in order for health research to have a greater impact on patient and population health outcomes, including: (1) the needs of patients and populations determine the research agenda, (2) the research agenda addresses contextual and implementation issues, including the development of delivery and accountability systems, and (3) the research agenda determines the research methods rather than the methods determine the research agenda. Kottke and colleagues also note that the goal should be to optimize practice through research, which was the approach we took in designing PFH-2. This approach also addresses significant concerns raised within implementation science about having an increased emphasis on external validity and moving away from artificial comparisons that have little bearing on real-world care delivery [55-58]. In the interests of maximizing fidelity and ensuring a minimum intervention dose delivered, we did provide print materials to participants in the Web condition who did not access the website after 3 months. Although this means that the Web condition was not "pure", as some participants had access to both the Web and print interventions, we felt that ethically it was important

to ensure that patients received the intervention content in some form, when we knew that participants had not received it via the Web.

Limitations

Study limitations should be noted. The response rate was impacted by stringent IRB requirements regarding patient contact and release of contact information to the coordinating center, which may impact on generalizability of findings. Cessation outcomes were self-reported, which is typical in population-level and Web-based studies such as this, but still a limitation [59]; the bogus pipeline procedure, a well-accepted strategy for increasing the accuracy of self-report, was used [17,18]. It is possible that self-report at the point of evaluation of study eligibility introduced a sampling bias. However, participants were not aware of the eligibility requirements at the time of recruitment, which minimized the likelihood of bias. Further, smokers who did not accurately report their smoking status would likely report this same inaccuracy in the context of their health care, and thus would avoid exposure to this type of intervention. Therefore, any reporting bias would not likely effect the outcome evaluation.

There are several important strengths to note. A population-based approach was used in conducting this study, identifying all potential smokers within several different survivorship programs in the United States and Canada, which contributes to the external validity of the findings. Data were conservatively analyzed using multiple imputation methods for missing data. This study builds on the previous effective PFH intervention and was designed to determine how best to deliver that intervention in a more scalable format. The study design emphasized external validity and maximizing generalizability of study findings.

Conclusions

Smoking cessation among childhood and young adult cancer survivors is critical. Effective evidence-based programs should be integrated into primary and survivorship care delivery on an on-going and routine basis. This study demonstrated that it was possible to achieve equivalent cessation rates with tailored and targeted Web- and print-based materials designed specifically for survivors and that these cessation rates were equivalent to those found with a more labor-intensive telephone-based intervention. These findings suggest that survivorship programs have flexibility in the format in which cessation services are delivered without sacrificing effectiveness. Patients who reported more frequent use of the materials had better cessation rates, suggesting that developing methods for increasing patient engagement in print and Web-based interventions might optimize outcomes. Future research should examine such approaches to increase engagement and should also evaluate whether survivorship-focused interventions are critical or if survivors experience equal benefits from robust Web-based interventions available to the general public.

Acknowledgments

This study was supported by grants from the National Cancer Institute, (5R01CA106914-5 and 1K05 CA124415). The authors would like to acknowledge Martha Zorn for her work on the data management and Nancy Klockson for participation in the manuscript preparation.

Authors' Contributions

All authors made substantial contributions to the development of this manuscript and have given final approval of the version to be published. KE was the study PI and oversaw all aspects of the study conceptualization, design, implementation, analysis, and reporting. EP was the study biostatistician and provided input to the study design, data collection procedures, and completed all analyses. KSH provided critical input into and oversaw study implementation procedures and drafted the Methods section of this paper. JF, JO, DH, MG, LD, VT, and JdM all contributed to the study design and implementation, participated in data analysis, and contributed to the Introduction, Results, and Discussion sections of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [60].

[PDF File (Adobe PDF File), 997KB - jmir_v15i11e218_app1.pdf]

References

1. Friedman DL, Whitton J, Leisenring W, Mertens AC, Hammond S, Stovall M, et al. Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. J Natl Cancer Inst 2010 Jul 21;102(14):1083-1095 [FREE Full text] [doi: 10.1093/jnci/djq238] [Medline: 20634481]
2. Hewitt M, Weiner S, Simone J. Childhood cancer survivorship; improving care and quality of life. Washington, DC: The National Academies Press; 2003.

3. Meacham LR, Sklar CA, Li S, Liu Q, Gimpel N, Yasui Y, et al. Diabetes mellitus in long-term survivors of childhood cancer. Increased risk associated with radiation therapy: a report for the childhood cancer survivor study. Arch Intern Med 2009 Aug 10;169(15):1381-1388 [FREE Full text] [doi: 10.1001/archinternmed.2009.209] [Medline: 19667301]
4. Mulrooney DA, Yeazel MW, Kawashima T, Mertens AC, Mitby P, Stovall M, et al. Cardiac outcomes in a cohort of adult survivors of childhood and adolescent cancer: retrospective analysis of the Childhood Cancer Survivor Study cohort. BMJ 2009;339(Dec 08 1):b4606-b4606. [doi: 10.1136/bmj.b4606]
5. Ries L, Melbert D, Krapcho M, Mariotto A, Miller B, Feuer E, et al. SEER Cancer Statistics Review, 1975-2004. Bethesda, MD: National Cancer Institute; 2007.
6. Emmons K, Li FP, Whitton J, Mertens AC, Hutchinson R, Diller L, et al. Predictors of smoking initiation and cessation among childhood cancer survivors: a report from the childhood cancer survivor study. J Clin Oncol 2002 Mar 15;20(6):1608-1616. [Medline: 11896111]
7. Emmons KM, Butterfield RM, Puleo E, Park ER, Mertens A, Gritz ER, et al. Smoking among participants in the childhood cancer survivors cohort: the Partnership for Health Study. J Clin Oncol 2003 Jan 15;21(2):189-196. [Medline: 12525509]
8. Frobisher C, Winter DL, Lancashire ER, Reulen RC, Taylor AJ, Eiser C, British Childhood Cancer Survivor Study. Extent of smoking and age at initiation of smoking among adult survivors of childhood cancer in Britain. J Natl Cancer Inst 2008 Aug 6;100(15):1068-1081 [FREE Full text] [doi: 10.1093/jnci/djn210] [Medline: 18664655]
9. Emmons KM, Puleo E, Park E, Gritz ER, Butterfield RM, Weeks JC, et al. Peer-delivered smoking counseling for childhood cancer survivors increases rate of cessation: the partnership for health study. J Clin Oncol 2005 Sep 20;23(27):6516-6523. [doi: 10.1200/JCO.2005.07.048] [Medline: 16116148]
10. Emmons KM, Puleo E, Mertens A, Gritz ER, Diller L, Li FP. Long-term smoking cessation outcomes among childhood cancer survivors in the Partnership for Health Study. J Clin Oncol 2009 Jan 1;27(1):52-60 [FREE Full text] [doi: 10.1200/JCO.2007.13.0880] [Medline: 19047296]
11. de Moor JS, Puleo E, Butterfield RM, Li FP, Emmons KM. Availability of smoking prevention and cessation services for childhood cancer survivors. Cancer Causes Control 2007 May;18(4):423-430. [doi: 10.1007/s10552-006-0110-y] [Medline: 17297556]
12. Nagler RH, Puleo E, Sprunck-Harrild K, Emmons KM. Internet use among childhood and young adult cancer survivors who smoke: implications for cessation interventions. Cancer Causes Control 2012 Apr;23(4):647-652 [FREE Full text] [doi: 10.1007/s10552-012-9926-9] [Medline: 22370697]
13. de Moor JS, Puleo E, Ford JS, Greenberg M, Hodgson DC, Tyc VL, et al. Disseminating a smoking cessation intervention to childhood and young adult cancer survivors: baseline characteristics and study design of the partnership for health-2 study. BMC Cancer 2011;11:165 [FREE Full text] [doi: 10.1186/1471-2407-11-165] [Medline: 21569345]
14. Prochaska J, DiClemente C. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
15. Prokhorov AV, Yost T, Mullin-Jones M, de Moor C, Ford KH, Marani S, et al. "Look at your health": outcomes associated with a computer-assisted smoking cessation counseling intervention for community college students. Addict Behav 2008 Jun;33(6):757-771 [FREE Full text] [doi: 10.1016/j.addbeh.2007.12.005] [Medline: 18280668]
16. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
17. Murray DM, Perry CL. The measurement of substance use among adolescents: when is the 'bogus pipeline' method needed? Addict Behav 1987;12(3):225-233. [Medline: 3661275]
18. Murray DM, O'Connell CM, Schmid LA, Perry CL. The validity of smoking self-reports by adolescents: a reexamination of the bogus pipeline procedure. Addict Behav 1987;12(1):7-15. [Medline: 3565116]
19. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
20. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. Psychosom Med 1979 May;41(3):209-218 [FREE Full text] [Medline: 472086]
21. Vigilante KC, Flynn MM, Affleck PC, Stunkle JC, Merriman NA, Flanigan TP, et al. Reduction in recidivism of incarcerated women through primary care, peer counseling, and discharge planning. J Womens Health 1999 Apr;8(3):409-415. [Medline: 10326995]
22. Tyc VL, Hadley W, Crockett G. Predictors of intentions to use tobacco among adolescent survivors of cancer. J Pediatr Psychol 2001 Mar;26(2):117-121 [FREE Full text] [Medline: 11181887]
23. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 1997 Jul;12(7):439-445 [FREE Full text] [Medline: 9229283]
24. Yuan Y. Multiple imputation using SAS software. Journal of Statistical Software 2011 Dec;45(6):1-25.
25. Graham AL, Cobb NK, Papandonatos GD, Moreno JL, Kang H, Tinkelman DG, et al. A randomized trial of Internet and telephone treatment for smoking cessation. Arch Intern Med 2011 Jan 10;171(1):46-53 [FREE Full text] [doi: 10.1001/archinternmed.2010.451] [Medline: 21220660]

26. 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]
27. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
28. 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]
29. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
30. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
31. Anhøj J, Jensen AH. Using the internet for life style changes in diet and physical activity: a feasibility study. J Med Internet Res 2004 Sep 8;6(3):e28 [FREE Full text] [doi: 10.2196/jmir.6.3.e28] [Medline: 15471754]
32. Wu RC, Delgado D, Costigan J, Maciver J, Ross H. Pilot study of an Internet patient-physician communication tool for heart failure disease management. J Med Internet Res 2005;7(1):e8 [FREE Full text] [doi: 10.2196/jmir.7.1.e8] [Medline: 15829480]
33. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
34. Muñoz RF, Barrera AZ, Delucchi K, Penilla C, Torres LD, Pérez-Stable EJ. International Spanish/English Internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 Sep;11(9):1025-1034 [FREE Full text] [doi: 10.1093/ntr/ntp090] [Medline: 19640833]
35. Pike KJ, Rabius V, McAlister A, Geiger A. American Cancer Society's QuitLink: randomized trial of Internet assistance. Nicotine Tob Res 2007 Mar;9(3):415-420. [doi: 10.1080/14622200701188877] [Medline: 17365773]
36. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
37. Graham AL, Cha S, Papandonatos GD, Cobb NK, Mushro A, Fang Y, et al. Improving adherence to web-based cessation programs: a randomized controlled trial study protocol. Trials 2013;14:48 [FREE Full text] [doi: 10.1186/1745-6215-14-48] [Medline: 23414086]
38. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
39. U.S. Department of Health and Human Services. Washington, DC; 2012. URL: http://www.healthypeople.gov/2020/default.aspx [accessed 2013-08-21] [WebCite Cache ID 6J2jpWnQF]
40. Centers for Disease ControlPrevention (CDC). State medicaid coverage for tobacco-dependence treatments --- United States, 2009. MMWR Morb Mortal Wkly Rep 2010 Oct 22;59(41):1340-1343 [FREE Full text] [Medline: 20966897]
41. Land T, Warner D, Paskowsky M, Cammaerts A, Wetherell L, Kaufmann R, et al. Medicaid coverage for tobacco dependence treatments in Massachusetts and associated decreases in smoking prevalence. PLoS One 2010;5(3):e9770 [FREE Full text] [doi: 10.1371/journal.pone.0009770] [Medline: 20305787]
42. Land T, Rigotti NA, Levy DE, Paskowsky M, Warner D, Kwass JA, et al. A longitudinal study of medicaid coverage for tobacco dependence treatments in Massachusetts and associated decreases in hospitalizations for cardiovascular disease. PLoS Med 2010;7(12):e1000375 [FREE Full text] [doi: 10.1371/journal.pmed.1000375] [Medline: 21170313]
43. Fiore MC, Jaén CR. A clinical blueprint to accelerate the elimination of tobacco use. JAMA 2008 May 7;299(17):2083-2085. [doi: 10.1001/jama.299.17.2083] [Medline: 18460668]
44. U.S. Department of Health and Human Services PHS. Clinical Practice Guideline, Number 18: Smoking Cessation. In: Agency for Health Care Policy and Research. Washington, DC: US Government Printing Office; 1996.
45. Estabrooks PA, Boyle M, Emmons KM, Glasgow RE, Hesse BW, Kaplan RM, et al. Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. J Am Med Inform Assoc 2012;19(4):575-582 [FREE Full text] [doi: 10.1136/amiajnl-2011-000576] [Medline: 22511015]
46. Campo RA, Rowland JH, Irwin ML, Nathan PC, Gritz ER, Kinney AY. Cancer prevention after cancer: changing the paradigm--a report from the American Society of Preventive Oncology. Cancer Epidemiol Biomarkers Prev 2011 Oct;20(10):2317-2324 [FREE Full text] [doi: 10.1158/1055-9965.EPI-11-0728] [Medline: 21908726]
47. Goldstein AO, Ripley-Moffitt CE, Pathman DE, Patsakham KM. Tobacco use treatment at the U.S. National Cancer Institute's designated Cancer Centers. Nicotine Tob Res 2013 Jan;15(1):52-58. [doi: 10.1093/ntr/nts083] [Medline: 22499079]
48. Park EB, Emmons KM, Malloy NW, Seifer E. A qualitative exploration of health perceptions and behaviors among adult survivors of childhood cancers. J Cancer Educ 2002;17(4):211-215. [doi: 10.1080/08858190209528840] [Medline: 12556058]

49. Kadan-Lottick NS, Robison LL, Gurney JG, Neglia JP, Yasui Y, Hayashi R, et al. Childhood cancer survivors' knowledge about their past diagnosis and treatment: Childhood Cancer Survivor Study. JAMA 2002 Apr 10;287(14):1832-1839. [Medline: 11939869]
50. Bashore L. Childhood and adolescent cancer survivors' knowledge of their disease and effects of treatment. J Pediatr Oncol Nurs 2004;21(2):98-102. [Medline: 15125553]
51. Zeltzer LK, Recklitis C, Buchbinder D, Zebrack B, Casillas J, Tsao JC, et al. Psychological status in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. J Clin Oncol 2009 May 10;27(14):2396-2404 [FREE Full text] [doi: 10.1200/JCO.2008.21.1433] [Medline: 19255309]
52. Wenninger K, Helmes A, Bengel J, Lauten M, Völkel S, Niemeyer CM. Coping in long-term survivors of childhood cancer: relations to psychological distress. Psychooncology 2013 Apr;22(4):854-861. [doi: 10.1002/pon.3073] [Medline: 22461240]
53. Kazak AE, Derosa BW, Schwartz LA, Hobbie W, Carlson C, Ittenbach RF, et al. Psychological outcomes and health beliefs in adolescent and young adult survivors of childhood cancer and controls. J Clin Oncol 2010 Apr 20;28(12):2002-2007 [FREE Full text] [doi: 10.1200/JCO.2009.25.9564] [Medline: 20231679]
54. Kottke TE, Solberg LI, Nelson AF, Belcher DW, Caplan W, Green LW, et al. Optimizing practice through research: a new perspective to solve an old problem. Ann Fam Med 2008;6(5):459-462 [FREE Full text] [doi: 10.1370/afm.862] [Medline: 18779551]
55. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [Medline: 12893608]
56. Green LW, Glasgow RE, Atkins D, Stange K. Making evidence from research more relevant, useful, and actionable in policy, program planning, and practice slips "twixt cup and lip". Am J Prev Med 2009 Dec;37(6 Suppl 1):S187-S191. [doi: 10.1016/j.amepre.2009.08.017] [Medline: 19896017]
57. Glasgow RE, Green LW, Klesges LM, Abrams DB, Fisher EB, Goldstein MG, et al. External validity: we need to do more. Ann Behav Med 2006 Apr;31(2):105-108. [doi: 10.1207/s15324796abm3102_1] [Medline: 16542124]
58. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006 Mar;29(1):126-153. [doi: 10.1177/0163278705284445] [Medline: 16510882]
59. Velicer WF, Prochaska JO, Rossi JS, Snow MG. Assessing outcome in smoking cessation studies. Psychol Bull 1992 Jan;111(1):23-41. [Medline: 1539088]
60. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BFR: baseline feedback report
IES: Impact of Events Scale
IRB: Institutional Review Board
PFH: Partnership for Health
PFH-2: Partnership for Health-2
PHS: Public Health Service

Edited by G Eysenbach; submitted 18.01.13; peer-reviewed by J Studts, D Abrams; comments to author 26.03.13; revised version received 21.05.13; accepted 04.07.13; published 05.11.13.

Please cite as:
Emmons KM, Puleo E, Sprunck-Harrild K, Ford J, Ostroff JS, Hodgson D, Greenberg M, Diller L, de Moor J, Tyc V
Partnership for Health-2, A Web-Based Versus Print Smoking Cessation Intervention for Childhood and Young Adult Cancer Survivors: Randomized Comparative Effectiveness Study
J Med Internet Res 2013;15(11):e218
URL: http://www.jmir.org/2013/11/e218/
doi:10.2196/jmir.2533
PMID:24195867

Original Paper

Effects of a Web-Based Intervention on Physical Activity and Metabolism in Older Adults: Randomized Controlled Trial

Carolien A Wijsman[1], MD; Rudi GJ Westendorp[1,2], MD, PhD; Evert ALM Verhagen[3], PhD; Michael Catt[4], BSc; P Eline Slagboom[5], PhD; Anton JM de Craen[1,6], PhD; Karen Broekhuizen[7], PhD; Willem van Mechelen[3], MD, PhD; Diana van Heemst[1,6], PhD; Frans van der Ouderaa[2,6], PhD; Simon P Mooijaart[1,6,7], MD, PhD

[1]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, Netherlands

[2]Leyden Academy of Vitality and Ageing, Leiden, Netherlands

[3]Department of Public and Occupational Health, EMGO+Institute, VU University Medical Center, Amsterdam, Netherlands

[4]Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, United Kingdom

[5]Department of Medical Statistics, Molecular Epidemiology Section, Leiden University Medical Center, Leiden, Netherlands

[6]Netherlands Consortium for Healthy Ageing, Leiden, Netherlands

[7]Institute for Evidence-Based Medicine in Old Age, IEMO, Leiden, Netherlands

Corresponding Author:
Carolien A Wijsman, MD
Department of Gerontology and Geriatrics
Leiden University Medical Center
PO Box 9600
Leiden, 2300 RC
Netherlands
Phone: 31 717370238
Fax: 31 715266881
Email: c.a.wijsman@lumc.nl

Abstract

Background: Lack of physical activity leads to detrimental changes in body composition and metabolism, functional decline, and increased risk of disease in old age. The potential of Web-assisted interventions for increasing physical activity and improving metabolism in older individuals holds great promise but to our knowledge it has not been studied.

Objective: The goal of our study was to assess whether a Web-based intervention increases physical activity and improves metabolic health in inactive older adults.

Methods: We conducted a 3-month randomized, waitlist-controlled trial in a volunteer sample of 235 inactive adults aged 60-70 years without diabetes. The intervention group received the Internet program Philips DirectLife, which was directed at increasing physical activity using monitoring and feedback by accelerometer and digital coaching. The primary outcome was relative increase in physical activity measured objectively using ankle- and wrist-worn accelerometers. Secondary outcomes of metabolic health included anthropometric measures and parameters of glucose metabolism.

Results: In total, 226 participants (97%) completed the study. At the ankle, activity counts increased by 46% (standard error [SE] 7%) in the intervention group, compared to 12% (SE 3%) in the control group ($P_{\text{difference}}$<.001). Measured at the wrist, activity counts increased by 11% (SE 3%) in the intervention group and 5% (SE 2%) in the control group ($P_{\text{difference}}$=.11). After processing of the data, this corresponded to a daily increase of 11 minutes in moderate-to-vigorous activity in the intervention group versus 0 minutes in the control group ($P_{\text{difference}}$=.001). Weight decreased significantly more in the intervention group compared to controls (−1.5 kg vs −0.8 kg respectively, *P*=.046), as did waist circumference (−2.3 cm vs −1.3 cm respectively, *P*=.036) and fat mass (−0.6% vs 0.07% respectively, *P*=.025). Furthermore, insulin and HbA1c levels were significantly more reduced in the intervention group compared to controls (both *P*<.05).

Conclusions: This was the first study to show that in inactive older adults, a 3-month Web-based physical activity intervention was effective in increasing objectively measured daily physical activity and improving metabolic health. Such Web-based interventions provide novel opportunities for large scale prevention of metabolic deregulation in our rapidly aging population.

Trial Registration: Dutch Trial Registry: NTR 3045; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3045 (Archived by WebCite at http://www.webcitation.org/6KPw52dCc).

(J Med Internet Res 2013;15(11):e233) doi:10.2196/jmir.2843

KEYWORDS

physical activity; Internet; accelerometry; aging; metabolism; self-monitoring

Introduction

Lack of physical activity is perhaps one of the greatest risk factors for contemporary societal health problems. Insufficient physical activity contributes to obesity and has been associated with increased risks of cardiovascular disease [1,2], diabetes mellitus [3], cognitive decline [4,5], and premature mortality [6,7]. For the older age groups, this is of particular relevance as physical activity decreases with age, while the prevalence of metabolic disease and its complications increase as a function of age. Intervention studies directed at increasing physical activity have been shown to be effective and may improve metabolism [8], including the older populations [9,10]. However, most of these interventions have used face-to-face communication, making them costly and time-consuming, thus hampering the potential of implementation as preventive programs at a larger scale.

Modern technologies, such as Internet and email, provide interactive ways to administer digital coaching and feedback on physical activity and have potential for wide-scaled implementation. A recent meta-analysis on Web-based physical activity intervention studies showed promising results in increasing daily physical activity [11]. Many of these, however, used either small study populations, included study populations under the age of 60, or did not measure physical activity objectively. Moreover, it is unclear what the effect is of such Web-based physical activity interventions on outcomes of metabolic health in old age [12,13].

Since 2005, Internet use has doubled in 65-75 year olds in the Netherlands [14]; 70% of this age group is familiar with using the Internet and therefore provides a target for Web-based interventions. In this randomized controlled trial, we examined whether a 3-month Web-assisted intervention directed at increasing daily physical activity was effective in 60-70 year old inactive individuals. The intervention comprised an Internet program aimed at increasing physical activity, which focused on effective components of health behavior change including self-monitoring by accelerometer and goal setting with the help of digital coaching. Furthermore, we studied the effect of this intervention on metabolic health including anthropometric measures and markers of glucose and lipid metabolism.

Methods

Study Design and Participants

The study recruited participants aged 60-70 years from the region of Leiden, Netherlands. The recruitment strategy included advertisements in local newspapers and press notification, directing participants motivated to increase physical activity to the study website where they completed an online questionnaire that assessed the following inclusion and exclusion criteria: (1) age between 60 and 70 years, (2) no history of diabetes or use of glucose lowering medication, (3) absence of disability impeding increase in physical activity, and (4) possession and use of personal computer with Internet connection. If all of the above criteria were met, potential participants filled out an email address where they subsequently received a questionnaire asking for current physical activity and personal data such as full name and address. The presence of an inactive lifestyle was then assessed by a self-reported physical activity questionnaire: general practice physical activity questionnaire, GPPAQ [15]. This yielded four categories of physical activity: inactive, moderately inactive, moderately active, and active. We defined inactive as having less than 3 hours of exercise and cycling combined weekly, corresponding to the inactive, moderately inactive, or moderately active category. Participants in the active category of GPPAQ did not meet inclusion criteria for our definition of an inactive lifestyle.

Subjects who met all inclusion criteria received detailed study information in print. If willing to participate, participants received an online questionnaire on education, smoking status, medical history, and medication use, and visited the study center at baseline and after 3 months (13 weeks). At the baseline visit, subjects were randomly assigned to the intervention group or a waitlist control group (in which the participants received the intervention program after 3 months when the study ended) by the study physician or research nurse. Randomization was performed by a computerized program for intervention versus waitlist control in a ratio of 1:1, with a block size of 12. Stratification was performed by gender. Concealment of treatment allocation was ensured by randomizing at the end of the first study visit, after all baseline measurements and instructions at the study center were completed. Blinding of this intervention was not possible and therefore not applied. Written informed consent was obtained from all subjects. The study was approved by the medical ethical committee of Leiden University Medical Center, Netherlands. An independent physician was available for questions regarding study information.

Sample Size Calculation

The sample size of the study was based on the assumption of a mean 10% higher increase (SD 25%) in daily physical activity counts as measured using accelerometers on the ankle and wrist in the intervention group compared to the control group during a 3-month period. For this effect size with a power of 0.80 at alpha 0.05 (two-sided), we calculated a sample size of 198 participants for the intention-to-treat (ITT) analysis. Based on an estimated drop-out rate of 15%, we aimed to include 232 participants and stopped after successful inclusion of 235 participants.

Intervention

Subjects in the intervention group received a commercially available Web-based physical activity program (DirectLife, Philips, Consumer Lifestyle, Amsterdam) directed at increasing daily physical activity. The DirectLife program is based on the stages of change and I-change health behavior change models [16,17] and takes into account the individual's current activity level. It then provides a personal goal. Briefly, DirectLife consists of three elements: (1) an accelerometer-based activity monitor, (2) a personal website, and (3) a personal e-coach, who provides regular updates of the individual's physical activity status by email and gives advice to increase physical activities (Multimedia Appendix 1). By means of these elements, the program aims to increase awareness about one's own physical activity behavior, to give feedback on recent actual physical activity, and to provide support to make sustainable changes in physical activity behavior. The activity monitor of DirectLife is based on the Tracmor tri-axial accelerometer and has been validated against doubly labeled water for the estimation of total daily life energy expenditure [18]. The DirectLife monitor is the consumer version of the Tracmor. Intervention group participants received the program, including the accelerometer, directly after randomization at the first study visit. They then received a link by email for registration and access to the Web program. Participants of the program were instructed to continuously wear the activity monitor throughout the day to measure daily physical activity (Multimedia Appendix 2). Data were uploaded through an Internet connection to the database of the commercial provider. After an initial 8-day "assessment period" starting 1 week after the study visit, in which the current level of daily activity was measured, a target was set to increase the level of daily activity during a 12-week Web-based interactive coaching program. Participants were given a target for daily activity, which increased weekly, and data from the accelerometer were used for regular feedback (Multimedia Appendix 3). Coaching included general recommendations on physical activities, and coaches were available for further questions and advice by email correspondence. All participants were in contact with one of the digital coaches available for the DirectLife program during the entire study period.

The control group was placed on a 3-month waiting list, after which they received access to the intervention program at the end of the study. No specific instructions regarding daily physical activity were given.

Measurements

Baseline Questionnaire

Enrollment and follow-up took place from November 2011 to August 2012. In preparation for the first visit to the study center, all participants completed a Web-delivered questionnaire on education, smoking status, and medical history, including medication use. Education was categorized as low (primary education and lower vocational education), intermediate (secondary education and intermediate vocational education), or high (high vocational education and university).

Primary Outcome

At baseline and 3-month follow-up, daily physical activity was measured during 7 days following the visit at the study center, using an ankle- and wrist-worn tri-axial accelerometer (GENEActiv, Kimbolton, Cambs, United Kingdom). Wear was started on a random weekday, and GENEActive monitors were returned after 7 days by standard mail. We chose to assess the primary outcome using accelerometers other than the one included in the intervention program in order to avoid interpretation of the intervention as an outcome. Both GENEActiv monitors were worn 24 hours per day on the right side. The GENEActiv wrist accelerometer provides a simple summary statistic of total activity counts has been validated for measuring daily physical activity against doubly labeled water [19]. We chose to additionally assess total activity counts using an ankle accelerometer as we hypothesized that this location would be more sensitive to walking and cycling behavior [20,21], the latter being a very frequent activity in our target population in the Netherlands. Primary outcome was the individual's relative change in activity counts after the intervention compared to baseline, measured at wrist and ankle. As a derivative outcome, we calculated from the wrist accelerometer the minutes per day spent in moderate or vigorous activity, which has been validated against indirect calorimetry [22].

Measurement frequency was set at 85.7 Hz, and raw acceleration values in "g" were recorded continuously on each axis over 7 consecutive days. Prior to processing, data were plotted for visual identification of non-wear and device faults. Non-wear was determined visually using thresholds of movement in combination with self-reported non-wear from participants. Short periods of non-wear (eg, bathing) were accepted, and data for these periods were not imputed. Accelerometer data from participants contributing to GENEActiv data for 5 days or more within the 7-day period were included in the analysis. Data from each axis were processed by a high pass RC filter (fc=0.27 HZ) before computation of the resultant acceleration for each recorded time point: $R=(x^2 + y^2 + z^2)^{0.5}$. The average of each 24-hour integral of these values over the first 5 days was used as the average daily activity count for the assessment period. As a measure of physical activity, we used total activity counts recorded at both the ankle and the wrist accelerometer independently. Data collected from the right wrist of each participant over 5 days of continuous movement monitoring were processed to yield activity counts for successive 1-minute epochs and classified according to the appropriate MET cut-off points according to the method of Esliger [22], to establish the average number of minutes (epochs) spent daily in moderate and vigorous physical activity. Outcome assessment was done by an independent researcher who was blind to study arm allocation (MC).

Secondary Outcomes

Body height was measured without shoes using a stadiometer. Body weight was assessed at both visits without shoes using a measurement scale. Waist circumference was obtained in a standing position halfway between the anterior superior iliac

spine and the lower rib. Hip circumference was measured halfway between the trochanter major and the iliac crest.

Lean body mass and body fat percentage were assessed by bio-electrical impedance (BIA) analysis using a commercial portable device with hand-to-foot single frequency measurement (Biostat 1500, Euromedix, Leuven, Belgium). Blood pressure was measured twice at each visit using a hand-held sphygmomanometer after 5 minutes of lying down. The mean of the two consecutive measurements was used. Heart rate was measured at the wrist after at least 5 minutes of lying down. Grip strength was measured to the nearest kilogram three times using a Jamar hand dynamometer (Sammons Preston, Inc, Bolingbrook, IL, United States) with the dominant hand. The highest value was used for analysis. Framingham risk scores were calculated using NIH criteria [23].

Biochemical Assessments

Fasting blood samples were drawn from each participant at both visits in the morning. Samples were transferred to the lab within 2 hours, aliquotted, and frozen at −80°C. All serum measurements were performed in one batch after completion of the entire study with fully automated equipment. Fasting glucose, cholesterol, HDL-cholesterol, and triglyceride levels were determined using the Modular P2 analyzer (Roche, Almere, Netherlands), and fasting serum insulin using immunoassay by Immulite 2500 (DPC, Los Angeles, CA, United States). Glycated hemoglobin was determined by high-performance liquid chromatography (Primus Ultra2, Trinity Biotech Company, Kansas City, MO, United States). C-reactive protein (hsCRP) was determined using a high-sensitive immunoassay (COBAS integra, Roche, IN, United States). Low density lipoprotein (LDL) cholesterol was calculated using the Friedewald formula in participants without hypertriglyceridemia [24].

Statistical Analyses

Differences between baseline and follow-up within groups were tested using a paired sample Student *t* test of the means. For skewed variables, ln (natural logarithm) transformation was used. The effect of the intervention on physical activity was assessed by an unpaired 2-sided *t* test, comparing the relative change in daily physical activity counts between the intervention group and control group. For relative change in moderate-to-vigorous physical activity, a nonparametric test was used due to skewness of data. The effect of the intervention on secondary outcomes was assessed using an unpaired 2-sided *t* test, comparing the change in the secondary outcome between the intervention group and control group. Primary analyses were performed by ITT principle. Our study did have one follow-up measurement only, and loss to follow-up was very low. We therefore did not use imputation to replace our data, and participants from whom data were lost were not in the ITT analysis. For per-protocol analysis, we included in the intervention group only those participants who finished the 12-week plan of the intervention program, as assessed by uploaded accelerometer data of the participant in week 12 of the DirectLife intervention program. All analyses were performed with SPSS version 20.0. Statistical significance was accepted at $P<.05$.

Results

Participant Characteristics

Figure 1 shows the inclusion flow chart of participants. A total of 631 participants completed the questionnaire on the study website. Of those, 235 participants (37%) met inclusion criteria and were randomized: 119 in the intervention arm, 116 in the control arm. Nine participants did not complete the study: 5 and 4 participants in the intervention and control group, respectively. Final analyses of outcomes therefore included 114 participants in the intervention group and 112 control participants.

Baseline characteristics of randomized participants are shown in Table 1. The study groups were similar for all parameters. The majority of participants were male and middle or highly educated. A substantial number of the study participants used antihypertensive medication: 46% and 38 % of participants in intervention and control group, respectively. On average, participants were overweight with a mean BMI of 28.9 kg/m^2 (SD 4.7) and 29.l kg/m^2 (SD 4.7) in the intervention and control group, respectively. There was no significant difference in baseline activity level between groups.

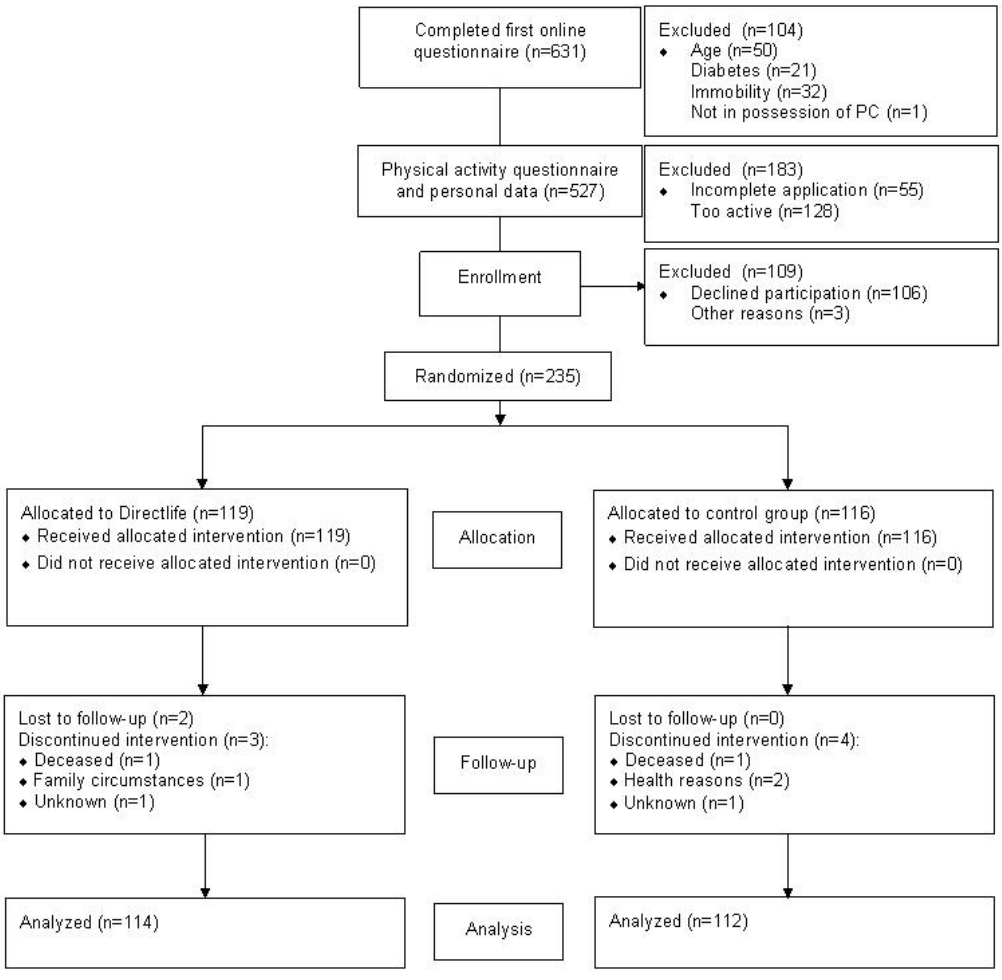
Table 1. Baseline characteristics of study participants (data are presented as medians with interquartile range [IQR] when skewed; alcohol use was calculated only in those who reported drinking alcohol [n=102 for intervention and n=101 for control]).

Characteristics	Intervention (n=119)	Control (n=116)
Demographics		
Female sex, n (%)	47 (39.5)	49 (42.2)
Age, yrs (mean, SD)	64.7 (3.0)	64.9 (2.8)
Degree of self-reported activity, n (%)		
Moderately active	41 (34.5)	48 (41.4)
Moderately inactive	36 (30.3)	34 (29.3)
Inactive	42 (35.3)	34 (29.3)
Level of education, n (%)		
Low	7 (5.9)	2 (1.7)
Intermediate	45 (37.8)	46 (39.7)
High	66 (55.5)	67 (57.8)
Current smoking	7 (5.9)	8 (6.9)
Alcohol use, units/wk (mean, SD)	12.0 (8.0)	10.2 (9.4)
Medical history and medication use, n (%)		
Coronary heart disease[a]	14 (11.8)	16 (13.8)
Arrhythmia	10 (8.4)	14 (12.1)
Lung emphysema	3 (2.5)	5 (4.3)
Stroke	6 (5.0)	3(2.6)
Malignancies	17 (14.3)	17 (14.7)
Thyroid disease	8 (6.7)	8 (6.9)
Antihypertensive use	55 (46.2)	44 (37.9)
Statin use	31 (26.1)	24 (20.7)
Anticoagulant use	25 (21.0)	20 (17.2)
Psychotropic use	11 (9.2)	16 (13.8)
Physical activity		
Ankle monitor, counts/day (mean, SD)	3.68×10^5 (1.72×10^5)	3.78×10^5 (1.84×10^5)
Wrist monitor, counts/day (mean, SD)	3.51×10^5 (1.10×10^5)	3.39×10^5 (1.16×10^5)
MVPA[b], min/day, median (IQR)	16.8 (7.8-26.4)	14.4 (8.2-32.0)
Clinical parameters, mean (SD)		
Height (cm)	173.6 (9.9)	172.1 (9.3)
Weight (kg)	87.4 (15.8)	86.3 (15.8)
BMI[c] (kg/m^2)	28.9 (4.7)	29.1 (4.7)
Waist circumference (cm)	102.3 (13.1)	101.4 (12.3)
Hip circumference (cm)	109.1 (9.1)	108.7 (8.9)
Waist/hip ratio	0.93 (0.08)	0.93 (0.08)
Fat percentage (%)	36.5 (7.6)	36.4 (8.1)
Systolic blood pressure (mmHg)	146 (18)	145 (17)
Diastolic blood pressure (mmHg)	86 (9)	86 (11)
Resting heart rate (beats/min)	72 (10)	71 (11)
Grip strength (kg)	37.5 (10.2)	37.9 (10.4)

Characteristics	Intervention (n=119)	Control (n=116)
Framingham 10-year CVD[d] risk %	11.9 (7.2)	11.3 (7.5)
Biochemistry, mean (SD)		
Fasting venous glucose (mmol/L)	5.7 (0.7)	5.7 (0.7)
Fasting insulin (mU/L) (median, IQR)	11.5 (8.1–16.9)	10.8 (7.0–15.8)
HbA1c (%)	5.4 (0.3)	5.4 (0.3)
HOMA[e] index (median, IQR)	2.8 (2.0–4.4)	2.6 (1.7–4.3)
Total cholesterol (mmol/L)	5.7 (1.1)	5.8 (1.0)
HDL[f] cholesterol (mmol/L)	1.5 (0.5)	1.4 (0.4)
Triglycerides (median, IQR), (mmol/L)	1.5 (1.1–2.0)	1.4 (1.1–2.0)
LDL[g] cholesterol (mmol/L)	3.6 (1.0)	3.6 (0.9)
Total/HDL cholesterol ratio	4.2 (1.3)	4.3 (1.3)
C-reactive protein (median, IQR), (mg/L)	1.6 (0.8–3.1)	1.4 (0.8–4.1)

[a]Coronary heart disease: myocardial infarction/ angina pectoris.
[b]MVPA=moderate-to-vigorous physical activity.
[c]BMI=body mass index.
[d]CVD=cardiovascular disease.
[e]HOMA=homeostatic model assessment.
[f]HDL=high density lipoprotein.
[g]LDL=low density lipoprotein.

Figure 1. Consort flowchart of participants.



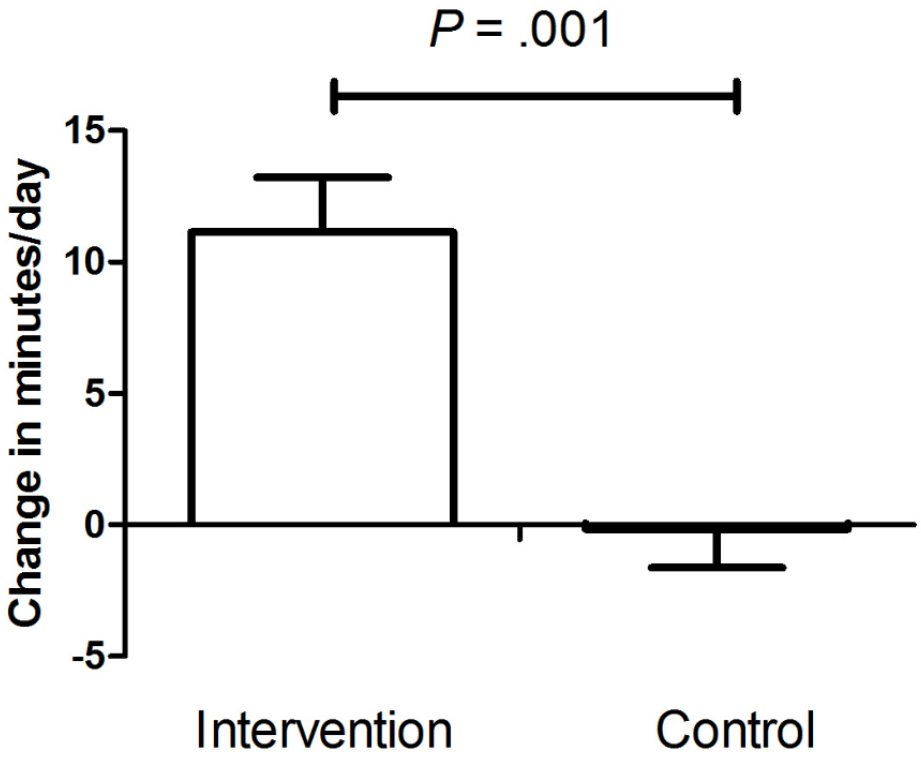
Adherence to the Intervention Program

All 114 participants who completed the study in the intervention arm received the intervention program during the visit to the study site at baseline. Of these, 109 participants (95.6%) started the intervention program after completing the first assessment week of the intervention program. In total, 104 participants (91.2%) completed the 12-week intervention program.

Primary Outcome: Physical Activity

Accelerometer data were available for 107 intervention and 109 control participants for ankle monitors, and 108 and 105 intervention and control participants for wrist monitors, respectively. After 13 weeks, daily physical activity as measured by the ankle accelerometer increased by 46% (SE 7%, P<.001) in the intervention group, compared to 12 % (SE 3%, P<.001) in the control group ($P_{difference}$<.001). Daily physical activity measured by the wrist accelerometer increased by 11% (SE 3%, P<.001) in the intervention group, and by 5% (SE 2%, P=.027) in the control group ($P_{difference}$=.11). In the intervention group, there was a mean increase of 11.1 minutes per day (SE 2.1) spent in moderate-to-vigorous activity, compared to a mean decrease of 0.1 minutes (SE 1.5) in the control group (P for relative difference=.001) (Figure 2). In the per-protocol analysis, taking into account only those 91% (n=104) of participants who completed the intervention phase of the DirectLife program, results did not change.

Figure 2. Change in daily physical activity expressed as moderate-to-vigorous physical activity measured at the wrist.



Secondary Outcomes: Parameters of Metabolic Health

Changes in parameters of metabolic health within and between groups are shown in Table 2 and Figure 3 (data represent mean changes with SE; P value for difference between intervention and control group using an unpaired 2-sided t test). A significant effect of the intervention on weight loss was seen with a mean change of −1.49 kg (SE 0.26) in the intervention group compared to −0.82 kg (SE 0.21) in the control group ($P_{difference}$=.046). Likewise, waist circumference decreased more in the intervention vs control group: −2.33 SE 0.36 cm vs −1.29 SE 0.34 cm, $P_{difference}$=.036. Fat percentage also decreased more in the intervention vs control group: −0.64 SE 0.23 vs 0.07 SE 0.22, $P_{difference}$=.025. Glucose metabolism improved significantly; HbA1c decreased more in the intervention group (−0.05%, SE 0.01) compared to the control group (−0.01%, SE 0.01, $P_{difference}$=.049). Similarly, fasting ln insulin levels (−0.16, SE 0.04 vs −0.04 SE 0.04, $P_{difference}$=.037) improved significantly in the intervention group compared to controls, as did the HOMA-index (−0.20 SE 0.05 vs −0.06 SE 0.04, $P_{difference}$=.035). Total cholesterol, LDL, and triglycerides levels improved in the intervention group, but these differences were not significant between groups. A decrease of more than 2 mmHg in systolic blood pressure was observed in both groups ($P_{difference}$=.83) and a decrease in resting heart rate of over 5 beats per minute in the intervention group and almost 4 beats per minute in the control group ($P_{difference}$=.15). No significant change in grip strength was seen in either of the groups. In the per-protocol analysis, taking into account only those 91% of participants who completed the intervention phase, differences did not materially change.

Table 2. Changes in parameters of metabolic health in study participants at follow-up (data represent mean differences with standard error [SE]).

	Intervention (n=114)		Control (n=112)		
	Mean difference (SE)	*P* value for change[a]	Mean difference (SE)	*P* value for change[a]	*P* value between groups[b]
Clinical parameters					
Weight, kg	−1.49 (0.26)	<.001	−0.82 (0.21)	<.001	.046
BMI[c], kg/m^2	−0.50 (0.09)	<.001	−0.29 (0.07)	<.001	.068
Waist circumference, cm	−2.33 (0.36)	<.001	−1.29 (0.34)	<.001	.036
Hip circumference, cm	−1.75 (0.33)	<.001	−1.12 (0.28)	<.001	.14
Waist/hip ratio	−0.008 (0.004)	.032	−0.001 (0.003)	.81	.16
Fat percentage, %	−0.64 (0.23)	.003	0.07 (0.22)	.91	.025
Lean mass, kg	−0.37 (0.19)	.32	−0.57 (0.20)	.033	.46
Systolic blood pressure (mmHg)	−2.73 (1.35)	.046	−2.30 (1.39)	.10	.83
Diastolic blood pressure (mmHg)	1.10 (0.78)	.16	0.10 (0.82)	.91	.38
Resting heart rate (beats/min)	−5.49 (0.94)	<.001	−3.68 (0.85)	<.001	.15
Grip Strength (kg)	0.31 (0.38)	.41	0.00 (0.34)	1.00	.50
Framingham 10-year risk %	−0.54 (0.33)	.11	−0.01 (0.31)	.98	.24
Biochemistry					
Fasting glucose venous (mmol/L)	−0.20 (0.05)	<.001	−0.13 (0.04)	.001	.31
Ln fasting insulin (mU/L)	−0.16 (0.04)	<.001	−0.04 (0.04)	.33	.037
HbA1c (%)	−0.05 (0.01)	<.001	−0.01 (0.01)	.25	.049
Ln HOMA[d]-index	−0.20 (0.05)	<.001	−0.06 (0.04)	.15	.035
Total cholesterol (mmol/L)	−0.25 (0.06)	<.001	−0.18 (0.05)	.001	.36
HDL[e] cholesterol (mmol/L)	−0.008 (0.02)	.71	−0.04 (0.02)	.033	.27
Ln Triglycerides (mmol/L)	−0.10 (0.03)	.003	−0.06 (0.02)	.023	.30
LDL[f] cholesterol (mmol/L)	−0.17 (0.04)	<.001	−0.11 (0.04)	.012	.37
Chol/HDL ratio	−0.20 (0.07)	.008	−0.05 (0.05)	.32	.10
Ln C-reactive protein (mg/L)	−0.12 (0.08)	.11	−0.11 (0.09)	.24	.92

[a]*P* values within group.

[b]*P* value between groups.

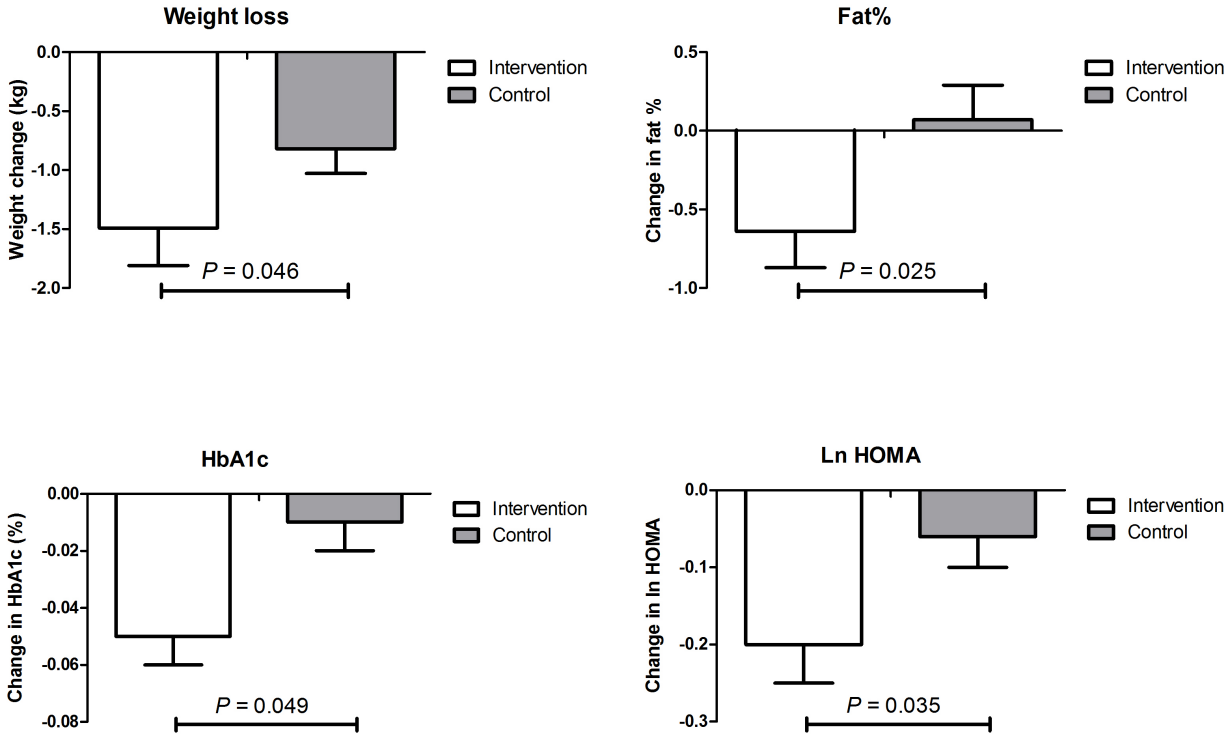
[c]BMI=body mass index.

[d]HOMA=homeostatic model assessment.

[e]HDL=high-density lipoprotein.

[f]LDL=low-density lipoprotein (Friedewald).

Figure 3. Changes in a selection of metabolic parameters.



Discussion

Principal Findings

In this study, we describe the effect of a Web-assisted physical activity intervention. We found a significant increase in daily physical activity in the intervention group compared to the control group. Furthermore, we showed that the intervention resulted in a significant improvement in body composition and parameters of glucose metabolism compared to the control group.

To our knowledge, our study is one of the first to show that a Web-based intervention can increase daily physical activity in older people from the general population. Attempts have been made to incorporate the Internet as an additional tool (eg, to generate computer-tailored advice) or intervention strategy (eg, forum, e-buddy) to promote physical activity in the elderly [25,26]. A recently published study showed the effects of a Web-based intervention in seniors, albeit with use of self-reported outcomes [27]. Furthermore, a Web intervention has been used in an elderly study population of 50 patients with chronic obstructive pulmonary disease (COPD) [28].The intervention was primarily directed at COPD self-management, used self-reported exercise as secondary outcome, and compared two different interventions instead of intervention vs a control group, making it difficult to compare their results with our study. Thus far, our study is the first to objectively assess the effects of a Web-based intervention on increasing physical activity in an older population. It is striking that—among an older and inactive population—a Web-based intervention was so effective. The baseline metabolic condition of our participants suggested that we included a study population with indication of the presence of metabolic syndrome: a large proportion of our study sample was overweight or obese, had high waist circumference, and/or used antihypertensive medication. This population, we believe, is particularly relevant for interventions aimed at prevention of age-related cardiovascular and metabolic diseases.

The increase in daily physical activity differed between accelerometer measurement locations; a more pronounced increase was found when using the ankle monitor. At the wrist, the increase in daily physical activity was smaller and not significantly different between groups when looking at total activity counts, but was statistically significant when assessing the increase in validated number of minutes spent in moderate-to-vigorous activity. The larger difference seen between groups in the ankle monitor compared to the wrist monitor is in line with our hypothesis that the ankle location is more sensitive to detect differences in daily physical activity such as cycling behavior. The Dutch population has the highest bicycle use worldwide, and it is likely that especially cycling and walking were stimulated in the inactive study population. In the control group, we also found a small but significant increase in daily physical activity as measured by wrist and ankle accelerometers, but not when assessing the number of minutes spent in moderate-to-vigorous activity. We expected an increase in physical activity in the control group as well, due to an increased awareness of physical activity because of the repeated trial-related measurements [29]. This finding illustrates the need for performing studies in a controlled manner and strengthens our finding that, despite this observation in the controls, the intervention group increased daily physical activity significantly more. Furthermore, there is a need for better understanding the measurements of physical activity patterns from the accelerometer to further underpin the behavior changes associated with such interventions.

Few data exist on the effect of Web interventions directed at physical activity on metabolic health in the general population. Three previous smaller Web-based intervention trials studied outcomes related to metabolic health in populations without chronic disease. Bosak et al studied the effect of a 6-week Web intervention directed at increasing physical activity in 22 participants (mean age 50 years) with metabolic syndrome but found no significant improvements in physical activity, fitness, or lipid levels [30]. After a 16-week Web-based physical activity intervention in inactive adults (mean age 41 years), Carr et al [12] found some improvement in triglycerides levels in the intervention group (n=14) only, but not in any other markers of body composition, lipid, or glucose metabolism, and differences were not compared to the control group (n=18). Hurling et al studied the effects of a 9-week Web-based intervention on anthropometric outcomes and blood pressure in 77 subjects (healthy adults with a mean age of 40 years) [13] and found a significant decrease in body fat between groups, but no effects regarding systolic or diastolic blood pressure. In our study, we also found no significant effect on systolic blood pressure and markers of glucose metabolism. Part of the explanation of why our results differ from the previous studies is that we studied a large sample of older participants with overweight and an inactive lifestyle, likely resulting in a higher burden of metabolic derangement. This has resulted in a higher statistical power to detect differences in health compared to the smaller studies. Alternatively, our intervention may have been especially effective since it was able to deliver personalized feedback on physical activity levels, thereby stimulating behavior change. Finally, we may have selected participants that were more motivated than the participants of the other studies, possibly as the result of chance or our selection process.

Limitations and Strengths

In our study, we used a waitlist control group. In general, a potential bias in this design is the existence of attention bias, meaning that results could be achieved due to the extra attention given to the intervention group compared to the control group, instead of intervention aspects such as goal setting and self-monitoring. However, the attention given to the intervention group in our study comprised emailing and contact with the coach, which was an essential part of the intervention program under study. Therefore, we believe that the intervention was effective in increasing physical activity and improving metabolic health.

The present study shows the large potential of using Web-assisted interventions for increasing physical activity and increasing metabolic health in a very relevant and aging population. We were able to include over 200 highly motivated participants and improve their metabolic profile within 3 months. However, it is unclear whether compliance can be sustained and whether long-term positive effects can be expected. The very few studies that reported on longer term follow-up and showed a significant increase in PA also after a shorter follow-up period, suggested that physical activity may increase further after 12 compared to 6 months [31]. This study, however, was performed in a primary care setting and used non-Web-based digital intervention methods such as face-to-face counseling. Evidence for long-term effectiveness of Web interventions is therefore required.

Our study population consisted of highly educated and motivated participants, which may hamper the generalizability of our results. Future study should assess the effects of Web-based interventions in elderly in a primary care setting using a population that better represents the general population. A drawback of the present study was that we did not record any dietary behaviors. It is possible that changes in diet account for a proportion of the observed beneficial effects on metabolic health. On one hand, it would be interesting to study which dietary changes are associated with increasing metabolic health, and insight in such behavior could increase the potential to increase effectiveness of such the intervention by specific coaching on this subject. On the other hand, a potential implicit role for dietary factors in the observed effect in the present study does not mitigate the relevance of these results. Another drawback is the fact that we did not have data on longer term follow-up.

The main strength of our study was the use of objectively measured daily physical activity. The majority of studies directed at increasing physical activity used self-reported physical activity measures, which could have resulted in an overestimation of the effect size. More recent studies used pedometers, which are unable to assess all types of physical activity and to give direct feedback to the wearer. With the use of tri-axial accelerometers, outcome assessment was blinded for participants as well as for study physicians and nurses. Another strength of the study was the use of a home-based intervention. This minimized the need for face-to-face contact and subsequently may have explained the low drop-out rate and high adherence to the intervention [6].

Conclusions

Our results show that using a Web-based intervention in older inactive people at risk for metabolic disease increases daily physical activity and improves metabolic health after 3 months. High retention rates in the intervention group were found, indicating that this Web-based intervention was feasible for use in an older population. Our findings show the large potential of Web-based interventions for large scale prevention of metabolic deregulation in a rapidly aging population.

Acknowledgments

We thank Marjan van der Elst, Robert du Puy, Leonieke ten Brinke, David Vroege, Marja Kersbergen, and Margo van Schie for their valuable contributions. This project was financially supported by Philips Consumer Lifestyle, and the Netherlands Genomics Initiative/Netherlands Organization for scientific research (NGI/NWO; 05040202 and 050-060-810). The funders had no role in the design and performance of the study, nor in the analyses or interpretation of the data or in the drafting of the manuscript.

CAW and SPM take full responsibility for the integrity of the data. Details on the CONSORT-EHEALTH checklist can be found in Multimedia Appendix 4.

Authors' Contributions

The authors contributed in the following ways: CAW, RGJW, EV, AJMdC, WvM, FvdO, and SPM designed the study. RGJW, FvdO, and SPM acquired funding. MC calculated accelerometer data. CAW and SPM coordinated participant recruitment, inclusion, and retention. CAW, AJMdC, and SPM performed statistical analyses. CAW, RGJW, EV, MC, PS, AJMdC, WvM, KB, DvH, FvdO, and SPM were involved in interpretation of the results and drafting of the manuscript. All authors reviewed the final version of the manuscript and agreed to its submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of DirectLife intervention program and accelerometer.

[PNG File, 529KB - jmir_v15i11e233_app1.png]

Multimedia Appendix 2

Example of DirectLife program and physical activity output.

[PNG File, 158KB - jmir_v15i11e233_app2.png]

Multimedia Appendix 3

Screenshot of a personal plan.

[PNG File, 452KB - jmir_v15i11e233_app3.png]

Multimedia Appendix 4

CONSORT-EHEALTH checklist V1.6.2 [32].

[PDF File (Adobe PDF File), 993KB - jmir_v15i11e233_app4.pdf]

References

1. Shortreed SM, Peeters A, Forbes AB. Estimating the effect of long-term physical activity on cardiovascular disease and mortality: evidence from the Framingham Heart Study. Heart 2013 May;99(9):649-654. [doi: 10.1136/heartjnl-2012-303461] [Medline: 23474622]
2. Tanasescu M, Leitzmann MF, Rimm EB, Hu FB. Physical activity in relation to cardiovascular disease and total mortality among men with type 2 diabetes. Circulation 2003 May 20;107(19):2435-2439 [FREE Full text] [doi: 10.1161/01.CIR.0000066906.11109.1F] [Medline: 12719277]
3. Manson JE, Nathan DM, Krolewski AS, Stampfer MJ, Willett WC, Hennekens CH. A prospective study of exercise and incidence of diabetes among US male physicians. JAMA 1992 Jul 1;268(1):63-67. [Medline: 1608115]
4. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. Total daily physical activity and the risk of AD and cognitive decline in older adults. Neurology 2012 Apr 24;78(17):1323-1329. [doi: 10.1212/WNL.0b013e3182535d35] [Medline: 22517108]
5. Yaffe K, Barnes D, Nevitt M, Lui LY, Covinsky K. A prospective study of physical activity and cognitive decline in elderly women: women who walk. Arch Intern Med 2001 Jul 23;161(14):1703-1708. [Medline: 11485502]
6. Matthews CE, George SM, Moore SC, Bowles HR, Blair A, Park Y, et al. Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. Am J Clin Nutr 2012 Feb;95(2):437-445 [FREE Full text] [doi: 10.3945/ajcn.111.019620] [Medline: 22218159]
7. Shaw BA, Agahi N. A prospective cohort study of health behavior profiles after age 50 and mortality risk. BMC Public Health 2012;12:803 [FREE Full text] [doi: 10.1186/1471-2458-12-803] [Medline: 22989155]
8. Hillsdon M, Foster C, Thorogood M. Interventions for promoting physical activity. Cochrane Database Syst Rev 2005(1):CD003180. [doi: 10.1002/14651858.CD003180.pub2] [Medline: 15674903]
9. Dubbert PM, Cooper KM, Kirchner KA, Meydrech EF, Bilbrew D. Effects of nurse counseling on walking for exercise in elderly primary care patients. J Gerontol A Biol Sci Med Sci 2002 Nov;57(11):M733-M740. [Medline: 12403802]
10. Fujita K, Nagatomi R, Hozawa A, Ohkubo T, Sato K, Anzai Y, et al. Effects of exercise training on physical activity in older people: a randomized controlled trial. J Epidemiol 2003 Mar;13(2):120-126 [FREE Full text] [Medline: 12675121]

11. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
12. Carr LJ, Bartee RT, Dorozynski C, Broomfield JF, Smith ML, Smith DT. Internet-delivered behavior change program increases physical activity and improves cardiometabolic disease risk factors in sedentary adults: results of a randomized controlled trial. Prev Med 2008 May;46(5):431-438. [doi: 10.1016/j.ypmed.2007.12.005] [Medline: 18207228]
13. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
14. Central Bureau of Statistics. ICT use in the Netherlands. 2013. URL: http://www.cbs.nl/nl-NL/menu/themas/vrije-tijd-cultuur/publicaties/artikelen/archief/2013/2013-3834-wm.htm [accessed 2013-07-23] [WebCite Cache ID 6IKRldGtM]
15. Physical Activity Policy and Health Improvement Directorate. 2013. The general practice physical activity questionnaire (GPPAQ): a screening tool to assess adult physical activity level, within primary care URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_063812 [accessed 2013-07-23] [WebCite Cache ID 6IKT5JweI]
16. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994 Jan;13(1):39-46. [Medline: 8168470]
17. De Vries H, Mudde AN, Dijkstra A, Willemsen MC. Differential beliefs, perceived social influences, and self-efficacy expectations among smokers in various motivational phases. Prev Med 1998;27(5 Pt 1):681-689. [doi: 10.1006/pmed.1998.0344] [Medline: 9808799]
18. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. Obesity (Silver Spring) 2010 Sep;18(9):1845-1851. [doi: 10.1038/oby.2010.34] [Medline: 20186133]
19. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One 2011;6(7):e22922 [FREE Full text] [doi: 10.1371/journal.pone.0022922] [Medline: 21829556]
20. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. Med Sci Sports Exerc 2012 Apr;44(4):733-741. [doi: 10.1249/MSS.0b013e3182351913] [Medline: 21904249]
21. Karabulut M, Crouter SE, Bassett DR. Comparison of two waist-mounted and two ankle-mounted electronic pedometers. Eur J Appl Physiol 2005 Oct;95(4):335-343. [doi: 10.1007/s00421-005-0018-3] [Medline: 16132120]
22. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Med Sci Sports Exerc 2011 Jun;43(6):1085-1093. [doi: 10.1249/MSS.0b013e31820513be] [Medline: 21088628]
23. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation,Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002 Dec 17;106(25):3143-3421 [FREE Full text] [Medline: 12485966]
24. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972 Jun;18(6):499-502 [FREE Full text] [Medline: 4337382]
25. Hageman PA, Walker SN, Pullen CH. Tailored versus standard internet-delivered interventions to promote physical activity in older women. J Geriatr Phys Ther 2005;28(1):28-33. [Medline: 16236225]
26. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. Health Psychol 2011 Jul;30(4):442-452. [doi: 10.1037/a0023579] [Medline: 21639638]
27. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]
28. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
29. van Sluijs EM, van Poppel MN, Twisk JW, van Mechelen W. Physical activity measurements affected participants' behavior in a randomized controlled trial. J Clin Epidemiol 2006 Apr;59(4):404-411. [doi: 10.1016/j.jclinepi.2005.08.016] [Medline: 16549263]
30. Bosak KA, Yates B, Pozehl B. Effects of an Internet physical activity intervention in adults with metabolic syndrome. West J Nurs Res 2010 Feb;32(1):5-22. [doi: 10.1177/0193945909333889] [Medline: 19357421]
31. Petrella RJ, Koval JJ, Cunningham DA, Paterson DH. Can primary care doctors prescribe exercise to improve fitness? The Step Test Exercise Prescription (STEP) project. Am J Prev Med 2003 May;24(4):316-322. [Medline: 12726869]
32. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BIA: bio-electrical impedance
BMI: body mass index
COPD: chronic obstructive pulmonary disease
GPPAQ: general physician physical activity questionnaire
HbA1c: glycated hemoglobin
HDL: high-density lipoprotein
HOMA: homeostatic model assessment
ITT: intention to treat
LDL: low-density lipoprotein
MET: metabolic equivalentœ
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SE: standard error

Edited by G Eysenbach; submitted 23.07.13; peer-reviewed by E van Sluijs, E Stanmore, M Kwan; comments to author 29.08.13; revised version received 12.09.13; accepted 18.09.13; published 06.11.13.

Please cite as:
Wijsman CA, Westendorp RGJ, Verhagen EALM, Catt M, Slagboom PE, de Craen AJM, Broekhuizen K, van Mechelen W, van Heemst D, van der Ouderaa F, Mooijaart SP
Effects of a Web-Based Intervention on Physical Activity and Metabolism in Older Adults: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e233
URL: http://www.jmir.org/2013/11/e233/
doi:10.2196/jmir.2843
PMID:24195965

Original Paper

Impact of Seasonality on Recruitment, Retention, Adherence, and Outcomes in a Web-Based Smoking Cessation Intervention: Randomized Controlled Trial

Amanda L Graham[1,2], PhD; Sarah Cha[1], MSPH; Nathan K Cobb[3,4,5], MD; Ye Fang[1], MS; Raymond S Niaura[1,2,3], PhD; Aaron Mushro[6], MS

[1]Schroeder Institute for Tobacco Research and Policy Studies, Legacy, Washington, DC, United States

[2]Department of Oncology, Georgetown University Medical Center / Cancer Prevention and Control, Lombardi Comprehensive Cancer Center, Washington, DC, United States

[3]Department of Health, Behavior and Society, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[4]Department of Pulmonary, Critical Care, and Sleep Medicine, Georgetown University Medical Center, Washington, DC, United States

[5]MeYou Health, Boston, MA, United States

[6]Marketing Department, Legacy, Washington, DC, United States

Corresponding Author:
Amanda L Graham, PhD
Schroeder Institute for Tobacco Research and Policy Studies
Legacy
1724 Massachusetts Avenue, NW
Washington, DC, 20036
United States
Phone: 1 202 454 5938
Fax: 1 202 454 5785
Email: agraham@legacyforhealth.org

Abstract

Background: Seasonal variations in smoking and quitting behaviors have been documented, with many smokers seeking cessation assistance around the start of the New Year. What remains unknown is whether smokers who are recruited to cessation treatment trials during the New Year are as motivated to quit, or as likely to enroll in a research trial, adhere to a research protocol, and benefit from a cessation intervention compared to those who are recruited during other times of the year.

Objective: The objective of this study was to determine whether smokers recruited during the New Year period differ on measures of motivation and desire to quit, recruitment and retention rates, website utilization rates, and short-term cessation outcomes compared to smokers recruited at other times.

Methods: Participants were current smokers who had registered on a free Web-based cessation program (BecomeAnEX.org) and were invited to participate in a clinical trial. The New Year period was defined according to a clear peak and drop in the proportion of visitors who registered on the site, spanning a 15-day period from December 26, 2012 to January 9, 2013. Two other 15-day recruitment periods during summer (July 18, 2012 to August 1, 2012) and fall (November 7, 2012 to November 21, 2012) were selected for comparison. Data were examined from 3 sources: (1) a Web-based clinical trials management system that automated the recruitment and enrollment process, (2) self-report assessments at baseline and 3 months postrandomization, and (3) online tracking software that recorded website utilization during the first 3 months of the trial.

Results: Visitors to BecomeAnEX during the New Year period were more likely to register on the site than smokers who visited during summer or fall (conversion rates: 7.4%, 4.6%, 4.9%, respectively; $P<.001$), but there were no differences in rates of study acceptance, consent, randomization, 3-month follow-up survey completion, or cessation between the 3 periods. New Year participants were older, more educated, more likely to be employed full time, and more likely to have a relationship partner compared with participants recruited at other times during the year, but did not differ on measures of motivation and desire to quit.

Conclusions: Smokers visiting a Web-based cessation program during the New Year period were more likely to register for treatment and differ on several demographic variables, but showed similar patterns of treatment engagement, retention, follow-up, and short-term cessation outcomes compared with participants who visited the site during other periods of the year. These results

allay scientific concerns about recruiting participants during this time frame and are reassuring for researchers conducting Web-based cessation trials.

Trial Registration: ClinicalTrials.gov ID: NCT01544153; http://clinicaltrials.gov/ct2/show/NCT01544153 (Archived by WebCite at http://www.webcitation.org/6KjhmAS9u).

(J Med Internet Res 2013;15(11):e249) doi:10.2196/jmir.2880

KEYWORDS

seasonal variation; smoking cessation; Internet; research subject recruitment

Introduction

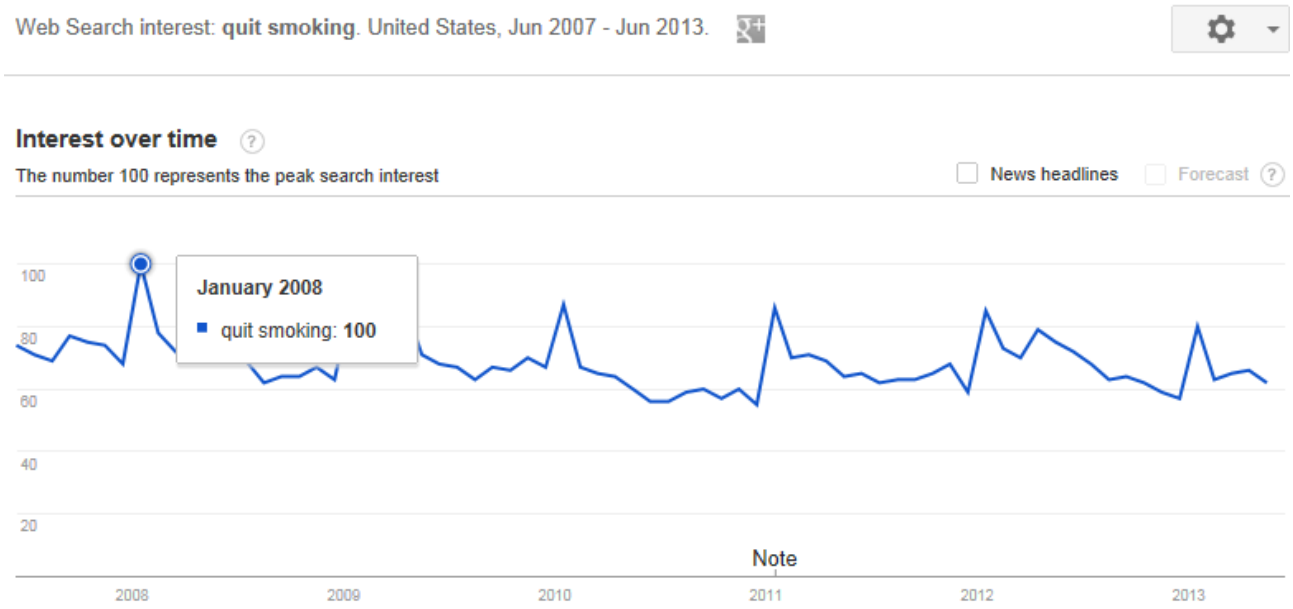
Seasonal variations across a number of smoking and quitting behaviors have been documented. Most smokers express a desire to quit [1] and many make a quit attempt around the start of the New Year [2-6]. Reports have shown that sales of cigarettes are at their lowest during January and February [7,8] and sales of nicotine replacement therapies are at their highest January through March [9]. Seasonal variations in motivational stage of change among callers to a state quitline have also been documented [10] with callers in December and January being more likely to have recently quit than callers during other months. Internet search queries also provide evidence of the seasonal variations in smoking cessation, with clear peaks observed in the use of "quit smoking" as a search term at the beginning of each calendar year. Figure 1 shows the relative use of the search term "quit smoking" in Google search engine queries over 6 years in the United States as reported by Google Trends [11], the public database of Google queries.

A greater number of smokers quitting around the New Year may mean a greater pool of potential research participants for smoking cessation trials. However, the effects of seasonality on research recruitment and retention have not been documented. Specifically, it is unknown whether smokers who are invited to cessation treatment trials during this period of time are as likely to enroll, to adhere to a research protocol, and to benefit from a cessation intervention. Smokers who elect to quit around the New Year may differ from those who quit during other times of the year on factors such as motivation, desire, confidence, or other factors that relate to trial participation, engagement, and cessation outcomes.

These are important questions to address from both a pragmatic and a scientific standpoint. From a pragmatic standpoint, conducting trial recruitment during the New Year holiday may have staffing and cost considerations for all aspects of a trial. Research staff may be needed to field study inquiries, conduct eligibility screening, administer assessments, and manage study communications; intervention staff may be required to orient new participants to the trial and begin intervention delivery. Given the potential for higher recruitment volume during this period, staffing increases may be required. From a scientific standpoint, if participants enrolled during this time are less likely to adhere to research protocol (ie, lower rates of intervention adherence, lower retention at follow-up) because of a more transient commitment to quitting, this could have important implications for treatment trials. Lower retention rates (ie, higher loss to follow-up) would result in a higher proportion of participants counted as smokers in intention-to-treat analyses, which may artificially deflate overall abstinence rates and perceived effectiveness of an intervention. Lower rates of intervention adherence among participants recruited during the New Year could influence metrics of intervention feasibility and receptivity as well as cessation outcomes.

We sought to examine these questions about the impact of seasonality on smoking cessation treatment trials in the context of an ongoing randomized trial of a Web-based cessation intervention. We extracted a subset of participants recruited during a 15-day window that spanned the 2013 New Year and compared them to participants recruited during 2 other 15-day periods in 2012 on recruitment and retention rates, baseline characteristics, website utilization, and cessation outcomes. Our a priori hypotheses were that New Year participants would differ on baseline measures of motivation and desire to quit consistent with a more transient commitment to quitting, and would have lower recruitment and retention rates, lower website utilization rates, and poorer cessation outcomes compared to participants enrolled during other periods.

Figure 1. Use of search term "quit smoking" in Google search engine queries relative to the total number of Google searches between June 2007 and June 2013 in the United States as reported in Google Trends.



Methods

Study Overview

The full study protocol has been published elsewhere [12]. Briefly, this is an ongoing Web-based randomized trial to compare the efficacy of an interactive, evidence-based, smoking cessation website alone and in conjunction with (1) a theory-driven, empirically informed social network intervention designed to integrate participants into an online community, and (2) access to a free supply of nicotine replacement therapy (NRT) products. The study uses a 2×2 factorial design to compare the following treatment conditions: (1) website, (2) website+social network intervention, (3) website+NRT, and (4) website+social network intervention+NRT. A total of 4000 participants will be randomized by the end of the study. Follow-up assessments are administered at 3- and 9-months postrandomization; 30-day point prevalence abstinence is the primary outcome of the parent trial. Study eligibility criteria are current smoking, age 18 years or older, and US residence. Exclusion criteria are contraindications to NRT (pregnant or breastfeeding, recent cardiac problems, current NRT use). Randomization is stratified by gender and baseline motivation.

Recruitment

The study is conducted within BecomeAnEX.org, a free, publicly available, evidence-based intervention developed in accordance with the 2008 US Department of Health and Human Service's Clinical Practice Guidelines [13]. The site was developed by Legacy, a nonprofit organization that develops smoking prevention and cessation programs, in collaboration with the Mayo Clinic Nicotine Dependence Center [14]. A national multichannel media campaign that included television, radio, and outdoor and online advertising was launched in 2008 to promote the website [15]. The present implementation of this campaign relies on various forms of online advertising, including social media, search engine marketing, and large targeted ad networks for display advertising. Search engine advertising targets keywords related to BecomeAnEX (eg, quit smoking, stop smoking) and display advertising targets males and females aged between 25 and 54 years.

Participants are recruited to the trial immediately following registration on BecomeAnEX. The entire recruitment and enrollment process is automated using a Web-based clinical trials management system. Individuals that indicate current smoking (every day/some days) during registration are invited to the study. Interested individuals complete online eligibility screening; eligible individuals provide online informed consent and contact information, including an email address that is used to send a link to the online baseline assessment. Participants are randomized to treatment upon completion of the baseline survey. No incentive is provided for enrollment in the study. Recruitment volume is capped at a maximum of 10 new participants per day to ensure a manageable workload for intervention and research staff throughout the study period. Once 10 individuals are randomized, no new registered users are invited for the remainder of the 24-hour period. Recruitment began in March 2012. As of October 30, 2013, 3602 participants have been randomized.

We defined the New Year period based on a clear peak and drop in the number of individuals that registered on BecomeAnEX between December 1, 2012 and January 31, 2013. The average conversion rate of unique visitors to registrants each day from December 1 through December 25 was approximately 4.7%. This proportion increased almost 2-fold on December 26, 2012 to 8.2% and stayed elevated through January 9, 2013, at an average daily conversion rate of 7.4%. Thus, we selected this 15-day period as our New Year period. For comparison, we selected 2 other 15-day periods during the year based on several criteria: (1) similar marketing and promotion approach, (2) variations in season (ie, summer, fall), (3) same span of days of the week (Wednesday to Wednesday), and (4) roughly similar number of participants randomized during the designated time period. Based on these factors, 2 separate 15-day periods were

selected for comparison: 1 during the summer (July 18, 2012 to August 1, 2012) and 1 during the fall (November 7, 2012 to November 21, 2012). We deliberately selected the fall period to include another popular quitting holiday, the American Cancer Society's Great American Smokeout, which falls on the third Thursday of November (November 15, 2012). Inclusion of this time frame enabled us to compare participants enrolling during the New Year to participants potentially enrolling in response to another seasonal trigger for cessation.

Interventions

Participants in all 4 treatment groups had full access to the BecomeAnEX website which provides assistance setting a quit date, assessment of motivation and nicotine dependence, problem-solving/skills training to enhance self-efficacy for quitting, assistance in selecting and using US Food and Drug Administration (FDA)-approved pharmacotherapies, and social support through a large online community [14,15]. Participants randomized to receive the social network intervention received proactive communications from established members of the BecomeAnEX community (integrators). Within 24 hours after a new participant joined the study, the integrators posted a public message on the new member's profile page to welcome them to the site, encourage them to fill out their profile, or comment on some aspect of an existing profile. Participants randomized to receive NRT products from the study were mailed a free 4-week supply of the NRT product of their choice (patch, gum, or lozenge) within 3 days of randomization. The NRT is provided as an over-the-counter product (ie, with no additional support or guidance provided) to parallel the experience participants would have if they purchased NRT on their own.

Data Collection

Data are obtained through 3 sources: (1) a Web-based clinical trials management system that automated the recruitment and enrollment process, (2) self-report assessments at baseline, 3-, and 9-months postrandomization, and (3) online tracking software that records utilization of BecomeAnEX. Our analyses of smoking outcomes focus on the 3-month follow-up because this is typically when treatment utilization and intervention effects are the strongest. Telephone follow-up by professional telephone interviewers blinded to treatment condition for online nonresponders is used to maximize follow-up rates. Participants are reimbursed via Amazon or PayPal for survey completion (US $20 for Web survey, US $15 for phone survey). Individual level tracking metrics of BecomeAnEX utilization are recorded using Adobe/Omniture SiteCatalyst [16] software.

Measures

Overview

The following measures from the parent trial were examined for these analyses.

Sociodemographic Variables

Participants reported age, gender, race, ethnicity, marital status, employment, education, and type of Internet connection.

Smoking Variables

At baseline, participants completed the Fagerström Test for Nicotine Dependence (FTND) [17] and also reported their confidence and desire to quit smoking (1=not at all, 5=very much), the current number of cigarettes smoked per day, the number of quit attempts in the previous year, and motivation to quit [18].

Psychosocial Variables

The appraisal and belonging subscales of the 12-item Interpersonal Support Evaluation List (ISEL) [19] were used to measure perceived availability of social resources at baseline. The appraisal subscale measures the perceived availability of someone to talk to about one's problems; the belonging subscale assesses the perceived availability of people with whom to engage in activities. Perceptions of cessation-related social support are measured at baseline and follow-up with a 6-item version of the Partner Interaction Questionnaire [20,21] that assesses receipt of positive behaviors (supportive of cessation) and negative behaviors (harmful to cessation) from an individual who has followed the participant's efforts to quit smoking.

Treatment Adherence

Website utilization during the first 3 months of the study was extracted from the BecomeAnEX database and included the following metrics: number of log-ins, minutes spent using the site during each visit/session, number of pages viewed during each visit/session, and the number of blog posts read and made. Website utilization was recorded using Adobe/Omniture SiteCatalyst. Each page view by a participant was recorded into a relational database, and page views were grouped into sessions. The duration of a session was defined as the time elapsed between the first page view and the last page view in a given session. If a user did not view a new page for more than 30 minutes, the system marked them as inactive and their next return visit created a new session. At each follow-up, participants reported use of NRT and prescription cessation medications (eg, Chantix, bupropion).

Three-Month Outcome Measures

Smoking outcomes examined in these analyses included self-reported point prevalence abstinence (30 day and 7 day) measured at 3 months. We also examined the number of quit attempts reported at 3 months.

Statistical Analyses

The effects of recruitment phase (New Year, summer, fall) on recruitment metrics, baseline characteristics, treatment utilization, and outcome measures were evaluated via chi-square tests for proportions or 1-way ANOVA, depending on whether the metrics were proportions or continuous variables. Significant omnibus tests were followed by unadjusted pairwise comparisons. Analyses were conducted on the full sample of participants recruited in each phase (ie, collapsed across treatment groups).

Results

Recruitment and Retention by Recruitment Phase

Table 1 shows the Consolidated Standards of Reporting Trials (CONSORT) metrics of each recruitment phase. We examined advertising expenditures during the 2 weeks before each recruitment phase in addition to the recruitment phase itself. For the New Year, summer, and fall phases, expenditures totaled US $57,508, US $48,632, and US $38,374, respectively. The conversion rate of unique visitors to new registered users during the New Year period (7.4%) was significantly higher than both summer (4.6%) and fall (4.9%) periods; summer and fall conversion rates did not differ. Among new registered users during the New Year period, 868 were invited to participate in the study. Of these, 44.1% (383/868) accepted the invitation and completed eligibility screening; 67.1% (257/383) were eligible, 93.0% (239/257) consented, and 52.9% of those eligible (136/257) completed the baseline assessment and were randomized to treatment. Among the New Year participants, 58.8% (80/136) completed the 3-month follow-up survey. With the exception of conversion rate, there were no significant differences between recruitment phases noted for any of the CONSORT metrics.

Baseline Characteristics by Recruitment Phase

New Year participants differed from participants recruited during other time periods on age, education, employment, and marital status (see Table 2). New Year participants were older (mean 43.2, SD 12.3 vs mean 39.1, SD 13.3; *P*=.01) and more likely to be employed full time (58.9%, 79/136 vs 40.6%, 43/106; *P*=.01) compared to summer participants. New Year participants were more likely to have attended college than both summer and fall participants (New Year: 80.9%, 110/136; summer: 66.0%, 70/106; fall: 68.7%, 68/99; *P*=.02). Both New Year and summer participants were more likely to have a spouse/partner compared to fall participants (New Year: 63.2%, 86/136; summer: 65.1%, 69/106; fall: 47.5%, 47/99; *P*=.02). Summer and fall participants differed on Internet access (*P*=.004), but were not different from New Year enrollees.

Treatment Utilization Metrics by Recruitment Phase

Among the various utilization metrics we examined (Table 3), only the number of total Web pages viewed differed significantly between the 3 groups of participants: page views was higher among New Year participants (median 57, IQR 20-57) than both summer (median 29, IQR 13-59; *P*=.002) and fall enrollees (median 36, IQR 19-69; *P*=.004). Summer and fall participants did not differ on page views.

Smoking Outcomes by Time of Enrollment

There were no significant differences in any of the smoking outcomes we examined by recruitment phase. Using intention-to-treat analyses, 30-day point prevalence abstinence was 11.8% (16/136), 15.1% (16/106), and 17.2% (17/99), and 7-day point prevalence abstinence was 16.2% (22/136), 18.9% (20/106), and 22.2% (22/99) for New Year, summer, and fall participants, respectively. Using responder-only analysis, 30-day point prevalence abstinence was 20% (16/79), 23% (16/68), and 32% (17/53), and 7-day point prevalence abstinence was 28% (22/79), 29% (20/68), and 41% (22/53) for New Year, summer, and fall participants, respectively. There was no difference in the number of quit attempts reported by the 3 groups at the 3-month follow-up (Table 4).

Table 1. Recruitment and retention metrics by recruitment phase.

Recruitment and retention metrics	New Year	Summer	Fall	*P* value
Total advertising expenditure (US $)	57,508	48,632	38,374	—
Unique visitors, n	32,853	30,605	22,017	—
New registered users, n	2424	1404	1079	—
Conversion rate, %	7.4	4.6	4.9	.001
Invited to study, n	868	792	594	—
Accepted invitation, n (% of invited to study)	383 (44.1)	325 (41.0)	270 (45.5)	.21
Eligible, n (% of accepted invitation)	257 (67.1)	218 (67.1)	186 (68.9)	.88
Consented, n (% of eligible)	239 (93.0)	205 (94.0)	176 (94.6)	.77
Randomized, n (% of eligible)	136 (52.9)	106 (48.6)	99 (53.2)	.79
Completed 3-month follow-up, n (% randomized)	80 (58.8)	69 (65.1)	54 (54.5)	.30

Table 2. Baseline characteristics by recruitment phase.

Baseline characteristics	New Year n=136	Summer n=106	Fall n=99	*P* value
Demographic variables				
Age (years), mean (SD)	43.2 (12.3)	39.1 (13.3)	40.6 (13.1)	.02
Sex (female), n (%)	97 (71.3)	76 (71.7)	68 (68.7)	.87
Race, n (%)				.78
Non-white	22 (16.2)	14 (13.2)	16 (16.2)	
White	114 (83.8)	92 (86.8)	83 (83.8)	
Ethnicity (Hispanic), n (%)	8 (5.9)	3 (2.8)	3 (3.0)	.45
Education, n (%)				.02
High school or less	26 (19.1)	36 (34.0)	31 (31.3)	
Some college or more	110 (80.9)	70 (66.0)	68 (68.7)	
Employment status, n (%)				.02
Full time	79 (58.9)	43 (40.6)	52 (52.5)	
Not full time	57 (41.9)	63 (59.4)	47 (47.4)	
Marital status, n (%)				.02
Partner	86 (63.2)	69 (65.1)	47 (47.5)	
No partner	50 (36.8)	37 (34.9)	52 (52.5)	
Smoking variables				
Cigarettes per day, mean (SD)	17.3 (8.1)	16.4 (7.6)	16.2 (8.3)	.54
Motivation to quit, [a] n (%)				.58
Next 30 days	116 (85.9)	87 (82.1)	86 (87.9)	
Next 6 months	19 (14.1)	19 (17.9)	13 (13.1)	
Desire to quit, mean (SD)	4.6 (0.6)	4.6 (0.6)	4.6 (0.6)	.94
Confidence in quitting, mean (SD)	3.3 (1.0)	3.3 (1.0)	3.4 (1.2)	.81
Quit attempts past year, mean (SD)	2.4 (3.0)	1.8 (2.6)	2.6 (3.3)	.10
FTND,[b] mean (SD)	5.6 (2.2)	5.2 (2.2)	5.0 (2.1)	.08
Psychosocial variables				
Partner Interaction Questionnaire, mean (SD)				
Positive subscale	9.1 (2.9)	9.8 (2.3)	9.3 (3.0)	.52
Negative subscale	6.8 (4.2)	5.9 (4.1)	7.1 (4.0)	.25
ISEL, [c] mean (SD)				
Appraisal subscale	8.2 (3.4)	8.1 (3.1)	8.4 (3.3)	.77
Belonging subscale	7.8 (2.9)	7.9 (3.1)	8.4 (3.0)	.29
Internet access, [d] n (%)				.02
High speed/broadband	108 (80.0)	92 (88.5)	72 (72.7)	
Mobile device	27 (20.0)	12 (11.5)	27 (27.3)	

[a]Motivation to quit excluded 1 participant who reported no plans to quit smoking in New Year group.

[b]FTND: Fagerström Test for Nicotine Dependence.

[c]ISEL: Interpersonal Support Evaluation List.

[d]Internet access: n=3 cases dropped that reported using a dial-up connection (summer: n=2; New Year: n=1; fall: n=0).

Table 3. Treatment utilization metrics by recruitment phase.

Treatment utilization metrics	New Year n=136	Summer n=106	Fall n=99	*P* value
Log-ins, median (IQR)[a]	2 (1-4)	2 (1-4)	2 (1-5)	.63
Return visits, n (%)				.73
None	50 (36.8)	48 (45.3)	42 (42.4)	
1	28 (20.6)	18 (17.0)	17 (17.2)	
≥2	58 (42.6)	40 (37.7)	40 (40.4)	
Time on site, median (IQR)	41 (20.5-86)	29 (13-59)	40 (16.5-64)	.11
Page views, median (IQR)	57 (20-57)	29 (13-59)	36 (19-69)	.02
Blogs read, n (%)				.71
None	95 (69.9)	72 (67.9)	71 (71.7)	
1	9 (6.6)	12 (11.3)	9 (9.1)	
≥2	32 (23.5)	22 (20.8)	19 (19.2)	
Blog posts, n (%)				.45
None	117 (86.0)	95 (89.6)	89 (89.9)	
1	3 (2.2)	4 (3.8)	4 (4.0)	
≥2	16 (11.8)	7 (6.6)	6 (6.1)	
Any NRT use (yes), n (%)	48 (60.0)	37 (53.6)	36 (66.7)	.34

[a]12 missing values (New Year: n=7; summer: n=3; fall: n=2).

Table 4. Smoking outcomes by recruitment phase.

Smoking outcomes	New Year	Summer	Fall	*P* value
30-day point prevalence abstinence, n (%)				
Intention-to-treat	16 (11.8)	16 (15.1)	17 (17.2)	.33
Responder only	16 (20.0)	16 (23.2)	17 (31.5)	.31
7-day point prevalence abstinence, n (%)				
Intention-to-treat	22 (16.2)	20 (18.9)	22 (22.2)	.27
Responder only	22 (27.5)	20 (29.0)	22 (40.7)	.23
Quit attempts, mean (SD)	3.3 (4.1)	4.3 (8.6)	3.7 (4.7)	.58

Discussion

Principal Findings

The results of this study indicate that smokers visiting a Web-based cessation program during the New Year period are more likely to register for treatment and differ on several demographic variables, but show similar patterns of treatment engagement, retention, and short-term cessation outcomes compared with participants who visit the site during other periods of the year. Our hypotheses that New Year participants would differ on measures of motivation and desire to quit were not supported, and there were no differences on any of the smoking variables we examined. In addition, our hypotheses about lower retention rates, website utilization rates, and cessation outcomes were also not supported. Follow-up rates were comparable across all 3 periods, and smokers recruited during the New Year period quit at the same rate as smokers recruited at other times during the year. These results mitigate scientific concerns about recruiting participants during this time frame and are reassuring for researchers conducting Web-based cessation trials.

Our findings that New Year participants were older, more educated, more likely to be employed full time, and more likely to have a relationship partner may suggest that smokers with greater resources are more affected by the seasonal trends of quitting around the New Year. Alternatively, these differences may be a function of differential message exposure: older employed individuals may have been more likely to be impacted by the BecomeAnEX online advertising campaign, or reminded of cessation through workplace wellness programs or other promotional activities. Although not significant, smokers recruited during the New Year period also had a higher level of nicotine dependence and a higher number of previous quit attempts at baseline, also suggesting that seasonal trends may serve as a cue to action for more dependent and motivated

smokers. We do not have an explanation for the finding that New Year participants viewed more website pages than participants in other recruitment phases did, especially because no other metric of engagement or utilization was significantly different.

To our knowledge, this is the first study to examine demographic, utilization, and outcome patterns of smokers recruited to a Web-based intervention during the New Year period, although there is interest in seasonal patterns of behavior. Delnevo and colleagues [10] noted seasonal patterns in motivation for quitting among callers to a quitline and discussed the implications for planning, promoting, and evaluating telephone quitlines. Using Internet search query data, Ayers et al [22] documented seasonality in searches for mental health information, with increases in information seeking that corresponded to patterns of seasonal affective disorder. The lack of previous publications in this area may be because dramatic increases in recruitment are to some degree unique to the online environment and Web-based studies that are capable of enrolling a large number of trial participants in a relatively short period of time.

Our findings add to the small but growing literature on recruitment methods for Web-based tobacco interventions [23-31]. Most studies to date have focused on comparisons between online recruitment methods (eg, online banner ads, search engine advertising) and more traditional recruitment methods (eg, newspaper ads, targeted mailings) [24,25,32], evaluation of different Internet-based methods [28,29], or the use of offline methods (eg, physician referral) to drive tobacco users to Web-based interventions [30,31]. The primary endpoints of interest in most studies are baseline participant characteristics and recruitment yield and/or efficiency. Heffner et al [25] evaluated the impact of Web-based and traditional recruitment methods on 3-month data retention and 30-day point prevalence smoking abstinence at the 3-month outcome assessment in a cessation trial and found no differences by recruitment method. Our findings regarding the impact of seasonality are consistent with previous studies that have demonstrated some differences in the types of participants recruited to Web-based tobacco interventions, but no differences in their participation in or outcomes from such trials.

Limitations

Several limitations to this study should be noted. We examined variations in participant characteristics for a single cessation website as part of an ongoing randomized trial. This site exists in a larger ecosystem of promotion, advertising, and branding as part of the national BecomeAnEX campaign, which has been in existence since 2008. Our results are likely related to the specific strategy employed by BecomeAnEX, which is largely online advertising in its present implementation. Other advertising and promotional strategies of different Web-based interventions could yield different results. In addition, our automated titration of recruitment volume should be noted when considering the pragmatic implications of our results. Although the number of visitors to BecomeAnEX was higher during the New Year period and a higher proportion of them registered to become members, the number of participants recruited to our clinical trial during different periods throughout the year has remained relatively constant because of the daily cap we have on enrollment. This cap is designed to maintain a consistent volume of participants for our research and intervention staff to manage. If this cap were not in place, we may have seen a higher number of participants invited to the study and differences in the proportion of participants accepting or declining the study invitation. This may have important pragmatic considerations for other Web-based trials that have human involvement, but we believe this is unlikely to have affected the other metrics we examined (ie, follow-up rates, cessation outcomes). The daily cap on recruitment may also have affected statistical power. The response rate to the 3-month follow-up is lower than desired despite numerous online and offline strategies to reach participants, but is comparable to or higher than other Internet studies [33].

Conclusions

Internet interventions for health behavior change are characterized by their ability to recruit broadly and provide treatment at scale. Secular or temporal variations, such as the New Year holiday, and the associated media attention to smoking cessation and resolution making can result in large-scale swings in the number of individuals arriving at Web-based cessation interventions. For interventions that can effectively capture and enroll those individuals, seasonal variations could dramatically increase recruitment efficiency for clinical trials.

Acknowledgments

Primary funding for this work was from the National Cancer Institute at the National Institutes of Health (1R01CA155489-01A1). The funding agency had no involvement in the conduct of the study or preparation of this manuscript.

Conflicts of Interest

A Graham, S Cha, Y Fang, R Niaura, and A Mushro are employees of Legacy, a nonprofit public health foundation that runs BecomeAnEX.org, an online tobacco cessation intervention. N Cobb is an employee of MeYou Health LLC, whose parent company's product line includes an online tobacco cessation intervention.

References

1. Centers for Disease Control and Prevention (CDC). Quitting smoking among adults--United States, 2001-2010. MMWR Morb Mortal Wkly Rep 2011 Nov 11;60(44):1513-1519 [FREE Full text] [Medline: 22071589]
2. Legacy. 2013 Jan 8. New Poll Finds Millions of Smokers Nationwide Resolving to Quit in the New Year URL: http://www.legacyforhealth.org/newsroom/press-releases/new-poll-finds-millions-of-smokers-nationwide-resolving-to-quit-in-the-new-year [accessed 2013-08-16] [WebCite Cache ID 6Iv9ScLHo]
3. Norcross JC, Mrykalo MS, Blagys MD. Auld lang syne: success predictors, change processes, and self-reported outcomes of New Year's resolvers and nonresolvers. J Clin Psychol 2002 Apr;58(4):397-405. [Medline: 11920693]
4. Norcross JC, Ratzin AC, Payne D. Ringing in the new year: the change processes and reported outcomes of resolutions. Addict Behav 1989;14(2):205-212. [Medline: 2728957]
5. Norcross JC, Vangarelli DJ. The resolution solution: longitudinal examination of New Year's change attempts. J Subst Abuse 1988;1(2):127-134. [Medline: 2980864]
6. Walsh RA, Paul CL. New Year quit smoking resolutions. Aust N Z J Public Health 2002 Apr;26(2):181. [Medline: 12054341]
7. Chandra S, Chaloupka FJ. Seasonality in cigarette sales: patterns and implications for tobacco control. Tob Control 2003 Mar;12(1):105-107 [FREE Full text] [Medline: 12612375]
8. Momperousse D, Delnevo CD, Lewis MJ. Exploring the seasonality of cigarette-smoking behaviour. Tob Control 2007 Feb;16(1):69-70 [FREE Full text] [doi: 10.1136/tc.2006.018135] [Medline: 17297078]
9. Chandra S, Gitchell JG, Shiffman S. Seasonality in sales of nicotine replacement therapies: patterns and implications for tobacco control. Nicotine Tob Res 2011 May;13(5):395-398. [doi: 10.1093/ntr/ntq258] [Medline: 21330268]
10. Delnevo CN, Foulds J, Vorbach U, Kazimir E. Seasonal variations in stage of change among Quitline clients. Tob Control 2006 Feb;15(1):70-71 [FREE Full text] [doi: 10.1136/tc.2005.013847] [Medline: 16436412]
11. Google. Trends URL: http://www.google.com/trends/ [accessed 2013-10-26] [WebCite Cache ID 6Kf2YhmeL]
12. Graham AL, Cha S, Papandonatos GD, Cobb NK, Mushro A, Fang Y, et al. Improving adherence to web-based cessation programs: a randomized controlled trial study protocol. Trials 2013;14:48 [FREE Full text] [doi: 10.1186/1745-6215-14-48] [Medline: 23414086]
13. Fiore M, Jaen CR, Baker TB, Bailey WC, Benowitz N, Curry SJ. Clinical Practice Guideline: Treating Tobacco Use and Dependence 2008 Update. Rockville, MD: US Department of Health and Human Services, Public Health Service; 2008. URL: http://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/tobacco/clinicians/treating_tobacco_use08.pdf [accessed 2013-10-31] [WebCite Cache ID 6KmkxW9oO]
14. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
15. McCausland JK, Curry LE, Mushro A, Carothers S, Xiao H, Vallone DM. Cases in Public Health Communication & Marketing. 2011. Promoting a Web-Based Smoking Cessation Intervention: Implications for Practice URL: http://sphhs.gwu.edu/departments/pch/phcm/casesjournal/volume5summer/peer-reviewed/V5_Case1PR.cfm [accessed 2013-08-16] [WebCite Cache ID 6Iv3IsLfE]
16. Adobe. 2012. Adobe Analytics: Marketing reports and analytics URL: http://www.adobe.com/solutions/digital-analytics/marketing-reports-analytics.html [accessed 2013-08-16] [WebCite Cache ID 6Iv2AztoH]
17. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [Medline: 1932883]
18. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [Medline: 1329589]
19. Cohen S, Hoberman HM. Positive events and social supports as buffers of life change stress. J Appl Social Psychol 1983 Apr;13(2):99-125. [doi: 10.1111/j.1559-1816.1983.tb02325.x]
20. Graham AL, Papandonatos GD. Reliability of internet- versus telephone-administered questionnaires in a diverse sample of smokers. J Med Internet Res 2008;10(1):e8 [FREE Full text] [doi: 10.2196/jmir.987] [Medline: 18364345]
21. Graham AL, Papandonatos GD, Bock BC, Cobb NK, Baskin-Sommers A, Niaura R, et al. Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation. Nicotine Tob Res 2006 Dec;8 Suppl 1:S49-S57 [FREE Full text] [Medline: 17491171]
22. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
23. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: a case of denominators. Nicotine Tob Res 2006 Dec;8 Suppl 1:S43-S48 [FREE Full text] [Medline: 17491170]
24. 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]

25. 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. [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]
26. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
27. McClure JB, Greene SM, Wiese C, Johnson KE, Alexander G, Strecher V. Interest in an online smoking cessation program and effective recruitment strategies: results from Project Quit. J Med Internet Res 2006;8(3):e14 [FREE Full text] [doi: 10.2196/jmir.8.3.e14] [Medline: 16954124]
28. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
29. 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]
30. Sadasivam RS, Cutrona SL, Volz E, Rao SR, Houston TK. Web-based peer-driven chain referrals for smoking cessation. Stud Health Technol Inform 2013;192:357-361. [Medline: 23920576]
31. Sadasivam RS, Kinney RL, Delaughter K, Rao SR, Williams JH, Coley HL, National Dental PBRN Group, QUIT-PRIMO Collaborative Group. Who participates in Web-assisted tobacco interventions? The QUIT-PRIMO and National Dental Practice-Based Research Network Hi-Quit studies. J Med Internet Res 2013;15(5):e77 [FREE Full text] [doi: 10.2196/jmir.2385] [Medline: 23635417]
32. 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]
33. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 1;20(3):568-576. [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FDA: US Food and Drug Administration
FTND: Fagerström Test for Nicotine Dependence
ISEL: Interpersonal Support Evaluation List
NRT: nicotine replacement therapy

Edited by G Eysenbach; submitted 16.08.13; peer-reviewed by J Bricker, N Zhang; comments to author 09.09.13; revised version received 17.10.13; accepted 21.10.13; published 07.11.13.

Please cite as:
Graham AL, Cha S, Cobb NK, Fang Y, Niaura RS, Mushro A
Impact of Seasonality on Recruitment, Retention, Adherence, and Outcomes in a Web-Based Smoking Cessation Intervention: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e249
URL: http://www.jmir.org/2013/11/e249/
doi:10.2196/jmir.2880
PMID:24201304

Original Paper

Directed Use of the Internet for Health Information by Patients With Chronic Kidney Disease: Prospective Cohort Study

Clarissa Jonas Diamantidis[1,2], MD, MHS; Wanda Fink[1], MS, RN; Shiming Yang[3], MS; Marni R Zuckerman[1], MA; Jennifer Ginsberg[1], MS; Peter Hu[4], PhD; Yan Xiao[5], PhD; Jeffrey C Fink[1,2,6], MS, MD

[1]Department of Medicine, Division of Nephrology, University of Maryland School of Medicine, Baltimore, MD, United States

[2]Department of Medicine, Division of Nephrology, Veterans Affairs Maryland Health Care System, Baltimore, MD, United States

[3]Department of Computer Science, University of Maryland Baltimore County, Baltimore, MD, United States

[4]Department of Anesthesiology, R Adams Cowley Shock Trauma Center, University of Maryland, Baltimore, MD, United States

[5]Baylor Health Care Quality Institute for Health Care Research and Improvement, Office of Patient Safety, Baylor Health Care System, Dallas, TX, United States

[6]Department of Epidemiology and Preventive Medicine, School of Medicine, University of Maryland, Baltimore, MD, United States

Corresponding Author:
Clarissa Jonas Diamantidis, MD, MHS
Department of Medicine
Division of Nephrology
University of Maryland School of Medicine
22 S Greene Street
Rm N3W143
Baltimore, MD,
United States
Phone: 1 410 328 5720
Fax: 1 410 328 5685
Email: cdiamantidis@medicine.umaryland.edu

Abstract

Background: Health information technology has become common in the care of patients with chronic diseases; however, there are few such applications employed in kidney disease.

Objective: The aim of the study was to evaluate the use of a website providing disease-specific safety information by patients with predialysis chronic kidney disease.

Methods: As part of the Safe Kidney Care (SKC) study, an educational website was designed to provide information on safety concerns in chronic kidney disease. Phase I study participants were provided a medical alert accessory with a unique ID number, the Safe Kidney Care website, and an in-person tutorial on the use of the Internet and accessing the SKC website at baseline. Participants were asked to visit the website and enter their unique ID as frequently as they desired over the next 365 days or until their annual follow-up visit, whichever occurred first. Participants' visits and dwell times on specific safety modules were tracked using embedded webpage PHP scripts linked to a MySQL database, enabling the collection of website usage statistics.

Results: Of 108 Phase I participants, 28.7% (31/108) visited the website from 1-6 times during the observation period (median follow-up 365 days). Median access time was 7 minutes per visit (range <1-46) and 13 minutes per person (range <1-123). The three most frequently visited pages were "Renal function calculator", "Pills to avoid", and "Foods to avoid". High school education and frequent Internet use were significantly associated with website entry (P=.02 and P=.03, respectively).

Conclusions: Preliminary results show general interest in a Web-based platform designed to improve patient safety in chronic kidney disease.

Trial Registration: Clinicaltrials.gov NCT01407367; http://clinicaltrials.gov/show/NCT01407367 (Archived by WebCite at http://www.webcitation.org/6KvxFKA6M).

(J Med Internet Res 2013;15(11):e251) doi:10.2196/jmir.2848

KEYWORDS

chronic kidney disease; health information technology; patient safety

Introduction

Health information technology is expected to play a growing role in medical care delivery over the coming years [1]. The Internet has become a major source of educational health materials for both patients and providers and mobile health applications are being developed for a wide variety of diseases and health conditions [2-7]. For health IT applications to have a beneficial impact on disease outcomes, they must be designed to match the predominant technologic proficiency (e-literacy) of the target population. While the use of computers, the Internet, and mobile devices continues to rise in the United States [8,9], many chronic disease populations are not the target of commercial IT developers. The population with chronic kidney disease is one such population, as it includes a high preponderance of individuals who are older, of lower socioeconomic status, and with lower health literacy [10,11]. Therefore, it is unknown whether health IT tools intended for chronic kidney disease can be effective.

The Safe Kidney Care (SKC) study was a prospective cohort designed to gauge adverse safety events in chronic kidney disease. In Phase I of this study, we set out to evaluate the acceptance and initial use of a health IT system including a device designed to increase recognition of chronic kidney disease (medical alert accessory) and linked to a website. The medical alert bracelet was devised for the purpose of alerting patients and providers of a patient's diagnosis of chronic kidney disease and directing these individuals to a website informing patients, family members, and providers about the unique patient safety concerns associated with chronic kidney disease management. We tracked the incidence of study participants' initial entry into the website over a one-year period and their prioritization of chronic kidney disease patient safety concerns.

Methods

Study Population

The SKC cohort study (Clinicaltrials.gov NCT 01407367) was approved by the University of Maryland School of Medicine Institutional Review Board and Veterans Affairs Maryland Health Care System Baltimore Research and Development committee. The primary objective of the SKC cohort study was to examine the relationship between chronic kidney disease recognition and patient safety. The objective of Phase I of the SKC cohort was to evaluate a medical alert accessory (American Medical ID, Houston, TX) noting the participant's diagnosis of chronic kidney disease and directing patients and providers to the Safe Kidney Care website offering information on common patient safety concerns pertinent to chronic kidney disease. To be eligible for the SKC cohort, participants needed two measures of renal function with an estimated glomerular filtration rate (eGFR) of less than 60 ml/min/1.73 m^2 at least 90 days apart and no more than 18 months prior to enrollment. Participants were excluded for Phase I if they were expected to reach end-stage renal disease or die within one year from enrollment or if they had skin sensitivity to silver or stainless steel. Primary sources for recruitment and enrollment were the University of Maryland Medical System and Baltimore Veterans Affairs Medical Center Early Renal Insufficiency disease management clinics.

Safe Kidney Care (SKC) Cohort Study Procedures

Upon enrollment into the SKC cohort, participants were seen at baseline and then annually in-center with a 6-month interim telephone follow-up visit. At the baseline visit, Phase I participants were provided with a fitted medical alert accessory of their choosing (bracelet or necklace) engraved with the message: "Decreased kidney function". For safe care, please visit the Safe Kidney Care website and a unique study number. This unique ID number did not contain any personally identifiable information. After baseline data collection and provision of the medical alert accessory, the study coordinator provided participants with a demonstration and short tutorial on accessing the Internet and SKC website, where to enter their unique ID number, and a tour of the website and its features. At completion of the tutorial, participants were encouraged to use the website, along with family members and providers. They were instructed to enter their unique ID number when accessing the website's patient portal so that their usage could be tracked, although this was not required for entrance into the website (Figure 1).

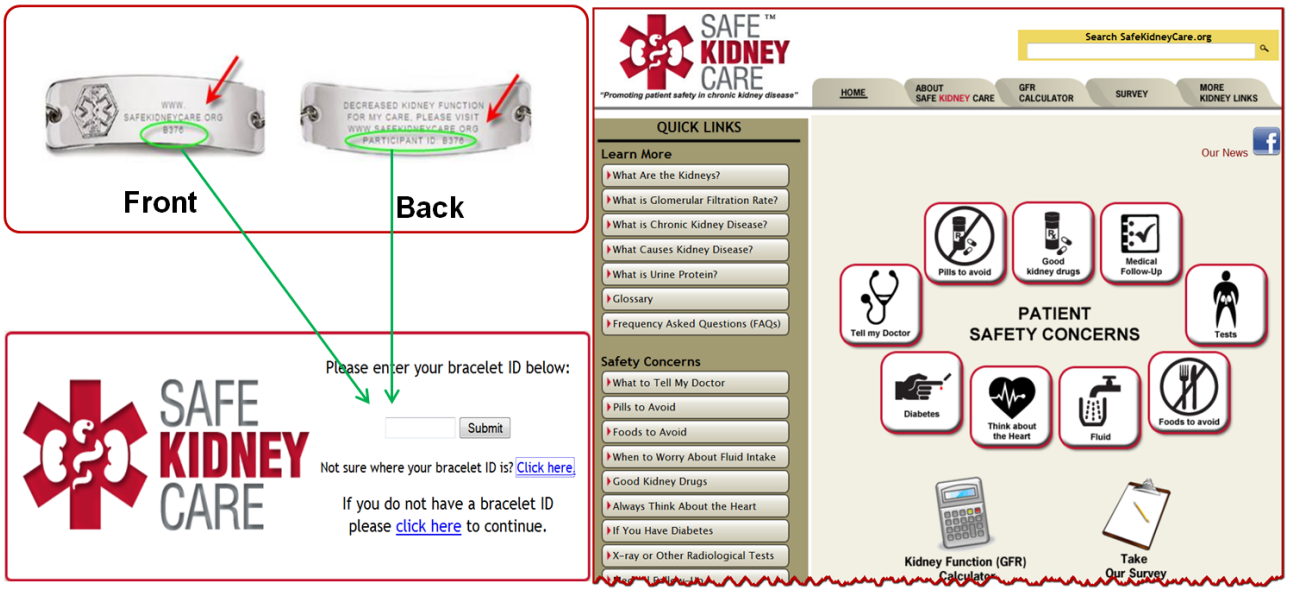
Safe Kidney Care Website Development

Prior to the commencement of enrollment into the SKC cohort, an extensive literature review was conducted to identify chronic kidney disease-pertinent patient safety concerns. Approximately 800 academic and community nephrologists were then invited to participate in a survey to prioritize identified candidate patient safety concerns and to suggest additional potential patient safety events not previously considered. The feedback by participating respondents (n=142) was then used as the basis for development of the SKC website, which covers the most commonly identified chronic kidney disease-specific patient safety concerns along with other general nephrology educational topics. The resulting modules were reviewed by an expert panel convened for the SKC project with additional modifications made based on their input. The website contains both a patient and family member portal, which provides module descriptions in lay terms, and a health care provider portal, which uses common medical terminology.

The resultant SKC website was designed using guidelines for color and content layout incorporated from the 508 Compliance and Disabilities Act and heuristic design [12,13]. Emphasis was placed on ease of readability, with text written to target a 6^{th} grade reading level in the patient portal. The primary placement of the "Safety Concerns" was in a circular distribution to avoid the implication of prioritization of topics. Images were designed to be static because of the anticipated limited access to Adobe Flash (or other multimedia platform). The website was designed to limit the need to scroll; however, on pages with a requirement for scrolling, the most important content was placed above the fold so as to ensure primary focus on the safety content. The capability to increase text size on every page was installed and displayed and meaningful, relatable icons were chosen for easy recognition of content topics (eg, picture of a stethoscope to represent "What to Tell My Doctor"). A quick-launch side panel, entitled "Learn More", included links to pages on the SKC

website which provide users with basic information about the function of the kidneys, what is meant by kidney disease, definition of the glomerular filtration rate (GFR), and a glossary of commonly used medication terms (eg, “blood pressure”, “dialysis”, and “nephrologist”). The website markup and content were constructed using HTML with embedded functions programmed using PHP programming language (Version 5.3.5). The design was conceived with the goal of optimal use on all Internet portals. Usability testing was conducted on a small representative sample and previously reported [14]. Modifications to the website were instituted based on feedback from usability participants.

Figure 1. Medical alert accessory and log-in page of Safe Kidney Care website patient portal.



Safe Kidney Care Website Surveillance

To evaluate the receipt of information designed for the targeted audience, we tracked activities on the SKC website of each bracelet ID assigned to study group participants, including page viewing frequency, browsing time for each page, and feedback. Due to flexibility and security concerns, our implemented computer routine was believed to surpass other available online tracking and analysis services (eg, Google Analytics), as our own implementation and hosting permitted precise tracking defined for the study purposes with data stored securely on our local server.

The website was programmed to track both registered and anonymous users through PHP scripts embedded in each webpage. The PHP scripts were linked with a database built on the MySQL server, to collect and store predefined website usage statistics. For each distinct participant log-in or anonymous guest visit, a unique session number was assigned to label all actions during the visit until the session expired or the visitor started a new session. Medical alert accessory unique ID, session number, log-in time, and IP address were recorded into a table for each independent visit. During a session, each click of an internal or external link was captured and a unique record was generated, including session number, page/link entry time, and page/link ID. With this information, it was possible to recreate the visiting pathway for each visit, to count total pages read in each session, and to calculate other pertinent statistics such as frequency of each page visited in each session, or number of visits from a registered user. It is generally difficult to determine the time when a visitor leaves a page because the visitor may leave the computer for an interruption and return after some time; hence, it was not possible to ascertain precisely how long a visitor focused on the content of the webpages. Nevertheless, the dwelling time for each page was estimated, under the assumption that a visitor would not spend an extended time on a single page but would continue to browse different pages. At the same time, a minimum length of time must also be spent on each page to conduct effective reading. Therefore, an effective dwelling length on a page was counted if the time difference between two distinct webpage entry times (within the same session) was longer than 5 seconds and less than 10 minutes.

Outcomes

The primary outcome in this analysis was the first entry into the SKC website with input of a unique ID number by SKC study participants at any time following completion of their baseline SKC visit until their first annual follow-up visit or 365 days of observation, whichever occurred first. Access of the health care provider portal was not analyzed for the purpose of this study. The number of days from baseline SKC visit was recorded, as was the length of dwell time on the website and on selected modules. Modules were ranked by frequency of selection if selected at least once during the participant’s initial visit on the website.

Covariates

Additional factors measured at baseline SKC visit and included in the analysis were demographic characteristics, socioeconomic status, education, health literacy and numeracy [15], and ease and frequency of use of the Internet. Baseline serum creatinine was measured and used to estimate the GFR with the abbreviated Modified Diet in Renal Disease equation [16].

Statistical Methods

For descriptive analyses, continuous variables were presented with mean and standard error (SE) and Student's *t* test for comparison across groups. Binomial and categorical variables were expressed as N (%) with comparisons made with the chi-square test. Stepwise logistic regression was used to examine factors that predict incident registration into the SKC website. All factors were used in a forward stepwise inclusion procedure with retention of only those factors with significance of $P<.05$.

Results

Enrollment commenced for Phase I of the SKC cohort on April 15, 2011 and the last of 108 participants completed the baseline visit on January 23, 2012. A medical alert bracelet was selected by 55 (50.9%, 55/108) of participants with 53 (49.1%, 53/108) choosing the medical alert necklace. The median observation time was 365 days (range 305-365). Table 1 shows the demographic and baseline characteristics of all participants. Mean age of SKC participants was 64 years (SD 11) with a mean estimated GFR of 42 ml/min/1.73m^2 (SD 14). Of note, 10 enrolled participants were estimated to have eGFRs >60 ml/min/1.73^2 at their baseline visit, despite two screening eGFRs that fit enrollment criteria of eGFR <60 ml/min/1.73m^2. The majority of participants were black, unmarried, with a household income of ≤ US$50,000 annually (72.2%, 78/108; 55.6%, 60/108; and 68.5%, 74/108, respectively). While 86.1% of participants (93/108) reported availability of computer access, 45.4% (49/108) estimated their Internet usage to be less frequent than weekly or daily.

Figure 2 shows the results of the SKC website usage at the end of observation for all participants. To date, 28.7% (31/108) of participants visited the website since study enrollment a median of 1 time (range 1-6). Median SKC website access time per visit was 7 minutes (range <1-46). Median cumulative SKC website access time per person was 13 minutes (range <1-123) with a cumulative access time by all 31 participants of 669 minutes. The three most frequently visited pages were "Renal function calculator", "Pills to avoid", and "Foods to avoid" (Table 2).

The adjusted proportion of individuals with SKC website log-in by baseline characteristic is shown in Table 3**.** After stepwise logistic analysis, pertinent and significant factors predictive of incident use of the SKC website were high school diploma (OR 3.22, 95% CI 1.18-8.81) and frequent use of the Internet (OR 3.31, 95% CI 1.13-9.72), P=.02 and P=.02, respectively.

Figure 2. Cumulative incidence of Safe Kidney Care website log-in.

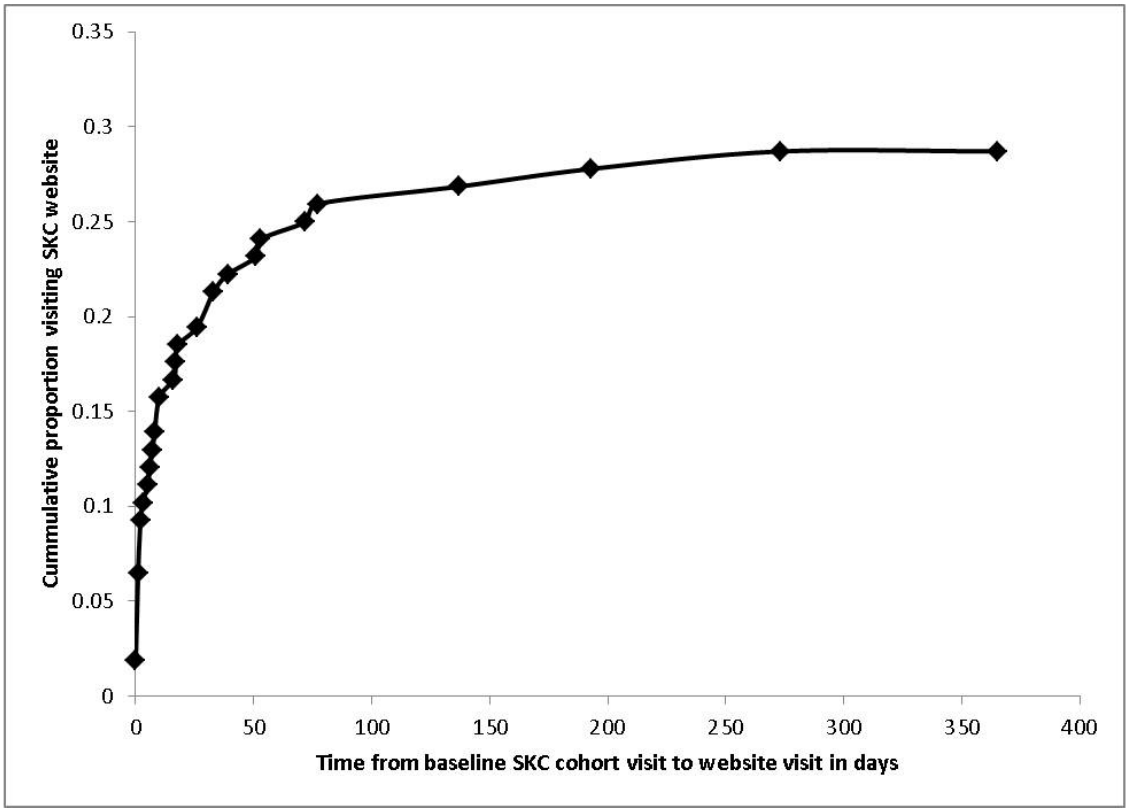


Table 1. Safe Kidney Care participants' baseline demographics.

Baseline characteristic		n (%)
Age		
	≤50	10 (9)
	51-64	46 (43)
	65+	52 (48)
Gender		
	Male	70 (65)
	Female	38 (35)
Race/ethnicity		
	Black	78 (72)
	Non-black	30 (28)
Annual household income [a]		
	≤$50,000	74 (69)
	>$50,000	34 (31)
Education		
	≤High school diploma	51 (47)
	>High school diploma	57 (53)
Marital status		
	Not currently married	60 (56)
	Currently married	48 (44)
Employment status		
	Not employed full-time	96 (89)
	Employed full-time	12 (11)
Baseline estimated glomerular filtration rate (mL/min/1.73m ^{2}) [b]		
	Stage 2 (≥60)	10 (9)
	Stage 3a (45-59)	33 (31)
	Stage 3b (30-44)	47 (44)
	Stage 4 or 5 (<30)	18 (17)
Health literacy [c]		
	Marginal/inadequate (TOFHLA<67)	32 (30)
	Adequate (TOFHLA≥67)	76 (70)
Computer access available		
	No	15 (14)
	Yes	93 (86)
Internet use		
	Rarely/never	49 (45)
	Often	59 (55)

[a]Five participants declined to answer income status.

[b]Ten participants had baseline estimated glomerular filtration rates (eGFR) ≥60 mL/min/1.73m^2 despite qualifying screening eGFRs of <60 mL/min/1.73m^2.

[c]Seven participants were unable to complete the abbreviated Test of Functional Health Literacy Assessment (TOFHLA) due to conditions such as blindness.

Table 2. Most frequently visited SKC webpages.

Webpage title	Total visits	Cumulative time (minutes)	Average time per visit (minutes)
Kidney Function (GFR) calculator	36	50	1
Pills to avoid	34	55	2
Foods to avoid	25	64	3
What to keep in mind about the “good” kidney drugs	17	58	3
When you need a special type of x-ray or other radiologic test…	17	28	2
What is glomerular filtration rate?	16	34	2
When to worry about your fluid intake	12	37	3
What is urine protein?	12	20	2
Frequently asked questions	10	33	3
Take our survey	10	28	3
What is chronic kidney disease (CKD) or weak kidneys?	10	26	3
If you have diabetes…	10	21	2
About the Safe Kidney Care website	10	19	2
What to tell my doctor, nurse, or pharmacist	10	17	2

Table 3. Predictors of Safe Kidney Care website log-in.

Baseline characteristic		Unadjusted proportion	Adjusted proportion (95% CI)
Age			
	≤50	30	Reference (ref)
	51-64	28	47 (13-84)
	65+	29	47 (13-84)
Gender			
	Male	27	ref
	Female	32	43 (20-69)
Race/ethnicity			
	Black	24	ref
	Non-black	40	31 (11-61)
Annual household income [a]			
	≤$50,000	20	ref
	>$50,000	47	40 (18-68)
Education			
	≤High school diploma	14	ref
	>High school diploma	42	28 (11-55)
Marital Status			
	Not currently married	23	ref
	Currently married	35	20 (8-44)
Employment status			
	Not employed full-time	26	ref
	Employed full-time	50	28 (8-65)
Baseline estimated glomerular filtration rate (mL/min/1.73m ^{2}) [b]			
	Stage 2 (≥60)	30	ref
	Stage 3a (45-59)	33	31 (7-72)
	Stage 3b (30-44)	32	27 (6-68)
	Stage 4 or 5 (<30)	11	9 (1-48)
Health literacy [c]			
	Marginal/inadequate (TOFHLA<67)	22	ref
	Adequate (TOFHLA≥67)	32	13 (4-36)
Computer access available			
	No	7	ref
	Yes	32	9 (1-52)
Internet use			
	Rarely/never	12	ref
	Often	43	36 (12-68)

[a]Five participants declined to answer income status.

[b]Ten participants had baseline estimated glomerular filtration rates (eGFR) ≥60 mL/min/1.73m^2 despite qualifying screening eGFRs of <60 mL/min/1.73m^2.

[c]Seven participants were unable to complete the abbreviated Test of Functional Health Literacy Assessment (TOFHLA) due to conditions such as blindness.

Discussion

Principal Findings

In this study, we have described the longitudinal access and usage of a Web-based reference on the unique concerns of patient safety in chronic kidney disease based on the provision of a medical alert accessory with a unique identifier and prompt for website use. The study evaluates the alignment of a commonly used medical alert device, designed to enhance awareness of chronic kidney disease, with a health IT portal offering disease-pertinent information to the target population. The results offer an assessment on the likelihood that individuals with chronic kidney disease and their family members will utilize this alert system intended to increase disease awareness and improve patient safety. The findings suggest a substantial proportion of the SKC Phase I sample, designed to be representative of the chronic kidney disease population, were capable and motivated to access the website to view the developed patient safety modules. The results reveal the population's prioritization of safety concerns specific to chronic kidney disease. Moreover, the report on dwell time provides indication of the importance users placed on the content.

Several reports show that patients with chronic disease who are empowered with IT tools for monitoring, training, and self-management have improved outcomes [2-7], yet few prior investigations evaluate the usage of these applications in general or those specifically designed for the chronic kidney disease population. While online educational materials designed to promote patient education and community awareness of kidney disease are available via high-quality programs such as the National Kidney Disease Education Program [17] and the National Kidney Foundation [18], the use of these materials by the target population remains unknown. Our study is unique in its demonstration of the feasibility of acceptance and utilization of a hybrid system combining a longstanding physical device (a variant of which has been most commonly utilized in kidney disease to preserve potential vascular access sites) [19] with its value enhanced by a usability-tested health IT portal [14] providing extensive information relevant to the chronic kidney disease population.

Introduction of online tools does not guarantee their usage by the target population and the results from this study support prior examinations that estimate use of online tools to be highly variable. Our findings align with studies in other chronic diseases, which reveal inconsistency in website use, but self-reported or measured use has been noted to range from approximately 10-40% of those queried [20-22]. Adults in the United States living with chronic disease are significantly less likely than healthy adults to have access to the Internet (62% vs 81%), but once online, those with chronic disease are more likely to use social media and online tools to share information and obtain support from their peers [23]. Individuals with chronic disease still report an overwhelming preference to receive health-related education and advice from a professional source such as a health care provider [24], which emphasizes the role of the Internet as an accessory to, rather than a replacement for, quality provider-patient interactions. The results presented here suggest information needs in the areas most frequently visited on the website, even for face-to-face encounters.

Limitations

The study presented has limitations related to the descriptive nature of the results, which may limit generalizability. While approximately 30% of participants visited the SKC website, 70% did not, which is an important finding. These individuals who did not access the website were not specifically asked to comment on reasons for their lack of participation (eg, lack of interest rather than access), which may not be captured in the demographic evaluation of computer and Internet access, but likely is highly relevant to future work using health IT applications in chronic kidney disease. The use of nephrologists only in identifying specific areas of safety concerns in chronic kidney disease was limited in that it did not include advanced practitioners or primary care providers, who play important roles in caring for individuals with chronic kidney disease and undoubtedly would have added value to the content. Further, we did not incorporate patients in the initial design of the website; however, usability testing was performed with such a population using an iterative process used to finalize the SKC website [14]. Finally, the relatively short follow-up of participants does not allow us to yet comment on the impact of the conjoined alert accessory and website resource on patient safety outcomes in chronic kidney disease.

Conclusions

Patient safety is a significant problem in kidney disease and online education tools such as the SKC website may serve as platform to educate individuals with chronic kidney disease about potential hazards related to their condition. Further examination is needed to assess individuals' long-term utilization and dissemination of such materials and the impact of such resources in preventing safety events in chronic kidney disease.

Acknowledgments

Dr Fink is funded for this work by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK084017) (PI: Jeff Fink). Medical alert bracelets were provided gratis by American Medical ID in Houston, TX, for use in this study [25].

Conflicts of Interest

C Diamantidis, W Fink, S Yang, M Zuckerman, J Ginsberg, P Hu, and Y Xiao have no conflicts to report. J Fink has received prior research funding from Amgen, Inc and honoraria from Sandoz, Inc and Amgen, Inc

References

1. Goodson JD. Patient Protection and Affordable Care Act: promise and peril for primary care. Ann Intern Med 2010 Jun 1;152(11):742-744. [doi: 10.7326/0003-4819-152-11-201006010-00249] [Medline: 20404263]
2. Gruman JC. Making health information technology sing for people with chronic conditions. Am J Prev Med 2011 May;40(5 Suppl 2):S238-S240. [doi: 10.1016/j.amepre.2011.02.003] [Medline: 21521600]
3. Cummings E, Turner P. Patient self-management and chronic illness: evaluating outcomes and impacts of information technology. Stud Health Technol Inform 2009;143:229-234. [Medline: 19380941]
4. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
5. Morren M, van Dulmen S, Ouwerkerk J, Bensing J. Compliance with momentary pain measurement using electronic diaries: a systematic review. Eur J Pain 2009 Apr;13(4):354-365. [doi: 10.1016/j.ejpain.2008.05.010] [Medline: 18603458]
6. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
7. Bennett GG, Warner ET, Glasgow RE, Askew S, Goldman J, Ritzwoller DP, Be Fit, Be Well Study Investigators. Obesity treatment for socioeconomically disadvantaged patients in primary care practice. Arch Intern Med 2012 Apr 9;172(7):565-574. [doi: 10.1001/archinternmed.2012.1] [Medline: 22412073]
8. Smith, A. Mobile Access 2010 URL: http://www.pewinternet.org/Reports/2010/Mobile-Access-2010 [accessed 2013-07-22] [WebCite Cache ID 6IJENyZkE]
9. Purcell, K. Half of adult cell phone owners have apps on their phones URL: http://pewinternet.org/Reports/2011/Apps-update.aspx?src=prc-headline [accessed 2013-07-22] [WebCite Cache ID 6IJET6CGv]
10. U.S. Renal Data System. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Disease, National Institute of Health. USRDS 2012 Annual Data Report URL: http://www.usrds.org/adr.aspx [accessed 2013-07-22] [WebCite Cache ID 6IJEWi7d3]
11. Devraj R, Gordon EJ. Health literacy and kidney disease: toward a new line of research. Am J Kidney Dis 2009 May;53(5):884-889. [doi: 10.1053/j.ajkd.2008.12.028] [Medline: 19324481]
12. Website Design 508 Compliance. URL: http://www.webconsuls.com/web-design [accessed 2013-07-22] [WebCite Cache ID 6IJEZf2pE]
13. Cyr DH, Head M, Larios H. Colour appeal in website design within and across cultures: A multi-method evaluation. International Journal of Human-Computer Studies 2010 Jan;68(1-2):1-21. [doi: 10.1016/j.ijhcs.2009.08.005]
14. Diamantidis CJ, Zuckerman M, Fink W, Hu P, Yang S, Fink JC. Usability of a CKD educational website targeted to patients and their family members. Clin J Am Soc Nephrol 2012 Oct;7(10):1553-1560 [FREE Full text] [doi: 10.2215/CJN.03690412] [Medline: 22798537]
15. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [Medline: 14528569]
16. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 1999 Mar 16;130(6):461-470. [Medline: 10075613]
17. National Kidney Disease Education Program (NKDEP): Improving the understanding, detection, and management of kidney disease. URL: http://nkdep.nih.gov/ [accessed 2013-07-22] [WebCite Cache ID 6IJElfLey]
18. National Kidney Foundation TM. URL: https://www.kidney.org/ [accessed 2013-07-22] [WebCite Cache ID 6IJEpje5R]
19. Vachharajani T. Medical alert bracelet: an effective way to preserve veins for future dialysis vascular access in patients with chronic kidney disease. Medscape J Med 2009;11(1):1. [Medline: 19295922]
20. Beckjord EB, Finney Rutten LJ, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
21. Fleisher L, Kandadai V, Keenan E, Miller SM, Devarajan K, Ruth KJ, et al. Build it, and will they come? Unexpected findings from a study on a Web-based intervention to improve colorectal cancer screening. J Health Commun 2012;17(1):41-53 [FREE Full text] [doi: 10.1080/10810730.2011.571338] [Medline: 22217118]
22. Zhang Y, Jones B, Spalding M, Young R, Ragain M. Use of the internet for health information among primary care patients in rural West Texas. South Med J 2009 Jun;102(6):595-601. [doi: 10.1097/SMJ.0b013e3181a52117] [Medline: 19434029]
23. Fox, S. Chronic Disease and the Internet URL: http://www.pewinternet.org/Reports/2010/Chronic-Disease.aspx [accessed 2013-07-22] [WebCite Cache ID 6IJEuAVm4]
24. Kuehn BM. Patients go online seeking support, practical advice on health conditions. JAMA 2011 Apr 27;305(16):1644-1645. [doi: 10.1001/jama.2011.509] [Medline: 21521841]
25. American Medical ID. American: Medical ID URL: http://www.americanmedical-id.com/ [accessed 2013-07-22] [WebCite Cache ID 6IJE4YxvD]

Abbreviations

eGFR: estimated glomerular filtration rate
SKC: Safe Kidney Care
TOFHLA: Test of Functional Health Literacy Assessment

Edited by G Eysenbach; submitted 31.07.13; peer-reviewed by J Nunes, J Davis, D Schatell; comments to author 19.09.13; revised version received 23.09.13; accepted 28.09.13; published 15.11.13.

Please cite as:
Diamantidis CJ, Fink W, Yang S, Zuckerman MR, Ginsberg J, Hu P, Xiao Y, Fink JC
Directed Use of the Internet for Health Information by Patients With Chronic Kidney Disease: Prospective Cohort Study
J Med Internet Res 2013;15(11):e251
URL: http://www.jmir.org/2013/11/e251/
doi:10.2196/jmir.2848
PMID:24240617

Original Paper

Effectiveness of a Web-Based Physical Activity Intervention in Patients With Knee and/or Hip Osteoarthritis: Randomized Controlled Trial

Daniël Bossen[1], MSc; Cindy Veenhof[1], PhD; Karin EC Van Beek[1], BSc; Peter MM Spreeuwenberg[1], MA; Joost Dekker[2], PhD; Dinny H De Bakker[1,3], PhD

[1]Netherlands Institute for Health Services Research (NIVEL), Utrecht, Netherlands

[2]EMGO Institute, Department of Rehabilitation Medicine & Department of Psychiatry, VU University Medical Center Amsterdam, Amsterdam, Netherlands

[3]Tranzo, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Daniël Bossen, MSc
Netherlands Institute for Health Services Research (NIVEL)
Otterstraat 118-124
Utrecht, 3500 BN
Netherlands
Phone: 31 302729649
Fax: 31 302729729
Email: d.bossen@nivel.nl

Abstract

Background: Patients with knee and/or hip osteoarthritis (OA) are less physically active than the general population, while the benefits of physical activity (PA) have been well documented. Based on the behavioral graded activity treatment, we developed a Web-based intervention to improve PA levels in patients with knee and/or hip OA, entitled “Join2move”. The Join2move intervention is a self-paced 9-week PA program in which the patient’s favorite recreational activity is gradually increased in a time-contingent way.

Objective: The aim of the study was to investigate whether a fully automated Web-based PA intervention in patients with knee and/or hip OA would result in improved levels of PA, physical function, and self-perceived effect compared with a waiting list control group.

Methods: The study design was a two-armed randomized controlled trial which was not blinded. Volunteers were recruited via articles in newspapers and health-related websites. Eligibility criteria for participants were: (1) aged 50-75 years, (2) self-reported knee and/or hip OA, (3) self-reported inactivity (30 minutes of moderate PA, 5 times or less per week), (4) no face-to-face consultation with a health care provider other than general practitioners, for OA in the last 6 months, (5) ability to access the Internet weekly, and (6) no contra-indications to exercise without supervision. Baseline, 3-month, and 12-month follow-up data were collected through online questionnaires. Primary outcomes were PA, physical function, and self-perceived effect. In a subgroup of participants, PA was measured objectively using accelerometers. Secondary outcomes were pain, fatigue, anxiety, depression, symptoms, quality of life, self-efficacy, pain coping, and locus of control.

Results: Of the 581 interested respondents, 199 eligible participants were randomly assigned to the intervention (n=100) or waiting list control group (n=99). Response rates of questionnaires were 84.4% (168/199) after 3 months and 75.4% (150/199) after 12 months. In this study, 94.0% (94/100) of participants actually started the program, and 46.0% (46/100) reached the adherence threshold of 6 out of 9 modules completed. At 3 months, participants in the intervention group reported a significantly improved physical function status (difference=6.5 points, 95% CI 1.8-11.2) and a positive self-perceived effect (OR 10.7, 95% CI 4.3-26.4) compared with the control group. No effect was found for self-reported PA. After 12 months, the intervention group showed higher levels of subjective (difference=21.2 points, 95% CI 3.6-38.9) and objective PA (difference=24 minutes, 95% CI 0.5-46.8) compared with the control group. After 12 months, no effect was found for physical function (difference=5 points, 95% CI −1.0 to 11.0) and self-perceived effect (OR 1.2, 95% CI 0.6-2.4). For several secondary endpoints, the intervention group demonstrated improvements in favor of the intervention group.

Conclusions: Join2move resulted in changes in the desired direction for several primary and secondary outcomes. Given the benefits and its self-help format, Join2move could be a component in the effort to enhance PA in sedentary patients with knee and/or hip OA.

Trial Registration: The Netherlands National Trial Register: NTR2483; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2483 (Archived by WebCite at http://www.webcitation.org/67NqS6Beq).

(J Med Internet Res 2013;15(11):e257) doi:10.2196/jmir.2662

KEYWORDS

physical activity; osteoarthritis; Web-based intervention; randomized controlled trial

Introduction

It has been recognized that regular physical activity (PA) positively impacts the severity and course of numerous chronic diseases [1,2]. Among patients with knee and/or hip osteoarthritis (OA), regular PA has proven to be beneficial in preserving physical function and reducing pain symptoms [3,4]. Improvement in physical function and reduction in pain are positively related to several psychological factors and thus may affect self-esteem, pain coping, and self-efficacy in patients with knee and/or hip OA [5,6]. However, due to pain and other symptoms, patients with OA are less physically active than the general population [7,8]. Therefore, PA as a nonpharmacological intervention has been advocated in the treatment of OA patients [9].

Since OA is mainly managed within primary care, general practitioners (GPs) are advised to stimulate patients to adopt and maintain higher levels of PA. In practice, however, a GP's ability to encourage physical exercise is limited by time constraints and lack of standard protocols [10-12]. At the same time, it is unlikely that patients with knee and/or hip OA receive help elsewhere, since patients are not referred to other health care professionals [13] and because people often view their peripheral joint pain as an inevitable part of aging [14]. Numerous patients lack knowledge and skills to modify their PA routines and have negative concerns (eg, fear of pain and catastrophizing thoughts) about the impact of PA on their joints [15,16].

In an attempt to promote a more physically active lifestyle among patients with knee and/or hip OA, effective PA interventions are needed. With the explosion of Internet accessibility, Web-based interventions seem to provide a novel medium to reach patients with knee and/or hip OA; 61% of Europeans and 79% of North Americans have Internet access [17]. In the Netherlands, 95% of adults (55-65 years) and 75% of older adults (65-75 years) have access to Internet in their home [18]. Web-based interventions are applications available through a website with the intent to enhance understanding of a health condition and to change health behavior. In particular, Web-based interventions with minimal human contact have the potential of high reach, low costs, and are accessible anytime and anywhere [19]. Previous Web-based interventions for inactive populations and patients with a chronic disease (eg, diabetes, cardiovascular diseases, and chronic obstructive pulmonary disease) have produced inconclusive findings [20-22].

To date, there are no Web-based PA interventions for patients with knee and/or hip OA that we know of. Given the advantages of the Internet, we developed "Join2move". The Join2move program differs from existing Web-based programs since it focuses on knee and hip OA and strategies to enhance PA despite the presence of pain. The design is inspired by a previously developed exercise program known as the behavior graded activity (BGA) program [23]. The BGA treatment is an exercise regimen based on operant behavior principles that stimulate OA patients to gradually increase their daily life activities for fixed time periods. In accordance with the BGA treatment, Join2move intervention is a 9-week PA program in which the patient's favorite recreational activity is gradually increased in a time-contingent way. The intensity of the modules is predetermined by the participants themselves. To investigate the effectiveness of Join2move, we compared the Web-based intervention versus no intervention. This study aimed to answer the following research question: "What is the short (3 months) and long-term (12 months) effectiveness of the Join2move intervention in patients with knee and/or hip OA in PA, physical function, and self-perceived effect in comparison with a waiting list control group?"

Methods

Study Design

This study was a two-armed, 12-month, randomized controlled trial (RCT) with continuous recruitment and data collection. Allocation ratio was 1:1 and enrollment started on January 3, 2011, and ended November 5, 2011. The trial is reported according to the CONSORT-EHEALTH checklist [24]. Ethics approval was obtained from the medical ethics committee of the VU University Medical Center, Amsterdam.

Participants

Patients with self-reported knee and/or hip OA were recruited through advertisements in Dutch newspapers and online on health-related websites. The advertisements briefly explained the purpose of the project and the beneficial health effects of PA. Interested individuals were referred to an open access study website and invited to complete an online eligibility questionnaire. Participants' email addresses were used to contact them for online follow-up questionnaires, and home addresses were used for sending an information letter, informed consent form, and accelerometer. The eligibility criteria for participants were: (1) aged 50-75 years, (2) self-reported OA in knee and/or hip, (3) self-reported inactivity (<30 minutes of moderate PA three or five times or less per week), (4) no face-to-face

consultation for OA with a health care provider, other than GP, in the last 6 months, (5) ability to access the Internet weekly, and (6) no contra-indications to exercise without supervision. Self-reported OA was determined by asking participants if they had a painful knee or hip joint and if a doctor or other health care provider had ever told them this was a result of OA. Contra-indication was determined by the PA-readiness questionnaire (PARQ) [25]. The PARQ questionnaire is designed to identify persons for whom increased PA may be contra-indicated. If patients filled out "no" to all questions, it was considered safe for the patients to engage Join2move. If participants answered "yes" to any of the seven PARQ questions, they were advised to see their GP before participation. Written medical clearance from a GP was not required.

Procedure

Interested patients who met the inclusion criteria were sent an invitation letter with informed consent. Once informed consent was obtained, participants were invited to fill out an online baseline questionnaire. When baseline assessments were completed, participants were randomly assigned to the intervention (n=100) or control group (n=99). For concealment, a researcher (CV), not involved in data collection, distributed sequentially numbered opaque sealed envelopes with allocation details. Each sealed envelope was opened after the participant had given their written consent to participate in the study. After randomization, all participants were informed through email of their group assignment. Participants in the intervention group received a username and password to log in. Due to the nature of the study (waiting list controlled), neither the study staff nor the participants were blinded to group allocation. To assess the effectiveness of the Join2move intervention, we conducted two post measurements at 3 months and 12 months. At these follow-up times, all participants received online questionnaires. In addition to the online questionnaires, a random subgroup from both groups (n=83) received and returned an accelerometer by post. The decision for sending accelerometers to a subgroup of participants was made based on time and cost savings. An email and telephone reminder was used when participants failed to complete their online questionnaire within 2 weeks. Apart from sending accelerometers and telephone reminders, the study used an automated design. There was no face-to-face contact with study subjects.

Development of the Intervention

Over the course of 1 year, a team of experts from the Netherlands institute for health services research (NIVEL) developed the program. During the development phase, an iterative design methodology [26] was used to test, analyze, and refine the Join2move intervention. We conducted a focus group (n=5), in home observations (n=4), a pilot study (n=20), and interviews (n=16). Furthermore, two usability methods (heuristic evaluation and a thinking aloud approach) were applied to determine the usability of the Web-based program. End-users (ie, patients with knee and/or hip OA) were involved continuously throughout the development process. The final version was used for the RCT study. No content changes were made during the trial period. Further details about the development are described elsewhere [27]. Participants involved in the focus group, pilot, and usability studies did not participate in the RCT study.

The Intervention

The Join2move intervention is based on a previously developed and evaluated BGA program for patients with knee and/or hip OA [23]. The BGA program incorporates a baseline test, goal setting, time-contingent PA objectives (ie, on fixed time points), and text messages to promote PA. An essential feature of the BGA program is the positive reinforcement of gradual PA, despite the presence of pain. The gradual increase in activities changes the perception that PA is related to pain and reinforces confidence to improve PA performance [28]. The Join2move intervention is a fully automated Web-based intervention that contains automatic functions (web-based text messaging and automatic emails) without human support. Screenshots illustrating different stages of the Join2move intervention are presented in Multimedia Appendix 1. Participants are initially presented with a homepage (see Figure 1). The password-secured PA program is available 24/7 from the homepage and is provided without charge. In keeping with the BGA treatment, the Join2move intervention is a self-paced 9-week PA program in which a patient's favorite recreational activity is gradually increased in a time-contingent way. In the first week of the program, users select a central activity such as cycling, walking, or gardening; perform a 3-day self-test; and determine a short-term goal for the next 8 weeks. Based on test performances and a short-term goal, 8 tailored weekly modules are automatically generated. Every week, new modules are posted on the password-secured website. Modules remain on the website for 1 week. After 7 days, users are presented with an evaluation form about pain and performance. Pain is assessed with a 10-point Numerical Rating Scale (0 is no pain, 10 is worst possible pain). Performance was measured by three items, namely: (1)"I completed the module as instructed", (2) "I did more than the instructed module", or (3) "I did less than the instructed module" due to "(a) time constraints, (b) weather conditions, (c) pain in my knee and/or hip, and (d) other physical complaints". Subsequently, tailored to the answers from the evaluation form, automated text-based messages were generated. Furthermore, if users indicated that a module was missed due to time constraints or weather conditions, they had the option to repeat the current module or to continue with the next module. If users indicated that a module was missed due to pain in knee/hip or other physical complaints, they had the ability to repeat the module (a maximum of three times), adapt the intensity of the module, or proceed with the next module. In addition to the weekly modules, information about OA, lifestyle, and videos are provided. Since personal messages are updated on a weekly basis, users are encouraged to log in once a week. Automatic emails are generated if participants do not log on to the website for two weeks. At the end of the program, the website presents a motivational message to perform regular PA in the future.

Figure 1. Join2move homepage.

Waiting List

In this study, we used a waiting list control group. The control group (as well as the intervention group) received a letter with information about the study, PA, and OA. During the follow-up period, participants from the control group had no contact with participants from the intervention group and no access to the Join2move intervention. After the follow-up period, patients in the waiting list group received access to the Join2move intervention.

Measures

Three online questionnaires (0, 3, and 12 months) were used for data collection and a subgroup of participants received an accelerometer to measure PA. Questionnaires were created by online survey experts from the NIVEL institute and tested among a pilot study of 20 participants prior to the RCT study [27]. All participants received an email with a URL link to an online questionnaire. We offered no incentives to complete questionnaires.

Demographic and Clinical Outcomes

Gender, education (low: primary and lower vocational education; middle: secondary and middle vocational education; high: higher vocational and university education), body height (centimeters), age (years), body weight (kilograms), location of OA complaints (knee, hip, or both), duration of OA complaints (years and months), and presence of comorbid conditions were obtained. Body mass index (BMI) was calculated as the weight in kilograms, divided by the height in meters squared.

Program Usage

Program usage was measured by the number of weekly modules completed. A module consisted of a text-based assignment and accompanying evaluation form, which was presented on the website for 7 consecutive days. Once a participant read the weekly assignments and filled out the evaluation form, the module was defined as completed and the user was automatically presented with a new module. In total, there were nine weekly modules that could have been opened by the participant. This was automatically registered. Adequate program use was defined if users completed at least 6 out of 9 modules. Intervention supplements (ie, videos and general information on the homepage) were not included in the adherence measure.

Primary Outcome Measures

Physical Activity

Self-reported PA was measured by the validated PA Scale for the Elderly (PASE) [29]. The PASE questionnaire is designed to assess PA patterns in older adults. The instrument consists of questions on household, leisure time, and work-related activities. The activities (assigned according to the level of intensity: light, moderate, and strenuous) are recorded as never, seldom (1-2 days/week), sometimes (3-4 days/week), or often (5-7 days/week). The amount of time spent in each activity is multiplied by its intensity. In addition to the PASE

questionnaire, assessment of PA was supported through ActiGraph GT3X tri-axial accelerometers [30]. A random subsample of participants from the intervention and control groups were invited to wear this accelerometer. In total, 83 accelerometers were distributed by post to 41 controls and 42 participants in the intervention group. Participants were instructed to wear the monitor on a belt around their waist for 5 consecutive days [31], except during sleeping, showering, or swimming. In addition, participants were requested to fill out a short activity diary. This diary contained questions about wearing time, unusual activities, and reasons for device removal. When accelerometers and diaries were returned by post, data were downloaded, processed, and subsequently analyzed. Participants with at least 10 hours of PA data for at least 4 valid days were included for further analysis. In order to determine the actual PA thresholds, the widely accepted thresholds by Freedson et al [32] were used: 0-99 counts for sedentary activities, 100-1951 for light PA, 1952-5724 moderate PA, 5725-9498 for vigorous PA, and 9499-max for very vigorous activities. The total time spent in light, moderate, and (very) vigorous PA was summed and subsequently divided by the number of days worn to compute the daily average time spent in total activity. For analysis, data were recorded at 1-minute intervals. Sequences of at least 60 minutes of zero counts were defined as non-wearing time. Although the accelerometer was tri-axial, only the vertical axis was used for analysis. This was decided since preprogrammed thresholds of the tri-axial model have yet to be determined [33].

Physical Function

Physical function was determined by a subscale of the Knee OA Outcome Score (KOOS) [34,35] and the Hip Injury OA Outcome Score (HOOS) [36,37]. The KOOS and HOOS are self-administered questionnaires to assess patients' opinions about their knee and/or hip-related problems according to five indicators on a 5-point Likert scale: (1) pain, (2) symptoms, (3) physical function, (4) sport and recreation function, and (5) quality of life.

Self-Perceived Effect

At 3 months and 12 months, self-perceived effect was assessed by a single question that asked participants about the degree of change since their previous assessment. We used a 7-point Likert scale ranging from "much worse" to "much better", with "about the same" located in the middle. The outcomes of self-perceived effect were dichotomized into "improved" (much better, better, and slightly better) and "not improved" (about the same, slightly worse, worse, much worse).

Secondary Outcomes

Pain and fatigue were assessed with a 10-point Numerical Rating Scale (0 is no pain/not tired and 10 is worst possible pain/very tired). OA-related symptoms, quality of life, and sport and recreation were measured with a subscale of the HOOS and KOOS. Anxiety and depression were evaluated by the 14-item Hospital Anxiety and Depression Scale (HADS) [38]. Self-efficacy for pain and other symptoms was evaluated by using the Arthritis Self-Efficacy Scale [39,40]. Active and passive pain coping were determined by the Pain Coping Inventory questionnaire [41]. Locus of control (people's belief that health is or is not determined by their behavior) was examined with the Multidimensional Health Locus of Control Scale [42].

Sample Size

Sample size calculations were performed. Since no previous research has provided adequate statistical information on PA, power calculations were based on physical function and self-perceived effect. We needed 200 patients with knee and/or hip OA in total to detect a small to medium effect (0.2-0.5) in the outcome measure physical functioning and self-perceived effect (25% difference). Conventional levels of statistical power (0.8) and level of statistical significance (P=.05) were used.

Statistical Analysis

Findings were analyzed using an intention-to-treat analysis. Complementary to the primary analysis, per-protocol analysis was employed using only adherent patients in the intervention group (at least 6 out of 9 modules completed) and the entire control group. A nonresponse analysis was carried out in order to examine differences among participants who completed the questionnaires and participants who did not. Furthermore, we compared primary baseline variables between the response and the nonresponse group in order to investigate selective attrition. A Generalized Estimating Equations (GEE) approach controlling for baseline values, age, OA location, and gender was used to analyze effects of the intervention on primary and secondary outcomes. An independent correlation structure was used to account for the within-subject correlations. Also, t tests and chi-square tests were used to compare baseline characteristics in the intervention and control group to perform nonresponse analysis and to determine selective attrition. Between-group effect sizes (ES) were calculated according to Cohen's d. Traditionally, ES of ≥0.8 are interpreted as "large" effects, effect sizes of 0.5 as "moderate", and effect sizes of ≤0.2 as "small" effects [43]. The effect size for self-perceived effect was given by odds ratios (OR). Since GEE analyses are tolerant to data missing, no imputation techniques were used [44].

Results

Participant Characteristics and Study Participation

Figure 2 depicts the flow of participants throughout the trial. In total, 581 persons were screened, 278 (47.8%, 278/581) were eligible, and 200 (71.9%, 200/581) consented to participate. Finally, a total of 99 participants were assigned to the control group, and 100 participants were allocated to the experimental group. With regard to the questionnaires, the overall response rate was 84.4% (168/199) after 3 months and 75.4% (150/199) after 12 months. With respect to the subgroup of participants who wore an accelerometer (n=83), the overall response rate was 72% (60/83) and 66% (55/83) after 12 months. Reasons for not participating in the follow-up surveys were health/medical issues (37%, 17/46), lack of motivation (15%, 7/46), personal/family reasons (13%, 6/46), other (13%, 6/46), and unknown reasons (22%, 10/46).

Table 1 presents participants' characteristics and primary outcome measures at baseline. Participants were predominantly

female (64.8%, 129/199), had knee OA (63.8%, 127/199), and no comorbidity (62.8%, 125/199). Mean age was 62 years (SD 5.7) and mean BMI was 27.6 (SD 4.5). Of the participants, 45.7% (91/199) had a high level of education and 9.0% (18/199) had OA symptoms for less than 1 year. Demographic baseline values were not statistically different between the two groups. Those who did not complete follow-up questionnaires were more likely to have at least one comorbidity (*P*=.01) than those who did. With respect to other baseline characteristics, no differences were found (data not shown). The subgroup of participants (n=83) who wore an accelerometer did not differ from the other participants (n=116) on baseline characteristics (data not shown).

Table 1. Baseline demographic and clinical characteristics.

Characteristic		Intervention, n=100	Control, n=99	*P* value
Gender, n (%)				
	Male	40 (40.0)	30 (30.3)	.15
	Female	60 (60.0)	69 (69.7)	
Age (years), mean (SD)		61 (5.9)	63 (5.4)	.05
BMI (kg/m^2), mean (SD)		27.6 (4.6)	27.5 (4.5)	.79
Location OA, n (%)				
	Knee	67 (67.0)	60 (60.6)	.37
	Hip	21 (21.0)	20 (20.2)	
	Both	12 (12.0)	19 (19.2)	
Duration of symptoms, n (%)				
	≤1 year	12 (12.0)	6 (6.1)	.33
	>1-3 years	28 (28.0)	27 (27.3)	
	>3-7 years	27 (27.0)	27 (27.3)	
	≥7 years	33 (33.0)	39 (39.4)	
Education				
	Low education	13 (13.0)	15 (15.2)	.36
	Middle education	36 (36.0)	43 (43.4)	
	High education	51 (51.0)	40 (40.4)	
Comorbidity, n (%)				
	None	65 (65.0)	60 (60.6)	.43
	One	19 (19.0)	16 (16.2)	
	Two or more	16 (16.0)	23 (23.2)	

Program Usage

Of the 100 participants who received a password and username to enroll, 94.0% (94/100) made a start with the first module and 6.0% (6/100) never logged in to their personal website. Figure 3 depicts an overview of the module completion rate. The first module was completed by 80.0% (80/100) of subjects. This ratio declined to 55.0% (55/100) during the second module. This percentage of completed modules remained steady up to the end of the program. Of the 94 participants who started the program, the average completion was 5.6 (SD 2.9) out of nine modules. Participants selected walking (46.0%, 46/100), cycling (32.0%, 32/100), Nordic walking (4.0%, 4/100), gardening (4.0%, 4/100), and other activities (8.0%, 8/100) as central activity; 6.0% (6/100) of the potential users never selected a central activity because they never logged in to their website. Since personal messages were updated on a weekly basis, patients had the opportunity to complete a module within 7 days. When a module was missed, users still had the ability to complete the next module. Finally, 19.0% (19/100) of the participants fulfilled all modules of the program, and 46.0% (46/100) reached the threshold of adherence (executed at least 6 out of 9 modules). The presence of comorbidity seemed a predictor for nonusage. Of the patients with an additional disease, 71% (25/35) were nonadherent to the Join2move intervention. This percentage was substantially lower among those without a comorbidity, namely 45% (29/65). Adverse events, such as extreme pain and injuries, were not reported during the intervention period.

Figure 2. Flow of participants throughout the trial.

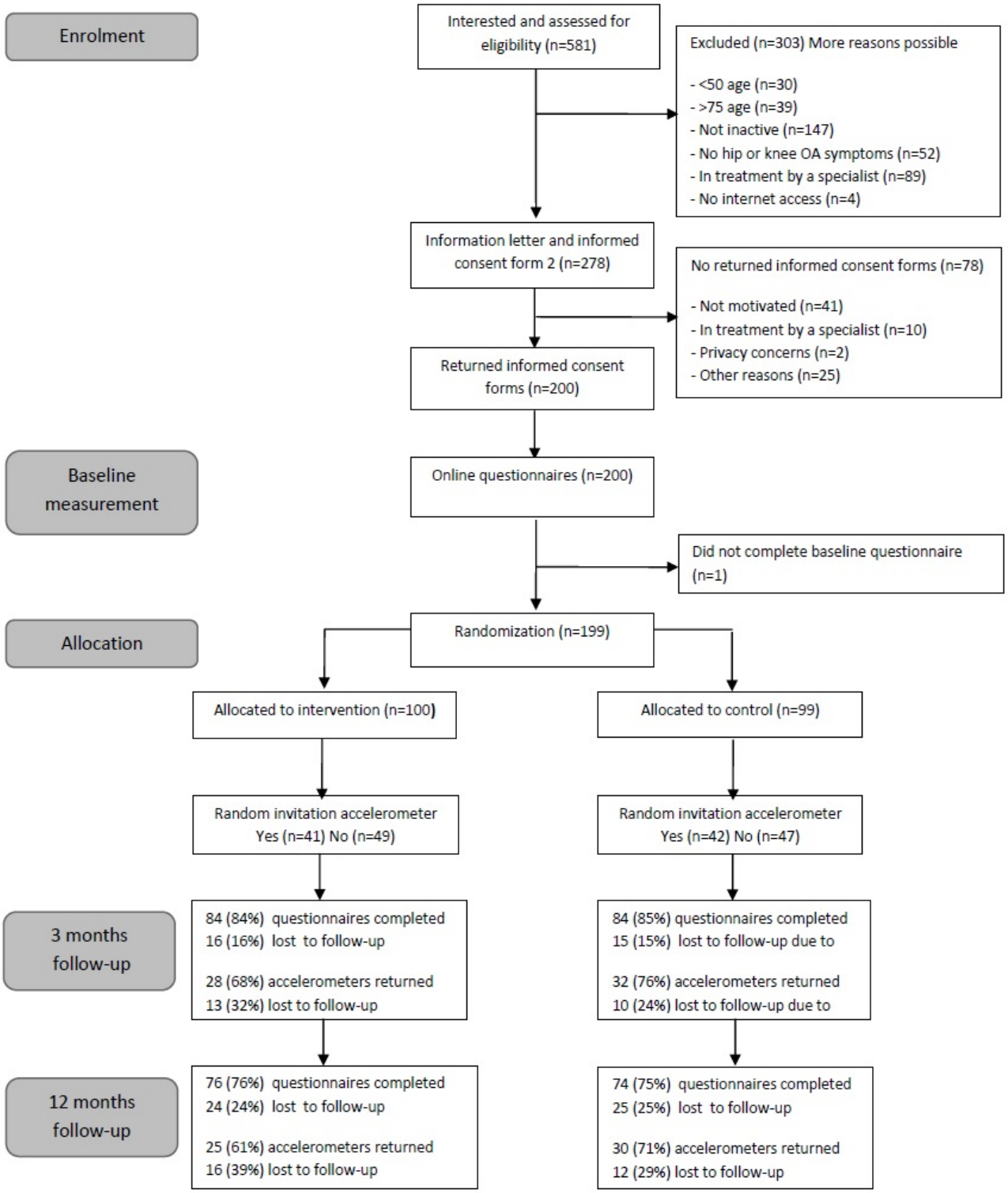
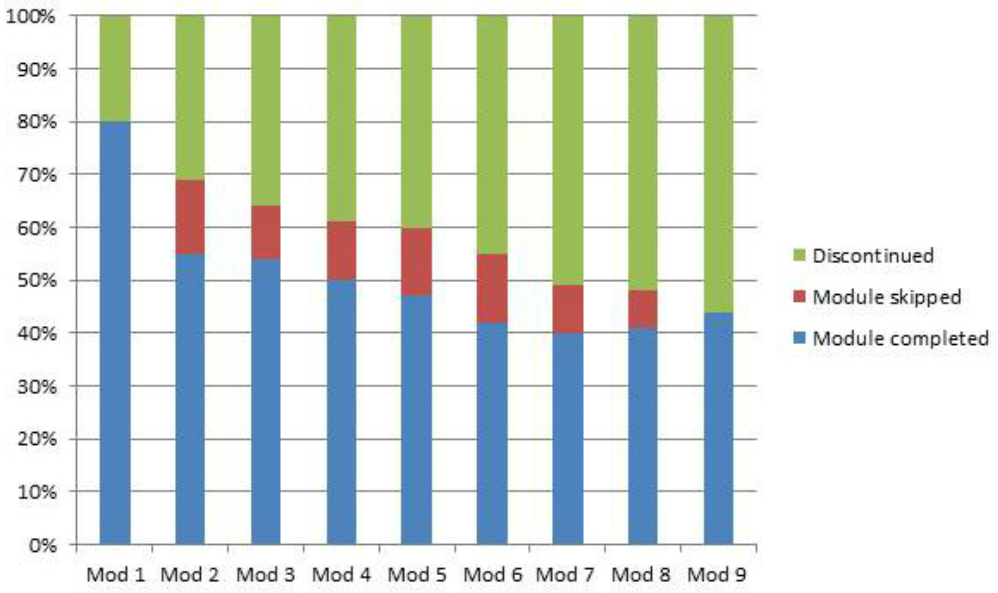


Figure 3. Module completion rate.



Primary Outcome Measures

Table 2 presents results of the primary outcome measures at 3 and 12 months. At 3 months, participants in the intervention group reported a significantly improved physical function status (P=.006, d=0.20) and a positive self-perceived effect (P<.001; OR 10.7, 95% CI 4.3-26.4). No effect was found for PA measured with the PASE questionnaire (P=.84, d=−0.01) and accelerometer (P=.83, d=0.02). After 12 months, the intervention group showed higher levels of subjective and objective PA (P=.02, d=0.18 and P=.045, d=0.19) compared with the control group. At 12 months, no effect was found for physical function (P=.10, d=0.17) and self-perceived effect (P=.50; OR 1.2, 95% CI 0.6-2.4). The accelerometer group (n=83) did not differ from the group who did not wear an accelerometer (n=118) with respect to short and long-term PASE scores (data not shown).

Secondary Outcome Measures

Table 3 presents results of the secondary outcome measures at 3 months and 12 months. At 3 months, we observed statistically significant differences between the intervention and control group with respect to pain (P=.002; d=−0.2), tiredness (P=.04, d=−0.16), and improvements in self-efficacy for pain (P=.008, d=0.17) in favor of the intervention group. Other secondary endpoints were not significantly different between the two groups. At 12 months, subjects in the intervention group reported less tiredness (P=.008; d=−0.22), better passive pain coping scores (P=.008, d=−0.18), and reduced anxiety levels (P=.007; d=−0.21) compared to those in the control group. Other secondary outcomes were not significantly different between the conditions at 12 months.

Per-Protocol Analyses

The per-protocol analysis—a comparison of the adherent patients in the intervention group (ie, participants who completed 6 out of 9 week modules) and the entire control group—yielded positive self-perceived effects in favor of the intervention group (data not presented). Higher levels of participation had no influence on other primary and secondary outcomes (data not presented).

Table 2. Primary outcome measures: improvements and differences between groups[a].

Outcome measures	n	Intervention, mean (95% CI)	n	Control, mean (95% CI)	Difference, I-C[b](95% CI)	ES	*P* value
Total PA, PASE (0-400)							
Baseline	100	163 (130-196)	97	160 (123-197)	—	—	—
3 months	85	162 (136-187)	79	163 (137-190)	−1.6 (−16.6 to 13.5)	−0.01	.84
12 months	74	174 (150-198)	71	153 (125-181)	21.2 (3.6-38.9)	0.18	.02
Total PA (accelerometer min/day)							
Baseline	39	369 (299-439)	40	395 (322-468)	—	—	—
3 months	27	361 (312-411)	30	358 (310-407)	3 (−26 to 32)	0.02	.83
12 months	24	361 (317-406)	28	338 (291-384)	24 (0.5-46.8)	0.19	.045
Physical functioning (0-100)							
Baseline	99	58.8 (51.5-66.0)	98	55.2 (47.9-62.5)	—	—	—
3 months	84	67.8 (59.2-76.4)	80	61.3 (52.7-69.9)	6.5 (1.8-11.2)	0.20	.006
12 months	75	67.9 (59.1-76.7)	72	62.9 (54.1-71.7)	5.0 (−1.0 to 11.0)	0.17	.1
Self-perceived effect (improved-not improved)							
3 months, n (%) improved	85	44 (44)	83	7 (7.1)	10.7[c] (4.3-26.4)	—	<.001
12 months, n (%) improved	76	34 (34)	74	27 (27.3)	1.2[c] (0.6-2.4)	—	.5

[a]For PA, physical functioning, and self-perceived effect, a higher score indicates an improvement. Results are based on GEE analyses and adjusted for corresponding baseline variables, age, OA location, and gender.

[b]I-C: difference between intervention and control group.

[c]odds ratio

Table 3. Secondary outcome measures: improvements and differences between groups[a].

Outcome measures	n	Intervention, mean (95% CI)	n	Control, mean (95% CI)	Difference, I-C[b](95% CI)	ES	*P* value
Sedentary intensity (accelerometer min/day)							
Baseline	39	571 (498-645)	40	555 (479-630)	—	—	—
3 months	27	508 (454-563)	30	540 (477-603)	−32 (−67.7 to 3.7)	−0.20	.08
12 months	24	514 (448-580)	28	531 (467-595)	−17 (−54.7 to 20.7)	−0.10	.38
Pain (0-10)							
Baseline	100	5.4 (4.2-6.5)	98	4.9 (3.7-6.1)			
3 months	85	3.5 (2.5-4.6)	81	4.5 (3.4-5.7)	−1 (−1.6 to −0.38)	−0.20	.002
12 months	76	3.5 (2.4-4.5)	71	3.8 (2.7-4.9)	−0.36 (−1.1 to 0.38)	−0.07	.33
Tiredness (0-10)							
Baseline	100	5.6 (4.3-6.9)	99	5.5 (4.3-6.8)	—	—	—
3 months	85	3.2 (2-4.4)	81	4.1 (2.9-5.3)	−0.84 (−1.6 to -0.06)	−0.16	.04
12 months	76	3 (1.9-4.2)	71	4.1 (3-5.2)	−1.15 (−1.9 to −0.28)	−0.22	.008
Symptoms (0-100)							
Baseline	100	68.2 (60.2-76.2)	99	70.9 (62.7-79.2)	—	—	—
3 months	85	67.4 (59.1-75.8)	80	64.3 (55.3-73.2)	3.1 (−1.3 to 7.6)	0.08	.16
12 months	76	65.7 (57.4-74.0)	71	62.8 (53.4-72.1)	3 (−2.1 to 8.1)	0.08	.25
Quality of life (0-100)							
Baseline	100	38 (30.6-45.5)	98	40.9 (33.6-48.2)	—	—	—
3 months	85	49.4 (41.7-57.0)	80	47.3 (39.4-55.1)	2.1 (−1.7 to 5.9)	0.06	.28
12 months	75	48.7 (40.8-56.6)	71	47.5 (39.3-55.6)	1.2 (−4.4 to 6.8)	0.03	.68
Sport/recreation (0-100)							
Baseline	88	27.6 (14.7-40.4)	78	27.6 (13.4-41.9)	—	—	—
3 months	58	42.6 (29.6-55.6)	55	42.6 (29-56.2)	0 (−8.0 to 8.1)	0	1
12 months	53	42.4 (28.1-56.8)	47	39.6 (25.6-53.5)	2.9 (−6.3 to 12.1)	0.08	.54
Self-efficacy pain (1-5)							
Baseline	100	4.1 (3.6-4.6)	97	3.8 (3.6-4.2)	—	—	—
3 months	85	4 (3.6-4.4)	79	3.7 (3.3-4.1)	0.31 (0.1-0.5)	0.17	.008
12 months	75	4 (3.6-4.4)	72	3.9 (3.5-4.3)	0.12 (−01 to 0.4)	0.06	.35
Self-efficacy other symptoms (1-5)							
Baseline	100	3.6 (3.1-4.1)	96	3.8 (3.4-4.3)	—	—	—
3 months	85	4 (3.7-4.4)	79	3.8 (3.7-4.4)	0.21 (0-0.4)	0.12	.07
12 months	75	4.1 (3.7-4.4)	72	3.8 (3.5-4.2)	0.23 (0-0.5)	0.20	.05
Active pain coping (0-4)							
Baseline	100	2.2 (2.0-2.4)	96	2.2 (2-2.4)	—	—	—
3 months	83	2 (1.9-2.2)	77	2 (1.8-2.2)	−0.02 (−0.1 to 0.1)	−0.02	.81
12 months	73	2 (1.8-2.2)	70	2 (1.8-2.2)	0 (−0.1 to 0.1)	0	.98
Passive pain coping (0-4)							
Baseline	100	1.8 (1.7-2.0)	96	1.8 (1.6-1.9)	—	—	—
3 months	83	1.7 (1.6-1.8)	77	1.7 (1.6-1.9)	−0.04 (−0.1 to 0.04)	0	.29
12 months	73	1.7 (1.5-1.8)	70	1.8 (1.7-1.9)	−0.12 (−0.2 to –0.03)	−0.18	.008

Outcome measures	n	Intervention, mean (95% CI)	n	Control, mean (95% CI)	Difference, I-C[b](95% CI)	ES	*P* value
Internal locus of control (6-36)							
Baseline	100	27.1 (25.1-29.2)	96	27.5 (25.2-29.8)	—	—	—
3 months	84	23.9 (21.9-25.8)	79	23.4 (21.3-25.6)	0.45 (−0.6 to 1.5)	0.06	.41
12 months	74	23.6 (21.7-25.6)	70	24 (21.7-26.2)	−0.3 (−1.5 to 0.9)	−0.05	.61
Powerful others locus of control (6-36)							
Baseline	99	17.4 (14.8-20.0)	96	18.8 (15.8-21.8)	—	—	—
3 months	83	16.5 (15.0-18.0)	79	16.1 (14.3-18.0)	0.37 (−0.8 to 1.5)	0.05	.53
12 months	73	15.2 (13.6-6.9)	70	16 (14.1-17.9)	−0.74 (−2.0 to 0.6)	−0.1	.26
Anxiety (0-21)							
Baseline	100	4 (2.5-5.6)	97	4.2 (2.6-5.9)	—	—	—
3 months	85	3.5 (2.5-4.5)	79	4.2 (3.1-5.2)	−0.64 (−1.3 to 0)	−0.15	.05
12 months	75	3.1 (2.0-4.3)	72	4.1 (2.9-5.2)	−0.9 (−1.6 to −0.2)	−0.21	.007
Depression (0-21)							
Baseline	100	4 (2.5-5.6)	96	4.2 (2.6-5.9)	—	—	—
3 months	85	2.6 (1.5-3.7)	78	3.2 (2.1-4.3)	−0.61 (−1.3 to 0.1)	−0.12	.09
12 months	75	2.4 (1.3-3.6)	72	3 (1.9-4.2)	−0.6 (−1.3 to 0.1)	−0.12	.09

[a]For symptoms, quality of life, sport and recreation, self-efficacy, active pain coping, and locus of control, a higher score indicates an improvement. For sedentary behavior, tiredness, pain, passive pain coping, anxiety and depression a lower score indicates an improvement. Results are based on GEE analyses and adjusted for corresponding baseline variables, age, OA location, and gender.

[b]I-C: difference between intervention and control group.

Discussion

Principal Findings

To date, unfortunately, a vast majority of patients with knee and/or hip OA remain sedentary and receive no help in the promotion of PA. Since a physically active lifestyle has been positively associated with physical function and pain [45], effective and accessible PA programs are needed. Findings from other Web-based PA interventions have been mixed [22,46,47]. With respect to the PASE questionnaire, this randomized controlled trial demonstrated that the Join2move intervention has the potential to improve PA behavior. Effect sizes for PA ranged between 0-0.19 and are congruent with findings from a meta-analysis that found an overall mean effect of 0.14 [22]. At 3 months and 12 months, PA scores in the intervention group increased with 1% (1 point) and 6% (11 points) compared to baseline. Objectively obtained PA yielded different patterns. The intervention group remained stable while the control group reported a PA reduction of 37 minutes after 3 months and 57 minutes after 1 year. A possible explanation, also highlighted in other studies [48,49] is that self-reports tend to overestimate follow-up PA levels when compared to objective monitoring by accelerometry. At the same time, accelerometer measurements are unable to register water activities such as swimming. Since swimming is a popular recreational activity for older adults in the Netherlands, underestimation of objective PA may have occurred.

Besides PA, we also found significant short-term improvements in the primary outcomes physical function and self-perceived effect. Over the long term, however, we found no significant effects for physical function and self-perceived effect. At 3 months and 12 months, physical function in the intervention group improved 15% (9 points) compared to baseline. According to a study by Roos et al [34], these values achieved the threshold of clinically meaningful improvement. Apart from the observed improvements in the primary outcome measures, we found beneficial effects for other physical (pain and fatigue) and psychological factors (self-efficacy, pain coping, and anxiety) in favor of the intervention group.

Since long-term follow-up studies demonstrated that effects of (Web-based) interventions are not sustained in the long term [21,22,50], we expected short-term rather than long-term PA effects. Surprisingly, we found only long-term effects in total PA. These results were confirmed by both self-reported and accelerometer data. Absence of short-term effects can be partly explained by improved self-reported PA outcomes in the control group. The potential presence of the so-called Hawthorne effect may have contributed to an overestimation of PA scores in the control group. Selective dropout, which may have enhanced the effects in the control group, was not found. A definitive explanation for the nonsignificant short-term differences remains unclear.

Several factors may have contributed to the success of the Join2move intervention. First, the program is the first Web-based PA intervention that focuses specifically on knee and hip OA. The intervention addresses how to perform PA despite the presence of pain. The gradual increase of activities changes the perception that physical movement is related to pain and reinforces confidence to improve PA performance [28]. This may have led to positive psychological and health

outcomes. Second, the Join2move intervention seeks to align with the day-to-day activities of people. Users perform common activities (eg, walking, cycling) that are easy to integrate in their daily routine. Third, over the course of 1 year, we systematically developed and evaluated the Join2move intervention. End-users considered the intervention as user-friendly and helpful [27], which is a prerequisite for effective Web-based interventions.

Nonusage attrition has been acknowledged as a common concern in the field of Web-based education [51]. In particular, interventions, such as Join2move, that use automatic functions with minimal human involvement suffer from substantial rates of nonusage. In this study, 94.0% of the participants actually started the program, 46.0% reached the adherence threshold of 6 out of 9 modules, and 19.0% finished all 9 weekly modules. When considered in light of other studies, these adherence rates can be interpreted as reasonable. Previous studies by Wanner et al [52] and Connon et al [53] showed that respectively 47% and 25% of the intervention subjects never logged in to their Web-based program. Similarly, in a Web-based intervention by Hansen and colleagues [54], only 7% of the participants used the program more than once. A possible explanation for the relatively high adherence rates could be that our program incorporated automatic email reminders and website refreshments. We believe, like others [55-58], that more advanced feedback systems and regular reminders will lead to even better rates of adherence. Therefore, future research should concentrate on which strategies can improve website usage. A second factor, which may have influenced our usage rates, is the recruitment strategy used in this study. Participants were self-selected volunteers who responded to advertisements. Since self-selected participants tend to be highly educated, healthy, and already motivated to change their PA behavior, it is presumed that they have better usage rates compared to those who do not elect to participate. For example, Hansen et al [54] attributed their poor usage rates to the non-self-selected sample. This suggests that Web-based interventions, especially those without supervision, could be most suitable for those who are already willing to change their PA levels. Details about the usage and nonusage of the Join2move are described in another publication [59].

With respect to dropout attrition, 9% (4/46) adherent and 48% (26/54) nonadherent subjects did not return at least one of the follow-up surveys. This is in line with the study by Eysenbach [51], indicating that dropout and nonusage attrition are linked to each other. Since dropout rates and demographics of dropouts were similar between conditions, it is not expected that this influenced the results of the study.

As there is no cure for OA, self-management is considered a key element in the nonpharmacological treatment of knee and/or hip OA [60,61]. Self-management aims to motivate OA patients to undertake changes necessary to improve physical and psychological well-being. Although the importance is generally acknowledged, provision of self-management is underutilized. Given the clinically relevant benefits and the self-help format, Join2move could be a key component in the effort to enhance self-management in sedentary patients with knee and/or hip OA. Considering the unique potential to reach large populations through Join2move, even the small effects observed in this study could have clinical public health consequences [19]. Besides the focus on outside-care populations, patients in a care setting may also benefit from Join2move. Therefore, future work should integrate and investigate Join2move in a health care environment.

Strengths and Limitations

First, the most important strengths are the design (ie, RCT) and the long-term character of the study. Second, we used both objective and subjective measures to assess PA. This study also has certain limitations that are important to acknowledge. First, participants were included on the basis of self-reported OA. Unfortunately, due to practical constraints, diagnosis was not confirmed through clinical tests or x-ray reports. In a previous pilot study [27], we verified self-reported OA through clinical tests. According to the American College of Rheumatology criteria [62,63], 80% had clinical knee or hip OA and 20% of the participants had no OA. These rates are in line with another validation study [64], reporting over 80% agreement between self-reported and clinically confirmed diagnoses. Although these rates are acceptable, it is presumed that we included false positive OA patients in our trial. Second, results could be biased by dropout of participants (15.6%, 31/199 at 3 months and 24.6%, 49/199 at 12 months). However, the nonresponse analysis showed similar baseline characteristics for responders, and nonresponders and dropouts were equally distributed between the intervention group and the control group. Third, with respect to the outcome variable PA, the study involved two different measures (questionnaires and accelerometers) on two occasions (3 months and 12 months). We acknowledge that this may have increased the possibility of Type I errors. Fourth, the representativeness was limited by the self-selected sample used in this study. Responders were predominantly healthy and highly educated patients. This widely recognized phenomenon is called “The inverse information law” [65]; Web-based interventions fail to reach those for whom PA behavior changes are most necessary [21,22,66-69]. In order to eliminate this issue, future studies should focus on how these specific groups could be involved in the field of Web-based education.

Conclusions

Health care providers, such as GPs and physical therapists, may play a pivotal role in the referral of patients to Web-based interventions. Furthermore, it will be important to translate Web-based interventions, such as Join2move, to other self-help formats (eg, videos, brochures, and self-help books).

Conflicts of Interest

The authors Daniel Bossen and Cindy Veenhof are the creators of the Join2move intervention.

Multimedia Appendix 1

Screenshots of the Join2move intervention.

[PDF File (Adobe PDF File), 654KB - jmir_v15i11e257_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.2 [24].

[PDF File (Adobe PDF File), 989KB - jmir_v15i11e257_app2.pdf]

References

1. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Conn VS, Hafdahl AR, Brown SA, Brown LM. Meta-analysis of patient education interventions to increase physical activity among chronically ill adults. Patient Educ Couns 2008 Feb;70(2):157-172 [FREE Full text] [doi: 10.1016/j.pec.2007.10.004] [Medline: 18023128]
3. Dunlop DD, Song J, Semanik PA, Sharma L, Chang RW. Physical activity levels and functional performance in the osteoarthritis initiative: a graded relationship. Arthritis Rheum 2011 Jan;63(1):127-136 [FREE Full text] [doi: 10.1002/art.27760] [Medline: 20862681]
4. Esser S, Bailey A. Effects of exercise and physical activity on knee osteoarthritis. Curr Pain Headache Rep 2011 Dec;15(6):423-430. [doi: 10.1007/s11916-011-0225-z] [Medline: 21956792]
5. van Dijk GM, Dekker J, Veenhof C, van den Ende CH, Carpa Study Group. Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. Arthritis Rheum 2006 Oct 15;55(5):779-785 [FREE Full text] [doi: 10.1002/art.22244] [Medline: 17013827]
6. Creamer P, Lethbridge-Cejku M, Hochberg MC. Determinants of pain severity in knee osteoarthritis: effect of demographic and psychosocial variables using 3 pain measures. J Rheumatol 1999 Aug;26(8):1785-1792. [Medline: 10451078]
7. Hootman JM, Macera CA, Ham SA, Helmick CG, Sniezek JE. Physical activity levels among the general US adult population and in adults with and without arthritis. Arthritis Rheum 2003 Feb 15;49(1):129-135 [FREE Full text] [doi: 10.1002/art.10911] [Medline: 12579604]
8. Shih M, Hootman JM, Kruger J, Helmick CG. Physical activity in men and women with arthritis National Health Interview Survey, 2002. Am J Prev Med 2006 May;30(5):385-393. [doi: 10.1016/j.amepre.2005.12.005] [Medline: 16627126]
9. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, Standing Committee for International Clinical Studies Including Therapeutic Trials ESCISIT. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003 Dec;62(12):1145-1155 [FREE Full text] [Medline: 14644851]
10. Cottrell E, Roddy E, Foster NE. The attitudes, beliefs and behaviours of GPs regarding exercise for chronic knee pain: a systematic review. BMC Fam Pract 2010;11:4 [FREE Full text] [doi: 10.1186/1471-2296-11-4] [Medline: 20082694]
11. Denoeud L, Mazières B, Payen-Champenois C, Ravaud P. First line treatment of knee osteoarthritis in outpatients in France: adherence to the EULAR 2000 recommendations and factors influencing adherence. Ann Rheum Dis 2005 Jan;64(1):70-74 [FREE Full text] [doi: 10.1136/ard.2003.015263] [Medline: 15608302]
12. Porcheret M, Jordan K, Jinks C, Croft P, Primary Care Rheumatology Society. Primary care treatment of knee pain--a survey in older adults. Rheumatology (Oxford) 2007 Nov;46(11):1694-1700 [FREE Full text] [doi: 10.1093/rheumatology/kem232] [Medline: 17938135]
13. Mamlin LA, Melfi CA, Parchman ML, Gutierrez B, Allen DI, Katz BP, et al. Management of osteoarthritis of the knee by primary care physicians. Arch Fam Med 1998 Nov;7(6):563-567. [Medline: 9821832]
14. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Ann Rheum Dis 2001 Feb;60(2):91-97 [FREE Full text] [Medline: 11156538]
15. Der Ananian C, Wilcox S, Saunders R, Watkins K, Evans A. Factors that influence exercise among adults with arthritis in three activity levels. Prev Chronic Dis 2006 Jul;3(3):A81 [FREE Full text] [Medline: 16776882]
16. Petursdottir U, Arnadottir SA, Halldorsdottir S. Facilitators and barriers to exercising among people with osteoarthritis: a phenomenological study. Phys Ther 2010 Jul;90(7):1014-1025 [FREE Full text] [doi: 10.2522/ptj.20090217] [Medline: 20466741]
17. Internet World Stats. World Internet Users and Population Stats URL: http://www.internetworldstats.com/stats.htm [accessed 2012-06-07] [WebCite Cache ID 68EtPhZnB]
18. CBS-Statline: CBS Statistics Netherlands, The Hague (2011). URL: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71098NED&D1=0-14,33-34,41-47,55-59,65-84,114-133&D2=12-13&D3=a&HDR=G1,G2&STB=T&VW=T [accessed 2013-10-08] [WebCite Cache ID 685vL3AsT]
19. 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]

20. Bossen D, Veenhof C, Dekker J, de Bakker D. The Effectiveness of Self-Guided Web-Based Physical Activity Interventions Among Patients with a Chronic Disease: A Systematic Review. J Phys Act Health 2013 Mar 14:1. [Medline: 23493018]
21. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
22. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
23. Veenhof C, Köke AJ, Dekker J, Oostendorp RA, Bijlsma JW, van Tulder MW, et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: A randomized clinical trial. Arthritis Rheum 2006 Dec 15;55(6):925-934 [FREE Full text] [doi: 10.1002/art.22341] [Medline: 17139639]
24. 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]
25. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
26. Nielsen J. Iterative user-interface design. Computer 1993 Nov;26(11):32-41. [doi: 10.1109/2.241424]
27. Bossen D, Veenhof C, Dekker J, de Bakker D. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013;13:61 [FREE Full text] [doi: 10.1186/1472-6947-13-61] [Medline: 23714120]
28. Fordyce WE, Fowler RS, Lehmann JF, Delateur BJ, Sand PL, Trieschmann RB. Operant conditioning in the treatment of chronic pain. Arch Phys Med Rehabil 1973 Sep;54(9):399-408. [Medline: 4729785]
29. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. J Clin Epidemiol 1993 Feb;46(2):153-162. [Medline: 8437031]
30. Actigraph. URL: http://www.actigraphcorp.com/products/ [accessed 2013-02-21] [WebCite Cache ID 6EbD4MMWg]
31. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011;8:62 [FREE Full text] [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]
32. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
33. ActiGraph. MeterPlus User's Manual, Version 4.3.0. San Diego State: MeterPlus; 2011.
34. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64 [FREE Full text] [doi: 10.1186/1477-7525-1-64] [Medline: 14613558]
35. de Groot IB, Favejee MM, Reijman M, Verhaar JA, Terwee CB. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. Health Qual Life Outcomes 2008;6:16 [FREE Full text] [doi: 10.1186/1477-7525-6-16] [Medline: 18302729]
36. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol 2003;32(1):46-51. [Medline: 12635946]
37. de Groot IB, Reijman M, Terwee CB, Bierma-Zeinstra S, Favejee MM, Roos E, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. Osteoarthritis Cartilage 2009 Jan;17(1):132. [doi: 10.1016/j.joca.2008.05.014] [Medline: 18620874]
38. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
39. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989 Jan;32(1):37-44. [Medline: 2912463]
40. Taal E, Rasker JJ, Wiegman O. Group education for rheumatoid arthritis patients. Semin Arthritis Rheum 1997 Jun;26(6):805-816. [Medline: 9213379]
41. Kraaimaat FW, Evers AW. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med 2003;10(4):343-363. [Medline: 14734263]
42. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Educ Monogr 1978;6(2):160-170. [Medline: 689890]
43. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J: L. Erlbaum Associates; 1988.
44. Twisk J, de Vente W. Attrition in longitudinal studies. How to deal with missing data. J Clin Epidemiol 2002 Apr;55(4):329-337. [Medline: 11927199]
45. Dunlop DD, Song J, Semanik PA, Sharma L, Chang RW. Physical activity levels and functional performance in the osteoarthritis initiative: a graded relationship. Arthritis Rheum 2011 Jan;63(1):127-136 [FREE Full text] [doi: 10.1002/art.27760] [Medline: 20862681]
46. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]

47. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
48. Hagströmer M, Oja P, Sjöström M. Physical activity and inactivity in an adult population assessed by accelerometry. Med Sci Sports Exerc 2007 Sep;39(9):1502-1508. [doi: 10.1249/mss.0b013e3180a76de5] [Medline: 17805081]
49. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S1-14. [Medline: 10925819]
50. Pisters MF, Veenhof C, van Meeteren NL, Ostelo RW, de Bakker DH, Schellevis FG, et al. Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review. Arthritis Rheum 2007 Oct 15;57(7):1245-1253 [FREE Full text] [doi: 10.1002/art.23009] [Medline: 17907210]
51. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
52. Wanner M, Martin-Diener E, Braun-Fahrländer C, Bauer G, Martin BW. Effectiveness of active-online, an individually tailored physical activity intervention, in a real-life setting: randomized controlled trial. J Med Internet Res 2009;11(3):e23 [FREE Full text] [doi: 10.2196/jmir.1179] [Medline: 19666456]
53. McConnon A, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The Internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007;7:206 [FREE Full text] [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
54. Hansen AW, Grønbæk M, Helge JW, Severin M, Curtis T, Tolstrup JS. Effect of a Web-based intervention to promote physical activity and improve health among physically inactive adults: a population-based randomized controlled trial. J Med Internet Res 2012;14(5):e145 [FREE Full text] [doi: 10.2196/jmir.2109] [Medline: 23111127]
55. 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]
56. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
57. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
58. 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]
59. 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]
60. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, American College of Rheumatology. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012 Apr;64(4):465-474. [Medline: 22563589]
61. Bennell KL, Hunter DJ, Hinman RS. Management of osteoarthritis of the knee. BMJ 2012;345:e4934. [Medline: 22846469]
62. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986 Aug;29(8):1039-1049. [Medline: 3741515]
63. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991 May;34(5):505-514. [Medline: 2025304]
64. March LM, Schwarz JM, Carfrae BH, Bagge E. Clinical validation of self-reported osteoarthritis. Osteoarthritis Cartilage 1998 Mar;6(2):87-93. [doi: 10.1053/joca.1997.0098] [Medline: 9692063]
65. Eysenbach G. Poverty, human development, and the role of eHealth. J Med Internet Res 2007;9(4):e34 [FREE Full text] [doi: 10.2196/jmir.9.4.e34] [Medline: 17951217]
66. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
67. Kelders SM, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
68. De Cocker K, Spittaels H, Cardon G, De Bourdeaudhuij I, Vandelanotte C. Web-based, computer-tailored, pedometer-based physical activity advice: development, dissemination through general practice, acceptability, and preliminary efficacy in a

randomized controlled trial. J Med Internet Res 2012;14(2):e53 [FREE Full text] [doi: 10.2196/jmir.1959] [Medline: 22532102]

69. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]

Abbreviations

BGA: behavior graded activity
BMI: body mass index
ES: effect size
GEE: generalized estimated equations
GP: general practitioner
HADS: Hospital Anxiety and Depression Scale
HOOS: hip OA outcome score
KOOS: knee OA outcome score
NIVEL: Netherlands Institute for Health Services Research
OA: osteoarthritis
OR: odds ratio
PA: physical activity
PARQ: Physical Activity Readiness Questionnaire
PASE: Physical Activity Scale for the Elderly
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 10.04.13; peer-reviewed by S Kelders, J Gao, L Brosseau; comments to author 15.06.13; revised version received 16.07.13; accepted 07.10.13; published 22.11.13.

Please cite as:
Bossen D, Veenhof C, Van Beek KEC, Spreeuwenberg PMM, Dekker J, De Bakker DH
Effectiveness of a Web-Based Physical Activity Intervention in Patients With Knee and/or Hip Osteoarthritis: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e257
URL: http://www.jmir.org/2013/11/e257/
doi:10.2196/jmir.2662
PMID:24269911

Original Paper

Diabetes Self-Management Smartphone Application for Adults With Type 1 Diabetes: Randomized Controlled Trial

Morwenna Kirwan[1], PhD; Corneel Vandelanotte[1], PhD; Andrew Fenning[1], PhD; Mitch J Duncan[1], PhD

Institute for Health and Social Science Research, Central Queensland University, North Rockhampton, Australia

Corresponding Author:
Morwenna Kirwan, PhD
Institute for Health and Social Science Research
Central Queensland University
Building 18, Bruce Highway
North Rockhampton, 4702
Australia
Phone: 61 0749306977
Fax: 61 0749306402
Email: m.kirwan@uws.edu.au

Abstract

Background: Persistently poor glycemic control in adult type 1 diabetes patients is a common, complex, and serious problem initiating significant damage to the cardiovascular, renal, neural, and visual systems. Currently, there is a plethora of low-cost and free diabetes self-management smartphone applications available in online stores.

Objective: The aim of this study was to examine the effectiveness of a freely available smartphone application combined with text-message feedback from a certified diabetes educator to improve glycemic control and other diabetes-related outcomes in adult patients with type 1 diabetes in a two-group randomized controlled trial.

Methods: Patients were recruited through an online type 1 diabetes support group and letters mailed to adults with type 1 diabetes throughout Australia. In a 6-month intervention, followed by a three-month follow-up, patients (n=72) were randomized to usual care (control group) or usual care and the use of a smartphone application (Glucose Buddy) with weekly text-message feedback from a Certified Diabetes Educator (intervention group). All outcome measures were collected at baseline and every three months over the study period. Patients' glycosylated hemoglobin levels (HbA1c) were measured with a blood test and diabetes-related self-efficacy, self-care activities, and quality of life were measured with online questionnaires.

Results: The mean age of patients was 35.20 years (SD 10.43) (28 male, 44 female), 39% (28/72) were male, and patients had been diagnosed with type 1 diabetes for a mean of 18.94 years (SD 9.66). Of the initial 72 patients, 53 completed the study (25 intervention, 28 control group). The intervention group significantly improved glycemic control (HbA1c) from baseline (mean 9.08%, SD 1.18) to 9-month follow-up (mean 7.80%, SD 0.75), compared to the control group (baseline: mean 8.47%, SD 0.86, follow-up: mean 8.58%, SD 1.16). No significant change over time was found in either group in relation to self-efficacy, self-care activities, and quality of life.

Conclusions: In adjunct to usual care, the use of a diabetes-related smartphone application combined with weekly text-message support from a health care professional can significantly improve glycemic control in adults with type 1 diabetes.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12612000132842; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12612000132842 (Archived by WebCite at http://www.webcitation.org/6Kl4jqn5u).

(J Med Internet Res 2013;15(11):e235) doi:10.2196/jmir.2588

KEYWORDS

type 1 diabetes; mobile health; mobile phone; text message; education

Introduction

Persistently poor glycemic control in adult type 1 diabetes patients is a common, complex, and serious problem initiating significant damage to the cardiovascular, renal, neural, and visual systems [1]. In many patients, glycosylated hemoglobin levels (HbA_{1c}) are unsatisfactory, with levels consistently above 8.0% [2]. In the pursuit of improving metabolic control, the importance of self-monitoring blood glucose is widely appreciated and recommended as a routine part of management in patients with type 1 diabetes [3]. There are a number of barriers to glycemic control in type 1 diabetes, including the fear of hypoglycemia and the demands of day-to-day management, in particular the need for frequent self-monitoring of blood glucose and regular adjustments in insulin dosing [4]. Additionally another difficulty is a patient's logbook, either paper-based or electronic, that a clinician is presented with at a consultation. Clinicians often face a lack of information on which to base their advice regarding their patient's self-care [5,6]. Utilizing mobile phone technology may help to overcome these difficulties.

The worldwide prevalence of mobile phones makes them a powerful platform for providing individualized health care delivered at the patient's convenience. Several reviews have documented the effectiveness, potential, and challenges in using mobile phones to improve health outcomes in diabetes [7-14]. Growing evidence suggests that utilizing mobile phones may improve diabetes self-management and clinical outcomes; however, this evidence is much stronger for type 2 populations than type 1 populations [9].

In recent years, mobile phones have improved dramatically in both design and function, from simple call and text devices to the more sophisticated mini-personal computers known as smartphones. Smartphone owners are now more prevalent within the overall population than owners of traditional mobile phones [15]. Smartphones allow individual users to install, configure, and run specialized applications on their phone. Increasing numbers of people are using these applications to self-manage chronic diseases [13]. For example, Chomutare and colleagues [16] identified that in 2011 there were more than 260 diabetes-related iPhone applications available for download from the Apple online store.

A small number of prototypes of type 1 diabetes smartphone applications have been developed and tested in clinical settings [2,17-21]. However, with the plethora of low-cost and free self-management diabetes applications currently available in online markets (Apple iPhone, Google Android, BlackBerry, and Nokia Symbian), it is pertinent to examine their effectiveness when integrated in secondary care [16]. Therefore, the aim of this study was to examine the effectiveness of a freely available smartphone application combined with text-message feedback to improve glycemic control and other diabetes-related outcomes in adult patients with type 1 diabetes in a two-group randomized controlled trial.

Methods

Design

The study, utilizing a two-arm (usual care and intervention) randomized controlled trial with measures at baseline, 3, 6, and 9 months (Figure 1), was conducted with the assistance of a Certified Diabetes Educator (CDE). Participants were recruited nationally by means of an invitation letter sent to type 1 diabetes patients registered with Diabetes Australia in New South Wales (n=3809) and Queensland (n=3207), as well as an advertisement in a type 1 diabetes national newsletter (Yada Yada newsletter) emailed to more than 5000 recipients and promotion in an online community forum (Reality Check Forum). Study inclusion criteria were: (1) aged 18-65 years, (2) diagnosed with type 1 diabetes >6 months, (3) HbA_{1c} >7.5%, (4) treated with multiple daily injections or insulin pump, and (5) own a smartphone (iPhone). Patients were excluded if they were pregnant or already using a smartphone application to self-manage their diabetes. This study was approved by Central Queensland University Human Research Ethics Board.

After confirming eligibility (via phone call) and obtaining written informed consent (via email) from the patient and their primary diabetes health care practitioner (general practitioner or endocrinologist), the study coordinator randomized patients using a freely available online randomization program. A permuted block randomization design method was used during the 3-month rolling recruitment to ensure roughly equal numbers of patients were allocated to each comparison group [22]. There was no face-to-face contact between the patients and research team at any point in the study, which allowed participants to live anywhere in Australia.

Intervention

Patients in both groups were asked to continue with their usual care, which included a visit to their primary diabetes health care practitioner every 3 months. Additionally, patients allocated to the intervention arm were given instructions to download the smartphone application named "Glucose Buddy". Glucose Buddy is a freely available diabetes self-management iPhone application that allows users to manually enter blood glucose levels, insulin dosages, other medications, diet (food item in grams), and physical activities (minutes) [16,23]. Users can also view their data on a customizable graph and export this information via email (Figures 2 and 3). Glucose Buddy was developed by SkyHealth LLC and was first made available on iTunes (Apple online store) in October 2008. Glucose Buddy has been reported to be the most downloaded diabetes management software on iOS, with downloads in excess of 100,000. There was no minimum amount of logging required and intervention patients were able to utilize the accompanying Glucose Buddy website to log diabetes parameters at their discretion.

The information logged in the Glucose Buddy application was reviewed by a CDE via a Web interface on a weekly basis. All patients in the intervention arm were sent a minimum of 1 personalized text-message communication per week for the first

6 months of the study. At the 6-month timeframe, all text-message communication ceased.

Figure 1. Participant flow. Note: Reason for subject "lost to follow-up" could not be determined as patients could not be re-contacted. CDE: Certified Diabetes Educator.

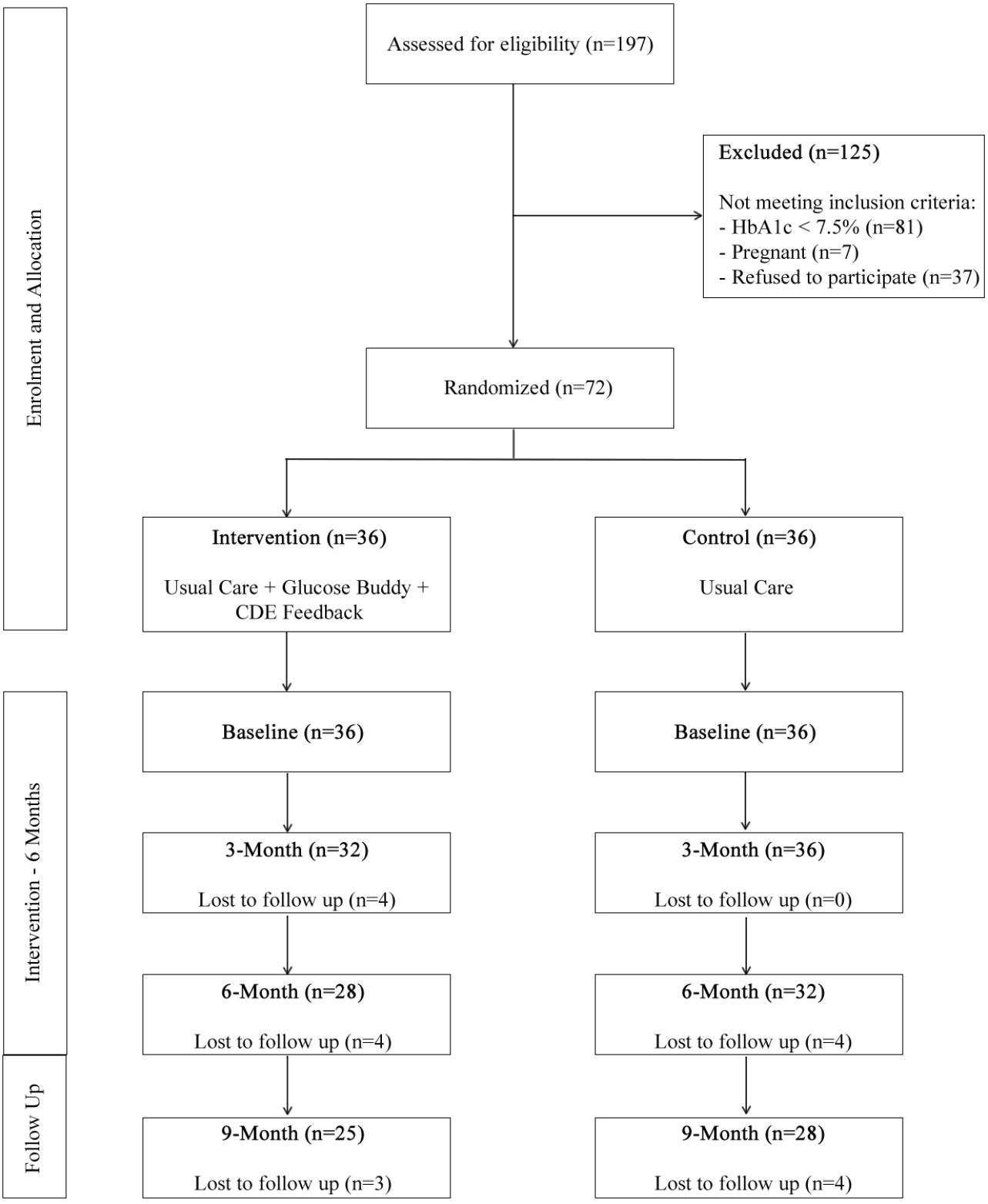


Figure 2. Glucose Buddy screenshots of the menu and adding a log.

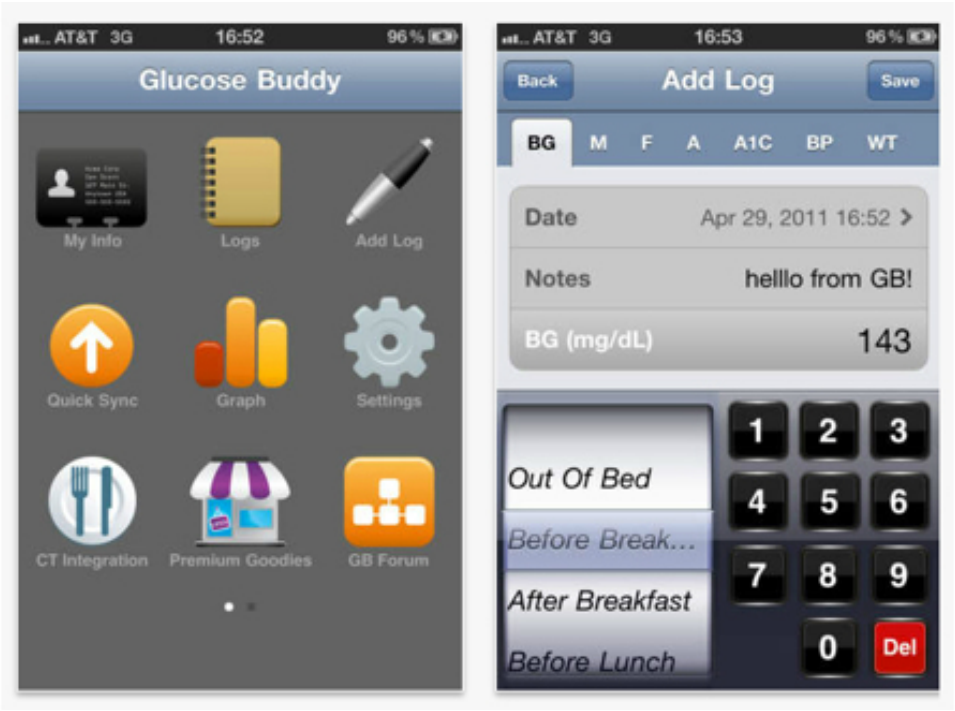


Figure 3. Glucose Buddy screenshots of the graphic display of logs and option to set up reminders.

Measures

All measures were collected at baseline and every 3 months over the 9-month study period for both groups (making 4 time points in total). The primary outcome measure was change in glycemic control assessed by HbA_{1c}, which was collected by a pathology lab at the request of the patients' general practitioner or endocrinologist as per usual care (every 3 months) and then forwarded to the research team. The secondary outcome measures, being diabetes-related self-efficacy, self-care activities, and quality of life, were collected via a Web-based survey. Details to access this survey were emailed to patients.

Diabetes-related self-efficacy was measured using the short form of the Diabetes Empowerment questionnaire (DES-SF). The DES-SF questionnaire measures eight conceptual dimensions relevant to the management of diabetes: (1) assessing the need for change, (2) developing a plan, (3) overcoming barriers, (4) asking for support, (5) supporting oneself, (6) coping with emotion, (7) motivating oneself, and (8) making diabetes care choices appropriate for one's priorities and circumstances. Patients respond to all items on a 5-point

Likert scale, with 1-strongly disagree to 5-strongly agree. Higher scores on DES-SF reflect a better self-efficacy. Research has shown that DES-SF is a valid and reliable measure of overall diabetes-related psychosocial self-efficacy [24,25].

The Summary of Diabetes Self-Care Activities (SDSCA) measure was used for assessing diabetes self-care behaviors. The SDSCA contains 6 scales: (1) general dietary behavior, (2) specific dietary behavior, (3) glucose monitoring, (4) exercise, (5) foot care, and (6) smoking. Only the first four listed scales were used in this study. Higher scores on the SDSCA reflect a greater number of days per week that self-care activities are carried out (range 0-7). Psychometric analyses support the construct validity and internal consistency reliability of SDSCA in an adult diabetes population [26].

The valid and reliable Diabetes Quality of Life (DQOL) scale [27] was used to assess the three aspects of quality of life directly related to diabetes: diabetes satisfaction, impact, and worry. All the subscales were included. Patients respond to all items on a 5-point Likert scale. A score of 1 indicates “always affected”, “always worried”, or “never satisfied”. A score of 5 represents “no impact”, “no worries”, or “always satisfied”. Higher scores on DQOL scales reflect better quality of life.

Intervention participants’ engagement with the intervention was also measured in terms of text-message communications with the CDE and the utilization of the Glucose Buddy application. Specifically, the number of text messages sent to patients and the number of text-message responses were collected as well as the number of logs (blood glucose, insulin, physical activity, and diet) entered by patients in the Glucose Buddy application.

Statistical Analysis

Demographic characteristics of participants and baseline data for all measures were compared between both study groups to detect differences at baseline using a series of independent sample *t* tests and chi-square tests. Logistic regression analyses were conducted to evaluate whether participant characteristics (age [years], duration of diabetes [months/years], gender, insulin pump use [Y/N], and baseline HbA_{1c}) were related to dropout (completed vs didn’t complete all assessments) during the study. This statistical method is common when evaluating the characteristics that may be related to attrition examined as a dichotomous outcome. Primary (HbA_{1c}) and secondary outcomes (diabetes-related self-efficacy, quality of life, and self-care) measures were analyzed using linear mixed effects models for repeated measures. Linear mixed model analysis allows for inclusion of cases with missing data, without replacement of missing values, and therefore includes all randomized patients. The linear mixed model analysis used group and the covariates of age (years), gender, and diabetes duration (years) as fixed effects. The “Type III Wald test” was used to test overall statistical significance of the effects. Linear regression analysis was conducted to analyze whether engagement in the study by the intervention group was predictive of change in HbA_{1c}; it allows to assess whether patients who engaged more in the intervention, in terms of text-message communications and logging parameters in the Glucose Buddy application, had a greater change in HbA_{1c}. Statistical significance was defined as $P<.05$ for all analysis and conducted using SPSS for Windows (Version 18.0).

The sample size was calculated on the expected difference in mean (1.5%) in the primary outcome variable (HbA_{1c}), and the logistically maximum available sample size was 36 patients per group based on part-time work status of the CDE. We allowed for a dropout of 11% (4 per group), consistent with dropout rates reported in recent reviews of similar studies [9,10], and variation in baseline (HbA_{1c}=1.80) similar to previous studies [28]. Based on these parameters and using an alpha of .05 and 90% power, the estimated sample size was 68 in total and subsequently increased to 72 in line with the maximum caseload of the CDE [29].

Results

Participant Characteristics

In total, 197 adults with type 1 diabetes registered their interest online or via phone call to the research team and were assessed for eligibility (Figure 1), with 125 excluded for not meeting the inclusion criteria. Seventy-two individuals were randomized to the two groups. Linear mixed model analysis allows for inclusion of cases with missing data, without replacement or imputation of missing values. Therefore, this analysis approach includes all available data of randomized patients at each time point as indicated in Figure 1. Table 1 provides an outline of the participant’s characteristics. Mean age of patients was 35.20 years (SD 10.43), 39% (28/72) were male, and patients had been diagnosed with type 1 diabetes for a mean of 18.94 years (SD 9.66). In total, 38% (27/72) of patients were using an insulin pump, with no significant difference between groups, χ^2_1=0.59, P=.81. The intervention group had a significantly higher (P=.02) baseline HbA_{1c} (mean 9.08, SD 1.18) than the control group (mean 8.47, SD 0.86) and reported a healthier diet (mean 3.56, SD 1.70 healthy days per week for the intervention group versus mean 2.60, SD 1.98 days for the control group, P=.03). There were significantly (P=.02) more females (75%, 27/36) in the control group. No other baseline differences were observed between groups. Dropout was 26% (11 males, 8 females, 19/72) with logistic regression analysis revealing no significant difference in age, gender, diabetes duration, insulin pump use, and baseline HbA_{1c} among those that completed the study and those that were lost to follow up.

Table 1. Participant characteristics at baseline.

		Overall, n=72	Control, n=36	Intervention, n=36	*P* value
Gender (M/F), n		28/44	9/27	19/17	.02
Age (years), mean (SD)		35.20 (10.43)	34.42 (10.26)	35.97 (10.67)	.51
Diabetes duration (years), mean (SD)		18.94 (9.66)	18.19 (9.77)	19.69 (9.64)	.53
Insulin pump use, n		27/72	13/36	14/36	.81
HbA $_{1c}$, mean (SD)					
	Total group	8.78 (1.07)	8.47 (0.86)	9.08 (1.18)	.02
	Male	8.79 (1.31)	8.17 (0.65)	9.10 (1.45)	.08
	Female	8.77 (0.90)	8.57 (0.91)	9.07 (0.84)	.08

Intervention Effects

As outlined in Table 2, the linear mixed model revealed that there was a significant interaction effect between groups for HbA_{1c} ($F_{1,246}$=20.07, P<.001). The intervention group had a significant decrease in HbA_{1c} (mean −1.10, SD 0.74, P<.001) over the 9-month study, compared to the control group which had a non-significant increase (mean 0.07, SD 0.99). There was a statistically significant change in the diabetes self-care measure of specific diet over time, but there was no difference between the two groups. No significant differences were observed for all other outcomes.

Engagement by Intervention Patients

Intervention patients' engagement with the Glucose Buddy application, in terms of the number of logs and text messages communicated between patients and the CDE, is outlined in Table 3. Over the 6-month intervention period, the CDE sent 1714 text messages in total, which equates to approximately 2 text messages per patient per week. Patients sent in total 559 text messages to the CDE over the 6-month period. The first month of the study was used for the CDE and the intervention group patients to establish a relationship—they never met in person. Thus, the text messages sent to patients (mean 9.75, SD 1.96) in the first month and those received by the CDE (mean 6.47, SD 3.92) are higher than the average of the other five months. Using the Glucose Buddy application, patients logged 24,720 diabetes parameters in total: 54.00% (13,349/24,720) of the logs related to blood glucose levels, 33.00% (8158/24,720) to insulin, 12.00% (2966/24,720) to diet, and 1.00% (247/24,720) to exercise. Linear regression analysis revealed no significant relationship between level of engagement and change in HbA_{1c} in the intervention, as measured by text messages sent to the patients, text messages received by the CDE, and the number of logs entered in the Glucose Buddy application.

Text Message Themes

The content of the text messages sent to and received from patients fell into four broad thematic categories covering feedback on logs, diabetes questions, educational tips, and positive reinforcement. Examples of text messages for each category are outlined in Table 4.

Costs Incurred

The Glucose Buddy application was freely available. A text-messaging software program was used to text message patients. Total cost over the study was $290.93 AUD; this equates to $8.08 per patient (n=36). The CDE spent on average 3 hours per week reviewing patients' logs and text messaging patients in the intervention group. This equated to 5 minutes per patient (n=36) per week (72 hours in total over 6-month period). A CDE hourly rate is approximately $28.85; thus, the total cost over the study was $2077.20 AUD.

Table 2. Results of linear mixed model analysis for diabetes glycemic control, self-efficacy, self-care, and quality of life measures.

		Overall[a]	Control[b]	Intervention[c]	Overall *F* statistic, (*df*=1,246)		
		Mean (SD)	Mean (SD)	Mean (SD)	Time	Group	Time x Group
HbA $_{1c}$ (%)							
	Baseline	8.78 (1.07)	8.47 (0.86)	9.08 (1.18)[d]			
	3-month	8.27 (0.86)	8.23 (0.89)	8.32 (0.84)	28.79[e]	10.55[d]	20.07[e]
	6-month	8.22 (0.91)	8.43 (1.00)	7.97 (0.73)			
	9-month	8.21 (1.05)	8.58 (1.16)	7.80 (0.75)			
SDSCA [f]							
General Diet							
	Baseline	3.81 (2.06)	4.19 (1.88)	3.42 (2.19)			
	3-month	4.10 (2.00)	3.97 (2.13)	4.23 (1.86)	5.30[d]	1.92	3.40
	6-month	4.36 (1.86)	4.16 (1.98)	4.59 (1.73)			
	9-month	4.37 (1.83)	4.14 (1.85)	4.62 (1.80)			
Specific Diet							
	Baseline	3.08 (1.89)	2.60 (1.98)	3.56 (1.70) [d]			
	3-month	3.10 (1.63)	3.00 (1.76)	3.22 (1.48)	0.52	6.90[d]	3.36
	6-month	3.68 (1.63)	3.56 (1.66)	3.80 (1.60)			
	9-month	3.93 (1.61)	4.05 (1.43)	3.80 (1.82)			
Exercise							
	Baseline	2.74 (2.09)	2.92 (2.12)	2.57 (2.08)			
	3-month	2.45 (2.12)	2.53 (2.31)	2.36 (1.91)	1.44	2.22	0.93
	6-month	2.83 (1.95)	3.06 (1.97)	2.55 (1.92)			
	9-month	2.96 (1.79)	2.82 (1.75)	3.12 (1.86)			
Glucose Testing							
	Baseline	5.46 (2.04)	5.51 (2.08)	5.40 (2.03)			
	3-month	5.88 (1.64)	5.75 (1.65)	6.02 (1.64)	3.65	1.42	1.86
	6-month	6.10 (1.56)	6.02 (1.67)	6.20 (1.46)			
	9-month	5.92 (1.62)	5.61 (1.95)	6.28 (1.06)			
DES-SF [g]							
	Baseline	3.62 (0.77)	3.62 (0.65)	3.62 (0.89)			
	3-month	3.78 (0.64)	3.70 (0.65)	3.88 (0.61)	0.01	0.16	0.001
	6-month	3.73 (0.77)	3.64 (0.81)	3.82 (0.73)			
	9-month	3.61 (0.72)	3.62 (0.74)	3.60 (0.71)			
DQOL [h]							
Satisfaction							
	Baseline	3.14 (0.61)	3.09 (0.55)	3.20 (0.66)			
	3-month	3.22 (0.61)	3.11 (0.54)	3.35 (0.67)	0.55	1.42	0.01
	6-month	3.32 (0.60)	3.23 (0.62)	3.43 (0.58)			
	9-month	3.35 (0.66)	3.29 (0.65)	3.42 (0.68)			
Impact							
	Baseline	3.70 (0.53)	3.66 (0.54)	3.75 (0.52)			

		Overall[a]	Control[b]	Intervention[c]	Overall *F* statistic, (*df*=1,246)		
		Mean (SD)	Mean (SD)	Mean (SD)	Time	Group	Time x Group
	3-month	3.81 (0.57)	3.75 (0.56)	3.89 (0.58)	0.58	0.71	0.07
	6-month	3.83 (0.61)	3.74 (0.61)	3.94 (0.62)			
	9-month	3.85 (0.53)	3.77 (0.55)	3.93 (0.52)			
Worry							
	Baseline	3.98 (0.66)	3.90 (0.78)	4.06 (0.52)			
	3-month	4.10 (0.70)	4.01 (0.77)	4.19 (0.61)	1.77	1.26	0.70
	6-month	4.16 (0.70)	3.99 (0.82)	4.35 (0.46)			
	9-month	4.15 (0.63)	3.99 (0.76)	4.34 (0.36)			

[a]Number of participants included in both groups at each time point is: baseline n=72, 3 month n=68, 6 month n=60, 9 month n=53.

[b]Number of participants included in the Intervention group at each time point is: baseline n=36, 3 month n=32, 6 month n=28, 9 month n=25.

[c]Number of participants included in the Control group at each time point is: baseline n=36, 3 month n=36, 6 month n=32, 9 month n=28.

[d]*P*<.05

[e]*P*<.001

[f]SDSCA: Summary of Diabetes Self Care Activities

[g]DES-SF: Diabetes Empowerment Scale

[h]DQOL: Diabetes Quality of Life

Table 3. Engagement with intervention.

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Average number of text messages sent to patients, mean (SD)	9.75 (1.96)	6.67 (1.47)	7.58 (2.12)	7.11 (1.91)	8.56 (2.26)	7.94 (2.52)
Median number of text messages sent to patients	9	7	7	7.5	8	8
Total number of text messages sent to patients	351	240	273	256	308	286
Average number of text messages received by CDE[a] , mean (SD)	6.47 (3.92)	2.36 (2.82)	2.03 (2.52)	1.11 (1.30)	2.22 (2.84)	1.33 (2.46)
Median number of text messages received by CDE	6	2	1	1	1.5	0
Total number of text messages received by CDE	233	85	73	40	80	48
Glucose Buddy Logs, mean (SD)	187.53 (137.41)	137.22 (143.39)	92.03 (109.66)	96.02 (129.06)	89.03 (107.14)	84.83 (153.47)
Glucose Buddy Logs, total N	6751	4940	3313	3457	3205	3054

[a]CDE: Certified Diabetes Educator

Table 4. Text messages.

Themes	Examples
Feedback on logs (blood glucose, insulin, diet)	*Hi <patient name>- Another pretty good week - just a bit concerned about some odd higher levels in the morning - looks like some of these are forgotten doses - would that be right? Otherwise all are getting better and no real hypos. Be aware you may need to tweak basal if those highs are not related to forgotten doses. Your thoughts?* [Sent from Certified Diabetes Educator]
	Hi Veronica- Yes I have been having my breakfast later after I arrive to work, I get distracted and then can't remember if I have had my insulin or not. I think I will just have a piece of fruit and insulin before I leave for work to prevent the high. I think the basal is ok but I will check over next couple of days, thanks for your help. [Response from Patient]
Diabetes questions	*Hi <patient name>- I've been thinking about your CHO ratio. Have you ever done a basal check - which is simply to just have basal insulin overnight & on waking up and having no CHO & omit fast acting insulin for breakfast or lunch & just check how your levels go? If basal is right you should stay pretty constant. It's a good place to start - if basal is right we can start on CHO ratio with novo?* [Sent from Certified Diabetes Educator]
	Hi Veronica, I have never heard of a basal check but I did this, as you said and you can see on my download I was 12 -2 hours after tea (before bed) and went down to 8 when I woke up. So I now understand, as you say I need to adjust my CHO ratio with the evening meals so my after tea BSL is lower. Will adjust and let you have a look at levels next week. [Response from Patient]
Educational tips	*Hi <patient name>- Great logging this week - just a few Xmas tips - don't forget to give extra fasting acting insulin for extra treats - remember extra alcohol will make you low so eat carbs if you drink! Today I will be standing by to answer Xmas queries so please feel free to send them through. Merry Xmas!* [Sent from Certified Diabetes Educator]
	I didn't realize that alcohol made my BSL lower. That explains a lot - I have had some bad hypos in the morning after a big night out. I usually give myself extra insulin for each beer I drink. Should I give any fast acting insulin for alcohol at all or just eat extra CHO serves? [Response from Patient]
Positive reinforcement	*Hi <patient name>- Great logging this week –WOW I'm thrilled with your HbA$_{1c}$& your positive comment about it being the best in 6 years. It is a testament to you because you are keeping such good track on your own adjustments and control. Looks like another great week again!* [Sent from Certified Diabetes Educator]
	Thanks, after I was first diagnosed, I kept a log but then I didn't bother to keep track of anything because I just always felt out of control. By getting direct feedback, I have understood more about the finer details about insulin/carbs/exercise and alcohol which has made me less reactive and think more about my adjustments. [Response from Patient]

Discussion

Principal Findings

In adjunct with usual care, use of the Glucose Buddy application combined with weekly text-message feedback from a CDE led to a significant decrease in HbA_{1c} in comparison to a control group receiving only usual care. While regression to the mean cannot be ruled out, these results suggest that the intervention was effective. Improvements in HbA_{1c} of this magnitude in type 1 diabetes patients have been found previously in a mobile phone study [2] but are rare [17,19,30,31]. All patients in our study had poorly controlled diabetes at baseline; however, the intervention group had a significantly higher HbA_{1c} at baseline (mean 9.08%, SD 1.18 vs mean 8.47%, SD 0.86) and thus had a greater potential to improve their glycemic control. However, a meta-analysis of mobile intervention studies on diabetes glycemic control demonstrated only a 0.3% improvement for type 1 patients [9]; this demonstrates the success of the current intervention (a decrease of 1.1% in the intervention group) despite the baseline differences observed between groups.

It is unclear what mediated the change in HbA_{1c} in the current study as our analysis did not show a significant association between Glucose Buddy application usage, text message communications, and change in HbA_{1c}. Those patients who used the Glucose Buddy application and text messaged the CDE most frequently did not have a greater change in HbA_{1c} than those who used them less. This is supported in previous research which has found no association between engagement (increased contact between patient and clinician) and clinical improvement [32-36]. Perhaps the improvement in glycemic control in our study may be attributed to offering both a smartphone application and a website to log parameters. Mulvaney and colleagues [10] identified in their systematic review that diabetes studies which included a mobile phone and Internet component showed a greater reduction in HbA_{1c} (0.7% vs 0.4%) when compared to studies with only a mobile phone component. Unfortunately, we do not know (due to software constraints) the extent to which each component was used by patients during our study. Alternatively, it may be that the user engagement/health status relationship in IT-based interventions is more complex than measured in the current study. The estimated sample size in the current study was based on logistical constraints and the ability to detect a change in the primary outcome; as such, the study may have been underpowered for the mediation analysis between platform usage and change in the primary outcome.

The current study did not find an improvement in self-efficacy in the intervention group. Previous Web and mobile phone diabetes studies have found positive changes in self-efficacy

[33,37-39]; however, this change in self-efficacy is not always correlated with a change in HbA_{1c} [38,39]. Perhaps the relation between self-efficacy and HbA_{1c} is less important than previously assumed [24]. There was no change in quality of life in our study within and between groups over time. This finding is supported by previous telemedicine studies in type 1 diabetes that were also unable to observe an improvement in quality of life despite improvements in HbA_{1c} [2,32,40-42]. This may be due to the fact that diabetes self-management remains a burden despite short-term improvements in glycemic control and effects on quality of life might only manifest themselves in the longer term [43]. Additionally, there was no change in self-care activities for either group over time. This was unexpected considering that our intervention group had a significant improvement in HbA_{1c}, which is traditionally correlated with an increase in frequency in blood-glucose testing [8,38,44,45].

A recent systematic review highlighted that the level of engagement of participants in mobile intervention diabetes studies is underreported [10]. Our intervention group (n=36) logged 24,720 parameters in the Glucose Buddy application over a 6-month period of time. This is comparative to Farmer and colleagues [35], whose intervention group (n=47) logged 29,795 blood glucose results over a 9-month period of time. Patients in our study limited their self-monitoring practices to those indicators with high importance to the self-management of their condition. For example, blood glucose was logged 54.00% of the time, whereas exercise was logged only 1.00% of time. Our intervention patients sent 559 text messages in total, equating to a mean of 15.5 (559/36) per participant over the entire intervention period. Similarly in a mobile phone-based type 1 diabetes study, intervention patients sent a total of 1180 text messages, a mean of 18.4 text messages per person [34]. The CDE in our study spent approximately 5 minutes per week per patient to monitor and provide feedback. This is comparative to time taken by clinical practitioners in Benhamou and Melki's [19] research who spent on average 4.5 minutes per week per patient.

Despite there being an overall high level of engagement by patients in this intervention, this did decrease over the study period. Keeping patients adherent to treatment and maintaining engagement over time are significant problems that have been documented in health care [46]. This is also true of behavioral interventions across varied intervention delivery modalities, especially the Internet [47-49]. Issues in patient engagement and frequency of contacts have also been noted in mobile research [11,50]. This may be attributed to actual decreases in self-management activities or may be due to individuals becoming more self-aware of their daily behaviors/ health status due to self-monitoring and education from CDE and as such becoming less reliant on the need to self-monitor via intervention platforms. This potential effect requires further examination.

Intervention patients in our study were not provided with any education or training on how to use the Glucose Buddy smartphone application or website. Previous diabetes telemedicine studies have documented education and training sessions provided to intervention patients on the use of the technology under investigation. These training sessions ranged from 15 minutes [51], 1 hour [39,44,52], 6 hours [53], and even 1 day [38,54]. Given the high usage of the application in combination with providing no training on how to use it, it is likely that the design and usability of the application were not barriers to usage. Research has shown that to be competitive and encourage uptake and long-term adoption, smartphone applications should be intuitive and user-friendly [55-60]. This is especially relevant when investigating the effectiveness of smartphone applications to improve self-care in type 1 diabetes patients; it is a disease with no endpoint, requiring long-term self-management that can only be facilitated by user-friendly applications.

There is much enthusiasm from researchers concerning real-time feedback using mobile technology to assist patients with diabetes self-management [2,61-63]. It has been espoused that the future role of smartphones in diabetes care relates to providing patients with sophisticated applications that automatically upload blood glucose levels from glucometers and provides systematic advice concerning insulin dosage—perhaps sending this information wirelessly to an insulin pump. Indeed, smartphone applications hold great potential for taking diabetes self-management to a new level. We do not dispute that automation of the process may decrease the burden on patients compared to manual entry. However, we would argue that there is still value in manual entry of glucose levels, insulin, diet, and physical activity, despite being onerous. Problem solving is a core component of diabetes self-management [64] and, if all facets of measurement are automated, this may actually result in less awareness and thus less reflection by the patient on their condition; this might paradoxically lead to poorer self-management and negative clinical outcomes in the long term, with the added risk that a failure of the automated technology (eg, empty battery) might lead to panic and wrong decisions taken by patients. Additionally, the development of decision-support systems [65], although designed using medical information and clinical guidelines, is focused on reducing the practitioner element in the feedback process. As outlined in a recent review on mobile intervention design in diabetes [10], those studies that had the greatest impact on HbA_{1c} made use of clinician involvement and feedback. The importance of the human element should not be discounted.

Limitations

There are limitations to our study that should be noted. First, this study was a randomized controlled trial with a small sample conducted over a short duration. Due to the dropout of patients, the study may not have been powered sufficiently to detect differences between groups for the secondary outcome measures. Second, there were differences in glycemic control and gender between groups at baseline. Third, although patients in the control group were instructed not to use any mobile applications to self-manage their diabetes during the study period, it is possible they did.

Conclusions

Despite these limitations, we did find that integrating a smartphone application into secondary care was effective in improving glycemic control in patients with type 1 diabetes.

Our findings can be applied to adults with poorly controlled type 1 diabetes that own a smartphone, though larger studies over a longer duration need to be conducted to validate our findings.

Acknowledgments

This study was funded by Central Queensland University, Australia. The authors thank Certified Diabetes Educator Veronica Mills (Queensland Health) and SkyHealth, the developers of Glucose Buddy application and website.

C Vandelanotte is supported by a National Health and Medical Research Council of Australia (#519778) and National Heart Foundation of Australia (#PH 07B 3303) postdoctoral research fellowship. M Kirwan is supported by a Queensland Government, Department of Tourism, Regional Development and Industry, SmartFutures PhD Scholarship.

Authors' Contributions

MK contributed to the conception and design, acquisition of data and analysis, interpretation of data, and the drafting of the article. CV and MJD contributed to the design and the revision of the article. AF contributed to the revision of the article. All authors gave final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [66].

[PDF File (Adobe PDF File), 1003KB - jmir_v15i11e235_app1.pdf]

References

1. Devries JH, Snoek FJ, Heine RJ. Persistent poor glycaemic control in adult Type 1 diabetes. A closer look at the problem. Diabet Med 2004 Dec;21(12):1263-1268. [doi: 10.1111/j.1464-5491.2004.01386.x] [Medline: 15569126]
2. Charpentier G, Benhamou PY, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar;34(3):533-539 [FREE Full text] [doi: 10.2337/dc10-1259] [Medline: 21266648]
3. Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
4. Aschner P, Horton E, Leiter LA, Munro N, Skyler JS, Global Partnership for Effective Diabetes Management. Practical steps to improving the management of type 1 diabetes: recommendations from the Global Partnership for Effective Diabetes Management. Int J Clin Pract 2010 Feb;64(3):305-315 [FREE Full text] [doi: 10.1111/j.1742-1241.2009.02296.x] [Medline: 20456170]
5. Franc S, Daoudi A, Mounier S, Boucherie B, Dardari D, Laroye H, et al. Telemedicine and diabetes: achievements and prospects. Diabetes Metab 2011 Dec;37(6):463-476 [FREE Full text] [doi: 10.1016/j.diabet.2011.06.006] [Medline: 21889388]
6. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002 May 18;324(7347):1193-1194 [FREE Full text] [Medline: 12016186]
7. Piette JD. Interactive behavior change technology to support diabetes self-management: where do we stand? Diabetes Care 2007 Oct;30(10):2425-2432. [doi: 10.2337/dc07-1046] [Medline: 17586735]
8. Montori VM, Helgemoe PK, Guyatt GH, Dean DS, Leung TW, Smith SA, et al. Telecare for patients with type 1 diabetes and inadequate glycemic control: a randomized controlled trial and meta-analysis. Diabetes Care 2004 May;27(5):1088-1094. [Medline: 15111526]
9. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
10. Mulvaney SA, Ritterband LM, Bosslet L. Mobile intervention design in diabetes: review and recommendations. Curr Diab Rep 2011 Dec;11(6):486-493. [doi: 10.1007/s11892-011-0230-y] [Medline: 21960031]
11. Wei J, Hollin I, Kachnowski S. A review of the use of mobile phone text messaging in clinical and healthy behaviour interventions. J Telemed Telecare 2011;17(1):41-48. [doi: 10.1258/jtt.2010.100322] [Medline: 21097565]

12. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
13. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
14. Lee E, Tatara N, Arsand E, Hartvigsen G. Review of mobile terminal-based tools for diabetes diet management. Stud Health Technol Inform 2011;169:23-27. [Medline: 21893707]
15. Smith A. Pew Internet & American Life Project. Washington, DC 20036: Pew Research Center; 2012 Mar 01. Nearly half of American adults are smartphone owners URL: http://pewinternet.org/Reports/2012/Smartphone-Update-2012.aspx [accessed 2013-10-14] [WebCite Cache ID 6KM9Vx7de]
16. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
17. Vahatalo M. Cellular phone transferred self blood glucose monitoring: Prerequisites for positive outcome. Practical Diabetes International 2004;21(5):192-194.
18. Gómez EJ, Hernando Pérez ME, Vering T, Rigla Cros M, Bott O, García-Sáez G, et al. The INCA system: a further step towards a telemedical artificial pancreas. IEEE Trans Inf Technol Biomed 2008 Jul;12(4):470-479. [doi: 10.1109/TITB.2007.902162] [Medline: 18632327]
19. Benhamou PY, Melki V, Boizel R, Perreal F, Quesada JL, Bessieres-Lacombe S, et al. One-year efficacy and safety of Web-based follow-up using cellular phone in type 1 diabetic patients under insulin pump therapy: the PumpNet study. Diabetes Metab 2007 Jun;33(3):220-226 [FREE Full text] [doi: 10.1016/j.diabet.2007.01.002] [Medline: 17395516]
20. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. 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]
21. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
22. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. Control Clin Trials 1988 Dec;9(4):327-344. [Medline: 3203524]
23. Rao A, Hou P, Golnik T, Flaherty J, Vu S. Evolution of data management tools for managing self-monitoring of blood glucose results: a survey of iPhone applications. J Diabetes Sci Technol 2010 Jul;4(4):949-957 [FREE Full text] [Medline: 20663461]
24. Anderson RM, Fitzgerald JT, Gruppen LD, Funnell MM, Oh MS. The Diabetes Empowerment Scale-Short Form (DES-SF). Diabetes Care 2003 May;26(5):1641-1642. [Medline: 12716841]
25. Anderson RM, Funnell MM, Fitzgerald JT, Marrero DG. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. Diabetes Care 2000 Jun;23(6):739-743 [FREE Full text] [Medline: 10840988]
26. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950 [FREE Full text] [Medline: 10895844]
27. The Diabetes Control and Complications Trial Research Group. Reliability and validity of a diabetes quality-of-life measure for the diabetes control and complications trial (DCCT). The DCCT Research Group. Diabetes Care 1988 Oct;11(9):725-732. [Medline: 3066604]
28. Istepanian RS, Zitouni K, Harry D, Moutosammy N, Sungoor A, Tang B, et al. Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. J Telemed Telecare 2009;15(3):125-128. [doi: 10.1258/jtt.2009.003006] [Medline: 19364893]
29. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant 2010 May;25(5):1388-1393 [FREE Full text] [doi: 10.1093/ndt/gfp732] [Medline: 20067907]
30. Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res 2007;9(5):e36 [FREE Full text] [doi: 10.2196/jmir.9.5.e36] [Medline: 18166525]
31. Rossi MC, Nicolucci A, Pellegrini F, Bruttomesso D, Bartolo PD, Marelli G, et al. Interactive diary for diabetes: A useful and easy-to-use new telemedicine system to support the decision-making process in type 1 diabetes. Diabetes Technol Ther 2009 Jan;11(1):19-24. [doi: 10.1089/dia.2008.0020] [Medline: 19132851]
32. Marrero DG, Vandagriff JL, Kronz K, Fineberg NS, Golden MP, Gray D, et al. Using telecommunication technology to manage children with diabetes: the Computer-Linked Outpatient Clinic (CLOC) Study. Diabetes Educ 1995;21(4):313-319. [Medline: 7621734]
33. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]

34. Franklin VL, Greene A, Waller A, Greene SA, Pagliari C. Patients' engagement with "Sweet Talk" - a text messaging support system for young people with diabetes. J Med Internet Res 2008;10(2):e20 [FREE Full text] [doi: 10.2196/jmir.962] [Medline: 18653444]
35. Farmer AJ, Gibson OJ, Dudley C, Bryden K, Hayton PM, Tarassenko L, et al. A randomized controlled trial of the effect of real-time telemedicine support on glycemic control in young adults with type 1 diabetes (ISRCTN 46889446). Diabetes Care 2005 Nov;28(11):2697-2702. [Medline: 16249542]
36. Rami B, Popow C, Horn W, Waldhoer T, Schober E. Telemedical support to improve glycemic control in adolescents with type 1 diabetes mellitus. Eur J Pediatr 2006 Oct;165(10):701-705. [doi: 10.1007/s00431-006-0156-6] [Medline: 16670859]
37. Howells L, Wilson AC, Skinner TC, Newton R, Morris AD, Greene SA. A randomized control trial of the effect of negotiated telephone support on glycaemic control in young people with Type 1 diabetes. Diabet Med 2002 Aug;19(8):643-648. [Medline: 12147144]
38. 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]
39. McCarrier KP, Ralston JD, Hirsch IB, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009 Apr;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]
40. Gómez EJ, Hernando ME, García A, Del Pozo F, Cermeño J, Corcoy R, et al. Telemedicine as a tool for intensive management of diabetes: the DIABTel experience. Comput Methods Programs Biomed 2002 Aug;69(2):163-177. [Medline: 12100795]
41. Bujnowska-Fedak MM, Puchała E, Steciwko A. Telemedicine for diabetes support in family doctors' practices: a pilot project. J Telemed Telecare 2006;12 Suppl 1:8-10. [doi: 10.1258/135763306777978551] [Medline: 16884563]
42. Maljanian R, Grey N, Staff I, Conroy L. Intensive telephone follow-up to a hospital-based disease management model for patients with diabetes mellitus. Dis Manag 2005 Feb;8(1):15-25. [doi: 10.1089/dis.2005.8.15] [Medline: 15722700]
43. Aalto AM, Uutela A, Aro AR. Health related quality of life among insulin-dependent diabetics: disease-related and psychosocial correlates. Patient Educ Couns 1997 Mar;30(3):215-225. [Medline: 9104378]
44. Forjuoh SN, Reis MD, Couchman GR, Ory MG. Improving diabetes self-care with a PDA in ambulatory care. Telemed J E Health 2008 Apr;14(3):273-279. [doi: 10.1089/tmj.2007.0053] [Medline: 18570552]
45. 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]
46. World Health Organization. 2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases. In: Noncommunicable diseases and mental health. Switzerland: WHO Press; 2008:2008-2013.
47. Couper MP, Peytchev A, Strecher VJ, Rothert K, Anderson J. Following up nonrespondents to an online weight management intervention: randomized trial comparing mail versus telephone. J Med Internet Res 2007;9(2):e16 [FREE Full text] [doi: 10.2196/jmir.9.2.e16] [Medline: 17567564]
48. Glasgow R, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
49. Glasgow RE, Eakin EG, Toobert DJ. How generalizable are the results of diabetes self-management research? The impact of participation and attrition. Diabetes Educ 1996;22(6):573-4, 581. [Medline: 8970287]
50. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
51. Hanauer DA, Wentzell K, Laffel N, Laffel LM. Computerized Automated Reminder Diabetes System (CARDS): e-mail and SMS cell phone text messaging reminders to support diabetes management. Diabetes Technol Ther 2009 Feb;11(2):99-106. [doi: 10.1089/dia.2008.0022] [Medline: 19848576]
52. Forjuoh SN, Reis MD, Couchman GR, Ory MG, Mason S, Molonket-Lanning S. Incorporating PDA use in diabetes self-care: a central Texas Primary Care Research Network (CenTexNet) study. J Am Board Fam Med 2007;20(4):375-384 [FREE Full text] [doi: 10.3122/jabfm.2007.04.060166] [Medline: 17615418]
53. Jansà M, Vidal M, Viaplana J, Levy I, Conget I, Gomis R, et al. Telecare in a structured therapeutic education programme addressed to patients with type 1 diabetes and poor metabolic control. Diabetes Res Clin Pract 2006 Oct;74(1):26-32. [doi: 10.1016/j.diabres.2006.03.005] [Medline: 16621113]
54. Ładyzyński P, Wójcicki JM, Krzymień J, Foltyński P, Migalska-Musiał K, Tracz M, et al. Mobile telecare system for intensive insulin treatment and patient education. First applications for newly diagnosed type 1 diabetic patients. Int J Artif Organs 2006 Nov;29(11):1074-1081. [Medline: 17160965]
55. Zhang D, Adipat B. Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. International Journal of Human-Computer Interaction 2005 Jul;18(3):293-308. [doi: 10.1207/s15327590ijhc1803_3]
56. Park Y, Chen JV. Acceptance and adoption of the innovative use of smartphone. Industrial Management & Data Systems 2007;107(9):1349-1365. [doi: 10.1108/02635570710834009]

57. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
58. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
59. 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 Jul 02;1(2):e8. [doi: 10.2196/mhealth.2526]
60. Tatara N, Årsand E, Skrøvseth SO, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013;1(1):1. [doi: 10.2196/mhealth.2432]
61. Franc S, Daoudi A, Mounier S, Boucherie B, Laroye H, Peschard C, et al. Telemedicine: what more is needed for its integration in everyday life? Diabetes Metab 2011 Dec;37 Suppl 4:S71-S77 [FREE Full text] [doi: 10.1016/S1262-3636(11)70969-7] [Medline: 22208714]
62. Farmer A, Gibson O, Hayton P, Bryden K, Dudley C, Neil A, et al. A real-time, mobile phone-based telemedicine system to support young adults with type 1 diabetes. Inform Prim Care 2005;13(3):171-177. [Medline: 16259856]
63. Curran K, Nichols E, Xie E, Harper R. An intensive insulinotherapy mobile phone application built on artificial intelligence techniques. J Diabetes Sci Technol 2010 Jan;4(1):209-220 [FREE Full text] [Medline: 20167186]
64. Mulvaney S. Improving patient problem solving to reduce barriers to diabetes self-management. Clinical Diabetes 2009;27(3):99-104.
65. Klonoff DC, True MW. The missing element of telemedicine for diabetes: decision support software. J Diabetes Sci Technol 2009 Sep;3(5):996-1001 [FREE Full text] [Medline: 20144411]
66. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

CDE: Certified Diabetes Educator
DES-SF: Diabetes Empowerment Scale
DQOL: Diabetes Quality of Life
HbA1c: glycosylated hemoglobin
SDSCA: Summary of Diabetes Self Care Activities

Edited by G Eysenbach; submitted 27.02.13; peer-reviewed by T Chomutare, CJ(Wu; comments to author 24.04.13; revised version received 22.08.13; accepted 15.09.13; published 13.11.13.

Please cite as:
Kirwan M, Vandelanotte C, Fenning A, Duncan MJ
Diabetes Self-Management Smartphone Application for Adults With Type 1 Diabetes: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e235
URL: http://www.jmir.org/2013/11/e235/
doi:10.2196/jmir.2588
PMID:24225149

Original Paper

Daily Text Messaging for Weight Control Among Racial and Ethnic Minority Women: Randomized Controlled Pilot Study

Dori M Steinberg[1], RD, PhD; Erica L Levine[1], MPH; Sandy Askew[1], MPH; Perry Foley[1], MPH, MSW; Gary G Bennett[1,2], PhD

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

[2]Department of Psychology and Neuroscience, Duke University, Durham, NC, United States

Corresponding Author:
Dori M Steinberg, RD, PhD
Duke Obesity Prevention Program
Duke Global Health Institute
Duke University
134 Trent Hall, Box 90519
310 Trent Drive
Durham, NC, 27708
United States
Phone: 1 919 613 5453
Fax: 1 919 249 2146
Email: dori.steinberg@duke.edu

Abstract

Background: Daily self-monitoring of diet and physical activity behaviors is a strong predictor of weight loss success. Text messaging holds promise as a viable self-monitoring modality, particularly among racial/ethnic minority populations.

Objective: This pilot study evaluated the feasibility of a text messaging intervention for weight loss among predominantly black women.

Methods: Fifty obese women were randomized to either a 6-month intervention using a fully automated system that included daily text messages for self-monitoring tailored behavioral goals (eg, 10,000 steps per day, no sugary drinks) along with brief feedback and tips (n=26) or to an education control arm (n=24). Weight was objectively measured at baseline and at 6 months. Adherence was defined as the proportion of text messages received in response to self-monitoring prompts.

Results: The average daily text messaging adherence rate was 49% (SD 27.9) with 85% (22/26) texting self-monitored behavioral goals 2 or more days per week. Approximately 70% (16/23) strongly agreed that daily texting was easy and helpful and 76% (16/21) felt the frequency of texting was appropriate. At 6 months, the intervention arm lost a mean of 1.27 kg (SD 6.51), and the control arm gained a mean of 1.14 kg (SD 2.53; mean difference –2.41 kg, 95% CI –5.22 to 0.39; P=.09). There was a trend toward greater text messaging adherence being associated with greater percent weight loss (r=–.36; P=.08), but this did not reach statistical significance. There was no significant association between goal attainment and text messaging adherence and no significant predictors of adherence.

Conclusions: Given the increasing penetration of mobile devices, text messaging may be a useful self-monitoring tool for weight control, particularly among populations most in need of intervention.

Trial Registration: Clinicaltrials.gov: NCT00939081; http://clinicaltrials.gov/show/NCT00939081 (Archived by WebCite at http://www.webcitation.org/6KiIIcnk1).

(J Med Internet Res 2013;15(11):e244) doi:10.2196/jmir.2844

KEYWORDS

self-monitoring; mHealth; text messaging; weight loss; black women

Introduction

Regular self-monitoring of diet and physical activity behaviors is one of the strongest predictors of weight loss success [1,2]. Self-monitoring likely improves weight loss outcomes by activating a host of self-regulatory mechanisms. Regular self-monitoring might enhance self-efficacy, improve accountability, and facilitate awareness of how behaviors affect weight [3]. Weight loss interventions have traditionally used paper-based self-monitoring methods. However, long-term adherence to this approach is poor [1,4], possibly because paper-based methods are time- and labor-intensive, require extensive numeracy and literacy skills, and can be perceived as burdensome [5]. What is concerning is that poor adherence adversely affects weight loss outcomes [1]. In contrast, electronic health (eHealth) self-monitoring approaches include features (eg, prompts, real-time data collection, data-driven feedback, asynchronous communication) that may decrease participant burden and increase adherence [6]. Indeed, evidence does show that electronic self-monitoring via either Web or mobile devices produces greater adherence over traditional paper-and-pencil methods [4,7,8] and improves attitudes toward self-monitoring behaviors [9].

Text messaging shows promise as an alternative eHealth self-monitoring approach [10] and offers several advantages compared to other eHealth modalities (eg, Web, interactive voice response). Data can be entered quickly on nearly all mobile phone platforms, making it portable, proximal to actual behaviors, and more accessible for providing tailored feedback [11]. Additionally, text messaging has been conventionally limited to 160 characters (approximately 15-20 words) per message, limiting the detail and cognitive load that is required. Thus, text messaging may be a viable and sustainable self-monitoring modality.

Text messaging has become ubiquitous [12], particularly among racial/ethnic minority groups [13]. Recent studies show that racial/ethnic minorities are more likely than white individuals to own mobile phones [13]. The high familiarity with and penetration of mobile technologies makes text messaging an ideal intervention platform among these populations. This is notable because we have few interventions that produce clinically meaningful weight loss outcomes among racial/ethnic minority populations [14-16], those with the highest rates of obesity. Black women, in particular, have alarmingly high rates of obesity as compared with other gender and racial/ethnic groups: 59% of black women are obese versus 36% in the general US population [17]. However, little is known about the use of text messaging for self-monitoring weight control behaviors in this population. The purpose of this pilot study was to evaluate the feasibility of daily text messaging for self-monitoring behavioral goals for weight loss among predominantly obese black women. Our secondary aim was to evaluate the effects of the intervention on weight change relative to an education control arm.

Methods

Participants

We recruited women aged 25-50 years, with a body mass index (BMI) greater than or equal to 25 kg/m^2. Other inclusion criteria were willingness to (1) come to all study assessments over 6 months, (2) use a personal cell phone to send and receive up to 5 texts per day for 6 months without compensation for the text messages, and (3) be randomized into either treatment arm. Exclusion criteria included pregnancy or planned pregnancy within the next 6 months and a history of myocardial infarction or stroke within the past 2 years.

Recruitment and Randomization

We partnered with a nonprofit church-based community wellness organization located in Raleigh, NC to recruit participants. The wellness organization provided the location for enrollment events and aided in recruitment by advertising the study in common spaces and during community meetings and church services. Additional recruitment was conducted in the surrounding community via flyers posted in neighborhood businesses and outreach to adults in the area who had expressed interest in weight loss research trials. Recruitment took place between June and September 2010.

Interested participants visited a study website to complete initial eligibility screening that assessed self-reported age, gender, height, weight, and race (American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or other Pacific Islander, white, or other). Eligible participants were then invited to an in-person enrollment event. Ineligible participants were directed to a website where they could access publicly available weight loss information. At the enrollment events, study staff obtained informed consent and collected baseline anthropometric and survey measures. Participants were then randomized to the intervention arm or the education control arm using a computer-generated algorithm. Because of the pilot nature of this study, participants were re-evaluated at 6 months, with no additional study visits. Participants received a US $35 gift card to a local store as an incentive for participation. The Duke University Institutional Review Board approved this study.

Intervention Arm

Overview

The intervention (Shape Plan) included daily tracking of tailored behavior change goals through text messaging and personalized daily and weekly feedback via text messaging and email, respectively. Participants also received information sheets about behavioral goals, a pedometer, 2 face-to-face group sessions, and a skills training DVD.

Behavior Change Goals

Behavior change goals were determined using the interactive obesity treatment approach (iOTA) [18-20], a theory-based approach whereby participants are assigned an individualized set of routine lifestyle behavior change goals and directed to change them to create an energy deficit sufficient to produce weight change. Lifestyle behavior change goals are assigned based on an algorithm that considers participants' need and

self-efficacy around changing behaviors, as well as the expected caloric deficit. The iOTA library (Textbox 1) contains 12 obesogenic behaviors framed as goals to create a caloric deficit for weight loss (eg, no sugary drinks, walking 7000 steps per day, no fast food) that were selected based on their (1) empirical support, (2) population relevance, (3) ease of self-monitoring, and (4) concreteness. Participants were assigned new goals at 3 months to introduce novelty, maintain motivation, and facilitate goal mastery.

At baseline, the iOTA algorithm ranked behavior change goals for intervention participants. Participants were instructed to self-monitor the 2 top goals daily for 12 weeks. All intervention participants also received a walking goal of at least 7000 steps every day. The physical activity goal increased based on participants' performance, up to 10,000 steps per day. The survey was re-administered at 3 months and updated goals were assigned using the same algorithm.

Textbox 1. List of iOTA behavioral goals.

- Walk 7000/8000/10,000 steps every day
- No sugary drinks
- Eat 5 or more fruits and vegetables every day
- No chips, cookies, or candy
- Switch to low-fat dairy
- No fast food
- Eat breakfast every day
- Watch no more than 2 hours of TV every day
- No fried food
- No snacks or dessert after dinner
- No more than 1 alcoholic drink per day
- Eat red meat no more than once a week

Information Sessions

At the baseline enrollment event, Shape Plan participants received a group-based orientation to the intervention led by community health educators experienced in delivering information on weight control. The orientation included a review of the iOTA goals, calorie balance, a demonstration of the text messaging self-monitoring and feedback, and an action planning session. Goal setting and text messaging monitoring began the following day. At 3 months, in an effort to reduce the number of face-to-face meetings, participants received a set of videos with skills training information on topics such as healthy eating patterns, eating cues, recognizing hunger, problem solving to meet goals, goal setting, exercise tips, and safety and action planning for the upcoming Shape Plan goals. At 6 months, participants received another hour-long group face-to-face session that focused on problem solving, assessment of overall progress, and tips for maintaining behavior changes.

Self-Monitoring and Feedback Through Text Messaging

The text messaging protocol included 1 daily morning text message at 8:00 am, which asked participants to report performance on their goals from the previous day (eg, "How many steps did you walk yesterday?";Figure 1). Although immediate self-monitoring was encouraged, participants could respond any time until 7:59 am the next day. Any responses received within the 24-hour window after the outbound text message was sent were counted as a successful self-monitoring response. A score was assigned to each goal based on self-monitoring data received from participants. For example, if a participant was working on the no sugary drinks goal, a score of 10 was given if the participant entered 0 drinks that day, a score of 5 if the participant reported 2 drinks, and a score of 1 if ≥5 drinks was entered. Conversely, if the goal was to eat 5 or more fruits and vegetables, a score of 10 was given if the participant entered ≥5 fruits and vegetables for that day, and a score of 1 if a participant reported eating 0 fruits and vegetables that day. A summary score was then calculated for all 3 goals together, with higher scores indicative of high overall goal attainment (range 1-10). A feedback message was sent via an automated system based on the summary score, along with relative feedback based on the previous day of self-monitoring (eg, "You're doing better than yesterday—great job!" or "You did worse than last time. Let's turn this slip around") and specific tips on how to change low-scoring goals (eg, "Try flavored seltzer water instead of regular soda" or "Try sliced bell peppers as a snack"). Messaging content was based on previous studies conducted using the iOTA approach in this population [19-22]. Rigorous testing of the logic was completed before the start of the intervention, and continuous quality checks were performed to ensure fidelity to protocol.

Additionally, Shape Plan participants received a weekly automated email on Sundays with a summary of their progress. Participants with at least 3 days of self-monitoring data received a weekly email with personalized feedback that included a summary of goal attainment and a graph of progress over the previous week. For participants with low adherence (3 or fewer texts in 1 week), the email did not include a summary, but rather acted as a prompt to improve adherence (eg, "We only received 2 text messages from you this week. In order for you to be most

successful losing weight in Shape Plan, you should track your numbers and send us a text every day").

Education Control Arm

To control for contact and isolate the behavior change goals, self-monitoring via text messaging and feedback, participants randomized to the education control arm received (1) 2 in-person group education sessions, one at baseline and another at 6 months; (2) a set of videos at 3 months that covered topics such as healthy eating patterns, eating cues, recognizing hunger, exercise recommendations, and how to read a nutrition facts food label; (3) pedometers; and (4) a "prescription" to walk 10,000 steps per day. Control arm participants received no text messaging during the study period, but had the option to receive a 3-month version of the text messaging intervention after the study was complete.

Measures

Demographic Characteristics

At baseline, a variety of sociodemographic variables were collected through an online survey to characterize the sample, including age, race/ethnicity, household income, education, marital status, and employment.

Self-Monitoring Adherence

A study database collected and stored text messaging self-monitoring data. Adherence was defined as the proportion of self-monitoring texts received of the number expected over the 6-month period (N=167). We examined total adherence and adherence by study week.

Anthropometrics

At baseline and 6 months, trained staff measured participant heights to the nearest 0.1 cm using a calibrated stadiometer (Seca 214). Weights were measured to the nearest 0.1 kg using an electronic scale (Seca Model 876) [23].

Program Satisfaction

At 6 months, intervention participants completed a 23-item online questionnaire to assess intervention satisfaction. Using a 4-point Likert scale with response options ranging from strongly agree (=4) to strongly disagree (=1), participants rated whether they found daily self-monitoring through text messaging to be easy, helpful overall, helpful for increasing daily steps, and important. Similarly, participants reported whether daily text messages were the appropriate frequency and whether they were satisfied with the feedback received via text messaging.

Statistical Analyses

We used chi-square tests and *t* tests to examine differences in baseline characteristics between study arms. Similar tests were used to describe the average rate of text messaging adherence and the proportion of participants who achieved various thresholds of self-monitoring adherence. We used intent-to-treat analyses, with baseline weight carried forward for missing data. The *t* tests were used to examine absolute weight change, percent weight loss, and BMI change between study arms, and ANOVA was conducted to examine goal attainment and weight change across tertiles of self-monitoring adherence. Pearson correlation tests were conducted to examine the overall association between text messaging adherence and goal attainment and weight loss. Analyses were conducted using SPSS ver 19 for Mac (IBM Corp, Armonk, NY, USA). All tests were 2-tailed and an alpha level <.05 was used to assess statistical significance.

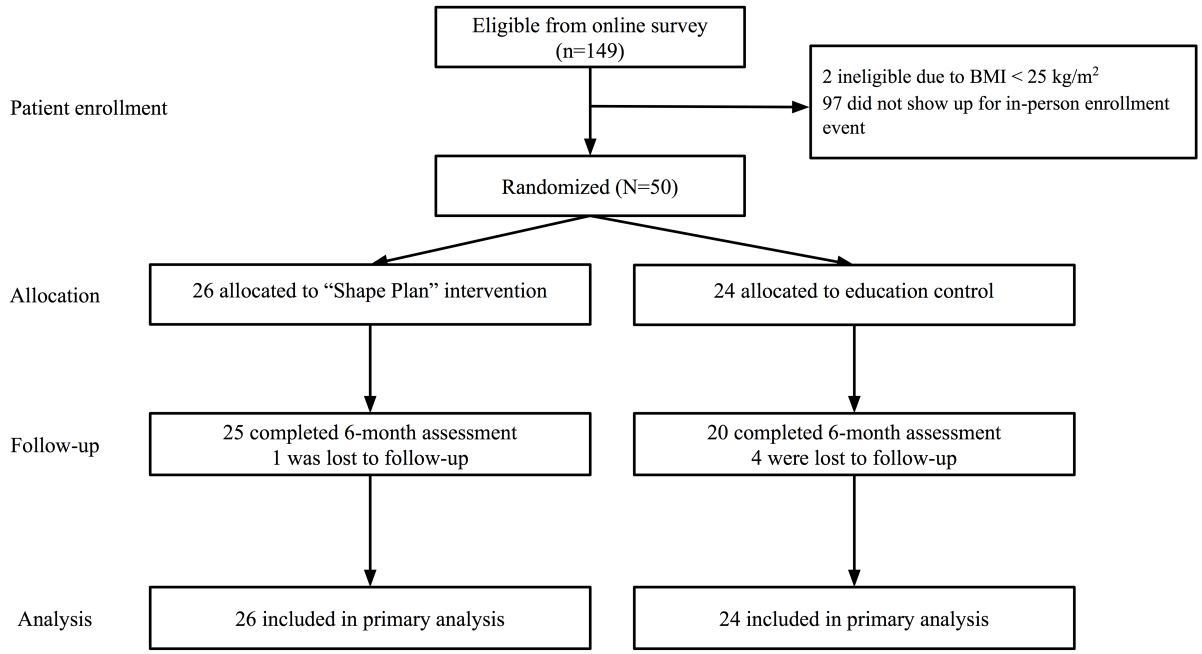
Figure 1. Screenshot of self-monitoring via text message with feedback.

Results

Enrollment and Retention

Figure 2 outlines the study enrollment and retention flow. A total of 149 individuals were deemed eligible from the online screener. The total number screened via the study website is not known. Of those initially eligible, 2 were ineligible because of BMI and 97 did not attend the enrollment event. In all, 50 participants were randomized to either the Shape Plan intervention (n=26) or the control (n=24) arm. At 6 months, 90% (45/50) of participants attended the 6-month follow-up assessment visit. There were no significant differences in attrition between study arms.

Figure 2. Flowchart of participant enrollment and retention in the study.



Baseline Characteristics

Mean age of participants was 38.3 years (SD 8.2). Participants were obese (BMI mean 35.8 kg/m^2, SD 6.1), and had an average weight of 99.1 kg (SD 20.0) (Table 1). Participants were predominantly black (82%, 41/50), employed (82%, 41/50), and college educated (64%, 32/50). Approximately one-third of the sample (32%, 16/50) had an annual income less than US $40,000 and 50% (25/50) were married or living with a partner. There were no significant differences between arms with regard to baseline sociodemographic characteristics.

Text Messaging Adherence

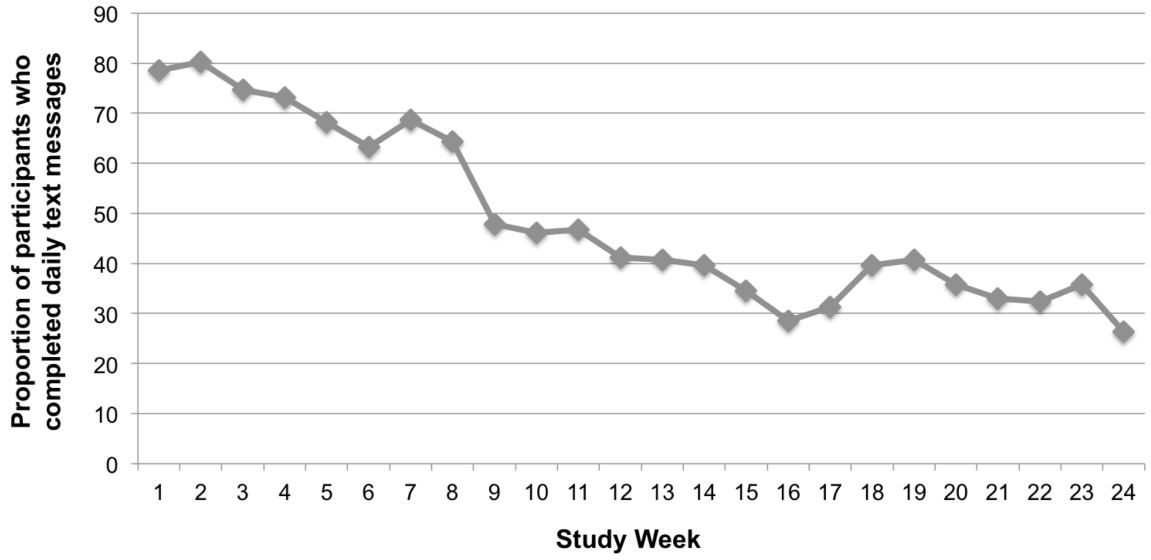
Figure 3 shows the text messaging self-monitoring completion rates over time by study week. Among all randomized intervention participants (n=26), the daily text messaging adherence rate was 49% (SD 28, IQR 27%-78%). Most participants (58%, 15/26) texted at least 3 days each week and 85% (22/26) texted at least 2 days per week, on average. There were no significant predictors of adherence. During the study, 8 participants demonstrated nonusage attrition [24], requesting to cease intervention participation before the end of the trial for a host of reasons (eg, cost of text messaging, change in phone service, not interested in continued participation). Excluding these 8 participants, the adherence rate was 54% (SD 25), and we similarly found no significant predictors of adherence.

The mean daily goal attainment score over the 6-month intervention period was 6.3 (SD 2.8, IQR 4.0-8.2), indicating moderate-high adherence to behavioral goals. Similarly, the mean number of steps reported was 4994 (SD 2741, IQR 3016-6489). There was no significant correlation between average goal attainment score and text messaging adherence (r=.24; P=.25) or mean number of steps reported and text messaging adherence (r=.33; P=.11). However, when examining tertiles of adherence, there was a trend toward greater adherence being associated with greater step counts, although this did not reach statistical significance (tertile 1: mean 3958, SD 2698; tertile 2: mean 4174, SD 1902; tertile 3: mean 6481, SD 2895; P=.09).

Table 1. Baseline characteristics of participants by study arm.

Characteristic	Total (N=50)	Control (n=24)	Shape Plan (n=26)
Marital status, n (%)			
Married or living with a partner	25 (50)	9 (38)	16 (62)
Divorced/separated or never married	25 (50)	15 (62)	10 (38)
Employment, n (%)			
Employed	41 (82)	19 (79)	22 (85)
Unemployed	9 (18)	5 (21)	4 (15)
Education, n (%)			
4 year college degree or higher	32 (64)	15 (62)	17 (65)
Less than a 4 year college degree	18 (36)	9 (38)	9 (35)
Race/ethnicity, n (%)			
Non-Hispanic black	41 (82)	18 (75)	23 (88)
Non-Hispanic other	9 (18)	6 (25)	3 (12)
Household income (US $), n (%)			
<40,000	16 (32)	9 (38)	7 (27)
40,000-69,999	18 (36)	7 (29)	11 (42)
≥70,000	16 (32)	8 (33)	8 (31)
Age (years), mean (SD)	38.3 (8.2)	39.0 (9.0)	37.6 (7.4)
Weight (kg), mean (SD)	99.1 (20.0)	96.0 (23.1)	102.0 (16.6)
BMI (kg/m^2), mean (SD)	35.8 (6.1)	34.6 (5.8)	36.9 (6.2)

Figure 3. Adherence to daily self-monitoring via text message by study week (n=26).



Intervention Satisfaction

At 6 months, most participants strongly agreed that texting was easy (70%, 16/23) and helpful (68%, 15/22), and 76% (16/21) either somewhat or strongly agreed that the text messages helped them increase the number of daily steps walked. Almost three-quarters of participants (71%, 15/21) reported that it took less than 3 minutes to reply to texts and most (79%, 15/19) responded to prompts for self-monitoring via text message in the morning. More than half (57%, 12/21) felt that receiving daily texts was very important and approximately three-quarters of participants (76%, 16/21) felt the frequency of texting was appropriate. Most (82%, 18/22) were satisfied with the feedback

content they received via text messaging and 62% (13/21) felt the feedback was very important.

Weight Change

Table 2 highlights changes in weight and BMI between study arms. At 6 months, using intent-to-treat analyses (N=50), participants randomized to the control arm gained a mean of 1.14 kg (SD 2.53), whereas intervention participants lost a mean of 1.27 kg (SD 6.51). The mean difference was –2.41 kg (95% CI –5.22 to 0.39; *P*=.09). This equates to a percent weight loss of 0.97% (SD 5.35) in the intervention arm, which was marginally higher than a gain of 1.32% (SD 2.77) in the control arm (mean difference –2.29, 95% CI –4.74 to 0.16; *P*=.06). There were no significant differences between arms with regard to changes in BMI. Similar results were found when we restricted the sample to study completers only (n=45). There were no significant correlations between adherence to text messaging and change in weight or BMI. However, there was a trend toward greater text messaging adherence being associated with greater percent weight loss (r = –.36; *P*=.08), but this did not reach statistical significance. Similarly, there were no differences in weight outcomes by tertiles of text messaging adherence (data not shown).

Table 2. Change in weight and body mass index (BMI) between baseline and 6 months by study group (N=50).

Weight change variables	Control, mean (SD) (n=24)	Shape Plan, mean (SD) (n=26)	Mean difference (95% CI)	*P* value
Change in weight (kg)	1.14 (2.53)	–1.27 (6.51)	–2.41 (–5.22, 0.39)	.09
Percent weight loss (%)	1.32 (2.77)	–0.97 (5.35)	–2.29 (–4.70, 0.12)	.06
Change in BMI (kg/m^2)	0.42 (0.90)	–0.47 (2.42)	–0.89 (–1.93, 0.15)	.09

Discussion

Principal Findings

In this pilot study, we found that daily text messaging for behavioral self-monitoring is both feasible and positively perceived. Approximately half of participants were fully adherent to daily self-monitoring through text messages during the 6-month study and 84% stayed active in the intervention throughout the study period. In contrast to previous studies that included paper-based self-monitoring modes [5], most participants felt that text messaging was helpful, easy, and important for achieving their behavior change goals. Text messaging has major advantages relative to other approaches. It has high familiarity and a greater potential for broad reach, particularly among racial/ethnic minority populations. Despite the intervention having only a marginal effect on weight change, these results indicate that text messaging may be a viable way to collect self-monitoring data and deliver intervention feedback and skills training.

The eHealth and mobile health (mHealth) approaches to self-monitoring offer numerous advantages over traditional approaches. Self-monitoring via text messaging is cheaper, easier to program, and more proximal to behavior changes as compared to paper-based and Web-based self-monitoring. Both eHealth and mHealth self-monitoring strategies seem not to exhibit the same steep decline typically seen in self-monitoring adherence [1,6,11] particularly when feedback is provided in response to self-monitoring [4,25]. The current pilot study provided immediate feedback in response to text messaging self-monitoring and found that most participants reported high satisfaction with this feedback. Providing feedback via text messaging may offer particular benefits over other modalities. However, feedback can only be sent when participants provide self-monitoring data; thus, finding ways to enhance adherence remains important.

What participants are self-monitoring may be as important to adherence as how they are self-monitoring. Typical self-monitoring behaviors include detailed reports of food intake, including portion size, and calorie and fat content. Although effective, self-monitoring of this type exhibits poor adherence [4,6]. Our pilot study, in contrast, included simple and discrete self-monitoring of behavior change goals associated with weight loss (eg, no sugary drinks, no fried food). Self-monitoring of specific behaviors rather than detailed dietary records may be less burdensome and more fitting for delivery via text messaging. Indeed, we found high satisfaction with our approach, adherence rates comparable to Web-based self-monitoring approaches [7,26], and a trend toward a positive association between adherence and weight loss. Whether text messaging for self-monitoring is more effective for weight change compared to other eHealth or paper-based approaches is not clear. Trials examining the comparative efficacy of these different modes are necessary.

To date, our weight loss outcomes are similar to those of other text messaging trials [27-29]. Most of these trials used text messaging to deliver skills training information and feedback via outbound messages. We are aware of only 1 trial that used text messaging to monitor behaviors and provide feedback. In their Text4Diet trial, Shapiro et al [29] tested an intervention that asked participants to self-monitor daily pedometer steps and weekly body weight. The intervention sent up to 4 text messages per day for 12 months that included personalized tips for eating behaviors, reminders, educational facts, motivational messages, knowledge-based questions, and feedback in response to self-monitoring data. After 12 months, the intervention arm lost 1.65 kg, but there were no significant differences between treatment arms.

Our pilot study used text messaging to collect self-monitoring data on diet and physical activity behaviors, but we did not gather data on body weight or include any type of coaching support. This may have affected the weight losses achieved because body weight self-monitoring has been shown to be

effective in the absence of self-monitoring of other behaviors [30]. Haapala and colleagues [28] emphasized calorie counting and used body weight to adjust calorie goals, which may have resulted in larger weight loss outcomes. By contrast, the Shape Plan study focused on engagement in behaviors known to produce a caloric deficit (eg, switching from high-fat dairy to low-fat dairy) rather than calorie counting to achieve weight loss. Long-term, we hypothesize that our approach may have greater potential for sustainability because calorie counting is cognitively complex and may not be suitable for populations with lower levels of education. However, this may come with some cost to the magnitude of weight loss.

Black women have the highest prevalence of obesity compared to any other group [17] and achieving clinically meaningful weight loss among this group has been particularly challenging [16]. Across numerous studies, black women achieved smaller weight losses compared to other groups and most did not lose more than 4 kg [15]. These results indicate a need for new obesity treatment approaches for black women that are relevant, effective, and sustainable. In the current study, we found that there was a trend toward a small weight loss among intervention participants and a small weight gain among control participants. Although our study focused on weight loss, this low-intensity approach may be helpful for staving off premenopausal weight gain that is common among black women and higher than other racial/ethnic groups [31,32]. Given the high satisfaction with text messaging for self-monitoring and comparable adherence ranges, future research might include this modality along with brief coaching calls in an effort to enhance weight loss outcomes.

Additionally, mobile phone use is high among blacks, and black individuals are more likely than white individuals are to use mobile phones to look for health or medical information [33]. Previous studies testing text messaging as a self-monitoring tool for weight loss included predominantly white samples with greater sociodemographic advantage. One exception is a recent trial that was short in duration, but found a positive association between text messaging for self-monitoring and walking among an older, black, and predominantly female sample [34]. These findings are consistent with our study and further confirm that this modality is feasible for behavior change, with the potential for broad reach.

Strengths and Limitations

This study has several strengths. Most of our sample (82%) were black women, which is a group typically underrepresented in weight control research. Although a strength of our study, the findings may not generalize to other populations and settings. The goal of this study was to test the feasibility of text messaging for self-monitoring through a low-intensity weight loss intervention among an understudied population. We isolated the impact of text messaging self-monitoring along with feedback with a control arm that received comparable group information sessions and the use of a pedometer. This study also has some limitations worth mentioning. A few participants (n=8) experienced barriers to participation, such as cost of text messaging and disconnected cell phone service, or they were no longer interested in participating. Although providing phones and/or text messaging plans may have enhanced our adherence rates, the use of personal cell phones provides insight into the "real world" feasibility of text messaging for weight control and improves the generalizability of the intervention.

Although comparable to other eHealth weight control interventions, higher adherence rates are needed to produce greater weight losses. This pilot study was low intensity and did not include any contact with study staff outside the assessment visits. Our main goal was to assess the feasibility of using text messaging for self-monitoring behavioral goals among a predominantly racial/ethnic minority population. To enhance adherence, future studies using this approach might also include daily tracking of weight and provide feedback on weight loss progress via text message. More frequent skills training through monthly videos may also boost adherence rates. Including elements of accountability and support (eg, monthly coaching calls with a lifestyle counselor) can also be an effective way to enhance adherence, but including these components will increase intensity and make it more difficult to ascertain the independent effects of the text messaging. Given that this was a pilot study, the small sample size limited our power to assess whether this intervention led to significantly greater weight loss as compared to an education control arm. Future studies should examine the efficacy of this approach with a larger sample size, longer duration, and multiple measures throughout the study period.

Conclusions

Text messaging holds promise as a self-monitoring modality for weight control, particularly among groups most at risk for obesity-related morbidities. Given that the majority of evidence indicates that greater adherence leads to better outcomes, our study suggests that mHealth-based approaches to self-monitoring may enhance engagement and reduce the burdens commonly associated with other modes. Our intervention was relatively low intensity and has greater potential for dissemination compared to higher intensity interventions. As technology penetration increases in the target population, the use of this modality will become increasingly relevant and helpful as an intervention tool for weight control.

Acknowledgments

The authors would like to thank the support received by Rachel Kroll Bordogna and staff at the Duke Obesity Prevention Program, particularly Michele Lanpher, Daniel Dix, and Jade Miller. We are also grateful to the community organization, Diversified Resources for Better Living, that was pivotal in helping recruit our participants. Lastly, we would like to thank the women participating in the Shape Plan. This trial was funded by grant K22CA126992 awarded to Dr Gary Bennett.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [35].

[PDF File (Adobe PDF File), 997KB - jmir_v15i11e244_app1.pdf]

References

1. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
2. Sarwer DB, von Sydow Green A, Vetter ML, Wadden TA. Behavior therapy for obesity: where are we now? Curr Opin Endocrinol Diabetes Obes 2009 Oct;16(5):347-352. [doi: 10.1097/MED.0b013e32832f5a79] [Medline: 19623061]
3. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behavior Therapy 1993 Jun;24(3):377-394. [doi: 10.1016/S0005-7894(05)80212-6]
4. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26. [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
5. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
6. Coons MJ, Roehrig M, Spring B. The potential of virtual reality technologies to improve adherence to weight loss behaviors. J Diabetes Sci Technol 2011 Mar;5(2):340-344 [FREE Full text] [Medline: 21527103]
7. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
8. Beasley JM, Riley WT, Davis A, Singh J. Evaluation of a PDA-based dietary assessment and intervention program: a randomized controlled trial. J Am Coll Nutr 2008 Apr;27(2):280-286. [Medline: 18689560]
9. Glanz K, Murphy S, Moylan J, Evensen D, Curb JD. Improving dietary self-monitoring and adherence with hand-held computers: a pilot study. Am J Health Promot 2006;20(3):165-170. [Medline: 16422134]
10. 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]
11. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. Health Informatics J 2012 Dec;18(4):235-250. [doi: 10.1177/1460458212442422] [Medline: 23257055]
12. Smith A. Mobile Access 2010. Washington, DC: Pew Internet & American Life Project; 2010 Jul 10. URL: http://www.pewinternet.org/Reports/2010/Mobile-Access-2010.aspx [accessed 2013-10-22] [WebCite Cache ID 6KZ9iKjD3]
13. Smith A. Technology trends among people of color. Washington, DC: Pew Internet & American Life Project; 2010 Sep 17. URL: http://www.pewinternet.org/Commentary/2010/September/Technology-Trends-Among-People-of-Color.aspx [accessed 2013-10-22] [WebCite Cache ID 6KZAiq09N]
14. Yancey AK, Kumanyika SK, Ponce NA, McCarthy WJ, Fielding JE, Leslie JP, et al. Population-based interventions engaging communities of color in healthy eating and active living: a review. Prev Chronic Dis 2004 Jan;1(1):A09 [FREE Full text] [Medline: 15634371]
15. Osei-Assibey G, Kyrou I, Adi Y, Kumar S, Matyka K. Dietary and lifestyle interventions for weight management in adults from minority ethnic/non-White groups: a systematic review. Obes Rev 2010 Nov;11(11):769-776. [doi: 10.1111/j.1467-789X.2009.00695.x] [Medline: 20059708]
16. Kumanyika SK, Gary TL, Lancaster KJ, Samuel-Hodge CD, Banks-Wallace J, Beech BM, et al. Achieving healthy weight in African-American communities: research perspectives and priorities. Obes Res 2005 Dec;13(12):2037-2047. [doi: 10.1038/oby.2005.251] [Medline: 16421334]
17. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA 2012 Feb 1;307(5):491-497. [doi: 10.1001/jama.2012.39] [Medline: 22253363]
18. Bennett GG, Herring SJ, Puleo E, Stein EK, Emmons KM, Gillman MW. Web-based weight loss in primary care: a randomized controlled trial. Obesity (Silver Spring) 2010 Feb;18(2):308-313. [doi: 10.1038/oby.2009.242] [Medline: 19696764]
19. Greaney ML, Quintiliani LM, Warner ET, King DK, Emmons KM, Colditz GA, et al. Weight Management Among Patients at Community Health Centers: The “Be Fit, Be Well” Study. Obesity and Weight Management 2009 Oct;5(5):222-228. [doi: 10.1089/obe.2009.0507]

20. Foley P, Levine E, Askew S, Puleo E, Whiteley J, Batch B, et al. Weight gain prevention among black women in the rural community health center setting: the Shape Program. BMC Public Health 2012;12:305 [FREE Full text] [doi: 10.1186/1471-2458-12-305] [Medline: 22537222]
21. Bennett GG, Warner ET, Glasgow RE, Askew S, Goldman J, Ritzwoller DP, Be Fit, Be Well Study Investigators. Obesity treatment for socioeconomically disadvantaged patients in primary care practice. Arch Intern Med 2012 Apr 9;172(7):565-574. [doi: 10.1001/archinternmed.2012.1] [Medline: 22412073]
22. Bennett GG, Foley P, Levine E, Whiteley J, Askew S, Steinberg DM, et al. Behavioral treatment for weight gain prevention among black women in primary care practice: A randomized controlled trial. JAMA Intern Med 2013 Aug 26:[Epub ahead of print]. [doi: 10.1001/jamainternmed.2013.9263]
23. CDC/National Center for Health Statistics. Centers for Disease Control and Prevention. 2007. 2007-2008 National Health and Nutrition Examination Survey (NHANES) URL: http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/current_nhanes_07_08.htm [accessed 2013-10-22] [WebCite Cache ID 6KZELg7J5]
24. 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]
25. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-Monitoring as a Mediator of Weight Loss in the SMART Randomized Clinical Trial. Int J Behav Med 2012 Sep 1:1-6. [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
26. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: A pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]
27. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
28. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
29. Shapiro JR, Koro T, Doran N, Thompson S, Sallis JF, Calfas K, et al. Text4Diet: a randomized controlled study using text messaging for weight loss behaviors. Prev Med 2012 Nov;55(5):412-417. [doi: 10.1016/j.ypmed.2012.08.011] [Medline: 22944150]
30. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and email. Obesity 2013 March 20 Epub ahead of print . [doi: 10.1002/oby.20396]
31. Ailshire JA, House JS. The Unequal Burden of Weight Gain: An Intersectional Approach to Understanding Social Disparities in BMI Trajectories from 1986 to 2001/2002. Soc Forces 2011 Dec;90(2):397-423 [FREE Full text] [doi: 10.1093/sf/sor001] [Medline: 23413316]
32. Lewis CE, Jacobs DR, McCreath H, Kiefe CI, Schreiner PJ, Smith DE, et al. Weight gain continues in the 1990s: 10-year trends in weight and overweight from the CARDIA study. Coronary Artery Risk Development in Young Adults. Am J Epidemiol 2000 Jun 15;151(12):1172-1181 [FREE Full text] [Medline: 10905529]
33. Fox S, Duggan M. Mobile Health 2012. Washington, DC: Pew Internet & American Life Project; 2012 Nov 08. URL: http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2013-10-22] [WebCite Cache ID 6KZAOgarg]
34. Kim BH, Glanz K. Text messaging to motivate walking in older African Americans: a randomized controlled trial. Am J Prev Med 2013 Jan;44(1):71-75. [doi: 10.1016/j.amepre.2012.09.050] [Medline: 23253653]
35. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BMI: body mass index
eHealth: electronic health
iOTA: interactive obesity treatment approach
mHealth: mobile health

Edited by G Eysenbach; submitted 23.07.13; peer-reviewed by K Wolin, B Davy; comments to author 19.09.13; revised version received 09.10.13; accepted 13.10.13; published 18.11.13.

Please cite as:
Steinberg DM, Levine EL, Askew S, Foley P, Bennett GG
Daily Text Messaging for Weight Control Among Racial and Ethnic Minority Women: Randomized Controlled Pilot Study
J Med Internet Res 2013;15(11):e244
URL: http://www.jmir.org/2013/11/e244/
doi:10.2196/jmir.2844
PMID:24246427

Original Paper

Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online

Felix Greaves[1,2], MBChB; Daniel Ramirez-Cano[2], PhD; Christopher Millett[1], PhD; Ara Darzi[2], MD; Liam Donaldson[2], MD

[1]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[2]Centre for Health Policy, Imperial College London, London, United Kingdom

Corresponding Author:
Felix Greaves, MBChB
Department of Primary Care and Public Health
Imperial College London
Charing Cross Hospital
London, W6 8RF
United Kingdom
Phone: 44 7866551172
Fax: 44 2075940584
Email: fg08@imperial.ac.uk

Abstract

Background: There are large amounts of unstructured, free-text information about quality of health care available on the Internet in blogs, social networks, and on physician rating websites that are not captured in a systematic way. New analytical techniques, such as sentiment analysis, may allow us to understand and use this information more effectively to improve the quality of health care.

Objective: We attempted to use machine learning to understand patients' unstructured comments about their care. We used sentiment analysis techniques to categorize online free-text comments by patients as either positive or negative descriptions of their health care. We tried to automatically predict whether a patient would recommend a hospital, whether the hospital was clean, and whether they were treated with dignity from their free-text description, compared to the patient's own quantitative rating of their care.

Methods: We applied machine learning techniques to all 6412 online comments about hospitals on the English National Health Service website in 2010 using Weka data-mining software. We also compared the results obtained from sentiment analysis with the paper-based national inpatient survey results at the hospital level using Spearman rank correlation for all 161 acute adult hospital trusts in England.

Results: There was 81%, 84%, and 89% agreement between quantitative ratings of care and those derived from free-text comments using sentiment analysis for cleanliness, being treated with dignity, and overall recommendation of hospital respectively (kappa scores: .40–.74, $P<.001$ for all). We observed mild to moderate associations between our machine learning predictions and responses to the large patient survey for the three categories examined (Spearman rho 0.37-0.51, $P<.001$ for all).

Conclusions: The prediction accuracy that we have achieved using this machine learning process suggests that we are able to predict, from free-text, a reasonably accurate assessment of patients' opinion about different performance aspects of a hospital and that these machine learning predictions are associated with results of more conventional surveys.

(J Med Internet Res 2013;15(11):e239) doi:10.2196/jmir.2721

KEYWORDS

Internet; patient experience; quality; machine learning

Introduction

Understanding patients' experience of health care is central to the process of providing care and is a fundamental pillar of health care quality [1,2]. Traditional measures of patient experience include surveys, and more recently, structured patient reported outcome measures. Such approaches ask specific and limited questions, are conducted infrequently, and are often

expensive to administer. Today's patients have begun to report their health care experience on the Internet in blogs, social networks, wikis, and on health care rating websites [3,4]. However, as this is largely unstructured, nonstandardized free-text information, it is not captured in a systematic way. This represents a missed opportunity for understanding patients' experience in an increasingly "connected" world. Survey data in the United States suggest that 85% of adults use the Internet [5], 25% have read someone else's experience about health on a website or blog, and 11% have consulted online reviews of hospitals or other medical facilities [6].

Outside health care, natural language processing of large datasets, including sentiment analysis and opinion mining, has been critical to understanding consumer attributes and behaviors, for example, in election forecasting [7,8]. Sentiment analysis enables the content of natural language—the words we write and speak—to be examined for positive and negative opinions and emotion [9]. If applicable to health care, these analytical methods could permit interpretation of textual information about patient experience on a huge scale. This information, because of its prose nature, has avoided the analytical spotlight of conventional quantitative analysis. Alemi and colleagues have proposed the use of sentiment analysis of comments as real-time patient surveys [10]. They have shown that patient comments about specific doctors could be attributed with reasonable accuracy to positive and negative sentiment. They further suggest that capture of sentiment analysis should be compared to traditional methods of assessing patient experience.

The Information Strategy for the National Health Service (NHS) in England states that sentiment analysis of data could be a novel source of information [11] that would be valuable for patients in facilitating choice of hospitals. We tested this assertion and furthered the work undertaken by Alemi and colleagues, by analyzing a large number of free-text comments on the main NHS website (NHS Choices). This website allows patients to describe their treatment experiences at all hospitals in England (and all other NHS-provided services). It is well used, with around a million hits a day. Over a 2-year period, each hospital had a mean average of 69 reviews [12]. These reviews include both free-text descriptions of general experience and Likert-scale ratings of particular aspects of care. This presents the opportunity for a natural experiment to assess the accuracy of sentiment analysis techniques (applied to the free-text comments) against the patients' own quantitative ratings. The NHS also measures patient experience via a national survey of hospital inpatients. If sentiment analysis techniques are to be considered as useful tools for assessing care quality, it is important to see whether there is an association with traditional measures of patient experience. We therefore compare our sentiment analysis findings to the national patient survey, at the hospital level.

Methods

Machine Learning From Patient Comments

We applied data processing techniques to all the online free-text comments about hospitals on the NHS Choices website in 2010. Our purpose was to test whether we could automatically predict patients' views on a number of topics from their free-text responses. A machine learning classification approach was chosen in which an algorithm "learns" to classify comments into categories from a given set of examples, using open-source Weka data mining software. This software has been extensively used in previous research and provides accurate classification results, including in health care [13-15]. To test the accuracy of the prediction, we compared our results to quantitative ratings provided by the same individual patients on a Likert scale. Free-text comments were examined in response to the questions: "What I liked", "What could have been improved", and "Any other comments". A prediction was then made about whether the patient would recommend the hospital or not, whether the hospital was clean or dirty, and whether they were treated with dignity and respect. The algorithm was trained using all comments and ratings about hospitals left on the NHS Choices website from 2008, 2009, and 2011 (13,802 in total) as a learning set. Data from 2010 were used to test the predicting accuracy of the process (6412 comments) because comparable patient experience survey data were available for that year. The validation set represents 31.7% of the total sample available. We performed a single round of cross validation. All comments left on the NHS Choices website were provided directly by the Department of Health in England but have subsequently been made publicly available [16].

Technical Aspects of the Machine Learning Approach

The machine learning approach had two components: (1) pre-processing, in which data from patient comments are split into manageable units to build a representation of the data [17], and (2) classification, in which an algorithm decides which category each comment falls into. A consistent set of methodologies was applied in our machine learning process, including a "bag-of-words" approach, "prior polarity", and "information gain".

In the "bag-of-words" approach, the total body of words analyzed (known as the corpora) is represented as a simplified, unordered collection of words [18]. For this analysis, unigrams (single elements or words) and bigrams (two adjacent elements in a string of tokens, in this case, a 2-word phrase) were used as the basic units of analysis. We extracted 5695 n-grams in total. Higher n-grams (longer phrases) could have been used, but the constraints were computer power and processing time. We also included our own classification of certain words in the machine learning approach, known as "prior polarity". The 1000 most common single words, and the 1000 most common 2-word phrases were extracted from the complete set of comments in the corpora. Two researchers independently rated the sentiment of each as positive, negative, either, or neutral separately for each of the three domains under consideration: (1) overall recommendation, (2) cleanliness, and (3) dignity. Where disagreements occurred, the sentiment was discussed and resolved between the 2 researchers. Kappa statistics for overall ratings were .76 for 1 word and .71 for 2-word phrases. For rating of dignity, they were .71 for 1 word and .70 for 2 words. For rating of cleanliness, they were .52 for 1 word and .48 for 2 words. For all of these calculations, $P<.001$.

A technique called "information gain" was used to reduce the size of the bag-of-words by identifying those words with the lowest certainty of belonging to a given class, and then removing them—this is an approach to feature selection [19]. This improved the computation time and also demonstrates the words with highest predictive accuracy.

A number of different technical approaches can be taken to classification in machine learning. We applied four different methods, to see which gave the quickest and most accurate results: (1) naïve Bayes multinomials (NBM) [20], (2) decision trees [21], (3) bagging [22], and (4) support vector machines [23]. Decision trees and bagging were carried out with REPTree in the Weka package. Support vector machines used an RBF Kernel. The accuracy of the prediction was compared with the patient's own quantitative rating by calculating, for each method, the accuracy (the percentage of correctly predicted observations from the total number of observations), the *F* measure (the harmonic mean of precision and recall), the Receiver Operating Characteristic (ROC), and the time taken to complete the task were calculated. To reduce computing processing time of the classification, we limited the words in the learning process to the top 10,000 words by frequency. All text was converted to lower case, and we removed all punctuation. Typographical errors and misspellings were not corrected.

Testing Prediction Accuracy

To obtain a score to predict sentiment analysis against, patient ratings left on the NHS Choices website on a Likert scale were converted into simple categories, either positive or negative about cleanliness and dignity, to simplify the prediction task. The website presents patients with five options to rate the cleanliness of a hospital: "exceptionally clean", "very clean", "clean", "not very clean", "dirty", and "does not apply". In this analysis, the first three options were grouped into a "clean" class and the "not very clean" and "dirty" into a "dirty" class. The website also asks patients to rate whether they were treated with dignity and respect by the hospital staff, with the options being "all of the time", "most of the time", "some of the time", "rarely", and "not at all". Again, the first three options were grouped, in this case into a "more dignity" class and the "rarely" and "not at all" into a "less dignity" class. Finally, the NHS Choices website asks all patients whether they would recommend the hospital or not.

Comparing Sentiment Analysis With the National Inpatient Survey

Having calculated the accuracy of our prediction algorithm, the results of the sentiment analysis were then compared with the national inpatient survey results for 2010. This is an annual national survey of randomly selected patients admitted to NHS hospitals in England, similar to the HCAHPS survey in the United States. The 2010 survey covered all 161 acute hospitals with adult services in England, involving 60,000 respondents nationally (response rate 50%). Patients were contacted via post between September 2010 and January 2011 if they had received overnight care in hospital in 2010 [24]. This survey includes both general and specific questions. In this study, we used only areas similar to the specific themes predicted from the NHS Choices data. The questions used were "In your opinion, how clean was the hospital room or ward that you were in?" (marked on a 4-point scale of "very" to "not at all"); "Overall, did you feel you were treated with respect and dignity while you were in the hospital?" (marked on a 3-point scale of "very" to "not at all"); and "Overall, how would you rate the care you received?" (marked on a 5-point scale from excellent to poor). The sentiment analysis ranking was compared with the patient survey ranking for each question at the hospital level, applying Spearman rank correlation, using Stata SE11 statistical software. We compared all 161 adult acute trusts in England.

Results

There was agreement between the patients' own quantitative rating of whether they would recommend their hospital and our prediction from sentiment analysis between 80.8% and 88.6% of the time (expressed as accuracy; Table 1), depending on the classification method used. Similarly, sentiment analysis agreed with whether the patient was treated with dignity and respect between 83.7% and 84.5% of the time, and agreed with whether the hospital was clean or not between 81.2% and 89.2% of the time. Table 2 shows the 10 words or 2-word phrases with the highest predictive accuracy.

NBM, bagging, and decision tree approaches to classification all produced similar measures of association, but the NBM algorithm, a first-order probabilistic model that uses word frequency information, performed the calculation faster (less than 0.2 seconds compared to hundreds of seconds for the other analyzed approaches). Of note, all algorithms tended to be worse at predicting cleanliness and SVM in particular. This may represent the limited language around cleanliness compared to the other opinions examined, or the more skewed results, with higher number of negative ratings.

On this basis, we choose to use NBM results for further comparison of our results with patient survey data. The relationship between the predictions of the NBM approach and the real ratings was reflected as Kappa statistics for interrater reliability of between .40 and .74 (P<.001 for all). We found significant, weak to moderate associations between machine learning predictions using NBM and quantitative responses from the national inpatient survey for the three categories examined: cleanliness, dignity, and overall recommendation (Spearman correlation coefficients between 0.37 and 0.51, P<.001 for all) (see Table 3). Rank correlations for overall recommendations of hospitals are displayed in Figure 1.

Table 1. Accuracy of different approaches to machine learning.

Question	Overall rating	Cleanliness	Dignity and respect
Naïve Bayes multinomial			
ROC	0.94	0.88	0.91
F measure	0.89	0.84	0.85
Accuracy (%)	88.6	81.2	83.7
Time (s)	0.11	0.05	0.06
Decision trees			
ROC	0.84	0.76	0.79
F measure	0.81	0.86	0.8
Accuracy (%)	80.8	88.4	83
Time (s)	552	206	332
Bagging			
ROC	0.89	0.83	0.87
F measure	0.82	0.87	0.85
Accuracy (%)	82.5	89.2	84.5
Time (s)	4871	2018	3164
Support vector machine			
ROC	0.79	0.53	0.6
F measure	0.84	0.84	0.8
Accuracy (%)	84.6	88.5	84.1
Time (s)	612	305	520

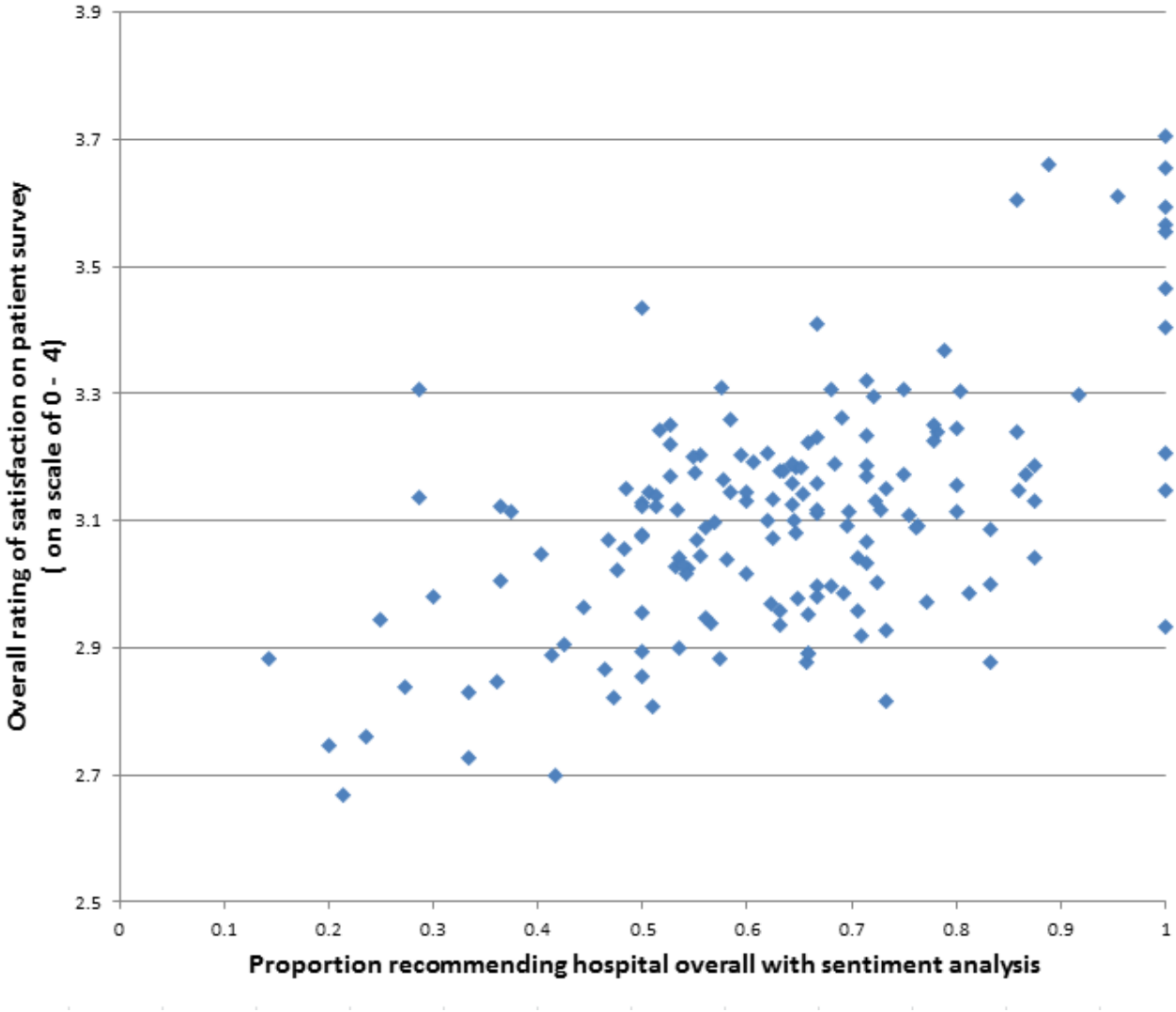
Table 2. The 10 one or two word phrases with the highest predictive accuracy for each topic.

Overall	Cleanliness	Dignity
told	dirty	rude
thank you	floor	told
left	left	left
rude	the floor	thank you
excellent	thank you	friendly
the staff	filthy	excellent
hours	bed	rude and
asked	patients	asked
was told	friendly	the staff
friendly	hours	staff

Table 3. Comparison of patient survey responses and machine learning prediction of comments at hospital trust level.

Patient survey question	Machine learning prediction	Spearman correlation coefficient	Probability
In your opinion, how clean was the hospital room or ward that you were in?	Machine learning prediction of comments about standard of cleanliness	0.37	*P*<.001
Overall, did you feel you were treated with respect and dignity while you were in the hospital?	Machine learning prediction of comments about whether the patient was treated with dignity and respect	0.51	*P*<.001
Overall, how would you rate the care you received?	Machine learning prediction of comments about whether the patient would recommend	0.46	*P*<.001

Figure 1. Comparison of the proportion recommending a hospital using sentiment analysis and traditional paper-based survey measures.



Discussion

Principal Findings

Our results reinforce earlier findings that sentiment analysis of patients' online comments is possible with a reasonable degree of accuracy [10] and that it is possible to identify salient aspects of reviews [25]. We have also shown that unstructured comments in free text, if processed appropriately, are associated with patient experience results from paper-based surveys conducted annually across all hospitals in England. This is in keeping with previous work that has demonstrated that there is a significant association between structured online ratings of care left on health care review websites and conventional surveys of satisfaction [12]. These results suggest a potential mechanism to make use of the large amounts of text on the Internet in which people describe their care, and that further exploration of the information contained within the free-text comments may be an important avenue for understanding patient experience, providing an additional source of information to complement traditional survey methods.

Strengths and Limitations

Sentiment analysis via a machine learning approach is only as good as the learning set that is used to inform it. By taking advantage of a complete national rating system over several years, we have been able to use many more ratings in this learning set than in other studies. Indeed, our learning set was more than 10 times larger than earlier work [9]. Moreover, in applications of sentiment analysis to health care data, researchers have had to train the system themselves by reviewing comments and ascribing characteristics to them, to allow the algorithm to learn. We used a large dataset that permitted us to directly compare free-text comments and quantitative ratings posted by the same patients, thus eliminating potential biases of reviewer assignment of comments. Similarly, the consistency of questions across the NHS inpatient survey and the questions asked on the NHS Choices website about health services allows for a direct comparison of patient comments and surveys that we believe has not previously been reported.

Online comments left without solicitation on a website are likely to have a natural selection bias towards examples of both good and bad care. It is likely that these online reviews are contributed more by those in particular demographic groups including younger and more affluent people [26]. Further, there are aspects

of patients' comments that are very hard for sentiment analysis to process. Irony, sarcasm, and humor, frequently adopted by English speakers when talking about their care, cannot be easily detected using this process. The use of prior polarity improved the results and mitigated some colloquial phrasing, but there were difficulties understanding those that depend on context. For example, phrases that cropped up repeatedly, such as "stank of urine" or "like an angel", could be easily characterized as negative or positive. The meaning of other frequently used phrases, however, was hard to establish without an understanding of their context. The best example of this was the phrase a "cup of tea". It was referred to in many different comments in these data, but without knowing the context, is it impossible to allocate it a direct sentiment. '"They didn't even offer me a cup of tea" is very different to "The nurse even made me a cup of tea". Our current algorithm could not yet make use of references to cups of tea or similar phrases that would be clear and obvious looked at by eye on a case-by-case basis. Future attempts to improve a natural language processing ability for patient experience would have to develop the capacity to accurately interpret this level of context-specific and idiomatic content. We appreciate that in this early exploratory work, we are not using the most state-of-the-art machine learning algorithms seen in other industries [27], or approaches to classification selection [28], but hope that further work might be able to adopt this.

Further Research Questions

Further research is needed to improve the performance of sentiment analysis tools, extending the process to other forms of free-text information on the Internet and exploring the relationships between views expressed by patients online and clinical health care quality. For example, several technical components might be added to improve the process, including the consideration of higher number n-grams (longer phrases) and refining contextual polarity (understanding what a word or phrase means given its context in a sentence). It would also be useful to compare the relatively simple techniques used in this analysis against other platforms and tools used for the sentiment analysis and opinion mining process, for example WordNet Affect [29] and SentiWordNet [30,31].

Policy Implications

Large amounts of data about the use of services are collected in digital form. An important strand of this is consumer opinion and experience. Today, many people express their views and share their experience of goods and services via the Internet and social media. Such data, converted to information, are essential in improving services, facilitating consumer choice and, in some sectors, exploring public accountability and value in the use of taxpayers' money.

By its nature, the information is highly personalized, idiosyncratic, and idiomatic. However, if it is to be useful, it must be analyzed in ways that are not solely reliant on someone reading individual contributions (although this is valuable to consumers) nor on pre-structured responses necessary to allow aggregation. A solution to the challenge of "big data" is to find automated methods for analyzing unstructured narrative commentary, which is a potentially rich source of learning. In this respect, health care is no different to many other industries although it has perhaps been slower than other sectors to recognize the importance of it.

As our confidence in techniques of data mining and sentiment analysis grow, information of this sort could be routinely collected, processed, and interpreted by health care providers and regulators to monitor performance. Moreover, information could be taken from a number of different text sources online, such as blogs and social media. If this information could be harvested from these locations and then processed into timely and relevant data, it could be a valuable tool for quality improvement. We have previously suggested that as the usage of rating websites, social networks, and microblogs increases [3,4], this free-text information represents a growing and largely untapped source of data that could be considered a "cloud of patient experience" [32]. Others, including Alemi, have discussed similar ideas, with Cambria and colleagues describing a notion of "crowd validation" of a health service [10,33]. This has the potential to provide up-to-date information about patient experience at lower cost than the traditional survey route. It might also allow the views of younger, more tech savvy groups—who are often poor responders to paper-based surveys—to be sampled. Eventually, there might even be the potential to develop a close to real-time early warning system for poor clinical care, if large enough amounts of data can be collected and prediction accuracy can be reliably reproduced. However, caution should be taken before placing too much faith in a quantitative approach, as the qualitative analysis of information of this sort is known to provide useful insights [34]. Qualitative and quantitative approaches should be seen as complementary.

Conclusions

This work demonstrates that sentiment analysis of patients' comments about their experience of health care is possible and that this novel approach is associated with patient experience measured by traditional methods such as surveys. This work adds to a growing body of literature opening up a new understanding of the patients' point of view of care from their postings online—on social networks, blogs, and rating websites. Although at an early and experimental stage, it presents future possibilities to understand health care system performance in close to real time. Bates and colleagues have described the confluence of patient-centered care and social media as a "perfect storm" that is likely to be of major value to the public and to health care organizations [35]. These early findings hint at how that might occur.

Acknowledgments

We would like to thank the team at NHS Choices, and John Robinson, Paul Nuki, and Bob Gann in particular, for providing access to their data. We thank Jane Lucas for reviewing the sentiment of words.

Dr Greaves was supported for this research by The Commonwealth Fund. The views presented here are those of the authors and should not be attributed to The Commonwealth Fund or its directors, officers, or staff. Dr Millett is funded by the Higher Education Funding Council for England and the National Institute for Health Research. The Department of Primary Care & Public Health at Imperial College is grateful for support from the National Institute for Health Research Biomedical Research Centre Funding scheme, the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care scheme, and the Imperial Centre for Patient Safety and Service Quality. The funding sources had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.

Conflicts of Interest

Professor Donaldson was Chief Medical Officer, England from 1997 to 2010. Professor Darzi was Parliamentary Under-Secretary of State (Lords) in the United Kingdom Department of Health from 2007 to 2009. The other authors declare no conflicts of interest.

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
2. Darzi A. High quality care for all: NHS Next Stage Review final report. London: Department of Health; 2008.
3. 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]
4. 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]
5. Demographics of internet users. 2012. Pew Research Center URL: http://pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx [accessed 2013-05-16] [WebCite Cache ID 6GfFDqBQr]
6. Fox S. The Social Life of Health Information. Washington, DC: Pew Research Center; 2011. URL: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx [accessed 2013-10-19] [WebCite Cache ID 6KUTtXtJH]
7. Tumasjan A, Sprenger T, Sandner P, Welp I. Predicting elections with Twitter: what 140 characters reveal about political sentiment. 2010 Presented at: Fourth International AAAI Conference on Weblogs and Social Media; 2010; Washington DC p. 178-185.
8. Holzinger A, Yildirim P, Geier M, Simonic KM. Quality-based knowledge discovery from medical text on the Web example of computational methods in Web intelligence. In: Pasi G, editor. Advanced Techniques in Web Intelligence - Quality-based Information Retrieval Studies in computational intelligence, Lecture Notes in Artificial Intelligence, LNAI. Berlin: Springer; 2013:145.
9. Pang B, Lee L. Opinion mining and sentiment analysis found. Trends Inf Retr 2008;2(1-2):1-138.
10. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4] [Medline: 22207014]
11. The Power of Information: Putting all of us in control of the health and care information we need. London: Department of Health; 2012. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf [accessed 2013-10-19] [WebCite Cache ID 6KUUB6fqh]
12. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
13. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. Bioinformatics 2004 Oct 12;20(15):2479-2481 [FREE Full text] [doi: 10.1093/bioinformatics/bth261] [Medline: 15073010]
14. Sigurdardottir AK, Jonsdottir H, Benediktsson R. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. Patient Educ Couns 2007 Jul;67(1-2):21-31. [doi: 10.1016/j.pec.2007.03.007] [Medline: 17420109]
15. Ivanciuc O. Weka machine learning for predicting the phospholipidosis inducing potential. Curr Top Med Chem 2008;8(18):1691-1709. [Medline: 19075775]
16. NHS Choices. 2013. NHS Choices datasets: Hospital comments and responses URL: http://www.nhs.uk/aboutNHSChoices/professionals/developments/Pages/NHSChoicesDatasets.aspx [accessed 2013-05-16] [WebCite Cache ID 6GfEQDA8O]
17. Petz G, Karpowicz M, Fürschuß H, Auinger A, Winkler S, Schaller S, et al. On text preprocessing for opinion mining outside of laboratory environments. In: Active Media Technology, Lecture Notes in Computer Science. Berlin: Springer; 2012:646.
18. Zhang Y, Jin R, Zhou ZH. Understanding bag-of-words model: a statistical framework. Int J Mach Learn & Cybe 2010;1(1):43-42.

19. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. 1997 Presented at: Fourteenth International Conference on Machine Learning; 1997; Nashville p. 412-420.
20. McCallum A, Nigam K. A comparison of event models for naive bayes text classification. 1998 Presented at: AAAI-98 Workshop on 'Learning for Text Categorization'; 1998; Madison.
21. Quinlan JR. Induction of decision trees. Mach Learn 1986;1(1):81-106.
22. Breiman L. Bagging predictors. Mach Learn 1996;24(2):123-140.
23. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273-297.
24. Acute Trusts: Adult Inpatients Survey (2010) Economic and Social Data Service. 2012. URL: http://discover.ukdataservice.ac.uk/catalogue?sn=6503 [accessed 2013-05-16] [WebCite Cache ID 6GfEDvzAK]
25. Brody S, Elhadad N. Detecting salient aspects in online reviews of health providers. AMIA Annu Symp Proc 2010 Nov:202-206.
26. 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]
27. Liu B. Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies. Seattle: Morgan & Claypool Publishers; 2012:1-167.
28. Demsar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 2006;7:1-30.
29. Strapparava C, Valitutti A. WordNet affect: an affective extension of WordNet. 2004 Presented at: 4th International Conference on Language Resources and Evaluation; 2004; Lisbon p. 1083-1086.
30. Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0. An enhanced lexical resource for sentiment analysis and opinion mining. 2010 Presented at: 7th Conference on International Language Resources and Evaluation; 2010; Valletta.
31. Esuli A, Sebastiani F. Sentiwordnet: a publicly available lexical resource for opinion mining. 2006 May Presented at: 5th Conference on Language Resources and Evaluation; 2006; Genoa.
32. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
33. Cambria E, Hussain A, Havasi C. Towards the crowd validation of the British National Health Service. 2010 Presented at: WebSci10: Extending the Frontiers of Society On-Line; 2010; Raleigh, NC.
34. 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;39(1):7-15. [Medline: 23367647]
35. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]

Abbreviations

HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
NBM: Naïve Bayes Multinomial
NHS: National Health Service
NHSC: NHS Choices
ROC: Receiver Operating Characteristic
SVM: Support Vector Machine

Edited by G Eysenbach; submitted 16.05.13; peer-reviewed by M Sokolova, E Yom-Tov, P Brooker, A Holzinger; comments to author 11.06.13; revised version received 10.07.13; accepted 29.08.13; published 01.11.13.

Please cite as:
Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L
Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online
J Med Internet Res 2013;15(11):e239
URL: http://www.jmir.org/2013/11/e239/
doi:10.2196/jmir.2721
PMID:24184993

Original Paper

Analysis of the Purpose of State Health Departments' Tweets: Information Sharing, Engagement, and Action

Rosemary Thackeray[1], MPH, PhD; Brad L Neiger[1], PhD; Scott H Burton[2], PhD; Callie R Thackeray[1]

[1]Brigham Young University, Department of Health Science, Provo, UT, United States

[2]Brigham Young University - Idaho, Department of Computer Science and Engineering, Rexburg, ID, United States

Corresponding Author:
Rosemary Thackeray, MPH, PhD
Brigham Young University
Department of Health Science
221 Richards Building
Provo, UT, 84602
United States
Phone: 1 801 422 1704
Fax: 1 801 422 0931
Email: rosemary_thackeray@byu.edu

Abstract

Background: Public health agencies are actively using social media, including Twitter. In the public health and nonprofit sectors, Twitter has been limited to one-way communication. Two-way, interactive communication on Twitter has the potential to enhance organizational relationships with followers and help organizations achieve their goals by increasing communication and dialog between the organization and its followers. Research shows that nonprofit organizations use Twitter for three main functions: information sharing, community building, and action.

Objective: It is not known whether state health departments are using Twitter primarily for one-way information sharing or if they are trying to engage followers to build relationships and promote action. The purpose of this research was to discover what the primary function of Twitter use is among state health departments in the United States and whether this is similar to or different from nonprofit organizations.

Methods: A complete list of “tweets” made by each state health department account was obtained using the Twitter application programming interface. We randomly sampled 10% of each state health department’s tweets. Four research assistants hand-coded the tweets’ primary focus (organization centric or personal health information centric) and then the subcategories of information dissemination, engagement, or action. Research assistants coded each tweet for interactivity, sophistication, and redirects to another website. Data were analyzed using SPSS version 20.

Results: There were 4221 tweets from 39 state health departments. There was no statistically significant difference in the number of tweets made by a state health department and the state population density (P=.25). The majority of tweets focused on personal health topics (69.37%, 2928/4221) while one-third were tweets about the organization (29.14% , 1230/4221). The main function of organization-based tweets was engagement through conversations to build community (65.77%, 809/1236). These engagement-related tweets were primarily recognition of other organizations’ events (43.6%, 353/809) and giving thanks and recognition (21.4%, 173/809). Nearly all of the personal health information-centric tweets involved general public health information (92.10%, 1399/1519) and 79.03% (3336/4221) of tweets directed followers to another link for more information.

Conclusions: This is the first study to assess the purpose of public health tweets among state health departments. State health departments are using Twitter as a one-way communication tool, with tweets focused primarily on personal health. A state health department Twitter account may not be the primary health information source for individuals. Therefore, state health departments should reconsider their focus on personal health tweets and envision how they can use Twitter to develop relationships with community agencies and partners. In order to realize the potential of Twitter to establish relationships and develop connections, more two-way communication and interaction are essential.

(J Med Internet Res 2013;15(11):e255) doi:10.2196/jmir.3002

KEYWORDS

social media; public health; engagement

Introduction

Individual and organizational use of social media is rapidly increasing. Researchers estimate that by 2017, the number of individuals and corporations who have social networking accounts will reach more than 4.8 billion [1]. In particular, Twitter is becoming ever more popular. A total of 18% of Internet users have a Twitter account [2]. Among Fortune 500 companies, 77% have Twitter accounts [3] and among Forbes' 200 largest charities, all use at least one form of social media, with 96% using Twitter [4]. Public health agencies are actively using social media, including Twitter [5-8]. Twitter has been used to disseminate information about diabetes [9], breast cancer [10], to communicate during a disaster [11], and to understand health-related trends and issues such as influenza [12], tobacco [13], prescription drug misuse [14], and suicide [15].

Fundamentally, organizations use social media sites to build relationships [16-20]. Twitter research shows that an organization's relationship with customers is influenced by the organization's interactivity or level of communication and contact [21]. An organization that includes social media as part of its communication strategy has the potential to increase the level of communication dialog with its customers. For example, social media can facilitate customers talking with each other, as well as customers talking with the organization [22]. When dialog and communication occur, relationships are built and these relationships are often tied to key organizational outcomes [23]. The Twitter homepage specifically states that Twitter allows businesses to connect in real-time to their customers to share product information, gather market research data, and develop relationships with both customers and partners [24]. Organizations that have relationships with customers can use Twitter to enhance brands, increase visibility, support customers, network, communicate internally, generate leads, and support other online presences [25].

Though use of Twitter has the potential to increase communication and dialog among organizations and its followers, in the nonprofit and public health sectors it has been limited to one-way communication [7,26], meaning a message is sent from one person or organization to a receiver with no expectation of a response. However, the value of social media is its ability to create two-way, interactive communication between two or more people. This two-way communication can enhance organizational relationships with followers and help organizations achieve their goals. On Twitter, interactive communication is achieved, in part, by the use of the @ symbol (which directs a message to a specific Twitter user), the user responding to public reply messages, the user asking for feedback in a "tweet", and the use of personal pronouns (eg, us, we, you) [27].

In a study of nonprofit organizations, Lovejoy and Saxton [28] identified three main functions for organizational Twitter use, namely: information sharing, community building, and action. Information sharing meant that the organization used Twitter to disseminate information including facts about the organization and its activities. Community building was a function served by tweets that aimed to build a community or network among followers through dialog and interactivity. Finally, the function of action was indicated by tweets that asked followers to do something for the organization. Lovejoy and Saxton found that nonprofit organizations primarily used Twitter as a one-way communication tool to convey information to their followers (58.6%); one-quarter of tweets were found to be community building and only 16% were action based. The authors concluded their study by calling for more research about how governments use social media.

This study expands the work of Lovejoy and Saxton by exploring the purposes for Twitter use among public health agencies, specifically state health departments. As reported earlier, though public health agencies use social media including Twitter, it is not known for what purpose they use Twitter and more particularly, if their purposes for using Twitter are similar to or different from nonprofit organizations. More specifically, it is not known whether state health departments are using Twitter primarily for one-way information sharing or if they are trying to engage followers to build relationships. Therefore, the purpose of this research was to answer the following two research questions: (1) What is the main function of state health department tweets—information, engagement, or action? and (2) Is the content of state health department tweets more often about the organization or about personal health?

Methods

This study involved state health departments in the United States. A list of all state health departments was obtained from the Association of State and Territorial Health Officers (ASHTO) website [29] and from the Centers for Disease Control and Prevention website [30].

State health departments represent all 50 US states [31], with 55% of state health departments classified as freestanding or independent (ie, the agency provides public health/health services only). The other 45% are categorized as super or umbrella agencies and provide additional services related to Medicaid and other public assistance, mental health, substance abuse, environmental protection, aging, child and family services, and so forth [31].

State health departments are granted legislative authority through codes and statutes to promote and protect public health and safety. These responsibilities are usually addressed by planning and implementing health promotion programs, enacting and enforcing laws and regulations, and providing access to primary care health services. State health departments also provide technical assistance to local health departments and nongovernmental agencies [31].

The existence of each state health department Twitter account was verified by three means: (1) visiting the state health department home page to ascertain the presence of a Twitter

icon, (2) searching on Twitter for the name of the state health department, and (3) performing a Google search for the state health department Twitter account. A complete list of tweets made by each state health department account was obtained using the Twitter application programming interface (API) during July 2012. Because the Twitter API limits the maximum number of tweets that can be retrieved to the most recent 3200 per account, this limit was considered a complete tweet list for any state health department account that exceeded 3200 tweets. Because the intent of this study was to analyze state health departments collectively, not individually, we randomly sampled 10% of each state health department's tweets.

We created a coding instrument based on Lovejoy and Saxton's original classifications [28]. Their coding classifications related exclusively to organization-focused tweets [28]. We expanded the classification by adding a second category that included a personal health information focus. This decision was made based on two factors. First, an initial analysis of the tweets showed there were several public health information-only tweets, and second, results from a related study showed personal health information tweets were common among public health agencies [8]. We recognize that at a basic level all tweets are focused on the organization and fulfilling its mission. However, for the purpose of this study, we identified organization-centric tweets as those with the purpose of building and strengthening the organization. In contrast, personal health information-centric tweets were focused on one-way information dissemination about health information.

Four research assistants hand-coded all tweets. Research assistants compared coding results and resolved any discrepancies. Discrepancies were resolved by discussing the issue and coming to a consensus. Discrepancies most often occurred because of a simple error related to data entry or a misinterpretation of the tweet.

The first step was to code each tweet according to its primary focus, whether it was an organization-centric tweet or a personal health information-centric tweet. Next, research assistants coded each organization-centric tweet to determine whether the primary function was information dissemination, engagement, or action. Information dissemination was defined as one-way sharing of information about the organization and its activities [28], and included events or services, news, facts, reports, or job announcements. Engagement tweets were posts that focused on building relationships and networks with followers [28]. These tweets included giving thanks and recognition for doing something for the organization; acknowledging another organization's events; responding to public reply messages; asking for a response or feedback; and asking for a follow, to become a fan, or to spread the word by retweeting the message. Action-based tweets represented those that encouraged the follower to do something for the organization [28], such as inviting followers to attend events or meetings, complete a survey, donate goods or money, volunteer time, or to participate in lobbying or advocacy.

Research assistants then coded each personal health information-centric tweet as one of two subcategories: information or action. Information-based tweets were one of three types: general public health information (eg, flooding can introduce impurities to both public and private drinking water sources), risk communication (eg, disease outbreaks or natural or man-made disasters), or public health reports (eg, Injury Prevention Policy Report Shows Arkansas Making Progress, More Work Needed). Action-based tweets encouraged individuals to participate in preventive health screenings (eg, Ask your health care provider for a group B strep test when you are 35-37 weeks pregnant), modifying one's lifestyle (eg, Portion control is a must, so keep a serving cup in your purse or briefcase for healthy meals throughout the work day), or encouraged individuals to learn more and increase their knowledge (eg, Does your child walk to school? Learn how you can help ensure safe walking routes in your community). Each classification category was mutually exclusive.

In addition, research assistants coded each tweet for the degree to which it was considered interactive, its level of sophistication, and if it redirected the follower to another site for more information. Interactivity was determined by the presence of (1) an @ reply symbol, signifying that the state health department was responding to a post made by another Twitter follower, (2) an @username, indicating that the state health department was directing its post to a specific user, and (3) the use of personal pronouns. Tweet sophistication was noted by whether or not it (1) was a truncated tweet, meaning that the state health department had posted the information on one platform (eg, Facebook) and it was automatically posted to the Twitter account as well, (2) was a retweet, and (3) included hashtags. Truncated tweets and retweets denote that the state health department was not developing content specifically for Twitter but was sharing other Twitter users' content. Hashtags, which are used to categorize tweets so users can easily follow topics posted to Twitter, are also reflective of a more advanced Twitter user. Finally, tweets that redirected followers to another source for more information signified that the state health department was using Twitter as a one-way communication tool to spread the word and link people to more information. Data were analyzed using SPSS version 20 [32].

Results

The final sample included 4221 tweets from 39 state health departments with a Twitter account. State health departments had a mean of 2033.2 (SD 1974.9) followers and were following a mean of 414.6 (SD 725.3) other Twitter users (Table 1).

Harris and colleagues found that local health departments that serve larger populations tweet more often than those serving smaller populations [9]. Therefore, we tested to see if there was a similar difference in the frequency of state health department tweeting based on a state's population density, which was defined as population per square mile of land area as identified in the 2010 census [33]. States were divided into three strata: low (less than 100 persons/square mile), medium (100-200 persons/square mile), and high (more than 200 persons per square mile). There were 16 states with low population density, 10 with medium, and 13 with high. The final number of tweets by population density included 1656 low (39.23%, 1656/4221), 860 medium (20.37%, 860/4221), and 1705 high (40.39%,

1705/4221). There was no statistically significant difference in the number of tweets made by a state health department and the state population density (P=.25). Therefore, we did no further analysis by population density.

Three-quarters (76.14%, 3214/4221) of tweets were original, meaning the state health department created and posted the content on Twitter. Only 4.45% (192/4221) of tweets were posted from another platform such as Facebook and 79.03% (3336/4221) of tweets directed followers to another link for more information. Roughly one-third (36.88%, 1557/4221) of tweets used personal pronouns. Hashtags were included 31.15% (1315/4221) of the time, while the @ symbol was used infrequently (6.99%, 295/4221).

The primary results of the study are presented in Table 2. The majority of tweets focused on personal health topics, while one-third were tweets about the organization. The main function of state health department organization-based tweets was engagement through conversations to build community. These engagement-related tweets were primarily recognition of other organization's events and giving thanks and recognition. Rarely was the state health department using Twitter to ask for feedback or suggestions, respond to public reply messages, ask for a response to a tweet, ask for a relationship such as becoming a follower, or to disseminate information by retweeting a post.

Just over one-quarter of organization-based tweets focused on sharing one-way information about the organization. These organization information-based tweets centered on events or services, job announcements, and facts. Only 6.67% (82/1230) of organization-based tweets related to asking followers to take action for the organization.

Personal health information-centric tweets were split nearly in half between information and action. Nearly all of the information-related tweets involved general public health information. Action-based tweets predominately encouraged followers to take action to learn more, followed by encouragement to take action to modify their lifestyle.

Table 1. Characteristics of state health department Twitter accounts.

State	Total tweets	Following	Followers	Date created
Alabama	1422	30	292	6/24/2011
Alaska	2029	130	2597	1/31/2009
Arizona	3199	162	4753	10/24/2008
Arkansas	677	338	1079	2/5/2010
California	2003	185	6612	4/21/2009
Colorado	348	530	812	7/7/2011
Connecticut	3199	4295	4361	4/27/2009
Delaware	2953	585	2070	6/15/2009
Florida	1164	830	1293	5/11/2011
Georgia	1544	684	710	6/27/2011
Hawaii	1604	928	1969	9/29/2009
Idaho	372	143	134	6/16/2011
Illinois	1707	526	591	9/4/2009
Indiana	143	71	87	5/18/2012
Iowa	1148	16	4535	4/30/2009
Kansas	1547	262	1370	9/1/2009
Kentucky	84	3	1643	4/28/2009
Louisiana	1030	1607	1915	8/10/2010
Maryland	983	437	1058	6/5/2009
Massachusetts	882	368	9546	3/11/2009
Michigan	823	1091	3820	7/16/2009
Minnesota	714	252	3830	3/18/2009
Mississippi	565	25	2719	9/2/2008
Missouri	1378	338	1440	9/23/2009
Nebraska	1819	664	1289	8/14/2009
New Hampshire	272	36	441	3/23/2009
New Jersey	198	18	546	2/14/2011
New Mexico	80	0	35	7/30/2010
New York	731	151	1064	3/17/2010
Ohio	668	331	1941	11/16/2009
Oklahoma	100	26	121	1/3/2012
Rhode Island	711	111	3128	4/25/2009
South Carolina	227	7	956	10/29/2010
Tennessee	2657	94	1822	10/23/2009
Utah	359	239	2853	4/28/2009
Vermont	795	99	1315	4/27/2009
Virginia	257	224	511	9/8/2010
Washington	1391	128	3692	7/23/2009
Wisconsin	436	209	347	5/4/2011

Table 2. State health departments' use of Twitter.

Tweet category	Tweet subcategory	n (%)
Organization-centric		1230/4221 (29.14%)
	Engagement to build community	809/1236 (65.77%)
	Recognition of other organization's events	353/809 (43.63%)
	Giving thanks and recognition	173/809 (21.38%)
	Ask for feedback or suggestions	28/809 (3.46%)
	Respond to public reply messages	67/809 (8.28%)
	Ask for response to a tweet	32/809 (3.96%)
	Ask for a relationship	29/809 (3.58%)
	Retweet a post	27/809 (3.34%)
	Other	100/809 (12.36%)
	Information about the organization	338/1230 (27.48%)
	Events or services	141/338 (41.72%)
	Job announcements	77/338 (22.78%)
	Facts	50/338 (14.79%)
	News	26/338 (7.69%)
	Reports	1/338 (0.30%)
	Other	44/338 (13.02%)
	Action	82/1230 (6.67%)
	Attend events	29/82 (35.37%)
	Attend meetings to provide input	6/82 (7.32%)
	Complete a survey	5/82 (6.10%)
	Donate goods or money	2/82 (2.44%)
	Volunteer time	13/82 (15.85%)
	Participate in lobbying and/or advocacy	20/82 (24.39%)
	Other	7/82 (8.54%)
Personal health information-centric		2928/4221 (69.37%)
	Information	1519/2928 (52.05%)
	Public health information	1399/1519 (92.10%)
	Risk communication	63/1519 (4.15%)
	Reports	27/1519 (1.78%)
	Other	30/1519 (1.97%)
	Action	1409/2926 (48.12%)
	Learn more	640/1409 (45.42%0
	Modify lifestyle	523/1409 (37.12%)
	Preventive health screenings	124/1409 (8.80%)
	Other	122/1409 (8.66%)
News		51/4221 (1.21%)
Miscellaneous		12/4221 (0.28%)

Discussion

Principal Findings

This study examined state health departments' Twitter posts to determine the main function of related tweets. Results show that the majority of tweets were about personal health and a limited number were about the state health department as an organization. These results are similar to those found in a study about local health departments' use of Twitter [8]. State health departments and other public health agencies may be unique, unlike other nonprofit organizations [28], when it comes to the main function for their social media use. State health departments do not appear to be using Twitter to build their organization and develop relationships with followers, but rather to disseminate health information.

Personal health information-centric tweets contained general public health information. This is similar to what was found among local health departments [8]. These results are comparable to a study in Australia that found government tweets were dominated by public health advice and nonspecific health conditions [34]. The predominant use of Twitter to share personal health information raises two primary questions: (1) Who are the state health departments' Twitter followers—individuals or other organizations? and (2) If followers are individuals, do they consider the state health department to be a primary source of health information? Although people do go online seeking health information [35], this health-seeking behavior is different from being a Twitter follower of a state health department, meaning one has opted in to receive regular updates. In a study among US adults about the perceived credibility of specific health information, physicians were rated as the most credible, followed by the Internet [36]. As far as credibility of online health information, one study found that perceived credibility was generally higher when the source was a specific website [37]. Therefore, general health information on a state health department Twitter account may not be perceived as highly credible.

In using Twitter, state health departments must understand the composition of their current followers and identify who it is they are trying to reach. If the state health departments' aim is to build a community of health-related organizations, their messages and strategy will be different than if they are aiming to attract individuals in their corresponding communities. Specifically, individual Twitter users tend to be younger, of Black or Hispanic background, have a college education, and have incomes over US$75,000 a year [2]. These may or may not be the individuals with whom the state health department is trying to cultivate relationships. Rather, state health departments may be more interested in using Twitter to develop relationships with community-based organizations and other agencies. Fostering online relationships with these agencies may result in offline collaborations. However, developing these online relationships will require concerted effort between Twitter users, as they do not appear to evolve naturally. For instance, in studying Twitter connections between state health departments, Harris and colleagues found that state health departments on Twitter tend to follow other state health departments in their region who are also on Twitter. However, the follow is not reciprocal, meaning they are not following each other back [38]. Reciprocal following builds the users network. It also allows followers to receive tweets from the other user, which are then more likely to be retweeted [39]. Additionally, social network analysis states that reciprocity indicates stronger ties among people [40]. Also, among individuals, the number of Twitter followers is linked to increased social capital [41]. There may also be a similar increase in strength of connections and social capital among connected state health departments and their followers.

The rate of tweeting information about the organization is substantially less than what was found by Lovejoy and Saxton [28] and among local health departments [8]. State health departments may be less concerned with promoting themselves as an agency and more focused on fulfilling one of public health's ten essential services: inform, educate, and empower [42]. However, state health departments may want to reconsider how they use Twitter and create ways to convey information about the organization in order to increase the community's awareness of the state health department, its purpose, priorities, and contribution to the state.

State health departments should continue to post engagement-related tweets that focus on recognizing other organizations' events or giving thanks and recognition. The rate of engagement-related tweets among state health departments was more than that found by Lovejoy and Saxton for nonprofit organizations [28] and among local health departments [8]. This indicates that state health departments are making an effort to reach out to other community organizations, which is a positive step toward building relationships with current or potential partners. Both information sharing and engagement with other organizations and partners can be particularly beneficial for a state health department when engaging in advocacy-related efforts, which have long been a core public health strategy and an essential public health service [42]. In fact, use of social media has been identified as critical to influencing advocacy and social movements [43-45].

Interestingly, although state health departments were posting engagement-related tweets to foster and build relationships, very few state health departments were asking followers to take action to benefit the organization. Researchers have proposed an engagement hierarchy between organizations and followers that progresses from low to medium to high with high engagement characterized by followers becoming involved with the organization as either partners in fulfilling the organization's goals or as direct recipients of the organization's programs and services [46]. The hierarchy posits that high engagement is the culminating and defining purpose of social media use in public health settings. This suggests that state health departments may want to consider using Twitter to recruit followers and foster relationships that will benefit organizational causes and programs.

The majority of the tweets were original, meaning the state health departments are investing effort in creating unique content. These results are similar to local health departments' use of Twitter [8]. This implies that the state health departments

have identified specific content they want to convey to their followers and are not re-posting random tweets. Original content may also be more likely to draw the interest and attention of followers as it suggests that the organization is tailoring its posts to the interests and needs of its followers.

Twitter is being used as a one-way communication tool. Though state health departments are trying to engage in conversation, most of these tweets were about recognition of other agencies' events and giving thanks. Rarely did state health departments attempt to engage followers in dialog by asking for a response, a retweet, and so forth. Furthermore, three-quarters of tweets included a link for where to go for more information. These results are similar to other research that showed a preponderance of hyperlinks included with tweets [26,47]. The emphasis on one-way communication is further evidenced by the lack of inclusion of the @ symbol in tweets, which would direct a message to a specific Twitter follower.

Limitations

The results should be interpreted with the following limitations in mind. First, we were able to sample only public tweets. It is possible that state health departments are responding to individuals through direct messages, which are private, but there is no way to assess that without having access to individual accounts. Second, we observed only one side of the potential dialogue between state health departments and their followers. That is, we were able to study posts that the state health department made, but were not able to analyze Twitter posts that were directed to the state health department from other Twitter users. Third, this study is about a specific social media application, Twitter, and it could be that some state health departments behave very differently on other applications. For example, on social networking sites such as Facebook, there may be more two-way communication or the purposes for posts may be different from Twitter. Last, although there were four research assistants coding the data, there is still the possibility of coder subjectivity in interpreting the main purpose of the tweet.

Conclusions

State health departments are using Twitter as a one-way communication tool, with tweets focused primarily on personal health. When tweeting about the organization, state health departments are trying to engage audiences through posts that focus on recognition of community and organizational events. There is potential for state health departments to use Twitter to develop relationships with other community agencies. To do so, state health departments should reconsider their focus on personal health tweets. To realize the potential of Twitter to establish relationships and develop connections with followers, more two-way communication and interaction are essential.

Acknowledgments

The authors acknowledge Victoria Doucette, Emily Christensen, and Jennifer Reese for assistance with coding the data; Dr Christophe G Giraud-Carrier for conceptual and technical assistance; and Mark and Christine Cronquist for financial assistance to hire research assistants.

Authors' Contributions

RT conceived of the study, participated its design and coordination, performed the statistical analysis, and drafted the manuscript. BN conceived of the study, participated the study design and coordination, and assisted with drafting the manuscript. SHB collected data and assisted with drafting the manuscript. CRT participated in the study design and coordination, coded the data, and assisted with drafting the manuscript.

Conflicts of Interest

None declared.

References

1. Radicati S, Levenstein J. Email statistics report-2013. URL: http://www.radicati.com/?=9659 [accessed 2013-10-01] [WebCite Cache ID 6K3BX1L8C]
2. Brenner J, Smith A. 72% of online adults are social networking site users. 2013. URL: http://pewinternet.org/Reports/2013/social-networking-sites/Findings/Twitter.aspx [accessed 2013-10-01] [WebCite Cache ID 6K3BaDIUb]
3. Barnes NG, Lescault AM, Wright S. Fortune 500 are bullish on social media: big companies get excited about Google+, Instagram, Foursquare and Pinterest. 2013. URL: http://www.umassd.edu/media/umassdartmouth/cmr/studiesandresearch/2013_Fortune_500.pdf [accessed 2013-11-01] [WebCite Cache ID 6KoQrw9KX]
4. Barnes NG. Social media usage now ubiquitous among US top charities, ahead of all other sectors. 2011. URL: http://www.umassd.edu/media/umassdartmouth/cmr/studiesandresearch/charity2010.pdf [accessed 2013-10-01] [WebCite Cache ID 6K3Dwhjyp]
5. Centers for Disease Control and Prevention. CDC eHealth metrics dashboard. 2012. URL: http://www.cdc.gov/metrics/socialmedia/index.html [accessed 2013-10-01] [WebCite Cache ID 6K3D6JPjt]
6. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013 Sep;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]

7. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
8. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
9. Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of twitter to disseminate diabetes information, United States. Prev Chronic Dis 2013;10:E70 [FREE Full text] [doi: 10.5888/pcd10.120215] [Medline: 23639765]
10. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13(1):508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
11. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
12. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. : Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; 27-31 July 2011; Edinburgh, Scotland p. 1568-1576.
13. Myslín M, Zhu SH, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
14. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res 2013;15(4):e62 [FREE Full text] [doi: 10.2196/jmir.2503] [Medline: 23594933]
15. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2013 Oct 11:1-9 [FREE Full text] [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
16. Phethean C, Tiropanis T, Harris L. Rethinking measurements of social media use by charities: a mixed methods approach. 2013 Presented at: Web Science; 2013 May 02 - 04; Paris, France p. 02-04.
17. McCorkindale T, DiStaso MW, Sisco HF. How millennials are engaging and building relationships with organizations on Facebook. The Journal of Social Media in Society 2013;2(1):66-87.
18. Men LR, Tsai WS. How companies cultivate relationships with publics on social network sites: Evidence from China and the United States. Public Relations Review 2012 Dec;38(5):723-730. [doi: 10.1016/j.pubrev.2011.10.006]
19. Bortree DS, Seltzer T. Dialogic strategies and outcomes: An analysis of environmental advocacy groups' Facebook profiles. Public Relations Review 2009 Sep;35(3):317-319. [doi: 10.1016/j.pubrev.2009.05.002]
20. Lapointe L, Ramaprasad J, Vedel I. Collaborating through social media to create health awareness. 2013 Presented at: 2013 46th Hawaii International Conference on System Sciences (HICSS): IEEE; January 7-10, 2013; Wailea, Maui. [doi: 10.1109/HICSS.2013.135]
21. Saffer AJ, Sommerfeldt EJ, Taylor M. The effects of organizational Twitter interactivity on organization–public relationships. Public Relations Review 2013 Sep;39(3):213-215. [doi: 10.1016/j.pubrev.2013.02.005]
22. Mangold WG, Faulds DJ. Social media: The new hybrid element of the promotion mix. Business Horizons 2009;52:357-365.
23. Bruning SD, Dials M, Shirka A. Using dialogue to build organization–public relationships, engage publics, and positively affect organizational outcomes. Public Relations Review 2008 Mar;34(1):25-31. [doi: 10.1016/j.pubrev.2007.08.004]
24. Twitter. About Twitter. 2013. URL: https://twitter.com/about [accessed 2013-10-01] [WebCite Cache ID 6K3BmF8Jf]
25. Dugan L. 7 ways businesses can use Twitter. 2012 URL: http://www.mediabistro.com/alltwitter/7-ways-businesses-can-use-twitter_b3994 [accessed 2013-10-01] [WebCite Cache ID 6K3BnanI4]
26. Waters RD, Jamal JY. Tweet, tweet, tweet: A content analysis of nonprofit organizations' Twitter updates. Public Relations Review 2011 Sep;37(3):321-324. [doi: 10.1016/j.pubrev.2011.03.002]
27. Kwon ES, Sung Y. Follow Me! Global Marketers' Twitter Use. Journal of Interactive Advertising 2011 Sep;12(1):4-16. [doi: 10.1080/15252019.2011.10722187]
28. Lovejoy K, Saxton GD. Information, community, and action: how nonprofit organizations use social media. Journal of Computer Mediated Communication 2012;17:337-353. [doi: 10.1111/j.1083-6101.2012.01576.x]
29. Association of State and Territorial Health Officials. About ASTHO. 2013 URL: http://www.astho.org/about/ [accessed 2013-10-01] [WebCite Cache ID 6K3BqGtab]
30. Centers for Disease Control and Prevention. 2011. Public health resources: state health departments URL: http://www.cdc.gov/mmwr/international/relres.html [accessed 2013-10-01] [WebCite Cache ID 6K3BsPDC2]
31. Association of State and Territorial Health Officials: 2011. ASTHO profile of state public health URL: http://www.astho.org/uploadedFiles/_Publications/Files/Survey_Research/ASTHO_State_Profiles_Single%5B1%5D%20lo%20res.pdf [accessed 2013-10-01] [WebCite Cache ID 6K3Bugr2E]
32. IBM. SPSS Statistics Version 20. 2013. URL: http://www-01.ibm.com/software/analytics/spss/products/statistics/ [accessed 2013-10-01] [WebCite Cache ID 6K3By7X8d]
33. US Department of Commerce. US Census. 2012. URL: http://www.census.gov/compendia/statab/2012/tables/12s0014.pdf [accessed 2013-10-01] [WebCite Cache ID 6K3C04Wjr]

34. Dumbrell D, Steele R. Twitter and health in the Australian context: What types of information are health-related organizations tweeting? 2013 Presented at: 2013 46th Hawaii International Conference on System Sciences (HICSS); January 7-10, 2013; Wailea, Maui, HI.
35. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
36. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
37. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Communication Research 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
38. Harris JK. The network of web 2.0 connections among state health departments: new pathways for dissemination. J Public Health Manag Pract 2013;19(3):E20-E24. [doi: 10.1097/PHH.0b013e318268ae36] [Medline: 23392206]
39. Hopcroft J, Lou T, Tang J. Who will follow you back? Reciprocal relationship prediction. New York: ACM; 2011 Presented at: 20th ACM International Conference on Information and Knowledge Management; 24-28 October 2011; United Kingdom. [doi: 10.1145/2063576.2063740]
40. Valente TW. Social Networks and Health: Models, Methods, and Applications. New York: Oxford University Press; 2010.
41. Hofer M, Aubert V. Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees. Computers in Human Behavior 2013 Nov;29(6):2134-2142. [doi: 10.1016/j.chb.2013.04.038]
42. Centers for Disease Control and Prevention. Public health system and the 10 essential public health services. 2013 URL: http://www.cdc.gov/nphpsp/essentialservices.html [accessed 2013-10-01] [WebCite Cache ID 6K3C2L8Ar]
43. Vicari S. Public reasoning around social contention: A case study of Twitter use in the Italian mobilization for global change. Current Sociology 2013 Apr 17;61(4):474-490. [doi: 10.1177/0011392113479747]
44. Gleason B. #Occupy Wall Street: Exploring informal learning about a social movement on Twitter. American Behavioral Scientist 2013 Mar 15;57(7):966-982. [doi: 10.1177/0002764213479372]
45. Guo C, Saxton GD. Tweeting social change: How social media are changing nonprofit advocacy. Nonprofit and Voluntary Sector Quarterly 2013 Jan 08:1. [doi: 10.1177/0899764012471585]
46. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
47. Lovejoy K, Waters RD, Saxton GD. Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. Public Relations Review 2012 Jun;38(2):313-318. [doi: 10.1016/j.pubrev.2012.01.005]

Abbreviations

API: application programming interface
ASHTO: Association of State and Territorial Health Officers

Edited by G Eysenbach; submitted 02.10.13; peer-reviewed by G Saxton, K Lovejoy; comments to author 23.10.13; revised version received 01.11.13; accepted 01.11.13; published 11.11.13.

Please cite as:
Thackeray R, Neiger BL, Burton SH, Thackeray CR
Analysis of the Purpose of State Health Departments' Tweets: Information Sharing, Engagement, and Action
J Med Internet Res 2013;15(11):e255
URL: http://www.jmir.org/2013/11/e255/
doi:10.2196/jmir.3002
PMID:24217361

Review

The Role of Social Media in Online Weight Management: Systematic Review

Tammy Chang[1], MD, MS, MPH; Vineet Chopra[1], MD, MSc; Catherine Zhang[1]; Susan J Woolford[1], MD, MPH

University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Tammy Chang, MD, MS, MPH
University of Michigan
1018 Fuller Street
Ann Arbor, MI, 48104-1213
United States
Phone: 1 7349987120
Fax: 1 7349987335
Email: tachang@med.umich.edu

Abstract

Background: Social media applications are promising adjuncts to online weight management interventions through facilitating education, engagement, and peer support. However, the precise impact of social media on weight management is unclear.

Objective: The objective of this study was to systematically describe the use and impact of social media in online weight management interventions.

Methods: PubMed, PsycINFO, EMBASE, Web of Science, and Scopus were searched for English-language studies published through March 25, 2013. Additional studies were identified by searching bibliographies of electronically retrieved articles. Randomized controlled trials of online weight management interventions that included a social media component for individuals of all ages were selected. Studies were evaluated using 2 systematic scales to assess risk of bias and study quality.

Results: Of 517 citations identified, 20 studies met eligibility criteria. All study participants were adults. Because the included studies varied greatly in study design and reported outcomes, meta-analysis of interventions was not attempted. Although message boards and chat rooms were the most common social media component included, their effect on weight outcomes was not reported in most studies. Only one study measured the isolated effect of social media. It found greater engagement of participants, but no difference in weight-related outcomes. In all, 65% of studies were of high quality; 15% of studies were at low risk of bias.

Conclusions: Despite the widespread use of social media, few studies have quantified the effect of social media in online weight management interventions; thus, its impact is still unknown. Although social media may play a role in retaining and engaging participants, studies that are designed to measure its effect are needed to understand whether and how social media may meaningfully improve weight management.

(J Med Internet Res 2013;15(11):e262) doi:10.2196/jmir.2852

KEYWORDS

Internet; systematic review; overweight; obesity; social media; weight loss

Introduction

Obesity is a major US public health problem that is associated with lower quality of life, stigma, medical complications, and higher health care costs [1-6]. Despite a decade of public awareness and attention, the prevalence of obesity continues to rise in some groups, a trend that reflects the complex nature of this disease and the diverse medical, social, and behavioral domains that underlie its management [7].

Over one-half of adults in the United States use social media platforms, such as Facebook, Twitter, MySpace, and LinkedIn [8]. The social support and feelings of interconnectedness individuals experience with social media help explain the prolific growth of these platforms in everyday life [9,10]. These domains are also relevant to the success of online weight-management interventions. Social media may represent a promising resource in combating obesity at a population level. Several properties of social media make it ideal for such purposes: (1) social media facilitates asynchronous communication and provides 24/7 access to participants; (2) it

overcomes barriers such as transportation and distance, allowing those with mobility, speech, or hearing problems to interact in online interventions; and (3) given the relative anonymity to discuss sensitive topics, social media is ideally suited for stigmatizing conditions such as obesity. However, despite these qualities, the precise implementation, effect, and benefit of social media in online weight-management interventions remains unknown.

For these reasons, we conducted a systematic review of the literature to understand whether and how online weight-management interventions have used social media to improve weight-related outcomes, such as weight loss, diet, and physical activity.

Methods

Data Sources and Search Terms

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations in conducting this systematic review [11]. With the assistance of a research librarian with experience in social media, we searched PubMed, PsycINFO, and EMBASE for articles written in English that reported outcomes associated with the use of social media in online interventions for weight management. Because the Medical Subject Heading (MeSH) term "social media" was not created until 2012, we developed a search strategy that included the following keywords to identify social media: social media, social technology, social network, online community, wiki, YouTube, Facebook, Myspace, Flickr, Twitter, and Delicious. MeSH terms and keywords to represent weight management included obesity, overweight, weight gain, weight loss, body mass index, diet, and physical activity. The full search criteria for PubMed is presented in Multimedia Appendix 1. Additional studies were identified through hand searches of electronically retrieved articles, review articles, and from a cited reference search (Web of Science and Scopus). No limits or filters were placed on search criteria; electronic searches were last updated on March 25, 2013.

Study Selection and Definitions

Studies were included if they were (1) randomized controlled trials (RCTs); (2) published in peer-reviewed literature; (3), reported weight-related outcomes, such as body mass index (BMI) or weight, dietary intake, or physical activity; and (4) included a social media component. As defined by Kaplan et al [12], we defined social media as Web-based applications that allow individuals to interact in a virtual community by exchanging user-generated information (eg, online discussion board, online bulletin board, chat room, online community). Weight-related outcomes included measures such as BMI, body weight, percent body fat, and waist and hip circumference. We defined devices that measured the intensity of physical activity as locomotion accelerometers, whereas pedometers were defined as devices that specifically measured step count [13].

Data Extraction

Two authors (TC and SW) independently abstracted variables by using a standardized template. Abstracted data included study variables (recruitment criteria, setting), participant variables (mean age, gender, mean BMI), intervention variables (brief description of weight-management intervention, intervention duration, type of social media used), outcome variables (eg, BMI, waist circumference, physical activity level, dietary intake), and quality variables (eg, data on randomization, control group, isolation of social media component). When encountered, discrepancies were resolved by consensus during a series of face-to-face and email discussions between 2 investigators (TC and SW).

Risk of Study Bias

The risk for bias in each RCT was assessed using the Jadad scale, which incorporates study domains including randomization, blinding, and description of withdrawals and dropouts [14]. Studies that received 4 or greater out of 5 possible points on the Jadad scale were considered as being at low risk of bias whereas scores of 2 and 3 or 0 and 1 were considered to be at moderate or high-risk of bias, respectively.

In addition, because our main interest was the effect of social media on online weight interventions, study quality was also rated using methodology developed by Norman et al [15]. Based on 9 methodological characteristics, this approach specifically evaluates the impact of technology (eg, social media) on specified outcomes of interest, thus allowing for a more precise approach to measuring these types of interventions. The Norman score also includes assessment of randomization, inclusion of a control group, pre-post test design, retention, baseline group equivalence, missing data, sample size calculations, and the validity of outcome measures. Each study was given 1 point for each criterion present with a maximum score of 9. Studies that scored 7 to 9 were considered high quality, studies that scored 5 to 7 were considered of moderate quality, and scores of <5 were considered poor quality.

Data Synthesis

Because the included studies varied greatly on study design, participants, measures, outcomes, and social media components, meta-analysis of interventions was not attempted or performed.

Results

Overview

In total, 517 studies were identified by our electronic searches. Following application of eligibility criteria, 20 studies [16-35] met our inclusion criteria for analysis (Figure 1). All 20 included studies involving adult populations and were published between 2001 and 2013. Studies were conducted in various parts of the world, including the United States (n=14), Australia (n=3), Canada (n=2), and the United Kingdom (n=1). Of the included studies, one study focused only on diet [16], 5 studies only on physical activity [17-21], 12 studies on both diet and physical activity [22-33], and 2 studies on weight maintenance after weight loss [34,35] (Table 1). Please see Multimedia Appendix 2 for a table of detailed study characteristics.

Figure 1. Study flow.

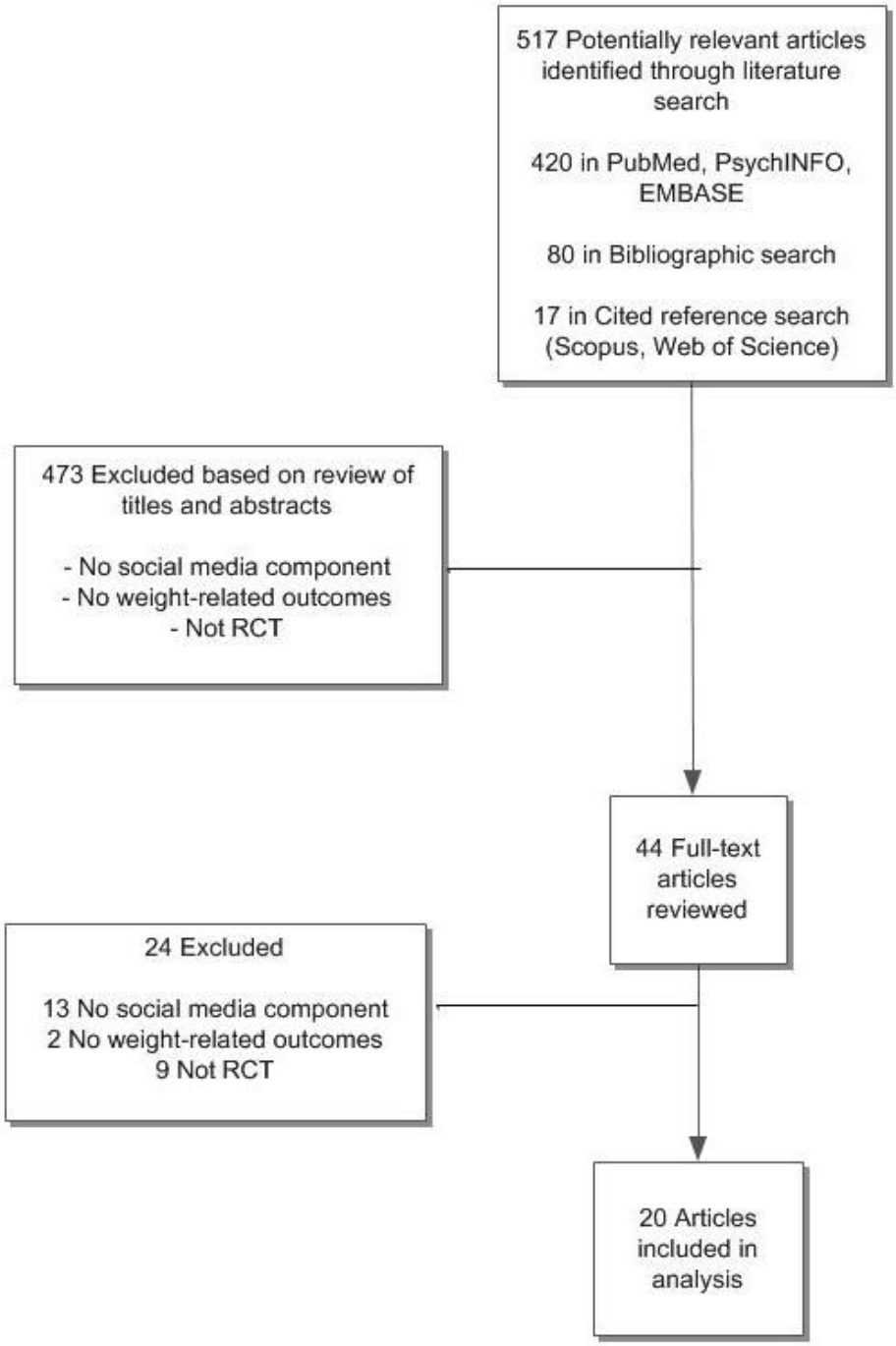


Table 1. Summary study characteristics.[a]

Source	Population					Type of social media	Primary outcomes	Results	Risk of bias (Jadad scale)	Quality score
	n	RR[b] (%)	Mean age	% Female	Mean BMI					
Diet										
Verheijden et al (2004) [16]	146	89	63	24.0	29.4	Bulletin board	Social support, BMI, waist-to-hip ratio, blood pressure, and cholesterol levels	No statistically significant differences in outcomes	Low	High
Physical activity										
Hurling et al (2007) [17]	77	100	40.4	66.2	26.3	Chat room–style message board	Change in moderate physical activity	Higher level of moderate physical activity and more percent body fat lost in the test group	Mod	High
Ferney et al (2009) [18]	106	87.7	52.1	71.7	NR	Bulletin board	Self-reported walking and physical activity	No statistically significant differences in outcomes	Mod	Mod
Liebreich et al (2009) [19]	49	89.8	54.1	59	33.9	Message board	Self-reported BMI, physical activity, and social cognitive measures	Significant improvement in total vigorous and moderate minutes of physical activity in intervention group	High	Mod
Richardson et al (2010) [20]	324	76.2	52	65	33.2	Online community with message board	Change in average daily step counts from baseline, valid days of pedometer data, and online community use	No statistically significant differences in outcomes	Mod	High
Cavallo et al (2012) [21]	134	89.6	NR	100	NR	Facebook	Self-report social support and physical activity	No statistically significant differences in outcomes	High	High
Diet and physical activity										
Tate et al (2001) [25]	91	71.4	40.9	89.0	29.0	Bulletin board	Body weight and waist circumference	Behavior therapy group lost more weight and had greater changes in waist circumference	Mod	Mod
Tate et al (2003) [26]	92	100	48.5	89.1	33.1	Message boards	Body weight, BMI, and waist circumference	Behavioral e-counseling group had greater reduction in weight, percentage of initial body weight, BMI, and waist circumference	Mod	High

Source	Population					Type of social media	Primary outcomes	Results	Risk of bias (Jadad scale)	Quality score
	n	RR[b] (%)	Mean age	% Female	Mean BMI					
Womble et al (2004) [32]	47	66.0	43.7	100	33.5	Online meetings, online bulletin board	Weight change	Manual group lost significantly more weight than the eDiets.com intervention	Mod	High
Tate et al (2006) [27]	192	80.7	49.2	84.4	32.7	Bulletin board	Weight loss, dietary intake, and physical activity	Weight losses were significantly greater in the human email-counseling group than computer-automated feedback or no counseling groups	Mod	High
Gold et al (2007) [28]	124	71.0	47.7	81.5	32.4	Discussion board, online chats/meetings	Change in body weight	VTrim group lost significantly more weight than the eDiets.com group at 6 months and maintained a greater loss at 12 months	Mod	Mod
Webber et al (2008) [22]	66	98.5	50	100	31.1	Separate message board for each website group, and online chat	Body weight	Minimal group lost more than the enhanced group	High	High
Morgan et al (2009) [29]	65	100	35.9	0	30.6	Bulletin board	Weight, waist circumference, BMI	Greater weight loss for the Internet group	Low	High
Sternfield et al (2009) [30]	787	69.8	40.1	78.1	in categories	Discussion board	Self-reported change in dietary intake and physical activity	Intervention group had increased physical activity, and increased consumption of fruits and vegetables	High	Mod
Harvey-Berino et al (2010) [31]	481	96.0	46.6	93	35.7	Chat rooms and a bulletin board	Body weight and BMI	Weight loss for InPerson was significantly greater than the Internet and Hybrid conditions	Mod	High
Turner-McGrievy et al (2011) [23]	96	89.6	42.9	75	32.5	Twitter	Body weight	No statistically significant differences in outcomes	Mod	High
Brindal et al (2012) [33]	8112	5.2	45.0	83	34.0	Social networking platform: friend networks, blogs, discussion forums, and news feeds	Body weight	No statistically significant differences in outcomes	Low	Mod

Source	Population					Type of social media	Primary outcomes	Results	Risk of bias (Jadad scale)	Quality score
	n	RR[b] (%)	Mean age	% Female	Mean BMI					
Napolitano et al (2013) [24]	52	96	20.5	86.5	31.4	Facebook	BMI	Facebook Plus group had significantly greater weight loss than Facebook and waiting list	High	High
Weight maintainence										
Harvey-Berino et al (2004) [34]	255	69	45.8	82	31.8	Chat room and bulletin board	Body weight, height, energy intake, and energy expended	No statistically significant differences in outcomes	Mod	High
Cussler et al (2008) [35]	161	69	48	100	31.1	Bulletin board and chat rooms	BMI, body fat percentage, and total body fat mass	No statistically significant differences in outcomes	High	Mod

[a]For detailed study characteristics, risk of bias, and quality scores, please see Multimedia Appendix 2.

[b]RR: response rate.

Diet Interventions

Only one study (n=146) [16] focused solely on a dietary intervention for weight management. This study tested whether Web-based nutrition counseling and a social support tool that included a bulletin board could improve weight outcomes. Low uptake of the Web-based intervention (24 of 73 participants) with limited posting on the bulletin board was reported. Messages on the bulletin board mostly contained requests for factual information directed to the research team with minimal participant interaction. The study found no significant differences between the intervention group and the usual care arm for any outcome [16].

Physical Activity Interventions

Five studies featured interventions targeting physical activity (n=690) [17-21]. These studies tested websites with a variety of other components. One study used accelerometers plus a website, one study used a pedometer plus a website, 2 interventions included only a website, and 1 used a website plus Facebook. Excluding the study that used Facebook, the social media component for all other studies in this category were message boards within the intervention website.

Only one study specifically isolated and measured the effect of the social media component, by including it in only 1 arm (online community) of the study [20]. Although this study found no difference in physical activity among the groups, the percentage of participants that completed the study and length of engagement was greater for those randomized to the social media component (ie, the online community).

Among the remaining 4 studies, 2 reported the usage of the social media component [17,18]. Within the 2 studies that reported social media use, Hurling et al [17] found that the chat room-style message board was the most frequently used component. In contrast, Ferney et al [18] reported that only 1 message was posted on their bulletin board and hypothesized that this was because of the small number of participants enrolled in the study (n=52). Although the study by Liebreich et al [19] did not report data on message board use, the authors theorized that the message board encouraged interactivity and, thus, adherence. With respect to weight outcomes, Cavallo et al [21] used Facebook as an adjunct for social support between participants and found no increased self-reported social support or physical activity. However, the remaining 3 studies showed higher levels of physical activity or greater maintenance of physical activity in participants in the intervention arms [17-19].

Diet and Physical Activity Interventions

Twelve studies featured interventions that included both diet and physical activity components (n=10,205) [22-33]. In addition to the typical online intervention and counseling, Tate et al [27] also included structured meals and meal replacements.

Most interventions in this category featured bulletin/message boards, chat rooms, or both as their social media component. Tate et al [27] created an online “ebuddy” tool that matched participants with others with similar characteristics across the country to gain support. In contrast, Turner-McGrievy et al [23] and Napolitano et al [24] used available mainstream social media, such as Twitter and Facebook, for education and to provide support to participants. Despite the use of social media in these 12 studies, no study uniquely isolated the effect of these platforms on participants; rather, the featured bulletin boards and chat rooms were embedded within a larger intervention.

Data regarding the frequency of use of the social media component were rarely reported, although when it was, use was low. For example, Tate et al [25] found that only 28% of participants ever posted a note to a bulletin board (range 1-7 postings per person) over 6 months. Examining the popularity of postings, Napolitano et al [24] found that less than one-quarter of the participants “liked” the study-related posts on Facebook.

Although the correlation between social media use and weight loss was generally positive, it was only reported in a few studies and could be because of greater adherence to the interventions overall. In the studies by Gold et al [28] and Webber et al [22], weight loss was correlated with bulletin board use in both arms. Likewise, Turner-McGrievy et al [23] found that the number of weight loss podcasts downloaded over 6 months was significantly correlated with weight loss in both arms of the study.

Although the influence of social media on weight-related measures was not specifically tested in any of these studies, findings were heterogeneous. For instance, 2 studies reported positive outcomes (greater weight loss, increased physical activity, increased consumption of fruits and vegetables, and marginally decreased sugar intake) in those randomized to interventions containing social media [29,30]. Conversely, 2 studies reported less weight loss in the study arm that included the social media component [31,32]. Other studies either had social media components in multiple arms of the study (n=7) [22-28] or showed no difference in weight outcomes (n=1) [33].

Weight Maintenance Interventions

Two studies (n=416) featured interventions focused on weight maintenance after weight loss. The social media components of the online weight maintenance interventions included both online bulletin boards and chat rooms. Overall, inclusion of social media did not result in differences in weight outcomes. In the study by Harvey-Berino et al [34], the arm with social media demonstrated no difference in perceived support compared to in-person therapy and it also had the highest rates of attrition. Interestingly, 100% of the participants within the social media arm in one study contributed to the bulletin board of the website, demonstrating high engagement with the social media component [35].

Risk of Study Bias

The median Jadad score overall was 2 out of 5 points (median 2, range 1-5) representing moderate risk for bias in the included studies (Table 1). Because many studies were unable to blind participants' and/or study coordinators' participation in social media, all but 2 studies had 2 of 5 points deducted for not describing a double-blinding process.

Using the scale developed by Norman et al [15], the median study score was 7 out of 9 (range 6-8, median 7) representing overall high study quality (Table 1). Only one study isolated social media in the design of their intervention, and 9 studies (45%) did not report a rationale for sample size. The median retention rate was 88.4% (range 5%-100%). Please see Multimedia Appendices 3 and 4 for detailed risk of bias and quality scoring data for each study.

Discussion

Principal Findings

In our systematic review of RCTs evaluating online weight-management interventions, we found that few studies implemented social media in a manner in which its impact could be measured and assessed. Therefore, the effect of social media is difficult to ascertain in the available literature. Our findings are consistent with previous systematic reviews on Internet-based behavioral interventions and electronic peer-to-peer support group interventions, which have found that the effect of the technology being studied was not isolated; thus, their effectiveness is not known [36-40]. Nevertheless, we found that contemporary studies continue to include online support-based behavioral interventions for weight management despite little evidence of their effectiveness.

However, some salient points emerged from the only study in our review that isolated its social media component from a broader intervention [20]. This study found no differences in physical activity outcomes between participants who had access to social media versus those who did not. Among those in the social media arms, greater use of the social media component was associated with improved weight-related outcomes. Therefore, for some people, social media components may be effective in promoting behavior change. Whether it would be effective just for those who are inclined to use it, or whether it would work broadly if one could encourage a wider group of participants to use it, is unknown. However, it appears that social media may fill a gap for some participants. Specifically, this study found that those with less baseline social supports (ie, family, friends) were more likely to use the social media component and that greater use of the social media component among this group was associated with lower dropout rates. This finding is consistent with other studies that suggested that use of social networking sites helped to satisfy the need for social support and connectedness [9,10].

We also observed that social media was incorporated into online interventions largely through the use of discussion boards and chat rooms. Mainstream social media platforms (eg, Facebook, Twitter) were used in only 15% of studies and mainly in more recent publications (2011-2013). This may indicate a move from program-specific, investigator-developed interventions to those that capitalize on media that participants already frequent. Furthermore, the extent of actual social media use in these studies was inconsistently reported and when reported, use was mostly low.

Why has social media not had as much uptake in weight-based interventions compared to other areas of life? One reason for this disparity may be the artificial nature of the types of social media (discussion boards and chat rooms) used on websites developed for weight management. The majority of current mainstream social media use relies on sophisticated, user-friendly, vibrant platforms that incorporate a rich, pleasing, graphical environment allowing for instantaneous transfer of information to a large community of users. Conversely, the components designed for weight-management studies may not have the same usability, access, or appeal. Furthermore, although the majority of Americans associate social media with positive terms such as good, great, fun, interesting, and convenient, the use of social media for weight management may diminish these positive feelings by associating its use with a health-specific and sensitive condition: weight management [41].

Studies often reported that social media components were included to encourage support from other participants and to

build community, although no study reported increased levels of social support after use of the social media components. A possible explanation relates to how social media has evolved over the years. Social media began as virtual communities and computer-mediated communication, which was based on the assumption that people participating would be using these platforms to connect with new people who shared similar interests or life experiences [42]. Current social networking sites can be distinguished from these early virtual communities by the fact that they are primary used for the conversion and maintenance of existing relationships into online ones [41,43]. Therefore, social support through social media platforms currently being employed by online interventions may simply be hampered as a result of this stranger phenomenon, a hypothesis supported by the fact that 57% of Americans explicitly report that they do not use social media to make new acquaintances [41]. One plausible strategy to overcome this weakness may be to supplement online interventions with face-to-face interventions. Incorporating this traditional way to cultivate relationships with the use of online social media is more in-line with how social media is used today.

Limitations

Our systematic review has some limitations. First, outcomes varied within the included studies so that studies could not be analyzed together or compared with one another. Second, most studies did not isolate the unique impact of social media on weight outcomes; thus, the role of social media in these interventions remains unknown. Third, risk of bias and study quality varied considerably within the included studies. Fourth, social media applications and platforms are evolving rapidly and it is possible, despite a rigorous search strategy, that studies of certain mobile devices with social media capabilities will be missed by our review. Finally, we limited our inclusion to RCTs only; other study designs may have been used to examine the use of this relatively novel technology in weight management.

Conclusions

Despite these limitations, our systematic review provides a comprehensive review of how social media is being used in online weight-management interventions to date. We found that social media is being incorporated in online weight-management interventions largely through message boards and chat rooms with unclear benefits. Although social media may play a role in retaining and engaging participants in online weight loss interventions, studies that are designed to measure the effect of social media are needed to understand whether and how social media may meaningfully improve weight management.

Acknowledgments

The authors would like to acknowledge Patricia Anderson (Senior Associate Librarian for Emerging Technologies, Health Sciences Libraries, University of Michigan, Ann Arbor, MI) for her assistance in the development of the search strategies of the medical literature. The authors would also like to thank Kyle Bevier for his work as a research assistant. Dr Chang receives salary support from the Robert Wood Johnson Foundation Clinical Scholars Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full PubMed search strategy.

[PDF File (Adobe PDF File), 3KB - jmir_v15i11e262_app1.pdf]

Multimedia Appendix 2

Detailed study characteristics.

[PDF File (Adobe PDF File), 133KB - jmir_v15i11e262_app2.pdf]

Multimedia Appendix 3

Jadad scale.

[PDF File (Adobe PDF File), 31KB - jmir_v15i11e262_app3.pdf]

Multimedia Appendix 4

Quality scores.

[PDF File (Adobe PDF File), 33KB - jmir_v15i11e262_app4.pdf]

References

1. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer- and service-specific estimates. Health Aff (Millwood) 2009;28(5):w822-w831 [FREE Full text] [doi: 10.1377/hlthaff.28.5.w822] [Medline: 19635784]
2. Woolford SJ, Gebremariam A, Clark SJ, Davis MM. Persistent gap of incremental charges for obesity as a secondary diagnosis in common pediatric hospitalizations. J Hosp Med 2009 Mar;4(3):149-156. [doi: 10.1002/jhm.388] [Medline: 19301381]
3. Woolford SJ, Gebremariam A, Clark SJ, Davis MM. Incremental hospital charges associated with obesity as a secondary diagnosis in children. Obesity (Silver Spring) 2007 Jul;15(7):1895-1901. [doi: 10.1038/oby.2007.224] [Medline: 17636109]
4. Freedman DS, Dietz WH, Srinivasan SR, Berenson GS. The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa Heart Study. Pediatrics 1999 Jun;103(6 Pt 1):1175-1182. [Medline: 10353925]
5. Tsiros MD, Olds T, Buckley JD, Grimshaw P, Brennan L, Walkley J, et al. Health-related quality of life in obese children and adolescents. Int J Obes (Lond) 2009 Apr;33(4):387-400. [doi: 10.1038/ijo.2009.42] [Medline: 19255583]
6. Hill AJ, Silver EK. Fat, friendless and unhealthy: 9-year old children's perception of body shape stereotypes. Int J Obes Relat Metab Disord 1995 Jun;19(6):423-430. [Medline: 7550528]
7. Fakhouri TH, Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity among older adults in the United States, 2007-2010. NCHS Data Brief 2012 Sep(106):1-8. [Medline: 23102091]
8. Madden M, Zickuhr K. 65% of online adults use social networking sites. Washington, DC: Pew Internet & American Life Project; 2011 Aug 26. URL: http://pewinternet.org/Reports/2011/Social-Networking-Sites.aspx [accessed 2013-07-26] [WebCite Cache ID 6IPBjkV5Y]
9. Manago AM, Taylor T, Greenfield PM. Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. Dev Psychol 2012 Mar;48(2):369-380. [doi: 10.1037/a0026338] [Medline: 22288367]
10. Köbler F, Riedl C, Vetter C, Leimeister JM, Krcmar H. Social connectedness on Facebook- An explorative study on status message usage. Proceedings of the Sixteenth Americas Conference on Information Systems, Lima, Peru, August 12-15 2010:1. [Medline: 20239982]
11. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
12. 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]
13. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S39-S49. [doi: 10.1249/MSS.0b013e3182399d8f] [Medline: 22157773]
14. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12. [Medline: 8721797]
15. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
16. Verheijden M, Bakx JC, Akkermans R, van den Hoogen H, Godwin NM, Rosser W, et al. Web-based targeted nutrition counselling and social support for patients at increased cardiovascular risk in general practice: randomized controlled trial. J Med Internet Res 2004 Dec 16;6(4):e44 [FREE Full text] [doi: 10.2196/jmir.6.4.e44] [Medline: 15631968]
17. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
18. Ferney SL, Marshall AL, Eakin EG, Owen N. Randomized trial of a neighborhood environment-focused physical activity website intervention. Prev Med 2009 Feb;48(2):144-150. [doi: 10.1016/j.ypmed.2008.10.022] [Medline: 19028519]
19. 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]
20. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
21. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
22. Webber KH, Tate DF, Michael Bowling J. A randomized comparison of two motivationally enhanced Internet behavioral weight loss programs. Behav Res Ther 2008 Sep;46(9):1090-1095. [doi: 10.1016/j.brat.2008.06.008] [Medline: 18675402]
23. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]

24. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]
25. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]
26. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
27. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
28. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: A pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]
29. Morgan PJ, Lubans DR, Collins CE, Warren JM, Callister R. The SHED-IT randomized controlled trial: evaluation of an Internet-based weight-loss program for men. Obesity (Silver Spring) 2009 Nov;17(11):2025-2032. [doi: 10.1038/oby.2009.85] [Medline: 19343018]
30. Sternfeld B, Block C, Quesenberry CP, Block TJ, Husson G, Norris JC, et al. Improving diet and physical activity with ALIVE: a worksite randomized trial. Am J Prev Med 2009 Jun;36(6):475-483. [doi: 10.1016/j.amepre.2009.01.036] [Medline: 19460655]
31. Harvey-Berino J, West D, Krukowski R, Prewitt E, VanBiervliet A, Ashikaga T, et al. Internet delivered behavioral obesity treatment. Prev Med 2010 Aug;51(2):123-128 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.018] [Medline: 20478333]
32. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018. [doi: 10.1038/oby.2004.124] [Medline: 15229342]
33. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: randomized trial. J Med Internet Res 2012;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
34. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329. [doi: 10.1038/oby.2004.40] [Medline: 14981225]
35. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. Obesity (Silver Spring) 2008 May;16(5):1052-1060. [doi: 10.1038/oby.2008.19] [Medline: 18309301]
36. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
37. 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]
38. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
39. Lau PW, Lau EY, Wong del P, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
40. van den Berg MH, Schoones JW, Vliet Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
41. Smith A. Why Americans use social media. Washington, DC: Pew Internet & American Life Project; 2011 Nov 15. URL: http://pewinternet.org/Reports/2011/Why-Americans-Use-Social-Media.aspx [accessed 2013-07-26] [WebCite Cache ID 6IPBXKOJd]
42. Wellman B, Salaff J, Dimitrova D, Garton L, Gulia M, Haythornthwaite C. Computer networks as social networks: Collaborative work, telework, and virtual community. Annu Rev Sociol 1996 Aug;22(1):213-238. [doi: 10.1146/annurev.soc.22.1.213]
43. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. J Comput-Mediat Comm 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]

Abbreviations

BMI: body mass index
MeSH: Medical Subject Heading
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: randomized controlled trial

Edited by G Eysenbach; submitted 26.07.13; peer-reviewed by L Lesser, D Clark; comments to author 10.09.13; revised version received 16.09.13; accepted 13.10.13; published 28.11.13.

Please cite as:
Chang T, Chopra V, Zhang C, Woolford SJ
The Role of Social Media in Online Weight Management: Systematic Review
J Med Internet Res 2013;15(11):e262
URL: http://www.jmir.org/2013/11/e262/
doi:10.2196/jmir.2852
PMID:24287455

Original Paper

MomMoodBooster Web-Based Intervention for Postpartum Depression: Feasibility Trial Results

Brian G Danaher[1], PhD; Jeannette Milgrom[2,3], PhD; John R Seeley[1], PhD; Scott Stuart[4], MD; Charlene Schembri[2], DClinPsy; Milagra S Tyler[1], MA; Jennifer Ericksen[2], MPsych; Whitney Lester[4], MA; Alan W Gemmill[2], PhD; Derek B Kosty[1], MS; Peter Lewinsohn[1], PhD

[1]Oregon Research Institute, Eugene, OR, United States

[2]Parent-Infant Research Institute, Heidelberg Repatriation Hospital, Heidelberg Heights, Victoria, Australia

[3]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia

[4]Depression and Clinical Research Center, University of Iowa Hospitals and Clinics, Iowa City, IA, United States

Corresponding Author:
Brian G Danaher, PhD
Oregon Research Institute
1776 Millrace Drive
Eugene, OR, 97403
United States
Phone: 1 541 484 2123 ext 2201
Fax: 1 541 484 1108
Email: briand@ori.org

Abstract

Background: Postpartum depression (PPD)—the most common complication of childbirth—is a significant and prevalent public health problem that severely disrupts family interactions and can result in serious lasting consequences to the health of women and the healthy development of infants. These consequences increase in severity when left untreated; most women with PPD do not obtain help due to a range of logistical and attitudinal barriers.

Objective: This pilot study was designed to test the feasibility, acceptability, and potential efficacy of an innovative and interactive guided Web-based intervention for postpartum depression, MomMoodBooster (MMB).

Methods: A sample of 53 women who satisfied eligibility criteria (<9 months postpartum, ≥18 years of age, home Internet access and use of personal email, Edinburgh Postnatal Depression Survey score of 12-20 or Patient Health Questionnaire score from 10-19) were invited to use the MMB program. Assessments occurred at screening/pretest, posttest (3 months following enrollment), and at 6 months follow-up.

Results: All six sessions of the program were completed by 87% (46/53) of participants. Participants were engaged with the program: visit days (mean 15.2, SD 8.7), number of visits (mean 20.1, SD 12.2), total duration of visits in hours (mean 5.1, SD 1.3), and number of sessions viewed out of six (mean 5.6, SD 1.3) all support high usage. Posttest data were collected from 89% of participants (47/53) and 6-month follow-up data were collected from 87% of participants (46/53). At pretest, 55% (29/53) of participants met PHQ-9 criteria for minor or major depression. At posttest, 90% (26/29) no longer met criteria.

Conclusions: These findings support the expanded use and additional testing of the MMB program, including its implementation in a range of clinical and public health settings.

Trial Registration: Clinicaltrials.gov NCT00942721; http://clinicaltrials.gov/ct2/show/NCT00942721 (Archived by WebCite at http://www.webcitation.org/6KjYDvYkQ).

(J Med Internet Res 2013;15(11):e242) doi:10.2196/jmir.2876

KEYWORDS

postpartum depression; Web-based intervention

Introduction

Postpartum depression has been defined to include any major or subsyndromal depression present at any time during the first year after delivery [1], and it is the most common complication of childbirth [2]. In terms of prevalence (10-20% of women) and severity, PPD lies between "baby blues" (less severe and quite common affecting 80% of women) and postpartum psychosis (more severe and less common affecting 0.1-0.2% of women) [1]. Left untreated, PPD has serious consequences [3]: for the mother, diminished well-being, feelings of failure, difficulties interacting with her infant, her family (partner's mental health, relationship problems); and for her infant, compromised cognitive and psychosocial development [4-8] and increased risk of mental health difficulties even in adolescence [9].

PPD is related to a range of biopsychosocial and cultural factors [10]. Previous episodes especially during pregnancy [11] or family history of mental health problems, low social and emotional support, drug and alcohol abuse, past or present abuse [12], and major life stressors are all major risk factors for PPD [13]. Less salient risk factors include marital relationship difficulties, low income, unemployment, and obstetric factors and complications [14]. Issues encountered during the reproductive year also increase risk for PPD, premature birth, and subsequent hospitalization of the infant (including care in a neonatal intensive care unit) [15-17]. Culturally and linguistically diverse women, especially refugees, asylum seekers, and immigrants have increased risk [18].

There are also significant economic and social costs (eg, loss of productivity for mother and father, health care costs, personal and broader social and economic costs) to the community. For example, Dagher et al [19] reported that PPD was related to increased health care services use, which translated into higher costs to providers. Research from the United Kingdom indicates that costs associated with PPD are higher in high-risk women [20]. An Australian analysis estimated the cost of PPD and anxiety in mothers delivering in 2012 to be $500 million by the time the children reach 2 years of age [21].

Given the limited empirical evidence supporting use of antidepressant medication with PPD [1,22], there is a significant need to develop effective psychosocial treatment approaches. Recent meta-analyses of psychosocial interventions for PPD concluded that they have a moderate beneficial effect [23,24]. Treatment modalities have included counseling [25], interpersonal psychotherapy (IPT) [26,27], and cognitive behavior therapy (CBT). A wealth of research supports the effectiveness of CBT interventions for depression in general [28,29], and perinatal depression specifically. About two-thirds of depressed individuals receiving CBT remit with treatment [30] and also have a reduced risk of relapse [29]. These benefits appear to accrue particularly in individuals with mild to moderate depression [31]. Milgrom and her colleagues created and conducted a series of successful trials [32,33] using a face-to-face individual PPD treatment program based on an adaptation of the Coping with Depression course by Lewinsohn [33,34] that also included elements of IPT (provided content on interpersonal relationships, an opportunity for partners to become involved and provide input, and a focus on infants).

Although recent trials [35,36] have demonstrated that, within a collaborative care model for depression, women suffering from PPD can be screened within a stepped-care treatment protocol during visits to their health care provider [36,37], the data indicate that, overall, fewer than 50% of postpartum women receive help for their depression [38-40]. Many of the available PPD treatment approaches are office-based, which reduces their practicality for new mothers. In addition, patient-level barriers to the uptake of treatment include travel requirements, childcare, stigma, feelings of failure, poor understanding of depression or what help is available, and safety concerns about using prescription medications [41-44]. Provider-level barriers that discourage physicians and medical/clinic staff from becoming more fully involved in PPD screening and treatment include their lack of knowledge and skills due to insufficient training regarding depression and mental health, their fear of liability, the dearth of mental health treatment resources and flexible referral systems, and inadequate reimbursement [45-50].

Web-based PPD treatment may reduce both patient- and provider-level barriers to treatment uptake and thus extend the reach of helpful treatments to underserved mothers suffering from depression. For example, Web-based treatments can reduce feelings of stigma because participation is relatively anonymous and can be completed in women's homes (thus avoiding travel) at times of their choosing without requiring childcare arrangements. Providers can recommend that women use an evidence-based Web-based PPD treatment program thus alleviating their concerns about training deficits and/or time required to provide treatment.

An increasing number of Web-based depression interventions have emerged [31,51-61]. The efficacy of these interventions has been demonstrated relative to control conditions in populations with elevated symptoms and, increasingly, in clinically diagnosed groups [62-66]. Face-to-face CBT has also been compared with Web-based CBT treatment. For example, Spek et al [60] found that both in-person and Web-based CBT interventions were superior to a waitlist control, that no significant differences were found between intervention modality, and that reductions in depressive symptoms were maintained at least 1 year after initiation of Web-delivered CBT. Similar results have emerged in other published comparisons [67,68]. Reviews of the available evidence [57,58,64,69] indicate that purely self-guided Web-based interventions benefit depressed individuals, but that effect sizes were enhanced when online programs were facilitated by a live coach [70]. Trained coaches have been shown to enhance the therapeutic alliance of Internet programs by providing low-intensity support [71] and increasing adherence to online mental health treatments [70,72].

Based on our review to date, there has been only one published report of the results of a Web-based depression intervention for postpartum women. In the randomized controlled trial (RCT) by O'Mahen et al based in England, 910 women with PPD symptoms (>12 on the Edinburgh Postnatal Depression Survey or EPDS [73]) were randomly assigned to either (1) a

full-featured 11-sesssion Web-based behavioral activation intervention (N=462) contextualized for PPD (NetMums) that also included access to features of the popular NetMums website/online community, special chat access to parent-supporters and specialist health visitors, or (2) a treatment as usual condition (N=448) [74,75]. Although there was notable attrition at the 15-week follow-up (61% attrition in the intervention and 64% in the control), results showed significant benefits to the intervention versus the control condition. Among completers at follow-up, there was clinically significant improvement among 61% of women in intervention versus 41% in the control.

The present study reports on the Web-based MomMoodBooster (MMB) program based on Milgrom's adaptation of the Coping With Depression Course (CWDC) [76] for postpartum depression [32,33,77] as well as an adaptation of the CWDC for Web-based delivery [78]. MMB was developed and pilot-tested by a multinational team from Oregon Research Institute (ORI), Parent-Infant Research Institute (PIRI) in Melbourne, Australia, and the Iowa Depression and Clinical Research Center (IDCRC). The Australian version of the program was localized for spelling (eg, it was rebranded to MumMoodBooster), word choices, and selected videos. We described the formative research foundation for this MMB in a previous paper [79]. This report describes the outcome results of a feasibility trial of the MMB program.

Methods

Participants and Procedures

Participants (N=53) were recruited from two different research sites (n=27 from our US site in Iowa and n=26 from our Australia site in greater Melbourne). Prospective participants were identified via birth records, nurse/health professional referrals, online advertisements, and news stories to local university and hospital settings.

Upon receipt of a referral or direct contact from a prospective participant, each woman was contacted by a member of the research team to explain the study and obtain informed consent for participation. During the initial contact, a preliminary check of eligibility criteria was conducted. Preliminary screening criteria included <9 months postpartum, ≥18 years of age, home Internet access and use of personal email, and an EPDS score [73] from 12-20 or a Personal Health Questionnaire (PHQ-9) score [80] from 10-19. These ranges were chosen to identify women with mild to moderately severe depression. Women satisfying initial eligibility criteria were then mailed a Participant Information and Consent Form for their signature.

Women meeting initial screening criteria then completed a phone-administered Structured Clinical Interview for DSM-IV Disorders (SCID) [81,82] and the Hamilton Rating Scale for Depression (HRSD) [83-86] to evaluate the following exclusion criteria: current diagnosis of substance abuse, bipolar disorder or psychotic depression, and/or current treatment for depressive symptoms including antidepressant medication or psychotherapy. A participant's endorsement of suicidal statements on assessments or to project staff triggered a suicide risk management protocol designed to determine the presence of current plans for self-harm, resulting in an offer of assistance and exclusion from participation in the study. Women who satisfied all inclusion and exclusion criteria were invited to participate in the study and were asked to complete the pretest assessment by visiting the secure research website. Women who did not meet eligibility criteria were offered treatment through the Infant Clinic (Australian site) and/or referral to other services as appropriate (US site). The research protocol and related informed consent procedures were reviewed and approved by the Human Research Ethics Committee of Austin Health in Australia and the Institutional Review Boards of both ORI and the University of Iowa.

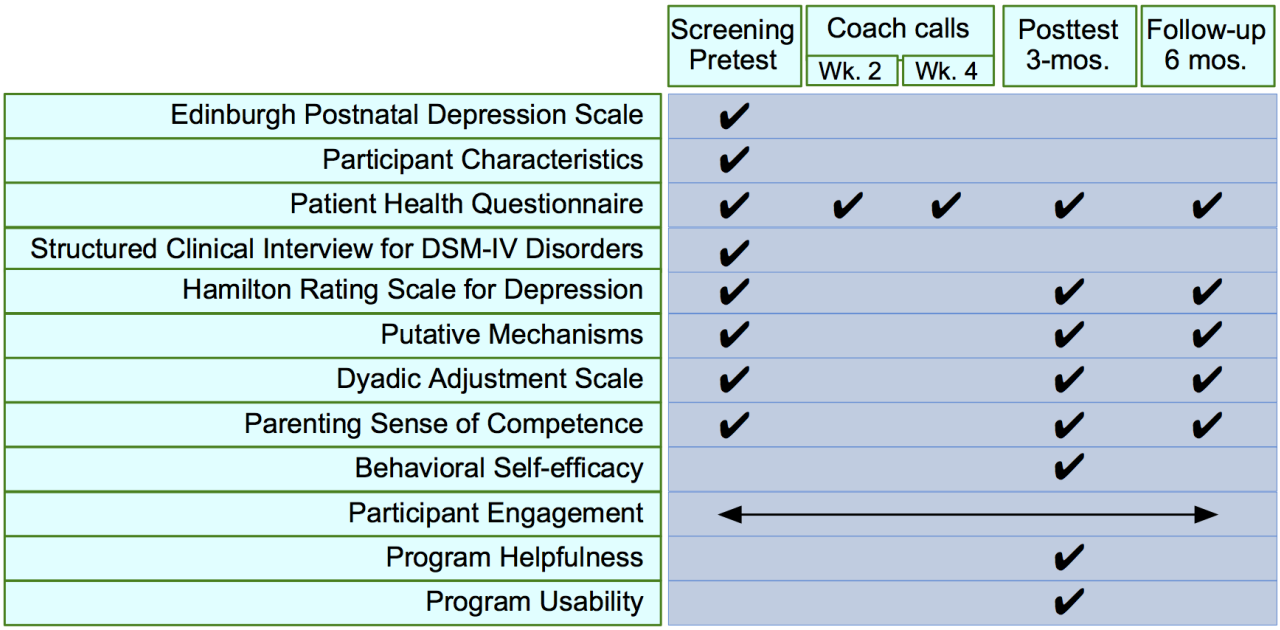
Following enrollment, participants worked through the MMB program and received weekly phone calls from a personal coach (psychologist or graduate research assistant at the Australian site or a research assistant at the US site) who encouraged participants to use the program, to practice the recommended strategies, and to report their mood levels on a PHQ-9 assessment. Every effort was made to use the same personal coach for each participant on each call. The program automatically sent email reminders to encourage participants to log into the program.

Measures

Overview

As described in Figure 1, assessments occurred at screening/pretest (corresponding to enrollment), a posttest (3 months following pretest), and follow-up (6 months following pretest). At posttest and follow-up, participants were asked to complete questionnaires both by visiting the secure website and completing another assessment by phone. By using the same phone assessor, we hoped to obtain a more sensitive measure of change. Expert phone assessors from our US research site provided assessors in our Australia site with systematic training (videoconferencing and reliability training using audio test cases) in the use of the SCID and HRSD. All phone-based assessments were recorded and reviewed.

Figure 1. Measures by assessment point.



Structured Clinical Interview for DSM-IV Disorders

Trained diagnostic interviewers conducted phone-based SCID interviews [81,82]. In order to minimize respondent burden, we used SCID Modules A-F, but we did not include the Somatoform, Eating Disorders, and Adjustment Disorder Modules.

Hamilton Rating Scale for Depression

Interviewers also administered the HRSD [27,83-86] by phone. Scoring is based on the sum of 24 items. The maximum overall score for the HRSD-24 is 69. For the current study, Cronbach alpha=.76.

Edinburgh Postnatal Depression Scale

During screening, participants were asked to complete the EPDS, a brief, simple self-rated, 10-item measure developed to screen for symptoms of postpartum depression [38,87]. Responses are rated from 0 to 3 and summed to yield the score with a maximum overall score of 30.

Participant Characteristics

We measured maternal age, delivery date and gestation, parity, education, history of previous treatment for depression, and household income.

Patient Health Questionnaire

Participants were asked to complete six separate PHQ-9 assessments from pretest, posttest, and follow-up. Personal coaches administered the PHQ-9 during phone calls that corresponded to Sessions 3 and 5 of the MMB program. These serial PHQ-9 assessments were used for program evaluation, to provide participants with a useful assessment of their status, and as an important safety check of participant status [80,88-90]. PHQ-9 scores showing a 5-point or greater escalation from pretest triggered a safety protocol, as did endorsement of the PHQ-9 suicidality item. The maximum overall score for the PHQ-9 is 27. For the current study, Cronbach alpha=.76. To evaluate the clinical significance of the intervention effects, we calculated the minimal clinically important difference (MCID [91]) based on Lowe et al [89], which represents a reduction in the PHQ-9 score of 5 points or greater. Thus, pretest-posttest changes on the PHQ-9 of ≥5 points represented a clinically important difference. For the current study, Cronbach alpha=.76.

Automatic Thoughts Questionnaire

Participants were asked to indicate how frequently over the previous week they had negative thoughts using the 30-item Automatic Thoughts Questionnaire (ATQ) [92,93] (eg, “My life is a mess”). Value options range from 0 to 4 (0=Not at all to 4=All of the time) with a maximum score of 120. For the current study, Cronbach alpha=.92.

Behavioral Activation for Depression Scale

We used the 25-item Behavioral Activation for Depression Scale (BADS) to measure changes in activation, avoidance/rumination, work/school impairment, and social impairment (eg, “I stayed in bed for too long even though I had things to do”) [94]. Value options range from 0 to 6 with a maximum score of 150. For the current study, Cronbach alpha=.83.

Dyadic Adjustment Scale

We assessed women’s relationships with their partners using the Dyadic Adjustment Scale-7 (DAS-7) [95], an abbreviated version of the Dyadic Adjustment Scale [96]. The general satisfaction score was calculated as the sum of all scores (maximum score=36). For the current study, Cronbach alpha=.85

Parenting Sense of Competence

We included the Parenting Sense of Competence (PSOC) efficacy scale [97] that asks the participant to describe her extent of agreement with 7 items designed to assess whether she is knowledgeable and competent in being a mother [98] (eg, “I honestly believe I have all the skills necessary to be a good mother to my baby”). Value options ranged from 1 to 6

(1=Strongly Disagree to 6=Strongly Agree) with a maximum score of 42. For the current study, Cronbach alpha=.90.

Behavioral Self-Efficacy

Based on the work of Bandura [99] and Maciejewski et al [100], we used 8 items to assess participant self-efficacy or confidence in being able to work with the program to reduce feelings of depression at pretest and posttest. The question asked was, "During the past week, including today, how confident are you in your ability to… (1) increase your daily pleasant activities?; (2) control your negative thinking?; (3) increase your positive thinking?; (4) get support when you need it?; (5) keep track of your mood?; (6) reduce tension using relaxation?; (7) set realistic goals for yourself?; and (8) manage your mood?" Value options ranged from 1 to 5 (1=Not At All Confident to 5=Very Confident). Self-efficacy score was computed as the mean across 8 items. For the current study, Cronbach alpha=.88.

Website Metrics

We used industry-standard website analytic tools and planned database flags recommended by Peterson [101] to track visit patterns including the date/time for each webpage viewed, which enabled us to unobtrusively measure visit frequency and duration. We also considered ways that participants were able to initiate interactions with the program (see Table 1) that shared similar characteristics, as in initiate interaction only (eg, play a video or tutorial), enter personal data into an activity (eg, typed in reasons into a list, completed a drag and drop activity, completed online activities as part of recommended homework), and personalized features of the program (eg, set goals for daily pleasant activities, updated tracking of mood and activities, uploaded personal pictures).

Personal Coach Call Metrics

Personal coaches also tracked the number and duration of calls with participants. After each call, personal coaches provided an impression of their working alliance with the participant (response options: 1=minimal, 2=partial, 3=good, 4=excellent) and the level of distraction during the call (response option: 1=none/limited, 2=some, 3=a lot).

Program Helpfulness (Self-Report by Phone at Posttest)

We used open-ended items to ask participants to identify aspects of the program that were most helpful and least helpful. We also asked participants if they would recommend the program to other depressed postpartum women.

Program Usability (Self-Report Online at Posttest)

We obtained a quantitative measure of usability by asking participants to complete our adapted version of the System Usability Scale (SUS) [102,103], a 10-item scale that asked the participant to rate the degree to which she agreed (1=Strongly Disagree to 5=Strongly Agree) with positive and negative descriptions of a Web-based program (eg, "I think that I would like to use this website frequently") [79]. The maximum score (indicating maximum usability) is 100. For the current study, Cronbach alpha=.80.

Table 1. Participant engagement activities in MomMoodBooster.

Activity	Function	Examples
List activities	Encouraged creation of personal lists to gain insight into their situation.	Lists of my pleasant activities, list of supporters, my reasons for wanting to feel better, my contributing factors, my high-tension situations, my warning signs.
Expand-collapse activities	Enabled exploration of additional detail on topics of interest.	FAQs, Myths & Facts, etc.
Drag & drop activity (see Figure 3)	Provided an interactive experience to more clearly distinguish between topics.	Activity focusing on the difference between extreme thoughts and everyday concerns.
Goal setting activity (see Figure 4)	Interactive series of steps to encourage selection of goals.	Activity designed to help the participant to choose (1) the number of pleasant activities to accomplish each day, and (2) which strategies to work on once the program had concluded.
Practice change activities	Homework tasks that were to be accomplished by each participant in their normal routine, the results of which could be shared with the personal coach.	Noticing and identifying a downward spiral, what started it and what happened; practice relaxation, making the most from pleasant activities by anticipating and savoring activities.
Online behavior tracking	Online tools used to capture participant data over time designed to encourage self-monitoring, to illuminate patterns, and to show progress.	Daily tracking of mood ratings and pleasant activities accomplished. These tracked data were also charted online.
Testimonial videos	Streaming videos of coping models who overcome barriers in order to make changes recommended in the program.	Other women's experiences; asking for help, not worrying, doing more fun activities, mood patterns, or managing stress.
Animated tutorials (see Figure 2)	Animations used to provide an explanation for underlying models for change.	Tutorials showed downward mood spirals and how they can be interrupted at critical choice points.
Personalizing pictures	Enabled participants to personalize the appearance of the program, to make it feel like "their own" website.	Women could add 10 pictures of their choice to personalize the webpages of the MMB program.

Figure 2. Animated tutorial engagement activity.

Figure 3. Drag & Drop engagement activity.

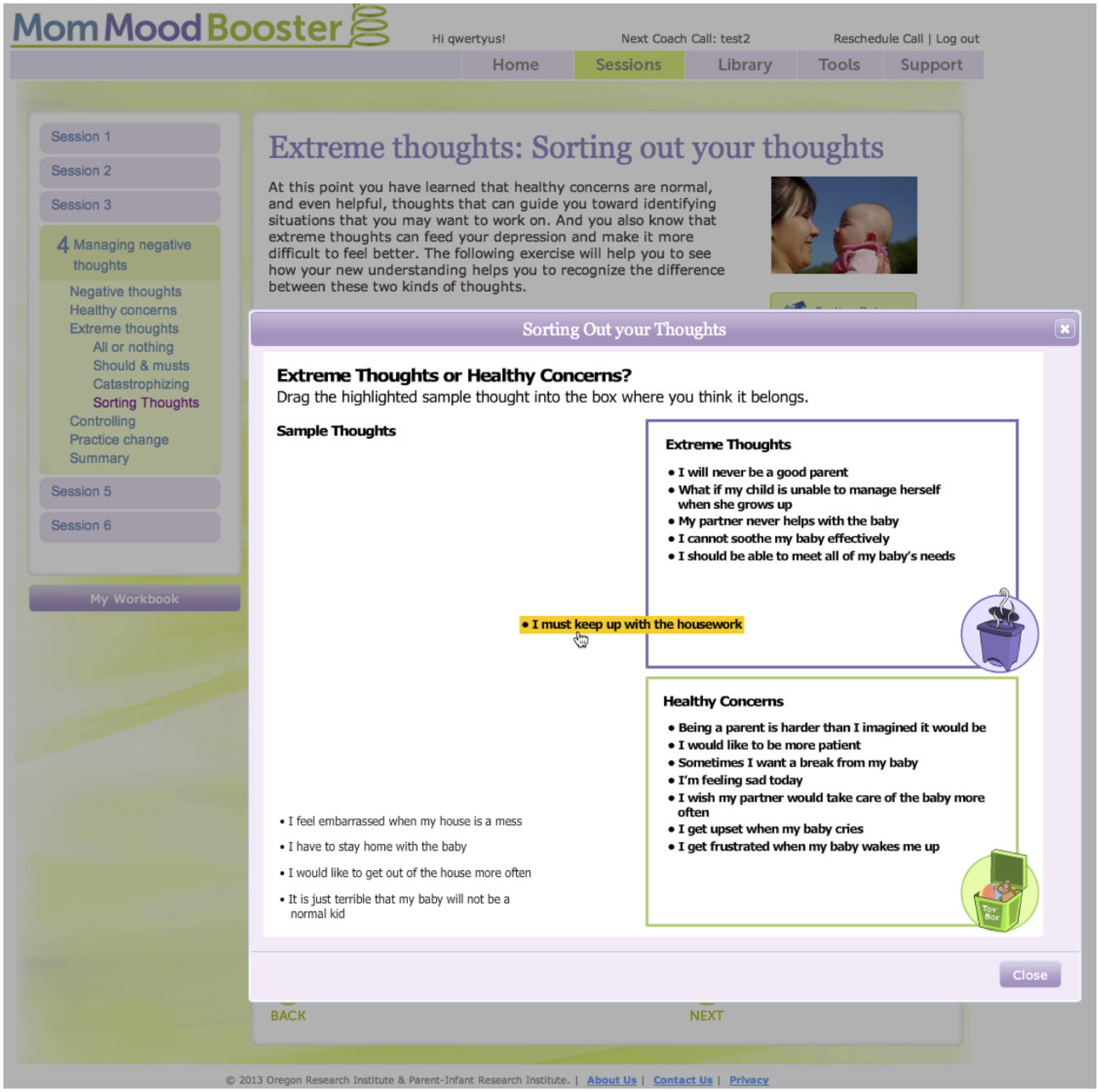


Figure 4. Goalsetting engagement activity.

MMB Intervention

Web-Based Content

We developed the MMB program using an iterative formative research process that included focus groups and usability testing [79]. A detailed description of this process and a schematic depiction of the MMB program is available in our earlier publication [79]. The MMB program also includes three complementary websites: (1) a personal coach portal to enable coaches to review the progress each participant has made going through the program, (2) a simplified Partner Support website designed to provide participant partners with information about PPD and an overview of MMB, and (3) an administrative website that enabled research staff to monitor completion of the assessments and other elements of the research project. The MMB program was designed to be fully scalable and to run on PC and Mac computers using various current browsers without plugins or applications. Java scripting, HTML+CSS, and Dynamic HTML were used to deliver interactive content.

The program consisted of the following six sequential sessions with each successive session becoming available weekly. Sessions were as follows: (1) Getting Started, (2) Managing Mood, (3) Increasing Pleasant Activities, (4) Managing Negative Thoughts, (5) Increasing Positive Thoughts, and (6) Planning for the Future. Each session opened with an autoplay host video that introduces the session goals. Webpages delivered content using text, programmed interactions, animations, and videos to present program content. Tunnel architecture [104] was used to guide participants through the six sessions. While each successive program session could be accessed weekly, the schedule was flexible in that participants could take an additional week to complete any session. Coach calls corresponded to each program week. Thus, it was possible for the MMB program to be completed in 6-12 weeks with 6-12 coach calls.

The program includes a number of features designed to encourage participant engagement and behavior change (see Table 1 and Figures 2-4). For example, each day the program encouraged participants to enter ratings of their mood and to note the number of pleasant activities they engaged in. They were also able to type in personal lists, view videos and animations, and access a library of relevant articles on communication skills, getting support, managing stress, managing time, solving problems, sleep and caring for baby, baby's needs, and relationship with partner.

Because social isolation and stigma are common in this population, MMB includes a private peer-based Web forum in which mothers can post messages as well as read and interact with the messages of other program participants. Finally, the

program could be used by participants to send an email invitation to their partner encouraging them to visit a separate MMB informational website that described postpartum depression, the MMB program, and the important role of partners play [105,106].

As research shows, receiving email reminders can help to encourage greater adherence with Web-based interventions [31], MMB participants were sent automated email reminders to encourage their engagement with the Web-based program as well as to prompt them to complete the online assessments. Participants were able to access the online program for 6+ months following enrollment.

Personal Coach Calls

The entire program was facilitated by a series of phone calls with a personal coach. All coaches were graduate research assistants or research psychologists who had received training in the content of the MMB program, their roles as coaches versus therapists/counselors, and in their data collection responsibilities. Coach training started with a guided tour of the MMB program and its coach portal that summarized participant use of the features of each session. This was followed by a videoconference that included review of the coach manual that contained detailed scripts for each call, the coach data collection responsibilities, and a discussion of the role of the coach and the rationale for making calls (ie, to provide a human voice behind the automated program, to help each participant problem solve possible barriers to using the program, and to encourage program use). All coach calls were audiorecorded, and a subset was selected to monitor fidelity of implementation as well as reliability of phone-administered assessments.

Statistical Analysis

Changes in PHQ-9 scores across time were evaluated using an unconditional growth model nesting repeated measures within individuals. This multilevel model includes time as the only predictor, coded as the number of weeks since the pretest assessment, and allows the number and spacing of measurement occasions to vary across persons [107]. A self-efficacy score was computed as the mean across 8 items. Pretest to 3-month posttest and pretest to 6-month follow-up comparisons on the ATQ, BADS, DAS-7, PSOC, and self-efficacy were evaluated using paired samples *t* tests.

All analyses involved an intent-to-treat approach whereby missing data were addressed in one of two ways recommended by Schafer and Graham [108]. We used a model-based maximum likelihood procedure in the analysis of PHQ-9 data in which parameter estimates were computed based on all available raw data. We used detailed data on participant engagement as person-level predictors of the linear and quadratic slope parameters specified in the unconditional PHQ-9 growth model. We also used the multiple imputation procedure in SPSS version 21 to account for missing data for our analysis of the ATQ, BADS, DAS-7, PSOC, and self-efficacy outcomes. Our multiple imputation procedure was fully conditional and used the iterative Markov chain Monte Carlo method to generate 20 complete datasets using all outcomes across time as predictors of missing values. The imputation model for each variable was a linear regression and included a constant term and main effects of predictor variables. Paired samples *t* tests were conducted for each measure across each of the 20 imputed datasets and reported pooled estimates in the results. To supplement tests of statistical significance, we computed partial point-biserial *r* as a measure of effect size in accordance with Rosenthal [109]. Partial point-biserial *r* was defined as $\sqrt{(t^2 / (t^2 + df))}$; small effect size=0.14, medium effect size=0.36, and large effect size=0.51 [110].

Results

Participant Characteristics and Study Attrition

Of the women who started the study, two were withdrawn because of concerns regarding self-harm and one woman withdrew of her own volition because she reported that she was feeling better and no longer wanted to be in the study. The resulting sample of 53 study participants had a mean age of 31.9 years (SD 5.1), a mean of 39.1 weeks (SD 2.4) gestation when their baby was born, and a mean number of 2.0 (SD 1.1) children. Mean baby age at the pretest was 5.5 months (SD 2.9). Participants were relatively well educated (14/53, 26% reported having graduate or postgraduate degrees) and 59% (31/53) reported annual family income of at least $60,000. Based on the SCID at screening, 49% (26/53) met criteria for DSM-IV major depressive disorder. Pretest characteristics of participants are presented in Table 2.

Of the remaining participants, 87% (46/53) completed all six sessions in the program. Posttest data were collected from 89% (47/53) on all key measures with the exception of the HRSD (45/53, 85%). Follow-up data at 6 months were collected from 87% of women (46/53). The extent to which attrition threatened the external validity of the study was evaluated using contingency table analyses and *t* tests. Overall attrition was 13% (7/53) from pretest to 6-month follow-up. Attrition was not associated with demographic characteristics or pretest values on the outcome measures. Given the minimal rates of missing data and the low likelihood of bias due to attrition, maximum likelihood estimation and multiple imputation procedures were appropriate for modeling potential intervention effects. Note that imputation was used to handle both types of missing data (ie, fully missing and “present” but with partial data).

Table 2. Selected participant characteristics at pretest (N=53).

Characteristics		n	%
Baby's gender			
	Male	28	53
	Female	25	47
Pregnancy was a multiple birth			
	Yes	2	4
	No	50	94
	No answer	1	2
Marital status			
	Married	43	81
	Widowed	1	2
	Divorced	1	2
	Separated	2	4
	Single	6	11
Education			
	< High school	5	9
	High school	6	11
	GED/certificate level	5	9
	Associates degree/advanced diploma	2	4
	Bachelor degree	17	32
	Master/graduate degree	5	9
	Doctoral/postgraduate degree	9	17
	Other	3	6
	No answer	1	2
Annual family income			
	Up to $20,000	4	8
	$20,001-$40,000	9	17
	$40,001-$60,000	5	9
	$60,001-$80,000	14	26
	>$80,000	17	32
	No answer	4	8

Primary Depression Outcomes

Patient Health Questionnaire Scores

As shown in Table 3, PHQ-9 scores decreased from pretest (mean 12.6, SD 4.1) to posttest (mean 5.0, SD 4.4) and the 6-month follow-up (mean 4.2, SD 3.9). Changes from pretest were statistically significant (P<.001) with large effects at posttest (partial r=.77) and 6-month follow-up (partial r=.82). In terms of clinical significance, at pretest, 55% (29/53) participants met PHQ-9 criteria for minor or major depression. At posttest, 90% (26/29) no longer met these PHQ-9 criteria. Results also indicated that 77% (36/47) of the participants experienced a minimal clinically important difference (ie, ≥5 point decrease) in their PHQ-9 depression scores from pretest to posttest.

Figure 5 depicts the observed and model-implied trajectory of PHQ-9 scores from pretest through the 6-month follow-up. A visual inspection of the data and a likelihood ratio (LR) test suggested that including a linear and quadratic growth parameter resulted in significantly better fit compared to a linear-only model (LR statistic with 2 degrees of freedom=36.79, P<.001). The statistical model that included linear and quadratic growth (-2 log-likelihood=1429.06, Akaike information criterion=1437.06, Bayesian information criterion=1451.25) implied an average pretest PHQ-9 score of 11.49 (SE 0.48), which decreased over time indicating a significant improvement in participant depression. Specifically, the model revealed a significant initial linear decrease (estimate=-0.79, SE 0.07, P<.001, partial r=.61) that significantly decelerated over time (estimate=0.02, SE 0.002, P<.001, partial r=.57). We also tested

for differential trajectories in PHQ-9 scores between US and Australian participants by adding the main effect of region and the time by region interactions to the unconditional growth model described earlier. None of these parameters were statistically significant (P>.50), suggesting similar PHQ-9 trajectories between US and Australian participants.

HRSD Scores

As noted in Table 3, HRSD scores also decreased from pretest (mean 16.9, SD 6.9) to posttest (mean 7.0, SD 5.6) and the 6-month follow-up (mean 6.6, SD 6.8). Changes from pretest were statistically significant (P<.001) with large effects at posttest (partial r=.75) and 6-month follow-up (partial r=.71).

Table 3. Outcome results (mean is pooled mean; SD is average standard deviation across 20 imputed datasets).

Measure	Pretest	Posttest (3 mos.)	Pretest compared to posttest			Follow-up (6 mos.)	Pretest compared to follow-up		
	Mean (SD)	Mean (SD)	*t* (df=52)	*P*	Partial *r*	Mean (SD)	*t* (df=52)	*P*	Partial *r*
PHQ-9[a,c]	12.6 (4.1)	5.0 (4.4)	8.66	<.001	.77	4.2 (3.9)	10.43	<.001	.82
HRSD[a,d]	16.9 (6.9)	7.0 (5.6)	8.28	<.001	.75	6.6 (6.8)	7.28	<.001	.71
ATQ[a,e]	23.7 (12.0)	11.2 (10.7)	6.29	<.001	.66	10.8 (13.9)	4.95	<.001	.57
BADS[b,f]	78.4 (18.4)	103.9 (19.3)	-8.73	<.001	.77	105.6 (22.3)	-7.13	<.001	.70
PSOC[b,g]	2.9 (1.1)	3.6 (1.0)	-5.63	<.001	.62	4.0 (1.0)	-5.44	<.001	.60
Self-efficacy[b]	1.6 (0.7)	2.4 (0.8)	-4.32	<.001	.51	2.6 (1.0)	-6.61	<.001	.68
DAS[b,h]	22.0 (6.8)	22.5 (7.1)	-0.40	.689	.06	24.0 (8.8)	-1.78	.077	.24

[a]Lower score is better.

[b]Higher score is better.

[c]PHQ-9—Patient Health Questionnaire.

[d]HRSD—Hamilton Rating Scale for Depression.

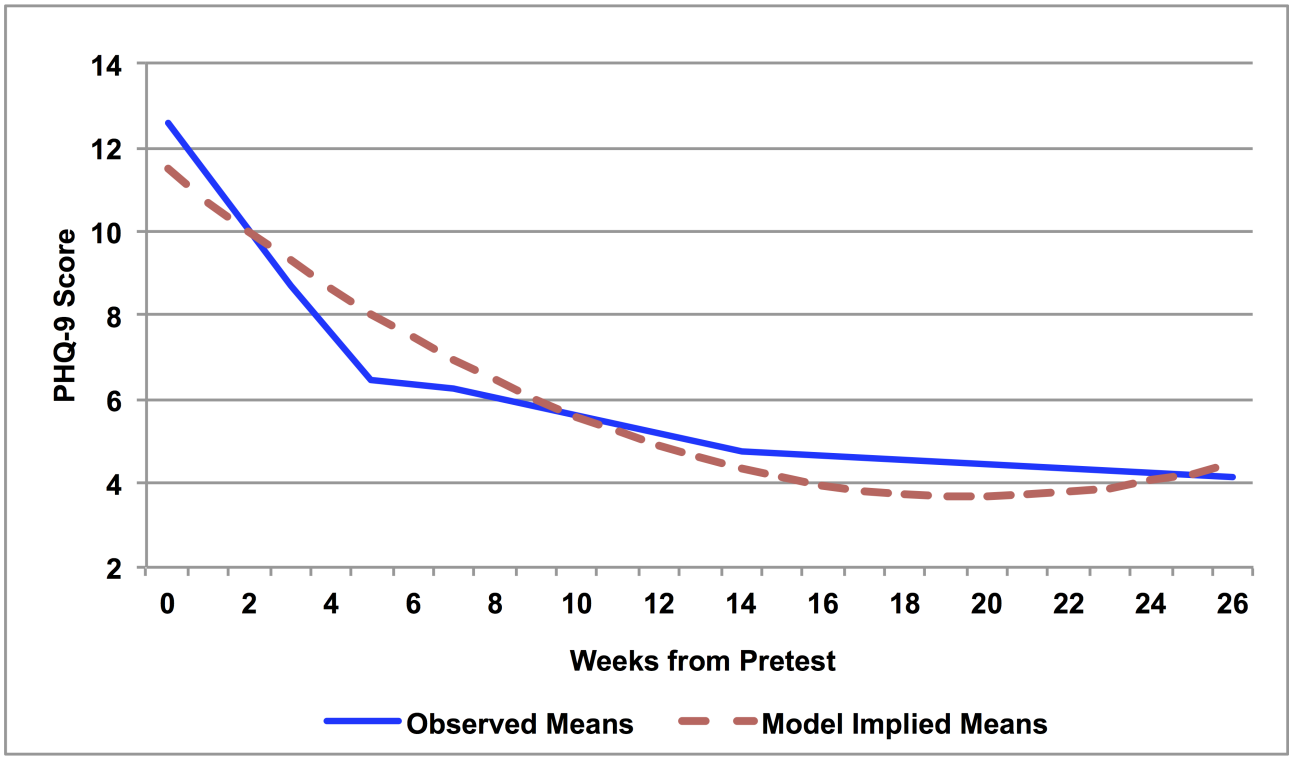
[e]ATQ—Automatic Thoughts Questionnaire.

[f]BADS—Behavioral Activation for Depression Scale.

[g]PSOC—Parenting Sense of Competence Scale.

[h]DAS—Dyadic Adjustment Scale.

Figure 5. Longitudinal trajectories of the Patient Health Questionnaire-9 scores from pretest through 6-month follow-up.



Secondary Outcomes and Putative Mechanisms of Change

Table 3 also provides descriptive statistics and the pretest to posttest and pretest to 6-month follow-up comparisons on the ATQ, BADS, DAS-7, PSOC, and self-efficacy measures. Statistically significant and large effects from pretest to posttest were obtained for the ATQ (partial *r*=.66), BADS (partial *r*=.77), PSOC (partial *r*=.62), and self-efficacy (partial *r*=.51). From pretest to 6-month follow-up, statistically significant and large effects were obtained on the ATQ (partial *r*=.57), BADS (partial *r*=.70), PSOC (partial *r*=.60), and self-efficacy (partial *r*=.68). Measures of the DAS from pretest to posttest and follow-up did not show significant change.

Website Engagement and Program Usability

Unobtrusive program use data indicated that mothers were engaged with the program: visit days (mean 15.2, SD 8.7); number of visits (mean 20.1, SD 12.2); total duration of visits in hours (mean 5.1, SD 1.3); and number of sessions viewed out of six (mean 5.6, SD 1.3). A total of 96% (51/53) of participants kept track of their daily mood ratings at least once (mean 38.1 days tracked, SD 25.6) and 92% (49/53) tracked the pleasant activities they wanted to accomplish each day (mean 29.2 days tracked, SD 23.7). The website forum provided the opportunity to post and view content; 38% (19/53) of mothers (mean 1.5, SD 2.9) posted forum content, and 74% (39/53) (mean 30.2, SD 26.8) viewed content. In addition, the MMB partner support website was accessed by 34% (18/53) of the participants' partners. Results on the System Usability Scale [111,112] administered at posttest provide a quantitative measure of program ease of use. The mean System Usability Scale score was 84.4 (SD 11.6, range 52.5-100), which translates to a usability grade of "A" for the MMB program.

We explored associations between engagement and trajectories of PHQ-9 scores within a series of conditional growth models. These models included program engagement indicators as a composite measure as well as separately. Overall, the statistical models implied that program engagement was associated with additional decreases in PHQ-9 from pretest through follow-up. Specifically, the composite measure of program engagement was significantly related to the PHQ-9 linear slope parameter (estimate=-0.24, *P*=.021, partial *r*=.37). The linear slope parameter was also significantly related to the number of visit days to the program (estimate=-0.02, *P*=.043, partial *r*=.31) and selected engagement activities (see Table 1): the proportion of list activities completed (estimate=-0.02, *P*=.006, partial *r*=.43) and the proportion of activities engaged in that involved entering personal data (estimate=-0.01, *P*=.008, partial *r*=.44). Interestingly, the overall duration of program use was not significantly associated with the trajectories of the PHQ-9 scores.

Personal Coach Calls

A total of 98% (52/53) of women agreed to receive personal coach calls. Coaches made a mean of 5.65 calls to each assigned participant (N=52; SD 1.58, Min=1; Max=9). Mean total contact duration per participant summed over all calls was 96.99 minutes (N=52; SD 49.21; Min=6.10; Max=212.07). Personal coaches reported, on average, that they had a good working alliance with participants (mean 3.07, SD 0.44) and reported low levels of distraction during the calls (mean 1.26, SD 0.35).

Participant Satisfaction

At posttest, participants reported being quite satisfied with MMB features (mean 3.3, SD 0.4 on a 4-point scale: Not at all satisfied to Very satisfied), and they rated personal coach calls as being helpful (mean 3.4, SD 0.9 on 4-point scale: Not at all helpful to Very helpful). Responses to open-ended questions about satisfaction are noted in Table 4.

Use of Other Programs

At posttest we also asked participants "Since you enrolled in the MomMoodBooster program 3 months ago, which of the following products or programs have you used to manage your mood?". A total of 30 out of 48 participants reported as follows: 12 (25.0%) read self-help books, 7 (14.6%) took medication for depression, 6 (12.5%) participated in an individual treatment program, 3 (6.3%) used hypnosis or acupuncture, 1 (2.1%) participated in a group treatment program, 1 (2.1%) participated in another Internet treatment program, and 30 (62.5%) participants indicated they had not participated in any other programs/products. Use of other programs for mood management was not significantly associated with the trajectories of the PHQ-9 scores.

Table 4. Participant comments on program satisfaction.

Question	Comments
Q1: In what ways did you find the Mum/MomMoodBooster program most helpful?	Support by phone, private time to do it
	Forced me to think about myself, focus on positive thinking helpful overall
	Valuable reassurance especially as can't get out
	Found it helpful in that feel more equipped to manage mood and emotions - online format is great as it allows easy access no matter what time of day
	Info was fine seemed very slow—sense of obligation was helpful—threat of phone coach calling forced to think about improving mood; to do list kind of person
	Phone calls to help keep you on track and tracking mood and activities so you can identify patterns
	Like how tasks were broken down into steps—strategies felt like they were achievable
	Gave permission to not have focus 100% be on the baby—to do something for self
Q2: In what ways did you find the personal coach calls to be helpful?	Really good at normalizing situations—also the flexibility of the coach (if baby cries, etc) was reassuring
	Reaffirmed things in the course, someone to talk to, to make sure you're on track —not isolated—good that someone was going to call—something to look forward to—someone was going to ask you how you're doing with program—motivated to do program—sharing
	Help me remember to log in
	"Personal" feeling rather than website but content nothing new/warm
	Felt someone was caring
	Calls tie the whole program together act as a "check-in" for how feeling, review the materials from session
	Makes you accountable—keep going with session—would be easy to leave it for next week if no coach calls—helpful to talk through the information and clarify certain points

Discussion

Principal Findings

Pilot study results described in this report—when combined with results of our formative research [79]—provide comprehensive evidence supporting MomMoodBooster, an innovative Web-based intervention for postpartum depression. Pilot study participants, a clinical sample of 53 women recruited from the United States and Australia, were very engaged with the MomMoodBooster program: 87% completed the 6-month follow-up assessment, they viewed an average of 5.6 out of the 6 sessions, spent an average of more than 5 hours using the program, and spent an average of more than 95 minutes on personal coach calls. Their average number of 20.1 program visits compares quite favorably to results reported for many other Web-based depression interventions [113]. Participants also reported positive ratings regarding program usability, which was mirrored in their favorable ratings and comments regarding the program, including coach calls.

It is important to note that the relationship between participant engagement in the program and depression outcomes warrants further analysis as our measure of program use duration was not significantly associated with improvement in depression as measured by trajectories of the PHQ-9 scores. Elsewhere [114] we have recommended that there may not be simple dose:response relationships between engagement and outcome and that composite measures incorporating several dimensions of program usage need to be explored in this regard.

Over the course of the program, participants showed significant improvements on clinician-rated HRSD and self-reported PHQ-9 assessments, and they sustained those improvements over the 6-month follow-up period. Fully 77% reported experiencing clinically important improvement in their PHQ-9 scores. Putative mechanisms of change showed corresponding improvements.

Our 13% participant attrition is slightly higher than what has been reported for telephone-delivered therapies [113], and notably lower than the 25% to 50% reported in face-to-face psychotherapy and the sizable attrition rates reported in self-help Internet interventions [115]. Importantly, attrition was much lower than the nearly 60% attrition rate reported in a published paper on a Web-based intervention with women with PPD [74] and a third of that reported by Milgrom in her group-based CBT intervention for postpartum depression [33], the treatment approach embodied in the MMB program.

We believe that the highly encouraging results for participants using the MMB program were associated with three factors: (1) our adaptation and contextualization to PPD of Lewinsohn's Coping with Depression Course, as embodied in Milgrom's work, (2) MMB's online engagement activities that encouraged participants to be actively involved, to spend time, and to follow treatment recommendations, and (3) personal coach calls that provided a key element of supportive accountability, which encouraged engagement and follow-through.

There are several study limitations that should also be noted. For example, we used a quasi-experimental design without a

randomized controlled condition, thus we were not able to control for potential threats to the internal validity such as biases due to selection or maturation effects. In addition, our relatively small sample size may limit the generalizability of the study findings. We also recruited a convenience sample, which may not be representative of depressed postpartum women, generally. In addition, participants were relatively well educated and had a relatively high socioeconomic status. Finally, we did not assess the maintenance of the treatment benefits beyond 6 months.

Next Steps

We agree with the conclusion expressed by Lewis et al [116]: "Given the time, cost, and childcare constraints of traditional interventions for postpartum depression, evaluations of new and innovative interventions are needed." Based upon the promising results of our pilot study, we believe that the Web-based MomMoodBooster program represents just such an innovative treatment option. Next steps worthy of consideration include additional research. For example, controlled research is needed to evaluate MMB compared to alternative approaches when implemented within extant treatment programs based in real-world settings, such as in telephone-administered treatment programs [117-119], depression care management programs [120], nurse home visitations to pregnant and postpartum women [121], and in depression treatment provided in physician offices [35,36]. MMB would seem to be particularly appropriate within a stepped-care model as it could offer a low-cost, high-reach option as a preliminary treatment step [36,37] and/or in conjunction with other, more intensive "high-touch" treatments. Additional research might also examine the role of the personal coach. For example, rather than using research staff as coaches, it would be useful to test the use of endogenous providers as coaches. And since the cost and feasibility of providing 6 scheduled personal coaching calls may limit implementation opportunities, additional research might consider ways to provide fewer coach calls or provide a stepped-care approach that would tailor calls to the expressed interests of the recipient.

It would also be helpful to determine how program content might be adapted and delivered to reach low-income and minority postpartum depressed women by accommodating cultural differences [119], learning styles, and preferences in terms of tools/platforms to access program content (eg, use of smartphones is closing the "digital divide" [122,123]).

Finally, MMB could be expanded to include content on antenatal depression and/or content to enhance mother:infant interactions, two under-recognized and often untreated problems [45, 46] that have profound effects on maternal and infant well-being and health. In addition to being a risk factor for PPD, antenatal depression is related to more frequent pre-eclampsia [124], preterm birth [125], low birth-weight [126], and adverse obstetric outcomes [127]. It also diminishes capacity for maternal self-care as it can be accompanied by inadequate nutrition, drug and alcohol abuse, and poor prenatal clinic attendance, all of which can further compromise the health of mother and baby [128,129]. Because research shows that treating postpartum depression does not improve poor mother:infant interactions, which results in risk to maternal and infant well-being and health [1,130,131], then additional program content might be included in MMB in order to address this important area.

Acknowledgments

The MomMoodBooster feasibility pilot study was supported by grant 5R01-MH084931 from the National Institute of Mental Health of the National Institutes of Health (Principal Investigators: Drs Brian G Danaher, Jeannette Milgrom, and Scott Stuart). MMB technology development was accomplished with assistance from Steve Christensen, Tom Jacobs, Ethan Sletteland, and Toan Tran from InterVision Media (Eugene, OR) and Timothy Woolley from IEQ Technology (Springfield, OR). We also acknowledge the important contributions of ORI colleagues: Edward Lichtenstein, Ryann Crowley, Coleen Hudkins, and Katie Clawson. Cartoons that enhanced the MomMoodBooster library articles and the Partner Support Program were drawn by Bev Aisbett.

Conflicts of Interest

None declared.

References

1. O'Hara MW, McCabe JE. Postpartum depression: current status and future directions. Annu Rev Clin Psychol 2013;9:379-407. [doi: 10.1146/annurev-clinpsy-050212-185612] [Medline: 23394227]
2. Beck CT. State of the Science on postpartum depression: what nurse researchers have contributed--part 1. MCN Am J Matern Child Nurs 2008;33(2):121-126. [doi: 10.1097/01.NMC.0000313421.97236.cf] [Medline: 18327112]
3. Field T, Diego M, Hernandez-Reif M. Prenatal depression effects and interventions: a review. Infant Behav Dev 2010 Dec;33(4):409-418 [FREE Full text] [doi: 10.1016/j.infbeh.2010.04.005] [Medline: 20471091]
4. Beck CT. The effects of postpartum depression on child development: a meta-analysis. Arch Psychiatr Nurs 1998 Feb;12(1):12-20. [Medline: 9489170]
5. Milgrom J, Westley D, Gemmill AW. The mediating role of maternal responsiveness in some longer term effects of postnatal depression on infant development. Infant Behavior and Development 2004 Dec;27(4):443-454. [doi: 10.1016/j.infbeh.2004.03.003]

6. Milgrom J, Ericksen J, McCarthy R, Gemmill AW. Stressful impact of depression on early mother–infant relations. Stress and Health 2006 Oct;22(4):229-238. [doi: 10.1002/smi.1101]
7. Kaplan PS, Danko CM, Diaz A, Kalinka CJ. An associative learning deficit in 1-year-old infants of depressed mothers: role of depression duration. Infant Behav Dev 2011 Feb;34(1):35-44. [doi: 10.1016/j.infbeh.2010.07.014] [Medline: 21071090]
8. Goodman SH, Gotlib IH. Risk for psychopathology in the children of depressed mothers: a developmental model for understanding mechanisms of transmission. Psychol Rev 1999 Jul;106(3):458-490. [Medline: 10467895]
9. Pawlby S, Hay DF, Sharp D, Waters CS, O'Keane V. Antenatal depression predicts depression in adolescent offspring: prospective longitudinal community-based study. J Affect Disord 2009 Mar;113(3):236-243. [doi: 10.1016/j.jad.2008.05.018] [Medline: 18602698]
10. Buultjens M, Murphy G, Robinson PH, Milgrom J. The perinatal period: A literature review from the biopsychosocial perspective. Clinical Nursing Studies 2013;1(3):19-31. [doi: 10.5430/cns.v1n3p19]
11. Leigh B, Milgrom J. Risk factors for antenatal depression, postnatal depression and parenting stress. BMC Psychiatry 2008;8:24 [FREE Full text] [doi: 10.1186/1471-244X-8-24] [Medline: 18412979]
12. Howard LM, Oram S, Galley H, Trevillion K, Feder G. Domestic violence and perinatal mental disorders: a systematic review and meta-analysis. PLoS Med 2013;10(5):e1001452 [FREE Full text] [doi: 10.1371/journal.pmed.1001452] [Medline: 23723741]
13. Milgrom J, Gemmill AW, Bilszta JL, Hayes B, Barnett B, Brooks J, et al. Antenatal risk factors for postnatal depression: a large prospective study. J Affect Disord 2008 May;108(1-2):147-157. [doi: 10.1016/j.jad.2007.10.014] [Medline: 18067974]
14. Robertson E, Grace S, Wallington T, Stewart DE. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-295. [doi: 10.1016/j.genhosppsych.2004.02.006] [Medline: 15234824]
15. Bergström E, Wallin L, Thomson G, Flacking R. Postpartum depression in mothers of infants cared for in a Neonatal Intensive Care Unit – Incidence and associated factors. Journal of Neonatal Nursing 2012 Aug;18(4):143-151. [doi: 10.1016/j.jnn.2011.11.001]
16. Vigod SN, Villegas L, Dennis CL, Ross LE. Prevalence and risk factors for postpartum depression among women with preterm and low-birth-weight infants: a systematic review. BJOG 2010 Apr;117(5):540-550. [doi: 10.1111/j.1471-0528.2009.02493.x] [Medline: 20121831]
17. Centers for Disease ControlPrevention (CDC). Prevalence of self-reported postpartum depressive symptoms--17 states, 2004-2005. MMWR Morb Mortal Wkly Rep 2008 Apr 11;57(14):361-366 [FREE Full text] [Medline: 18401329]
18. Collins CH, Zimmerman C, Howard LM. Refugee, asylum seeker, immigrant women and postnatal depression: rates and risk factors. Arch Womens Ment Health 2011 Feb;14(1):3-11. [doi: 10.1007/s00737-010-0198-7] [Medline: 21153849]
19. Dagher RK, McGovern PM, Dowd BE, Gjerdingen DK. Postpartum depression and health services expenditures among employed women. J Occup Environ Med 2012 Feb;54(2):210-215. [doi: 10.1097/JOM.0b013e31823fdf85] [Medline: 22267187]
20. Petrou S, Cooper P, Murray L, Davidson LL. Economic costs of post-natal depression in a high-risk British cohort. Br J Psychiatry 2002 Dec;181:505-512 [FREE Full text] [Medline: 12456521]
21. Beyondblue. Valuing perinatal health: The consequences of not treating perinatal depression and anxiety. 2012. URL: http://www.beyondblue.org.au/docs/default-source/8.-perinatal-documents/bw0079-report-valuing-perintal-health.pdf?sfvrsn=2 [accessed 2013-10-24] [WebCite Cache ID 6KmYOXRB3]
22. Howard LM, Hoffbrand S, Henshaw C, Boath L, Bradley E. Antidepressant prevention of postnatal depression. Cochrane Database Syst Rev 2005(2):CD004363. [doi: 10.1002/14651858.CD004363.pub2] [Medline: 15846711]
23. Cuijpers P, Brännmark JG, van Straten A. Psychological treatment of postpartum depression: a meta-analysis. J Clin Psychol 2008 Jan;64(1):103-118. [doi: 10.1002/jclp.20432] [Medline: 18161036]
24. Dennis CL, Hodnett E. Psychosocial and psychological interventions for treating postpartum depression. Cochrane Database Syst Rev 2007(4):CD006116. [doi: 10.1002/14651858.CD006116.pub2] [Medline: 17943888]
25. MacArthur C, Winter HR, Bick DE, Lilford RJ, Lancashire RJ, Knowles H, et al. Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women's physical and psychological health needs. Health Technol Assess 2003;7(37):1-98 [FREE Full text] [Medline: 14622490]
26. Stuart S, O'Hara MW, Gorman LL. The prevention and psychotherapeutic treatment of postpartum depression. Arch Womens Ment Health 2003 Aug;6 Suppl 2:S57-S69. [doi: 10.1007/s00737-003-0003-y] [Medline: 14615924]
27. O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. Arch Gen Psychiatry 2000 Nov;57(11):1039-1045. [Medline: 11074869]
28. Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. Arch Gen Psychiatry 1989 Nov;46(11):971-82; discussion 983. [Medline: 2684085]
29. NIMH. Breaking ground, breaking through: The strategic plan for mood disorders research of the National Institute of Mental Health. 2003. URL: http://www.nimh.nih.gov/about/strategic-planning-reports/breaking-ground-breaking-through--the-strategic-plan-for-mood-disorders-research.pdf [accessed 2013-10-24] [WebCite Cache ID 6KcLpVpzF]

30. Gloaguen V, Cottraux J, Cucherat M, Blackburn IM. A meta-analysis of the effects of cognitive therapy in depressed patients. J Affect Disord 1998 Apr;49(1):59-72. [Medline: 9574861]
31. 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]
32. Milgrom J, Martin PR, Negri LM. Treating postnatal depression: a psychological approach for health care practitioners. Chichester: Wiley; 1999.
33. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. Br J Clin Psychol 2005 Nov;44(Pt 4):529-542. [doi: 10.1348/014466505X34200] [Medline: 16368032]
34. Cuijpers P, Muñoz RF, Clarke GN, Lewinsohn PM. Psychoeducational treatment and prevention of depression: the "Coping with Depression" course thirty years later. Clin Psychol Rev 2009 Jul;29(5):449-458. [doi: 10.1016/j.cpr.2009.04.005] [Medline: 19450912]
35. Yawn BP, Dietrich AJ, Wollan P, Bertram S, Graham D, Huff J, TRIPPD practices. TRIPPD: a practice-based network effectiveness study of postpartum depression screening and management. Ann Fam Med 2012 Jul;10(4):320-329 [FREE Full text] [doi: 10.1370/afm.1418] [Medline: 22778120]
36. Gjerdingen D, Katon W, Rich DE. Stepped care treatment of postpartum depression: a primary care-based management model. Womens Health Issues 2008 Jan;18(1):44-52. [doi: 10.1016/j.whi.2007.09.001] [Medline: 18215764]
37. Pollack LO, Segre LS, O'Hara MW, Stuart S. Postpartum depression: the distribution of severity in a community sample. Arch Womens Ment Health 2011 Aug;14(4):363-364 [FREE Full text] [doi: 10.1007/s00737-011-0228-0] [Medline: 21713455]
38. Cox J, Holden J. Perinatal Mental Health: A Guide to the Edinburgh Postnatal Depression Scale. London: The Royal College of Psychiatrists; 2003.
39. Austin MP, Frilingos M, Lumley J, Hadzi-Pavlovic D, Roncolato W, Acland S, et al. Brief antenatal cognitive behaviour therapy group intervention for the prevention of postnatal depression and anxiety: a randomised controlled trial. J Affect Disord 2008 Jan;105(1-3):35-44. [doi: 10.1016/j.jad.2007.04.001] [Medline: 17490753]
40. Watt S, Sword W, Krueger P, Sheehan D, Ontario Mother & Infant Survey. A cross-sectional study of early identification of postpartum depression: implications for primary care providers from The Ontario Mother & Infant Survey. BMC Fam Pract 2002 Apr 11;3:5 [FREE Full text] [Medline: 11950393]
41. Gentile S. Use of contemporary antidepressants during breastfeeding: a proposal for a specific safety index. Drug Saf 2007;30(2):107-121. [Medline: 17253877]
42. Bilszta J, Ericksen J, Buist A, Milgrom J. Women's experience of postnatal depression - beliefs and attitudes as barriers to care. Aust J Adv Nurs 2010;27(3):44-54.
43. Murray L, Woolgar M, Murray J, Cooper P. Self-exclusion from health care in women at high risk for postpartum depression. Journal of Public Health 2003 Jun 01;25(2):131-137. [doi: 10.1093/pubmed/fdg028]
44. Byatt N, Biebel K, Friedman L, Debordes-Jackson G, Ziedonis D. Women's perspectives on postpartum depression screening in pediatric settings: a preliminary study. Arch Womens Ment Health 2013 Oct;16(5):429-432. [doi: 10.1007/s00737-013-0369-4] [Medline: 23812739]
45. Byatt N, Biebel K, Debordes-Jackson G, Lundquist RS, Moore Simas TA, Weinreb L, et al. Community mental health provider reluctance to provide pharmacotherapy may be a barrier to addressing perinatal depression: a preliminary study. Psychiatr Q 2013 Jun;84(2):169-174. [doi: 10.1007/s11126-012-9236-0] [Medline: 22941573]
46. Byatt N, Biebel K, Friedman L, Debordes-Jackson G, Ziedonis D, Pbert L. Patient's views on depression care in obstetric settings: how do they compare to the views of perinatal health care professionals? Gen Hosp Psychiatry 2013 Aug 19:436-449. [doi: 10.1016/j.genhosppsych.2013.07.011] [Medline: 23969144]
47. Byatt N, Simas TA, Lundquist RS, Johnson JV, Ziedonis DM. Strategies for improving perinatal depression treatment in North American outpatient obstetric settings. J Psychosom Obstet Gynaecol 2012 Dec;33(4):143-161. [doi: 10.3109/0167482X.2012.728649] [Medline: 23194018]
48. Olson AL, Kemper KJ, Kelleher KJ, Hammond CS, Zuckerman BS, Dietrich AJ. Primary care pediatricians' roles and perceived responsibilities in the identification and management of maternal depression. Pediatrics 2002 Dec;110(6):1169-1176. [Medline: 12456915]
49. Chaudron LH, Szilagyi PG, Campbell AT, Mounts KO, McInerny TK. Legal and ethical considerations: risks and benefits of postpartum depression screening at well-child visits. Pediatrics 2007 Jan;119(1):123-128 [FREE Full text] [doi: 10.1542/peds.2006-2122] [Medline: 17200279]
50. Leiferman JA, Dauber SE, Scott K, Heisler K, Paulson JF. Predictors of maternal depression management among primary care physicians. Depress Res Treat 2010;2010:671279 [FREE Full text] [doi: 10.1155/2010/671279] [Medline: 21152221]
51. 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]

52. 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]
53. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. J Med Internet Res 2002 Dec;4(3):E14 [FREE Full text] [doi: 10.2196/jmir.4.3.e14] [Medline: 12554545]
54. Galler JR, Harrison RH, Ramsey F, Forde V, Butler SC. Maternal depressive symptoms affect infant cognitive development in Barbados. J Child Psychol Psychiatry 2000 Sep;41(6):747-757. [Medline: 11039687]
55. 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]
56. 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 Jun;14(3):e68 [FREE Full text] [doi: 10.2196/jmir.1859] [Medline: 22738715]
57. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
58. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
59. Spek V, Nyklícek I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. J Affect Disord 2008 Jan;105(1-3):137-145. [doi: 10.1016/j.jad.2007.05.001] [Medline: 17543392]
60. Spek V, Cuijpers P, Nyklícek I, Smits N, Riper H, Keyzer J, et al. One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years. Psychol Med 2008 May;38(5):635-639. [doi: 10.1017/S0033291707002590] [Medline: 18205965]
61. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
62. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. Aust N Z J Psychiatry 2009 Jun;43(6):571-578. [doi: 10.1080/00048670902873722] [Medline: 19440890]
63. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
64. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
65. Wiersma JE, Titov N, Van Schaik DJ, Van Oppen P, Beekman AT, Cuijpers P, et al. Treating chronic symptoms of depression in the virtual clinic: findings on chronicity of depression in patients treated with internet-based computerized cognitive behaviour therapy for depression. Psychother Psychosom 2011;80(5):313-315. [doi: 10.1159/000323167] [Medline: 21720195]
66. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 7;192(11 Suppl):S4-11. [Medline: 20528707]
67. Hedman E, Andersson G, Andersson E, Ljótsson B, Rück C, Asmundson GJ, et al. Internet-based cognitive-behavioural therapy for severe health anxiety: randomised controlled trial. Br J Psychiatry 2011 Mar;198(3):230-236 [FREE Full text] [doi: 10.1192/bjp.bp.110.086843] [Medline: 21357882]
68. Andrews G, Davies M, Titov N. Effectiveness randomized controlled trial of face to face versus Internet cognitive behaviour therapy for social phobia. Aust N Z J Psychiatry 2011 Apr;45(4):337-340. [doi: 10.3109/00048674.2010.538840] [Medline: 21323490]
69. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
70. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
71. 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]
72. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. Drug Alcohol Depend 2009 Feb 1;100(1-2):178-181 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.09.015] [Medline: 19041197]
73. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [Medline: 3651732]

74. O'Mahen HA, Woodford J, McGinley J, Warren FC, Richards DA, Lynch TR, et al. Internet-based behavioral activation-treatment for postnatal depression (Netmums): a randomized controlled trial. J Affect Disord 2013 Sep 25;150(3):814-822. [doi: 10.1016/j.jad.2013.03.005] [Medline: 23602514]
75. O'Mahen H, Himle JA, Fedock G, Henshaw E, Flynn H. A pilot randomized controlled trial of cognitive behavioral therapy for perinatal depression adapted for women with low incomes. Depress Anxiety 2013 Jul;30(7):679-687. [doi: 10.1002/da.22050] [Medline: 23319454]
76. Lewinsohn PM. Control your depression. New York: Simon and Schuster; 1992.
77. Milgrom J, McCloud P. Parenting stress and postnatal depression. Stress Medicine 1996;12(3):177-186.
78. 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. [doi: 10.1037/a0028820] [Medline: 22663903]
79. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: Formative research and design of the MomMoodBooster Program. JMIR Res Protoc 2012 Nov;1(2):e18. [doi: 10.2196/resprot.2329] [Medline: 23612274]
80. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
81. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry 1992 Aug;49(8):624-629. [Medline: 1637252]
82. First MB. Structured clinical interview for DSM-IV axis I disorders SCID-I: clinician version, administration booklet. Washington, DC: American Psychiatric Press; 1997.
83. Hamilton M. Hamilton depression scale. In: Guy W, editor. ECDEU assessment manual for psychopharmacology. Rockville, MD: National Institute of Mental Health; 1976:179-192.
84. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967 Dec;6(4):278-296. [Medline: 6080235]
85. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [Medline: 14399272]
86. Farabaugh AH, Bitran S, Witte J, Alpert J, Chuzi S, Clain AJ, et al. Anxious depression and early changes in the HAMD-17 anxiety-somatization factor items and antidepressant treatment outcome. Int Clin Psychopharmacol 2010 Jul;25(4):214-217 [FREE Full text] [doi: 10.1097/YIC.0b013e328339fbbd] [Medline: 20400905]
87. Milgrom J, Ericksen J, Negri L, Gemmill AW. Screening for postnatal depression in routine primary care: properties of the Edinburgh Postnatal Depression Scale in an Australian sample. Aust N Z J Psychiatry 2005 Sep;39(9):833-839. [doi: 10.1111/j.1440-1614.2005.01660.x] [Medline: 16168042]
88. 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]
89. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of Affective Disorders 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8]
90. Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Arch Womens Ment Health 2012 Oct;15(5):367-374. [doi: 10.1007/s00737-012-0295-x] [Medline: 22983357]
91. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989 Dec;10(4):407-415. [Medline: 2691207]
92. Hollon SD, Kendall PC. Cognitive self-statements in depression: development of an automatic thoughts questionnaire. Cognit Ther Res 1980;4:383-395.
93. Kaufman NK, Rohde P, Seeley JR, Clarke GN, Stice E. Potential mediators of cognitive-behavioral therapy for adolescents with comorbid major depression and conduct disorder. J Consult Clin Psychol 2005 Feb;73(1):38-46. [doi: 10.1037/0022-006X.73.1.38] [Medline: 15709830]
94. Kanter JW, Mulick PS, Busch AM, Berlin KS, Martell CR. The Behavioral Activation for Depression Scale (BADS): psychometric properties and factor structure. J Psychopathol Behav Assess 2006 Oct 25;29(3):191-202. [doi: 10.1007/s10862-006-9038-5]
95. Hunsley M, Best M, Lefebvre D, Vito J. The Seven-Item Short Form of the Dyadic Adjustment Scale: further evidence for construct validity. The American Journal of Family Therapy 2001 Oct;29(4):325-335. [doi: 10.1080/01926180126501]
96. Spanier GB. Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. Journal of Marriage and the Family 1976 Feb;38(1):15-28. [doi: 10.2307/350547]
97. Johnston C, Mash EJ. A measure of parenting satisfaction and efficacy. Journal of Clinical Child Psychology 1989 Jun;18(2):167-175. [doi: 10.1207/s15374424jccp1802_8]
98. Gilmore L, Cuskelly M. Factor structure of the Parenting Sense of Competence scale using a normative sample. Child Care Health Dev 2009 Jan;35(1):48-55. [doi: 10.1111/j.1365-2214.2008.00867.x] [Medline: 18991983]

99. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
100. Maciejewski PK, Prigerson HG, Mazure CM. Self-efficacy as a mediator between stressful life events and depressive symptoms. Differences based on history of prior depression. Br J Psychiatry 2000 Apr;176:373-378 [FREE Full text] [Medline: 10827887]
101. Peterson ET. Web site measurement hacks. Sebastopol, California: O'Reilly; 2005.
102. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
103. Tullis T, Albert W. Measuring the user experience: Collecting, analyzing, and presenting usability metrics. Boston: Elsevier/Morgan Kaufmann; 2008.
104. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. J Med Internet Res 2005 May;7(2):e12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
105. Wee KY, Skouteris H, Pier C, Richardson B, Milgrom J. Correlates of ante- and postnatal depression in fathers: a systematic review. J Affect Disord 2011 May;130(3):358-377. [doi: 10.1016/j.jad.2010.06.019] [Medline: 20599275]
106. Solantaus T, Salo S. Paternal postnatal depression: fathers emerge from the wings. The Lancet 2005 Jun;365(9478):2158-2159. [doi: 10.1016/S0140-6736(05)66754-2]
107. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods, 2nd edn. Thousand Oaks, CA: Sage; 2002.
108. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods 2002;7(2):147-177. [doi: 10.1037//1082-989x.7.2.147]
109. Rosenthal R. Meta-analytic procedures for social research. Newbury Park: Sage Publications; 1991.
110. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J: L. Erlbaum Associates; 1988.
111. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
112. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Waltham, MA: Morgan Kaufmann; 2012.
113. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]
114. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39. [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
115. 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]
116. Lewis BA, Gjerdingen DK, Avery MD, Guo H, Sirard JR, Bonikowske AR, et al. Examination of a telephone-based exercise intervention for the prevention of postpartum depression: design, methodology, and baseline data from The Healthy Mom study. Contemp Clin Trials 2012 Nov;33(6):1150-1158. [doi: 10.1016/j.cct.2012.07.015] [Medline: 22890220]
117. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral Intervention Technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Jul;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
118. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 6;307(21):2278-2285. [doi: 10.1001/jama.2012.5588] [Medline: 22706833]
119. Baker-Ericzén MJ, Connelly CD, Hazen AL, Dueñas C, Landsverk JA, Horwitz SM. A collaborative care telemedicine intervention to overcome treatment barriers for Latina women with depression during the perinatal period. Fam Syst Health 2012 Sep;30(3):224-240. [doi: 10.1037/a0028750] [Medline: 22709321]
120. Logsdon MC, Foltz MP, Stein B, Usui W, Josephson A. Adapting and testing telephone-based depression care management intervention for adolescent mothers. Arch Womens Ment Health 2010 Aug;13(4):307-317 [FREE Full text] [doi: 10.1007/s00737-009-0125-y] [Medline: 20020164]
121. Izzo CV, Eckenrode JJ, Smith EG, Henderson CR, Cole R, Kitzman H, et al. Reducing the impact of uncontrollable stressful life events through a program of nurse home visitation for new parents. Prev Sci 2005 Dec;6(4):269-274. [doi: 10.1007/s11121-005-0010-5] [Medline: 16075193]
122. Lopez MH, Gonzalez-Barrera A, Patten E. Pew Hispanic Center. 2013. Closing the Digital Divide: Latinos and technology adoption URL: http://www.pewhispanic.org/files/2013/03/Latinos_Social_Media_and_Mobile_Tech_03-2013_final.pdf [accessed 2013-10-24] [WebCite Cache ID 6KcoHd5ev]
123. Smith A. Pew Research Center. 2013. Smartphone ownership: 2013 Update URL: http://www.pewinternet.org/~/media/Files/Reports/2013/PIP_Smartphone_adoption_2013.pdf [accessed 2013-10-25] [WebCite Cache ID 6KmZ8pyi2]
124. Kurki T, Hiilesmaa V, Raitasalo R, Mattila H, Ylikorkala O. Depression and anxiety in early pregnancy and risk for preeclampsia. Obstet Gynecol 2000 Apr;95(4):487-490. [Medline: 10725477]

125. Dayan J, Creveuil C, Marks MN, Conroy S, Herlicoviez M, Dreyfus M, et al. Prenatal depression, prenatal anxiety, and spontaneous preterm birth: a prospective cohort study among women with early and regular care. Psychosom Med 2006 Nov;68(6):938-946 [FREE Full text] [doi: 10.1097/01.psy.0000244025.20549.bd] [Medline: 17079701]
126. Orr ST, James SA, Blackmore Prince C. Maternal prenatal depressive symptoms and spontaneous preterm births among African-American women in Baltimore, Maryland. Am J Epidemiol 2002 Nov 1;156(9):797-802 [FREE Full text] [Medline: 12396996]
127. Chung TK, Lau TK, Yip AS, Chiu HF, Lee DT. Antepartum depressive symptomatology is associated with adverse obstetric and neonatal outcomes. Psychosom Med 2001;63(5):830-834 [FREE Full text] [Medline: 11573032]
128. Austin MP. Psychosocial assessment and management of depression and anxiety in pregnancy. Key aspects of antenatal care for general practice. Aust Fam Physician 2003 Mar;32(3):119-126. [Medline: 12666348]
129. Coverdale JH, McCullough LB, Chervenak FA, Bayer T. Clinical implications and management strategies when depression occurs during pregnancy. Aust N Z J Obstet Gynaecol 1996 Nov;36(4):424-429. [Medline: 9006826]
130. Milgrom J, Ericksen J, McCarthy R, Gemmill AW. Stressful impact of depression on early mother–infant relations. Stress and Health 2006 Oct;22(4):229-238. [doi: 10.1002/smi.1101]
131. Milgrom J, Newnham C, Anderson PJ, Doyle LW, Gemmill AW, Lee K, et al. Early sensitivity training for parents of preterm infants: impact on the developing brain. Pediatr Res 2010 Mar;67(3):330-335. [doi: 10.1203/PDR.0b013e3181cb8e2f] [Medline: 19952869]

Abbreviations

ATQ: Automatic Thoughts Questionnaire
BADS: Behavioral Activation for Depression Scale
CBT: cognitive behavioral therapy
CWDC: Coping With Depression Course
DAS-7: Dyadic Adjustment Scale (7-item version)
EPDS: Edinburgh Postnatal Depression Survey
HRSD: Hamilton Rating Scale for Depression
IDCRC: Iowa Depression and Clinical Research Center
IPT: interpersonal psychotherapy
MMB: MomMoodBooster (MumMoodBooster)
NHMRC: National Health and Medical Research Council (Australia)
ORI: Oregon Research Institute
PHQ-9: Patient Health Questionnaire (9-item version)
PIRI: Parent-Infant Research Institute (University of Melbourne)
PPD: postpartum depression
PSOC: Parenting Sense of Competence scale
SUS: System Usability Scale
SCID: Structured Clinical Interview for DSM-IV Disorders

Edited by G Eysenbach; submitted 14.08.13; peer-reviewed by L Walekr, D Linares; comments to author 19.09.13; accepted 08.10.13; published 01.11.13.

Please cite as:
Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, Ericksen J, Lester W, Gemmill AW, Kosty DB, Lewinsohn P
MomMoodBooster Web-Based Intervention for Postpartum Depression: Feasibility Trial Results
J Med Internet Res 2013;15(11):e242
URL: http://www.jmir.org/2013/11/e242/
doi:10.2196/jmir.2876
PMID:24191345

Original Paper

Comparing In-Person to Videoconference-Based Cognitive Behavioral Therapy for Mood and Anxiety Disorders: Randomized Controlled Trial

Daniel R Stubbings[1], PhD; Clare S Rees[1], PhD; Lynne D Roberts[1], PhD; Robert T Kane[1], PhD

School of Psychology and Speech Pathology, Faculty of Health Sciences, Curtin University of Technology, Perth, Australia

Corresponding Author:
Clare S Rees, PhD
School of Psychology and Speech Pathology
Faculty of Health Sciences
Curtin University of Technology
Kent Street, Bentley
Perth,
Australia
Phone: 61 8 9266 3442
Fax: 61 8 9266 3442
Email: C.Rees@curtin.edu.au

Abstract

Background: Cognitive-behavioral therapy (CBT) has demonstrated efficacy and effectiveness for treating mood and anxiety disorders. Dissemination of CBT via videoconference may help improve access to treatment.

Objective: The present study aimed to compare the effectiveness of CBT administered via videoconference to in-person therapy for a mixed diagnostic cohort.

Methods: A total of 26 primarily Caucasian clients (mean age 30 years, SD 11) who had a primary Diagnostic and Statistical Manual of Mental Disorders, 4th edition text revision (DSM-IV-TR) diagnosis of a mood or anxiety disorder were randomly assigned to receive 12 sessions of CBT either in-person or via videoconference. Treatment involved individualized CBT formulations specific to the presenting diagnosis; all sessions were provided by the same therapist. Participants were recruited through a university clinic. Symptoms of depression, anxiety, stress, and quality of life were assessed using questionnaires before, after, and 6 weeks following treatment. Secondary outcomes at posttreatment included working alliance and client satisfaction.

Results: Retention was similar across treatment conditions; there was one more client in the videoconferencing condition at posttreatment and at follow-up. Statistical analysis using multilevel mixed effects linear regression indicated a significant reduction in client symptoms across time for symptoms of depression ($P<.001$, $d=1.41$), anxiety ($P<.001$, $d=1.14$), stress ($P<.001$, $d=1.81$), and quality of life ($P<.001$, $d=1.17$). There were no significant differences between treatment conditions regarding symptoms of depression ($P=.165$, $d=0.37$), anxiety ($P=.41$, $d=0.22$), stress ($P=.15$, $d=0.38$), or quality of life ($P=.62$, $d=0.13$). There were no significant differences in client rating of the working alliance ($P=.53$, one-tailed, $d=–0.26$), therapist ratings of the working alliance ($P=.60$, one-tailed, $d=0.23$), or client ratings of satisfaction ($P=.77$, one-tailed, $d=–0.12$). Fisher's Exact P was not significant regarding differences in reliable change from pre- to posttreatment or from pretreatment to follow-up for symptoms of depression ($P=.41$, $P=.26$), anxiety ($P=.60$, $P=.99$), or quality of life ($P=.65$, $P=.99$) but was significant for symptoms of stress in favor of the videoconferencing condition ($P=.03$, $P=.035$). Difference between conditions regarding clinically significant change was also not observed from pre- to posttreatment or from pretreatment to follow-up for symptoms of depression ($P=.67$, $P=.30$), anxiety ($P=.99$, $P=.99$), stress ($P=.19$, $P=.13$), or quality of life ($P=.99$, $P=.62$).

Conclusions: The findings of this controlled trial indicate that CBT was effective in significantly reducing symptoms of depression, anxiety, and stress and increasing quality of life in both in-person and videoconferencing conditions, with no significant differences being observed between the two.

Trial Registration: Australian New Zealand Clinical Trials Registry ID: ACTRN12609000819224; http://www.anzctr.org.au/ACTRN12609000819224.aspx (Archived by WebCite at http://www.webcitation.org/6Kz5iBMiV).

(J Med Internet Res 2013;15(11):e258) doi:10.2196/jmir.2564

KEYWORDS

telepsychology; videoconferencing; cognitive behavioral therapy; anxiety; mood disorder

Introduction

Approximately 20-25% of people are likely to suffer from a mood disorder during their lifetime [1]. They are a leading cause of disability for people aged 15-44, and the direct medical costs associated with unipolar depression have been estimated to be AUD $27,237 per person per year [2]. Anxiety disorders are also common, with approximately 29% of people likely to meet the diagnostic criteria during their lifespan [1]. Cognitive behavioral therapy (CBT) is an empirically validated psychological treatment that can be used to effectively treat mood and anxiety disorders. It is a time-limited and structured therapy that aims to help clients identify and reality-test unhelpful cognitions and correct maladaptive behavior [3,4]. Meta-analysis research has indicated that CBT is effective in the treatment of disorders such as depression [5], obsessive-compulsive disorder [6], generalized anxiety disorder [7], social phobia [8], panic disorder [9], and posttraumatic stress disorder [10]. Strong evidence also exists indicating that CBT can be effectively used to treat client groups presenting with a mixed array of mood and anxiety disorders [11].

At present, there is limited access to empirically validated treatments [12], particularly in remote locations. One way of increasing access is to provide services via remote media, such as telepsychology, which is the provision of psychological services via videoconference [13-15]. There is a substantial amount of evidence supporting the use of CBT to treat mood and anxiety disorders in-person; thus, it is important to determine if these findings extend to CBT administered via videoconference. Providing mental health services via videoconference for rural and remote populations has been reported to be cost effective [16-18], have high satisfaction [19,20], be used to reduce referrals to in-patient clinics [21], and offer diagnosis results comparable outcomes to in-person [22,23]. Interaction via videoconference has also been shown to reduce depression and loneliness in elderly nursing home residents [24].

Some evidence exists indicating that manualized CBT provided via telepsychology for restrictive diagnostic profiles can be effective [25-28]. Bouchard et al [25] conducted a controlled trial comparing a manualized CBT treatment for panic and agoraphobia via videoconference (n=11) to in-person (n=10). The results indicated that 81% (9/11) of participants in the videoconferencing condition were panic-free at posttreatment and 91% (10/11) at 6-month follow-up. The differences between conditions with regards to client symptoms and quality of the working alliance were not significant. Mitchell et al [27] conducted a randomized controlled trial of CBT for the treatment of bulimia nervosa and compared the clinical outcomes obtained in-person to via videoconference. Six doctoral-level students administered therapy, and 128 participants were involved over a 4-year time period. Again, no significant differences were observed between the in-person and videoconferencing condition with regards to reduction in symptomology, therapeutic alliance, and treatment retention. Morland et al [28] conducted a study involving 125 participants with posttraumatic stress disorder randomly assigned to either in-person or videoconference-based group therapy for anger management. Similarly, the findings indicated that the clinical outcomes obtained via videoconference were not inferior to what were obtained in-person.

Some studies have been conducted that involve a mixed diagnostic client cohort. Day and Schneider [29] conducted a randomized controlled trial comparing the effectiveness of CBT when administered either in-person, via 2-way audio or via videoconference for a cohort of clients with a variety of presenting issues. All treatment sessions were conducted at the same location. The study involved 16 clinicians and 80 clients, and the most frequent presenting concerns pertained to body image, family relationships, self-esteem, and work/school issues. The intervention was effective in all conditions, and the results did not indicate any significant difference between treatment modalities. Despite these encouraging findings, the intervention included only five treatment sessions, psychiatric conditions were not diagnosed, reliable change and clinical significance were not analyzed, and a multivariate analysis was used to address a series of univariate research questions. Griffiths, Blignault, and Yellowlees [30] reported the details of a study involving 15 clients presenting with depression and/or anxiety who were treated with 6-8 sessions of CBT by their case managers. The pre-post treatment data indicated that the clients' mental health was significantly better after treatment, but there was no in-person control condition. In a more recent study involving an in-person randomized comparison condition, Dustan and Tooth [31] investigated clinical outcomes for 6 clients (3 in each condition) that presented with either an anxiety or mixed anxiety-depression disorder. Two trainee psychologists provided 6-8 sessions of CBT, and symptoms were measured pre-, post, and 1-month follow-up. All 3 clients in the videoconferencing condition showed significantly reduced symptomology at follow-up, but due to the small sample size, an analysis comparing in-person to videoconference-based outcomes could not be made. Given these limitations, further videoconference-based research is needed to extend in-person research pertaining to mood and anxiety disorders [11].

Research has indicated that when CBT is successful in reducing the symptoms of a psychological disorder, clients typically experience a concurrent increase in quality of life [32]. Changes in quality of life as a result of treatment administered via digital media have been addressed in some contexts [33], but research specifically pertaining to videoconferencing and quality of life is limited [34]. Hence further research is recommended to address this gap in the literature.

The present study aimed to investigate whether CBT administered via videoconference produces comparable clinical outcomes to treatment provided in-person. The treatment will be focused on addressing anxiety and mood disorders including participants that present with comorbid diagnosis. This study used a randomized, active control design, structured clinical

assessments, with symptoms measures administered at pretreatment, posttreatment, and at 6 weeks following treatment.

Methods

Participants

The Curtin University ethics committee approved this study and all participants provided signed consent. Participants were recruited via the Curtin University Psychology Clinic and were either self-referred or referred from community health agencies via telephone, letter, fax, or email. Participant recruitment began in January 2010 and ended in April 2011. The recruitment ended because no more time was available for this activity during the course of the degree. We recruited 29 participants but 3 did not meet the inclusion criteria. Inclusion criteria consisted of a primary diagnosis of a Diagnostic and Statistical Manual of Mental Disorders, 4th edition text revision (DSM-IV-TR) [35] Axis-I disorder, aged 18-65 years old, and living in Perth, Western Australia. Exclusion criteria included a DSM-IV-TR [35] diagnosis of anorexia, psychosis (past or present), or a personality disorder as the primary diagnosis, as well as any self-harm or suicidal behaviors currently receiving psychotherapy and/or involvement in legal proceedings. Figure 1 displays the participant flow over the course of the study. All assessments, treatment, and data collection were conducted at the university clinic.

The sample included 26 participants, aged 18-59 years (mean 30 years, SD 11 years). The videoconferencing condition included 6 males and 8 females; the in-person condition included 5 males and 7 females. An in-person clinical interview along with the Structured Clinical Interview for the DSM (SCID) [36,37] was used to screen the participants. The Interference of Severity Scale adapted from the Anxiety Disorders Interview Schedule [38] was used to determine which disorder should be the primary focus of treatment. The average number of presenting disorders in both conditions was 3, and there were only 2 clients who presented with only 1 diagnosis. The number of clients in each condition who met the criteria for various Axis-I and Axis-II disorders is displayed in Tables 1 and 2. No data were collected regarding clients' prior experience with computers and videoconferencing technology. Six participants reported that they were on antidepressants before starting the study. A record of medication usage by participants was not taken, but they were asked not to change their medication dosage during the course of the study.

Figure 1. The number of participants and their flow throughout the study.

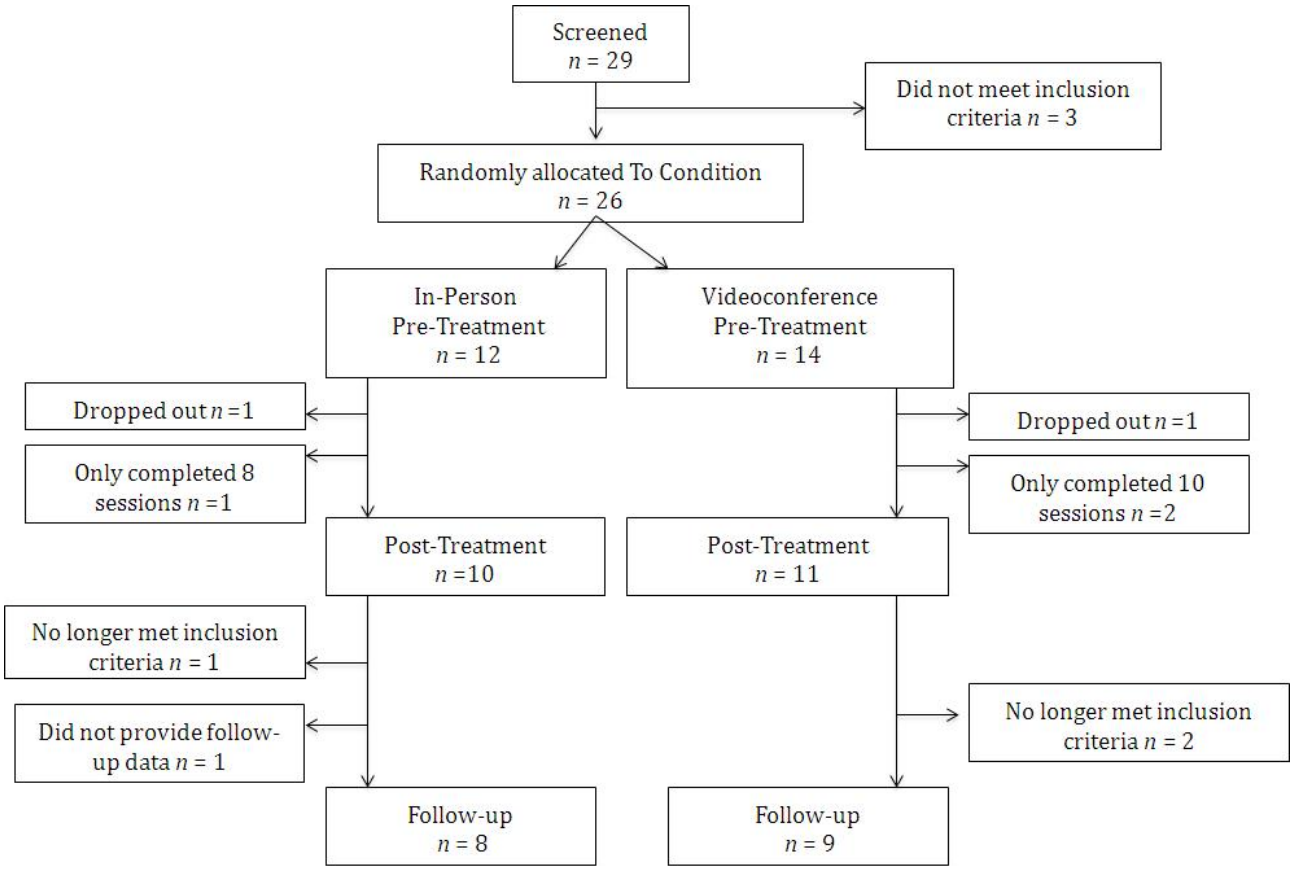


Table 1. The number of clients that met the criteria for Axis-I disorders at pretreatment.

	Primary Axis-I disorder		Comorbid Axis-I disorder	
	IP[a]	VC[b]	IP	VC
Axis-I mood disorders				
Major depressive disorder	5	1	3	4
Dysthymic disorder		1	3	3
Cyclothmic disorder				1
Axis-I anxiety disorder				
Panic with agoraphobia		1		
Panic without agoraphobia	1		2	
Social phobia		1	1	
Obsessive-compulsive disorder	4	8	1	1
Posttraumatic stress			2	
Generalized anxiety	2		3	2
Hypocondriasis		1	1	1
Other Axis-I disorders				
Adjustment disorder		1		
Eating disorder NOS (binge eating)			2	
Alcohol dependence				1
Cannabis dependence				1
Dysparenuia			1	
Impulse control disorder (not otherwise specified)				1
Attention-deficit hyperactivity			2	
Amphetamine-induced psychotic disorder				1
Substance-induced mood disorder with manic features				1

[a]IP=In-person.

[b]VC=Videoconference.

Table 2. The number of clients that met the criteria for Axis-II disorders at pretreatment.

Axis-II disorders	In-person	Videoconference
Schizotypal personality disorder		1
Borderline personality disorder	1	
Narcissistic personality disorder	1	1
Avoidant personality disorder	6	4
Obsessive-compulsive	5	9

Treatment Protocols

For each client, an individualized CBT formulation was devised. Manualized CBT interventions were used as a guide to planning and implementing treatment for specific disorders such as panic/agoraphobia [39], depression/dysthymia [3], hypochondriasis [40], generalized anxiety [41], and obsessive-compulsive disorder [42]. Other conditions were treated with standard CBT techniques such as psychoeducation, symptom monitoring, cognitive restructuring, and exposure exercises relevant to their presenting symptoms. The interventions were focused on addressing symptoms of the primary Axis-I diagnosis. Axis-II conditions were noted in the assessment sessions but were not targeted during the course of treatment.

Self-Report Measures

The Depression Anxiety and Stress Scale (DASS) [43] was used to measure global clinical symptoms. The Quality of Life Enjoyment and Satisfaction scale (QLES) [44] was used to measure changes in quality of life as a result of treatment. Given that a mixed diagnostic cohort was used, disorder specific measures for the primary presenting condition were also administered. These included the Beck Depression Inventory-II

[45], Obsessive-Compulsive Inventory [46], Health Anxiety Questionnaire [47], Penn State Worry Questionnaire [48], and the Anxiety Sensitivity Index [49]. The DASS, QLES, and disorder-specific measures were administered before treatment, after treatment, and 6 weeks after treatment had ended. The Working Alliance Inventory Short Form [50] was used to measure the strength of the working alliance from both the clients' and therapists' perspectives. Client satisfaction was measured using the shortened Client Satisfaction Questionnaire [51]. The Telehealth Satisfaction Questionnaire [52] was used to measure client satisfaction specifically with the technology. Both the Working Alliance Inventory Short Form and the Client Satisfaction Questionnaire were administered at posttreatment to all clients; the Telehealth Satisfaction Questionnaire was administered at posttreatment only to clients allocated to the videoconferencing condition. All questionnaires were completed by hand.

Apparatus

Three Mac computers were used during the course of this study. All three computers were connected via ethernet (100 Mbps/sec) within the same building. The program used to conduct the videoconferencing was iChat, versions 4.0.9 to 5.0.3. When the study was conducted, the interactions via iChat were encrypted, but this feature was removed in Mac OS X 10.7. Therefore iChat may no longer be suitable for future telepsychology research.

Procedure

The initial assessment consisted of a clinical interview, the SCID, and administration of the DASS and QLES. The therapist (primary author) who conducted the screening, diagnosis, and treatment was a provisionally registered (trainee) clinical psychologist doctoral student. Following the initial assessment, clients were randomly allocated to receive treatment either in-person or via videoconference. Simple random allocation was used to determine assignment to condition. This was achieved by generating a randomized list of binary numbers [53]. The generated list was calculated on the basis of 200 potential participants, two groups, and one repeat of randomization. Clients in the in-person condition began "treatment as usual". Clients in the videoconferencing condition were instructed to walk into the treatment room at the clinic during their allotted time and sit in front of the computer. After the initial diagnosis session, clients allocated to the videoconferencing condition had no other in-person contact with the treating therapist. Materials, such as new homework diaries, were placed in the treatment room before the client arrived. Participants in the videoconferencing condition shared the content of their homework diaries orally; therefore, they did not need to hand in material to the therapist. All of the client sessions were recorded so that the fidelity and credibility of the therapists' diagnosis and treatments could be monitored. The clinical research supervisor (the second author, a registered clinical psychologist) provided weekly supervision for a minimum of 1 hour to the practicing psychologist (primary author) throughout the duration of data collection. During supervision, videotaped sessions of all clients in the study were observed and checked for adherence to CBT protocols and that the majority of the session time adhered to the protocol. Clients who took part in the study were offered 12 weekly 1-hour sessions and an additional follow-up session 6 weeks after the 12th session. A fee for clinical services was not charged, and no reimbursement for their time was provided. In-session exposure activities were limited to tasks that could be accomplished within the treatment room. Sessions times were arranged directly with the therapist; therefore, there was no need for participants in either condition to interact with support staff such as a receptionist. Participants were aware that the therapist was located within the same building. The CONSORT-EHEALTH guidelines for improving and standardizing the reporting of Web-based and mobile health interventions were used as a guide throughout this research [54].

Research Design and Data Analysis

The research design included one nominal fixed effect (condition: in-person versus videoconference), one ordinal fixed effect (time: pre, post, follow-up), one nominal random effect (participant), and four scale outcomes (depression, anxiety, stress, and quality of life). The data generated by this design were analyzed with a multilevel mixed effects linear regression model [55] as implemented through the Generalized Linear Mixed Models procedure in the SPSS software, version 19 [56]. The analysis was "multilevel" in the sense that it was conducted within the context of a hierarchical data structure in which time was nested within participant. In order to optimize the likelihood of convergence, a separate analysis was run for each of the four outcomes. With a predicted large effect size of d=0.5, an alpha level of .05, desired power of 0.8, and a correlation of 0.6 between repeated measures, the estimated total sample size using G-Power [57] was 26 (13 participants per condition). The original intent was to have sufficient power to detect a small difference, which would have required 115 participants. However, this target became impractical given the time and resource constraints.

Results

Attrition

Of the clients eligible to take part in the study, 81% (21/26) completed the full course of treatment, and 65% (18/26) completed the follow-up data. Figure 1 shows the participant flow over the course of the study. In the videoconferencing condition, 2 of the participants completed only 10 sessions: the first, because they moved overseas, and the second stated that they had done all the change they felt capable of doing at that time. In the videoconferencing condition, 2 of the participants were no longer eligible for inclusion in the study at follow-up: the first, because they required ongoing treatment after the posttreatment data had been collected and the second decided to begin medication in the final week of treatment. In the in-person condition, 1 participant dropped out after nine sessions and no reason was given. Also in the in-person condition, 1 participant was no longer eligible to be included in the study at follow-up because they commenced additional ongoing psychotherapy treatment after the initial 12-weeks of treatment in the study. The analyses were conducted twice: once with only the participants that completed all 12 sessions and once with all the participants that completed eight or more sessions.

The results of the two analyses provided the same findings; therefore, participants who completed eight or more sessions were included in the final analyses.

Comparison of Participant Characteristics at Pretreatment

There was a small nonsignificant difference between the mean ages of participants in the in-person condition (mean 29.67, SD 9.31) and the videoconferencing condition (mean 31.93, SD 13.33) (t_{24}=–0.49, P=.63, two-tailed, 95% CI -11.73 to 7.21, d=–0.19). A Pearson chi-square test of contingencies indicated that gender was equally distributed across conditions χ^2_1=0.004 (N=26), P=.95, and the association between gender and condition was very small with phi=.01. The difference between conditions at pretest was not significant on the DASS depression subscale (t_{23}=0.05, P=.96, two-tailed, 95% CI -8.28 to 8.56, d=0.02), the anxiety subscale (t_{23}=0.37, P=.71, two-tailed, 95% CI -5.98 to 8.57, d=0.15), the stress subscale (t_{23}=0.19, P=.85, two-tailed, 95% CI -8.36 to 6.98, d=–0.08), or on the QLES (t_{24}=0.28, P=.78, two-tailed, 95% CI -0.47 to 0.62, d=0.10).

Primary Analyses

In order to reduce the chances of a Type-1 error, the Bonferroni correction was applied throughout the four analyses making the alpha level .0125. The descriptive statistics for the in-person and videoconferencing conditions are reported in Table 3. The interaction between time and condition was not significant for all three DASS subscales (Depression $F_{2,58}$=1.77, P=.18; Anxiety $F_{2,58}$=0.36, P=.7; and Stress $F_{2,58}$=4.19, P=.02), and the QLES ($F_{2,62}$=0.82, P=.45); therefore, the main effects can be interpreted without qualification. There was a significant main effect for time on all three DASS subscales (Depression $F_{2,58}$=14.47, P<.001; Anxiety $F_{2,58}$=9.34, P<.001; and Stress $F_{2,58}$=23.70, P<.001) and the QLES ($F_{2,62}$=10.64, P<.001). The effect sizes were large throughout (d=1.41, 1.14, 1.81, and 1.17 respectively). In contrast, there was no significant main effect for condition on any of the DASS subscales (Depression $F_{1,58}$=1.98, P=.16; Anxiety $F_{1,58}$=0.69, P=.41 and Stress $F_{1,58}$=2.11, P=.15) or the QLES ($F_{1,62}$=0.25, P=.62). The effect sizes ranged from small to medium (d=0.37, 0.22, 0.38, and 0.13 respectively). Figures 2-5 depict the upper and lower bound of the 95% confidence intervals around the mean scores at pre, post, and 6 weeks following treatment in both conditions for the three DASS subscales and the QLES.

Table 3. Descriptive statistics for the scale measures in the in-person and videoconferencing conditions.

Scale measure		In-person condition					Videoconferencing condition				
		n	Mean	SD	Min	Max	n	Mean	SD	Min	Max
DASS depression subscale											
	Pre	11[a]	18.36	10.27	4	32	14	18.14	10.15	6	42
	Post	10	14.2	9.59	6	34	13	8.46	7.62	0	28
	Follow-up	7	9.43	7.46	2	22	9	5.11	5.93	0	18
DASS anxiety subscale											
	Pre	11[a]	14.73	8.4	6	32	14	13.43	8.96	0	26
	Post	10	9.09	8.02	0	26	13	8.62	7.63	2	26
	Follow-up	7	7.75	6.63	0	18	9	6.22	6.28	0	16
DASS stress subscale											
	Pre	11[a]	23.45	8.72	10	36	14	24.14	9.56	8	38
	Post	10	18.55	12.3	0	42	13	13.23	8.81	2	30
	Follow-up	7	13.75	9.65	0	34	9	8.89	5.58	0	18
QLES											
	Pre	12	3.38	0.76	2.06	4.88	14	3.31	0.59	2.29	4.29
	Post	11	3.58	0.72	2.24	4.71	13	3.81	0.62	2.76	4.94
	Follow-up	8	3.85	0.76	2.65	5.88	9	4.15	0.51	3.41	4.88
Credibility of therapy		11	34.14	3.73	25	38	13	34.69	4.52	23	40
Working alliance: Client		11	6.14	0.45	5.5	7	12[b]	6.33	0.89	3.75	7
Working alliance: Therapist		11	5.89	0.41	4.92	6.417	13	5.74	0.83	3.58	6.75
Client satisfaction		11	93.21	6.37	78.13	100	13	94.23	10.04	65	100

[a]The DASS data for one of the participants in the in-person condition was removed because it was invalid at pretreatment.

[b]The Working Alliance-Client data for one of the participants in the videoconferencing condition was removed because it was invalid.

Figure 2. The change in symptoms of depression across time and condition.

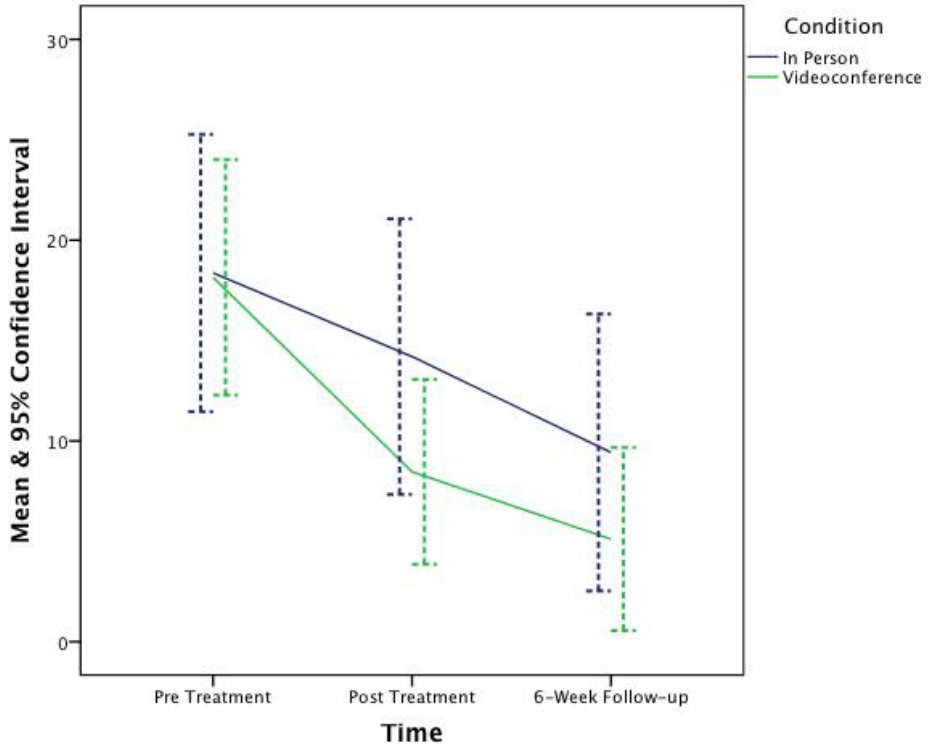


Figure 3. The change in symptoms of anxiety across time and condition.

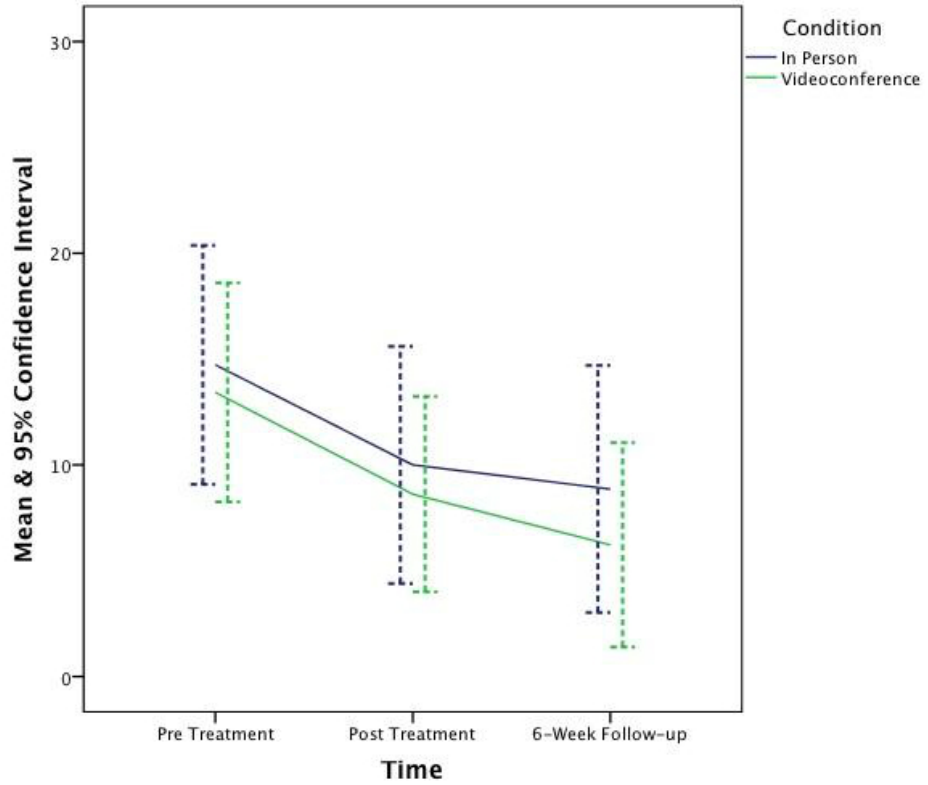


Figure 4. The change in symptoms of stress across time and condition.

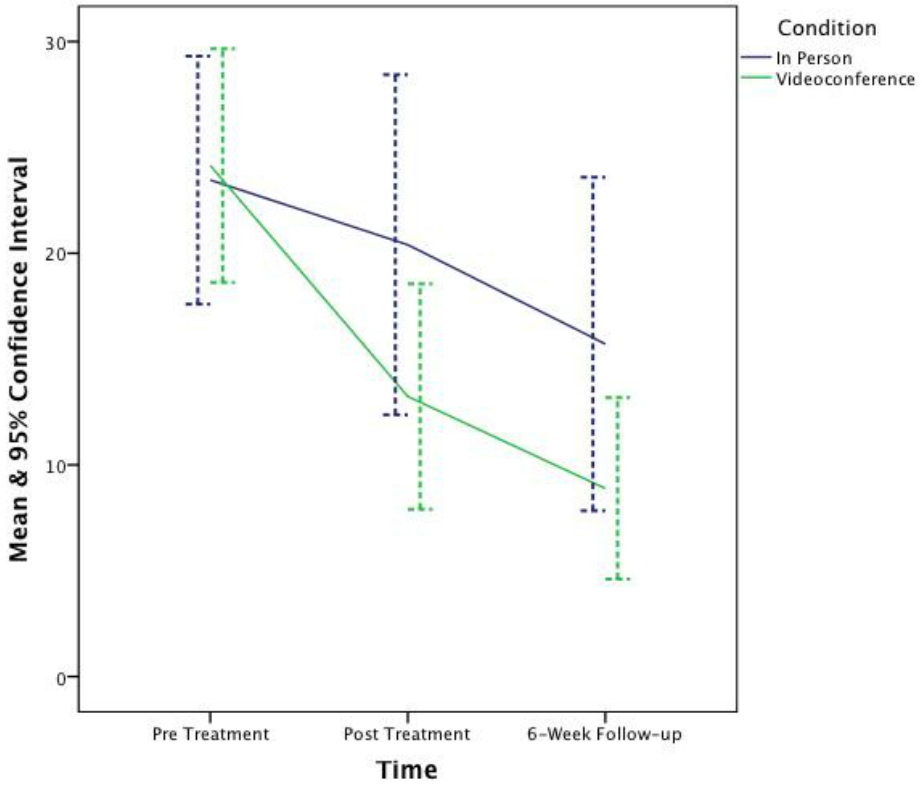
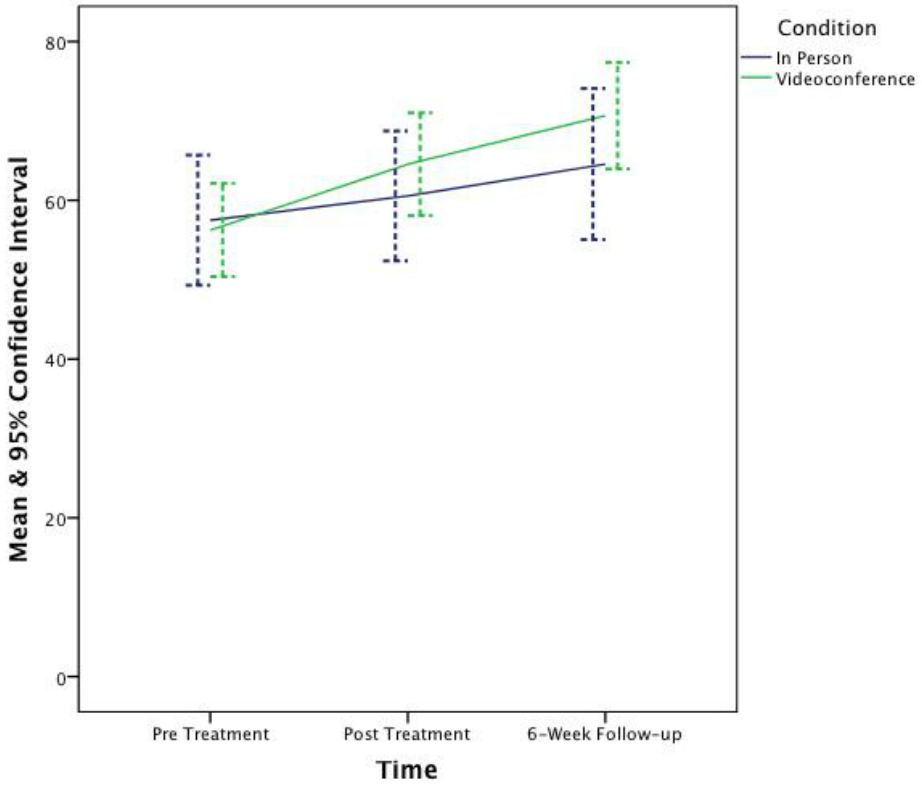


Figure 5. The change in quality of life across time and condition.



Secondary Analyses

The secondary analyses were administered only at posttest. There were no significant differences between conditions in client ratings of the Working Alliance Inventory Short Form (t_{21}=–0.63, P=.53, one-tailed, d=–0.26), or in therapist ratings (t_{22}=0.53, P=.60, one-tailed, d=0.23), and client ratings of the Client Satisfaction Questionnaire (t_{22}=–0.29, P=.77, one-tailed, d=–0.12). Clients in the videoconference-based condition who

completed 10 or more sessions were included in the Telehealth Satisfaction Questionnaire analysis (n=13). Scores ranged from 61.4 to 100 (maximum score is 100), with a mean of 90.44 (SD 11.11).

Reliable Change and Clinical Significance

Reliable change and clinical significance calculations [58] were conducted on the DASS subscales, the QLES, and disorder specific measures. Normative data were used to calculate the cutoff criteria [3,43-47,59-64]. Tables 4-6 display the proportion of clients in each condition that met the respective criteria for the three DASS subscales and the QLES respectively.

Fisher's Exact test [65] was used to compare the proportion of participants meeting the criteria for reliable change and clinical significance in each condition. With regards to reliable change, Fisher's Exact was not significant from pre- to posttreatment or from pretreatment to follow-up for the depression (*P*=.41, *P*=.26) and anxiety (*P*=.60, *P*=.99) DASS subscales, and the QLES (*P*=.65, *P*=.99). Fisher's Exact *P*, however, was significant for the stress DASS subscale from both pre- to posttreatment (*P*=.03) and from pretreatment to follow-up (*P*=.03) in favor of the videoconferencing condition. With regards to clinically significant change, the difference between conditions was not significant from pre- to posttreatment or from pretreatment to follow-up for the depression (*P*=.67, *P*=.30), anxiety (*P*=.99, *P*=.99), and stress (*P*=.19, *P*=.13) DASS subscales or the QLES (*P*=.99, *P*=.62). Table 7 displays the number of clients that upon discharge met the criteria for both reliable change (improved) and clinically significant change (recovered) for the various disorder specific measures administered in each condition. However, not all clients presented with severe symptoms at pretreatment. The proportion of clients whose symptoms were identified in the moderate to severe range at pretreatment are displayed in Table 8.

Table 4. Rates of clinical improvement from pre- to posttreatment for the DASS.

Clinical status	Pretreatment to posttreatment					
	Depression		Anxiety		Stress	
	IP[a]	VC[b]	IP	VC	IP	VC
Recovered	3/10	5/13	1/10	2/13	2/10	7/13
Improved	4/10	8/13	1/10	3/13	2/10	9/13
Unchanged	6/10	4/13	9/10	10/13	8/10	4/13
Deteriorated	None	1/13	None	None	None	None

[a]IP=In-person.

[b]VC=Videoconference.

Table 5. Rates of clinical improvement from pretreatment to follow-up for the DASS.

Clinical status	Pretreatment to follow-up					
	Depression		Anxiety		Stress	
	IP[a]	VC[b]	IP	VC	IP	VC
Recovered	3/7	7/9	2/7	2/9	2/7	7/9
Improved	4/7	8/9	2/7	2/9	2/7	8/9
Unchanged	4/7	1/9	5/7	7/9	5/7	1/9
Deteriorated	None	None	None	None	None	None

[a]IP=In-person.

[b]VC=Videoconference.

Table 6. Rates of clinical improvement from pre to posttreatment and follow-up for the QLES.

Clinical status	Pre to post		Pre to follow-up	
	IP[a]	VC[b]	IP	VC
Recovered	2/11	2/13	3/8	3/9
Improved	2/11	4/13	3/8	4/9
Unchanged	9/11	9/13	4/8	5/9
Deteriorated	None	None	1/8	None

[a]IP=In-person.

[b]VC=Videoconference.

Table 7. The number of clients in each condition identified as recovered on the disorder specific measures.

Condition	Primary diagnosis	Disorder specific measure	Change status	n
Videoconference				
	Obsessive- compulsive disorder	Obsessive-Compulsive Inventory	Recovered	4/9
			No change[a]	2/9
			No change	3/9
	Depression	Beck Depression Inventory-II	Recovered	1/2
			No change[a]	1/2
	Hypochondriasis	Health Anxiety Questionnaire	Recovered	1/1
	Generalized anxiety disorder	Penn State Worry Questionnaire	Recovered	1/1
In-person				
	Depression	Beck Depression Inventory-II	Recovered	2/4
			No change[a]	1/1
			No change	1/1
	Obsessive- compulsive disorder	Obsessive-Compulsive Inventory	Recovered	1/3
			No change[a]	1/3
			No change	1/3
	Generalized anxiety disorder	Penn State Worry Questionnaire	Recovered	2/3
			No change	1/3
	Panic disorder	Anxiety Sensitivity Index	Recovered	1/1

[a]Although some clients were not identified as recovered or improved on their disorder specific measure, they were identified as recovered on the DASS.

Table 8. The severity of client symptoms at pretreatment.

Symptom	In-person (n=12) n (%)		Videoconference (n=14) n (%)	
	Moderate-severe	Normal-mild	Moderate-severe	Normal-mild
Depression	7 (58)	5 (42)	7 (50)	7 (50)
Anxiety	8 (76)	4 (33)	9 (64)	5 (36)
Stress	10 (83)	2 (17)	11 (79)	3 (21)
Quality of life	10 (83)	2 (17)	11 (79)	3 (21)

Discussion

Principal Findings

The findings of this controlled trial indicate that CBT was effective in significantly reducing symptoms of depression, anxiety, and stress and increasing quality of life in both in-person and videoconferencing conditions. Furthermore, outcomes for the videoconferencing group were comparable to the in-person group, with no significant differences being observed between the two conditions. This finding is consistent with prior CBT meta-analysis research [66,67] and prior telepsychology research [18,25-27,68-71]. Although the sample was not large enough to permit a test of noninferiority, the 95% confidence intervals around the mean scores across time and condition suggest CBT treatment administered via videoconference is not inferior to treatment provided in-person. This finding is congruent with prior noninferiority analyses [28,70]. It is also important to note that the clinical outcomes at posttreatment and follow-up were better, albeit not significantly better, in the videoconferencing condition than in the in-person condition. Therefore, if this study had a greater power to detect an effect, it is unlikely this study would have shown videoconference-based CBT to be inferior to in-person.

Client and therapist ratings of the working alliance were high in both the in-person and videoconferencing conditions. The difference between conditions was not significant, and the effect size was trivial. These findings suggest that CBT treatment via videoconference does not compromise the working alliance, which is congruent with prior telepsychology research [26,72]. The results also indicated that client satisfaction was high and did not significantly differ between conditions, which again is congruent with prior research [19,20,73].

Strengths and Limitations

The inclusion of a mixed diagnostic client cohort with comorbid diagnosis (as measured in a structured clinical interview) is a

strength of this study. A client group with a range of disorders and comorbidities helps in generalizing the findings to real-world clinical practice populations and thus provides empirical support for the effectiveness of CBT-oriented telepsychology. Second, the same therapist treated all clients, and thus this variable was held constant. By including only 1 therapist in the study, the difference between conditions cannot be attributed to differences between therapists. A further strength of this study is that it was conducted on a high-speed local network. As a result of conducting the study on a system that is presently faster than the average Australian Internet speed, it may be a longer period of time before the conclusions of this study become outdated. Also, all of the participants in both conditions received treatment in the same building; consequently, this study was able to demonstrate that satisfaction remained high even when travel was not a contributing factor in the ratings of service satisfaction. No technological problems occurred during the study, and clients did not report any complaints about the technology.

Although this study has successfully extended previous telepsychology research [25-27,30,69,70], there are some limitations. First, because only 1 therapist was included in this study, it remains unclear as to whether or not the findings generalize to other therapists. Second, the diagnosis of clients in both conditions was conducted in-person by the same therapist who conducted the subsequent treatment. It is possible that rapport and a therapeutic alliance began in this initial in-person diagnosis session and influenced clinical outcomes. Third, the sample size was not large enough to conduct an analysis of noninferiority. The limited small sample size also weakens the generalizability of the findings. Fourth, there was a high risk of bias in the study given that the principal investigator was also responsible for treatment and assessing participants. However, clinical supervision was provided regarding both diagnosis and treatment throughout the study, and sessions were recorded and watched by the supervising clinician to temper this potential bias. Fifth, the diagnosis of participants was not evenly distributed across conditions, and as a result, the two groups might not have been equally matched. Sixth, as is the case in most research studies, the participants did not have to pay for the services, and this may have led them to provide socially desirable answers on the questionnaires. A replication study involving a fee for service would be able to clarify if this is an issue. Seventh, the average age of participants was 30; therefore, it is likely that most participants were already familiar with computer technology, which may have influenced their acceptance of the media.

Despite being a strength of the study, the inclusion of a mixed diagnostic cohort combined with a small sample size means there was a large degree of variability across the clients, which may have reduced the power of the study. Also, Bonferroni corrections were applied to reduce the risk of a Type-I error, further reducing the power of the study to detect significant interactions between time and condition. While providing promising results, a larger study is still needed to more thoroughly address the primary research question.

Not all outcomes in this study were favorable. Only 10% (1/10) of clients in the in-person condition and 33% (2/13) of participants in the videoconference condition were identified as meeting the criteria for a clinically significant reduction in anxiety. These results could be because 33% (4/12) and 36% (5/14) (in-person and videoconference respectively) of participants were identified on the DASS as having normal-to-mild levels of anxiety before beginning treatment. Alternatively, the therapist may not have been as competent at treating symptoms of anxiety compared to depression for which 43% (3/7) and 77% (7/9) of participants (in-person and videoconference respectively) were identified as meeting the criteria for clinically significant change. The presence of comorbid personality disorders may also have hampered treatment outcomes. In either case, a larger trial involving multiple therapists, systematic monitoring of treatment protocol adherence and a single diagnosis may help to overcome these issues in future research.

Some psychotherapy procedures via videoconference required creative modification. In-session exposure exercises via videoconference had to be limited to what could be conducted within the clinic room. For example a client working through exposure-response prevention activities pertaining to germs had to find items they found troubling that were in their treatment room as opposed to in-person sessions that could have the freedom to move outside of the treatment room. Another change was that cognitive restructuring worksheets had to be conducted on a digital Microsoft Word document on the clients' computer but controlled from the therapist's computer. These examples show that treatment via videoconference may alter how some interactions are conducted but does not prevent essential treatment components from being provided.

Conclusions

It is important to consider the findings of this study within the context of the social justice, financial, and practical issues facing remote psychological services in Australia. In Australia, there is a disparity in access to psychological services between people who live in cities and those who live in rural and remote areas [74]. Rural and remote populations have higher rates of accidents, suicide, and exposure to violence [75] as well as increased rates of risk factors known to play a role in the emergence of psychological dysfunction, including poor physical health, obesity, smoking, drug abuse, high blood pressure, and poor nutrition [76]. As for the indigenous populations in rural and remote communities, the state of affairs is considerably worse where the rates of mortality and morbidity are approximately 2-4 times higher than for nonindigenous Australians [77]. People in rural and remote Australia are in the highest need of specialized mental health services; however, due to the lack of specialized clinicians working in these areas, the accessibility of those services frequently range from little to none [76]. Hence, there is a substantial demand for videoconference-based psychological services. However, there is often insufficient technological infrastructure to provide reliable high-speed broadband Internet communications, particularly in remote areas where there can be less than one person per square kilometer [78]. It is hoped that with the rollout of the National Broadband Network in Australia more rural and remote areas will gain access to high-quality videoconferencing technology. Prior research [17,79] has indicated that providing

remote services via videoconference can be cost effective under some circumstances, but further research is needed to determine if it is also cost effective for small isolated indigenous communities. Further research is needed to identify the logistical challenges of providing videoconference-based CBT to rural and remote communities and to determine which method of service delivery is best suited to the needs of specific communities. Telepsychiatry services through medicare in Australia have increased from 15 consultations in 2002/3 to 2555 in 2010/11 [80]. This indicates that when available, both service providers and clients are willing to use the technology. Hopefully, this trend will extend to telepsychology services in the future as access to the technology increases.

In conclusion, this study provides support for the effectiveness of CBT via telepsychology for an adult client cohort and preliminary support of noninferiority. The results of this controlled trial provide important evidence to justify the greater use of videoconferencing to bridge the gap in service provision to populations who would otherwise not receive effective psychological treatments. Future research should continue to involve mixed diagnostic client cohorts with comorbid diagnoses and be directed at conducting large multisite randomized controlled trials that employ noninferiority design and analysis procedures.

Acknowledgments

This research was conducted as part of a Doctor of Philosophy (Clinical Psychology) degree and was not funded by an external body.

Conflicts of Interest

None declared.

References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
2. Parker G, McCraw S, Hadzi-Pavlovic D, Fletcher K. Costs of the principal mood disorders: a study of comparative direct and indirect costs incurred by those with bipolar I, bipolar II and unipolar disorders. J Affect Disord 2013 Jul;149(1-3):46-55. [doi: 10.1016/j.jad.2012.10.002] [Medline: 23141630]
3. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression. New York: Guilford Press; 1979.
4. Bennett-Levy J, Westbrook D, Fennell M, Cooper M, Rouf K, Hackmann A. Behavioural experiments: Historical and conceptual underpinnings. In: Bennett-Levy J, Butler G, Fennell M, Hackman A, Mueller M, Westbrook D, editors. Oxford guide to behavioural experiments in cognitive therapy. Oxford: Oxford University Press; 2004.
5. Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J Consult Clin Psychol 2010 Oct;78(5):668-680. [doi: 10.1037/a0020570] [Medline: 20873902]
6. Rosa-Alcázar AI, Sánchez-Meca J, Gómez-Conesa A, Marín-Martínez F. Psychological treatment of obsessive-compulsive disorder: a meta-analysis. Clin Psychol Rev 2008 Dec;28(8):1310-1325. [doi: 10.1016/j.cpr.2008.07.001] [Medline: 18701199]
7. Mitte K. Meta-analysis of cognitive-behavioral treatments for generalized anxiety disorder: a comparison with pharmacotherapy. Psychol Bull 2005 Sep;131(5):785-795. [doi: 10.1037/0033-2909.131.5.785] [Medline: 16187860]
8. Acarturk C, Cuijpers P, van Straten A, de Graaf R. Psychological treatment of social anxiety disorder: a meta-analysis. Psychol Med 2009 Feb;39(2):241-254. [doi: 10.1017/S0033291708003590] [Medline: 18507874]
9. Mitte K. A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. J Affect Disord 2005 Sep;88(1):27-45. [doi: 10.1016/j.jad.2005.05.003] [Medline: 16005982]
10. Bisson JI, Ehlers A, Matthews R, Pilling S, Richards D, Turner S. Psychological treatments for chronic post-traumatic stress disorder. Systematic review and meta-analysis. Br J Psychiatry 2007 Feb;190:97-104 [FREE Full text] [doi: 10.1192/bjp.bp.106.021402] [Medline: 17267924]
11. McEvoy PM, Nathan P. Effectiveness of cognitive behavior therapy for diagnostically heterogeneous groups: a benchmarking study. J Consult Clin Psychol 2007 Apr;75(2):344-350. [doi: 10.1037/0022-006X.75.2.344] [Medline: 17469892]
12. Shafran R, Clark DM, Fairburn CG, Arntz A, Barlow DH, Ehlers A, et al. Mind the gap: Improving the dissemination of CBT. Behav Res Ther 2009 Nov;47(11):902-909. [doi: 10.1016/j.brat.2009.07.003] [Medline: 19664756]
13. Nickelson DW. Behavioral Telehealth: Emerging Practice, Research, and Policy Opportunities. Behav Sci Law 1996;14(4):443-457. [doi: 10.1002/(SICI)1099-0798(199623)14:4<443::AID-BSL256>3.0.CO;2-G]
14. García-Lizana F, Muñoz-Mayorga I. Telemedicine for depression: a systematic review. Perspect Psychiatr Care 2010 Apr;46(2):119-126. [doi: 10.1111/j.1744-6163.2010.00247.x] [Medline: 20377799]

15. O'Reilly R, Bishop J, Maddox K, Hutchinson L, Fisman M, Takhar J. Is telepsychiatry equivalent to face-to-face psychiatry? Results from a randomized controlled equivalence trial. Psychiatr Serv 2007 Jun;58(6):836-843. [doi: 10.1176/appi.ps.58.6.836] [Medline: 17535945]
16. Vaitheswaran S, Crockett P, Wilson S, Millar H. Telemental health: videoconferencing in mental health services. Advances in Psychiatric Treatment 2012 Sep 03;18(5):392-398. [doi: 10.1192/apt.bp.111.008904]
17. Jong M. Managing suicides via videoconferencing in a remote northern community in Canada. Int J Circumpolar Health 2004 Dec;63(4):422-428. [Medline: 15709317]
18. Simpson S. Psychotherapy via videoconferencing: a review. British Journal of Guidance & Counselling 2009 Aug;37(3):271-286. [doi: 10.1080/03069880902957007]
19. Ruskin PE, Silver-Aylaian M, Kling MA, Reed SA, Bradham DD, Hebel JR, et al. Treatment outcomes in depression: comparison of remote treatment through telepsychiatry to in-person treatment. Am J Psychiatry 2004 Aug;161(8):1471-1476. [doi: 10.1176/appi.ajp.161.8.1471] [Medline: 15285975]
20. Sorvaniemi M, Ojanen E, Santamäki O. Telepsychiatry in emergency consultations: a follow-up study of sixty patients. Telemed J E Health 2005 Aug;11(4):439-441. [doi: 10.1089/tmj.2005.11.439] [Medline: 16149889]
21. Saurman E, Perkins D, Roberts R, Roberts A, Patfield M, Lyle D. Responding to mental health emergencies: implementation of an innovative telehealth service in rural and remote new South Wales, Australia. J Emerg Nurs 2011 Sep;37(5):453-459. [doi: 10.1016/j.jen.2010.11.005] [Medline: 21889653]
22. Jones BN, Johnston D, Reboussin B, McCall WV. Reliability of telepsychiatry assessments: subjective versus observational ratings. J Geriatr Psychiatry Neurol 2001;14(2):66-71. [Medline: 11419569]
23. Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol 2010 Apr;30(2):193-197. [doi: 10.1097/JCP.0b013e3181d20912] [Medline: 20520295]
24. Tsai HH, Tsai YF. Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents. J Med Internet Res 2011;13(4):e93 [FREE Full text] [doi: 10.2196/jmir.1678] [Medline: 22086660]
25. Bouchard S, Paquin B, Payeur R, Allard M, Rivard V, Fournier T, et al. Delivering cognitive-behavior therapy for panic disorder with agoraphobia in videoconference. Telemed J E Health 2004;10(1):13-25. [doi: 10.1089/153056204773644535] [Medline: 15104911]
26. Himle JA, Fischer DJ, Muroff JR, Van Etten ML, Lokers LM, Abelson JL, et al. Videoconferencing-based cognitive-behavioral therapy for obsessive-compulsive disorder. Behav Res Ther 2006 Dec;44(12):1821-1829. [doi: 10.1016/j.brat.2005.12.010] [Medline: 16466688]
27. Mitchell JE, Crosby RD, Wonderlich SA, Crow S, Lancaster K, Simonich H, et al. A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2008 May;46(5):581-592 [FREE Full text] [doi: 10.1016/j.brat.2008.02.004] [Medline: 18374304]
28. Morland LA, Greene CJ, Rosen CS, Foy D, Reilly P, Shore J, et al. Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: a randomized noninferiority trial. J Clin Psychiatry 2010 Jul;71(7):855-863. [doi: 10.4088/JCP.09m05604blu] [Medline: 20122374]
29. Day SX, Schneider PL. Psychotherapy using distance technology: A comparison of face-to-face, video, and audio treatment. Journal of Counseling Psychology 2002;49(4):499-503. [doi: 10.1037//0022-0167.49.4.499]
30. Griffiths L, Blignault I, Yellowlees P. Telemedicine as a means of delivering cognitive-behavioural therapy to rural and remote mental health clients. J Telemed Telecare 2006;12(3):136-140. [doi: 10.1258/135763306776738567] [Medline: 16638234]
31. Dunstan DA, Tooth SM. Treatment via videoconferencing: a pilot study of delivery by clinical psychology trainees. Aust J Rural Health 2012 Apr;20(2):88-94. [doi: 10.1111/j.1440-1584.2012.01260.x] [Medline: 22435769]
32. Diefenbach GJ, Abramowitz JS, Norberg MM, Tolin DF. Changes in quality of life following cognitive-behavioral therapy for obsessive-compulsive disorder. Behav Res Ther 2007 Dec;45(12):3060-3068. [doi: 10.1016/j.brat.2007.04.014] [Medline: 17559803]
33. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
34. Shepherd L, Goldstein D, Whitford H, Thewes B, Brummell V, Hicks M. The utility of videoconferencing to provide innovative delivery of psychological treatment for rural cancer patients: results of a pilot study. J Pain Symptom Manage 2006 Nov;32(5):453-461. [doi: 10.1016/j.jpainsymman.2006.05.018] [Medline: 17085271]
35. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV. Washington, DC: American Psychiatric Association; 2000.
36. First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. Structured clinical interview for DSM-IV Axis II personality disorders (SCID-II): [single-use packet]. Washington, DC: American Psychiatric Press; 1997.
37. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured clinical interview for DSM-IV axis I disorders SCID-I: clinician version, administration booklet. Washington, DC: American Psychiatric Press; 1996.

38. Brown TA, DiNardo PA, Barlow DH. Anxiety Disorders Interview Schedule Adult Version: Client Interview Schedule (Treatments That Work). Oxford: Oxford University Press; 1994.
39. Barlow DH, Craske MG. Mastery of Your Anxiety and Panic (MAP-3): Client Workbook for Agoraphobia. New York: Academic Press; 2000.
40. Taylor S, Asmundson GJG. Treating Health Anxiety: A Cognitive-Behavioral Approach. New York: The Guilford Press; 2004.
41. Craske MG, Barlow DH, O'Leary T. Mastery of Your Anxiety and Worry: Client Workbook (therapy works). New York: Harcourt Brace & Company; 1992.
42. Rees CS. Obsessive-compulsive disorder: A practical guide to treatment. Melbourne: IP Communications; 2009.
43. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney: Psychological Foundation; 1995.
44. Ritsner M, Kurs R, Gibel A, Ratner Y, Endicott J. Validity of an abbreviated quality of life enjoyment and satisfaction questionnaire (Q-LES-Q-18) for schizophrenia, schizoaffective, and mood disorder patients. Qual Life Res 2005 Sep;14(7):1693-1703. [Medline: 16119181]
45. Beck AT, Steer RA, Brown GK. BDI-II, Beck depression inventory: manual. San Antonio, Tex: Psychological Corp; 1996.
46. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The Obsessive-Compulsive Inventory: development and validation of a short version. Psychol Assess 2002 Dec;14(4):485-496. [Medline: 12501574]
47. Lucock MP, Morely S. The Health Anxiety Questionnaire. British Journal of Health Psychology 1996;1:137-150. [doi: 10.1111/j.2044-8287.1996.tb00498.x]
48. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [Medline: 2076086]
49. Taylor S, Cox BJ. An expanded anxiety sensitivity index: evidence for a hierarchic structure in a clinical sample. J Anxiety Disord 1998 Oct;12(5):463-483. [Medline: 9801964]
50. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. Psychological Assessment 1989;1(3):207-210. [doi: 10.1037//1040-3590.1.3.207]
51. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [Medline: 10259963]
52. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [Medline: 12641893]
53. Graphpad Software. 2013. Quick Calcs URL: http://graphpad.com/quickcalcs/randomize1.cfm [accessed 2013-06-13] [WebCite Cache ID 6HKoDLzvX]
54. 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]
55. Bryk AS, Raudenbush SW. Application of hierarchical linear models to assessing change. Psychological Bulletin 1987;101(1):147-158.
56. IBM Corp. SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp; 2010.
57. Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers 1996 Mar;28(1):1-11. [doi: 10.3758/BF03203630]
58. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
59. Bernstein A, Stickle TR, Zvolensky MJ, Taylor S, Abramowitz J, Stewart S. Dimensional, categorical, or dimensional-categories: testing the latent structure of anxiety sensitivity among adults using factor-mixture modeling. Behav Ther 2010 Dec;41(4):515-529. [doi: 10.1016/j.beth.2010.02.003] [Medline: 21035615]
60. Buckley TC, Parker JD, Heggie J. A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. J Subst Abuse Treat 2001 Apr;20(3):197-204. [Medline: 11516588]
61. Fresco DM, Mennin DS, Heimberg RG, Turk CL. Using the Penn State Worry Questionnaire to identify individuals with generalized anxiety disorder: a receiver operating characteristic analysis. J Behav Ther Exp Psychiatry 2003 Dec;34(3-4):283-291. [doi: 10.1016/j.jbtep.2003.09.001] [Medline: 14972674]
62. Jackson J, Kincey J, Fiddler M, Creed F, Tomenson B. Differences between out-patients with physical disease and those with medically unexplained symptoms with respect to patient satisfaction, emotional distress and illness perception. Br J Health Psychol 2004 Nov;9(Pt 4):433-446. [doi: 10.1348/1359107042304597] [Medline: 15509353]
63. Page AC, Hooke GR, Morrison DL. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in depressed clinical samples. Br J Clin Psychol 2007 Sep;46(Pt 3):283-297. [doi: 10.1348/014466506X158996] [Medline: 17697478]
64. Zlomke KR. Psychometric properties of internet administered versions of Penn State Worry Questionnaire (PSWQ) and Depression, Anxiety, and Stress Scale (DASS). Computers in Human Behavior 2009 Jul;25(4):841-843. [doi: 10.1016/j.chb.2008.06.003]
65. Joosse S, Hannemann J. Fisher's exact test. 2013. URL: http://in-silico.net/tools/statistics/fisher_exact_test [accessed 2013-06-13] [WebCite Cache ID 6HKsPJQcU]

66. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
67. Cuijpers P, van Straten A, Schuurmans J, van Oppen P, Hollon SD, Andersson G. Psychotherapy for chronic major depression and dysthymia: a meta-analysis. Clin Psychol Rev 2010 Feb;30(1):51-62. [doi: 10.1016/j.cpr.2009.09.003] [Medline: 19781837]
68. Grady B, Singleton M. Telepsychiatry "coverage" to a rural inpatient psychiatric unit. Telemed J E Health 2011 Oct;17(8):603-608. [doi: 10.1089/tmj.2011.0031] [Medline: 21939381]
69. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current Directions in Videoconferencing Tele-Mental Health Research. Clin Psychol (New York) 2009 Sep 1;16(3):323-338 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]
70. Frueh BC, Monnier J, Yim E, Grubaugh AL, Hamner MB, Knapp RG. A randomized trial of telepsychiatry for post-traumatic stress disorder. J Telemed Telecare 2007;13(3):142-147. [doi: 10.1258/135763307780677604] [Medline: 17519056]
71. Richards D, Viganó N. Online counseling: a narrative and critical review of the literature. J Clin Psychol 2013 Sep;69(9):994-1011. [doi: 10.1002/jclp.21974] [Medline: 23630010]
72. Simpson S, Bell L, Knox J, Mitchell D. Therapy via videoconferencing: a route to client empowerment? Clin. Psychol. Psychother 2005 Mar;12(2):156-165. [doi: 10.1002/cpp.436]
73. Kopel H, Nunn K, Dossetor D. Evaluating satisfaction with a child and adolescent psychological telemedicine outreach service. J Telemed Telecare 2001;7 Suppl 2:35-40. [Medline: 11747654]
74. Joshi H, Wiggins R, Bartley M, Mitchell R, Gleave S, Lynch K. Putting health inequalities on the map: does where you live matterwhy? In: Understanding health inequalities. Buckingham: Open University Press; 2000:156-169.
75. Australian Institute of Health and Welfare. Australia's Health 2006. 2013. URL: http://aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442453483 [accessed 2013-07-17] [WebCite Cache ID 6IAjwRuq6]
76. National Rural Health Alliance. Under pressure and under-valued: Allied health professionals in rural and remote areas. Canberra: NRHA; 2013. URL: http://ruralhealth.org.au/document/under-pressure-and-under-valued-allied-health-professionals-rural-and-remote-areas [accessed 2013-07-17] [WebCite Cache ID 6IAkH8nqK]
77. Australian Institute of Health and Welfare. The health and welfare of Australia's and Torres Strait Islander peoples. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442458617 [accessed 2013-07-17] [WebCite Cache ID 6IAki9PeB]
78. Australian Bureau of Statistics. Regional Population Growth, Australia, 2010-11. URL: http://www.abs.gov.au/ausstats/abs@.nsf/Products/3218.0~2010-11~Main+Features~Main+Features?OpenDocument [accessed 2013-07-17] [WebCite Cache ID 6IAl1zkTW]
79. Smith AC, Stathis S, Randell A, Best D, Ryan V, Bergwever E, et al. A cost-minimization analysis of a telepaediatric mental health service for patients in rural and remote Queensland. J Telemed Telecare 2007 Dec 01;13(8):79-83. [doi: 10.1258/135763307783247239]
80. Smith AC, Armfield NR, Croll J, Gray LC. A review of Medicare expenditure in Australia for psychiatric consultations delivered in person and via videoconference. J Telemed Telecare 2012 Apr;18(3):169-171. [doi: 10.1258/jtt.2012.SFT111] [Medline: 22362827]

Abbreviations

CBT: cognitive-behavioral therapy
DASS: Depression Anxiety and Stress Scale
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, 4th edition text revision
QLES: Quality of Life and Satisfaction Questionnaire

Edited by G Eysenbach; submitted 05.02.13; peer-reviewed by P Yellowlees, R Kok, E Borgmann; comments to author 11.06.13; revised version received 23.07.13; accepted 15.10.13; published 19.11.13.

Please cite as:
Stubbings DR, Rees CS, Roberts LD, Kane RT
Comparing In-Person to Videoconference-Based Cognitive Behavioral Therapy for Mood and Anxiety Disorders: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e258
URL: http://www.jmir.org/2013/11/e258/
doi:10.2196/jmir.2564
PMID:24252663

Original Paper

Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study

Andrey Zheluk[1], BAppSc, GradDip(History), MBA, GradDip(IT); Casey Quinn[2], BCom, MPhil, PhD; Daniel Hercz[3], BSc, MSc Biostat; James A Gillespie[1], BA(Hons), PhD

[1]Menzies Centre for Health Policy, The University of Sydney, University of Sydney NSW, Australia

[2]PRMA Consulting Ltd, New York, NY, United States

[3]University of Sydney, Sydney, Australia

Corresponding Author:
Andrey Zheluk, BAppSc, GradDip(History), MBA, GradDip(IT)
Menzies Centre for Health Policy
The University of Sydney
D02 Victor Coppleson Building
University of Sydney NSW, 2006
Australia
Phone: 61 2 9351 2818
Fax: 61 2 9351 5204
Email: andreyzheluk@gmail.com

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2013/11/e267/

Abstract

Background: Human immunodeficiency virus (HIV) is a serious health problem in the Russian Federation. However, the true scale of HIV in Russia has long been the subject of considerable debate. Using digital surveillance to monitor diseases has become increasingly popular in high income countries. But Internet users may not be representative of overall populations, and the characteristics of the Internet-using population cannot be directly ascertained from search pattern data. This exploratory infoveillance study examined if Internet search patterns can be used for disease surveillance in a large middle-income country with a dispersed population.

Objective: This study had two main objectives: (1) to validate Internet search patterns against national HIV prevalence data, and (2) to investigate the relationship between search patterns and the determinants of Internet access.

Methods: We first assessed whether online surveillance is a valid and reliable method for monitoring HIV in the Russian Federation. Yandex and Google both provided tools to study search patterns in the Russian Federation. We evaluated the relationship between both Yandex and Google aggregated search patterns and HIV prevalence in 2011 at national and regional tiers. Second, we analyzed the determinants of Internet access to determine the extent to which they explained regional variations in searches for the Russian terms for “HIV” and “AIDS”. We sought to extend understanding of the characteristics of Internet searching populations by data matching the determinants of Internet access (age, education, income, broadband access price, and urbanization ratios) and searches for the term “HIV” using principal component analysis (PCA).

Results: We found generally strong correlations between HIV prevalence and searches for the terms “HIV” and “AIDS”. National correlations for Yandex searches for “HIV” were very strongly correlated with HIV prevalence (Spearman rank-order coefficient [*rs*]=.881, *P*≤.001) and strongly correlated for “AIDS” (*rs*=.714, *P*≤.001). The strength of correlations varied across Russian regions. National correlations in Google for the term “HIV” (*rs*=.672, *P*=.004) and “AIDS” (*rs*=.584, *P*≤.001) were weaker than for Yandex. Second, we examined the relationship between the determinants of Internet access and search patterns for the term “HIV” across Russia using PCA. At the national level, we found Principal Component 1 loadings, including age (-0.56), HIV search (-0.533), and education (-0.479) contributed 32% of the variance. Principal Component 2 contributed 22% of national variance (income, -0.652 and broadband price, -0.460).

Conclusions: This study contributes to the methodological literature on search patterns in public health. Based on our preliminary research, we suggest that PCA may be used to evaluate the relationship between the determinants of Internet access and searches for health problems beyond high-income countries. We believe it is in middle-income countries that search methods can make the greatest contribution to public health.

(J Med Internet Res 2013;15(11):e256) doi:10.2196/jmir.2936

KEYWORDS

Russia; search engine; human immunodeficiency virus; surveillance

Introduction

Search Patterns in Health

Internet search patterns provide a low-cost, rapidly accessible data source for a range of health problems. Search patterns have been described as behavioral measures of an issue's importance to individuals [1]. If individual Internet users are concerned or interested in an issue, they are more likely to search for information related to that issue. The relative importance of an issue among populations of Internet users can thus be inferred from the volume of searches for a term or terms representing that issue. Since 2006, researchers have used search patterns to study a wide range of health problems, notably influenza [2-5], as well as undocumented adverse drug interactions [6,7], suicide-related information [8], and HIV (human immunodeficiency virus) [9]. Despite the widespread use of search patterns, researchers commonly suggest Internet users may not be representative of the entire population. Specific concerns include differences in access based on age [10], income and education [11], and gender [12]. This means the social, economic, and demographic status of Internet users may not fully reflect those of the population as a whole.

Determinants of Internet Access

The propensity to access the Internet varies between socioeconomic and demographic cohorts. The strongest determinants of Internet access are income and education. This finding is consistent in studies from the United States [13] and the European Union [14] and across middle-income countries [15]. Additionally, gender [16], English-language ability [17], broadband access price [18], urban location [19], ethnicity [20], and age [21] have also been reported as determinants of Internet use in both high- and middle-income countries. In summary, Internet users in both high- and middle-income countries are more likely to have higher incomes and higher levels of education.

The Digital Divide—Access and Use

Access to the Internet is an economic development policy issue. Telecommunications networks, including the Internet, are regarded as a catalyst for economic growth [22]. Since the early 2000s, the term "digital divide" has been widely used to describe differences in Internet access and use across socioeconomic gradients within and between countries [23]. In 2011, Hilbert reviewed international policy responses to the digital divide [24]. In his review, Hilbert proposes four classes of variables with which to analyze the digital divide. These classes are the unit of analysis (eg, individual, country), determinants of access (eg, income, education), the kind of technology (eg, cell phones, fixed broadband), and how individuals connect (ie, access vs effective use). Others have similarly argued that access to infrastructure inadequately describes the digital divide [25]. Basing their arguments on Roger's theory of diffusion of innovations, these authors suggest analysis of the digital divide should focus on effective use, incorporating technical competence, and individuals' adaptation of technology to meet their personal needs rather than access alone.

Use of the Internet for Health Information Seeking Online

The Internet is widely used for health information seeking in high-income countries. A 2013 study found 59% of all US adults searched for health information online, with 77% of these starting at search engines such as Google [26]. Equally, there is a general scholarly consensus that a digital divide applies to online health seeking behavior [27]. In 2006, Rice described the limited research into health-related Internet use across economic and demographic gradients in the United States [28]. More recent studies European [29] and US [30] studies suggest that income and education are the most important determinants of seeking health information online.

Search Patterns and Effective Use

Although a digital divide may exist, determining the sociodemographic profile of Internet users from search results is not straightforward. Aggregated Google search queries are the most commonly used data source for search studies but carry no demographic or economic information. In the case of disease surveillance, this means that groups with a significant disease burden, such as older or economically disadvantaged people without Internet access may be excluded from search results [31]. By contrast, health information seeking research is generally based on qualitative research and statistical surveys. This research generally includes demographic characteristics and covers issues such as health literacy [32] and behaviors following access to health information [33]. In summary, researchers have widely investigated the effective use of online health information in high-income countries. It is this research that provides the empirical foundation for a rich analysis of the relationship between health information seeking across economic and demographic gradients and patterns of online search.

Chronic Illness and Internet Use

Individuals with chronic health problems and disabilities are more likely to search for health information online. Online information seeking among people with chronic and terminal diseases has been widely researched [34,35]. Cancer information seeking in particular has attracted considerable research interest

due to its diversity, duration, and treatment complexity [36]. The management of HIV as a chronic illness has similarly attracted scholarly interest. Studies suggest PLHIV (people living with HIV) use the Internet extensively for health information. A 2006 US study found that 66% of PLHIV participants searched for health information at least half the time they were online [37]. Furthermore, PLHIV Internet users were more likely to be better educated, have higher incomes, exhibit greater knowledge of HIV disease processes, and adhere to medication [38,39]. In summary, while income and education are the most important determinants of health-related Internet use, individuals with chronic diseases may have a stronger incentive to use the Internet effectively.

Online Health Information Seeking in Middle-Income Countries

While research is limited, online health information seeking also appears to be important in middle-income countries. In 2011, the international health insurer Bupa surveyed online health information seeking among Internet users in 12 high- and middle-income countries [40]. The researchers found higher rates of health information seeking in middle income countries (China 94%, Thailand 93%, and Saudi Arabia 91%) than in high-income countries (Australia 77%, United Kingdom 70%, and Spain 71%). Similarly, a 2010 Bupa study found 95% of Russian Internet users sought advice on health, medicines, or medical conditions online [41]. Bupa researchers attributed the high rates of online health information seeking in middle-income countries to the high cost of medical consultations and concerns over service quality. While not peer reviewed, these Bupa surveys point to a particularly important role for health-related searches outside of high-income countries. Conversely, these studies investigated only the propensity to access health information among Internet users, leaving aside international comparisons of how effectively online health information is used across social and economic gradients. The relationship between the need for health information and access to the Internet was not investigated.

Search Studies in Middle-Income Countries

As recently as 2009, researchers suggested that Google Trends was unsuitable for disease surveillance outside of developed countries due to insufficient Internet access [42]. However, the rapid increase of Internet use in middle-income countries suggests otherwise. Internet use is forecast to grow considerably more quickly by 2015 in middle-income than high-income countries (see Table 1; [43]). Since 2009, studies from Southeast Asia [44], Latin America [45], Russia [46], and China [47] suggest that search pattern studies are increasingly regarded as valid and reliable methods of disease surveillance in middle-income countries.

The potential of search patterns to improve public health surveillance in middle-income countries is well documented. First, online surveillance offers immediate insights into the present status of disease. That is, online surveillance may "predict the present" [48] without the reporting lags associated with complicated reporting procedures in public health bureaucracies [44]. Second, online surveillance may overcome the weaknesses of traditional surveillance systems, such as poor sensitivity to new diseases [49] and the lack of skills and equipment required for early disease detection [50]. Third, searches may overcome underreporting gaps from the private sector and from individuals who do not seek formal medical care [51]. Fourth, online surveillance may improve transparency. Central or regional governments may wish to minimize reports of disease outbreaks that could affect tourism or political reputation [52] or what sensitive issues surveys may not reveal [53,54]. In summary, online surveillance has the potential to improve disease surveillance in populations bearing the greatest burden of disease.

Consistent with the aims of infodemiology, our exploratory study examined "the science of distribution and determinants of information...(on) the Internet, or in a population, with the ultimate aim to inform public health and public policy" [55]. We examined the relationship between Internet search patterns, disease prevalence, and the determinants of Internet access using the case of HIV in a middle-income country. Through investigating these relationships, we aimed to develop methods to complement traditional HIV surveillance in Russia and contribute to the science of health-related searches.

Table 1. Changes in Internet use in selected middle- and high-income countries (values indicate penetration in %, ie, number of users divided by population).

Country	2009 actual Internet use	2015 predicted Internet use
China	28	47
India	7	19
Brazil	33	74
Russia	31	55
Indonesia	12	37
United States	70	73
Japan	74	81

Methods

Summary

This exploratory study sought to determine if search methods can be used for disease surveillance in a large middle-income country with a dispersed population. We first assessed whether online surveillance is a valid and reliable method for monitoring HIV in the Russian Federation. Second, we analyzed the determinants of Internet access to determine the extent that they explain regional variations in searches for the Russian terms for "HIV" and "AIDS".

Google and Yandex Searches in Russia

Most search pattern studies have used Google Trends (or the defunct Google Insights for Search) as the data source. Google Trends has been deployed in studies of influenza [3], dengue [11], and HIV [9]. However, the structure of the Russian-language Internet market is unique. Whereas Google provided 84% of global Internet search queries in May 2011 [56], Google's market share in Russia was only 25% in 2010/2011 [57]. The largest search provider in Russia in 2011 was Yandex, with 60% market share. In 2011, Russia overtook Germany as the European country with the highest number of unique visitors online [58]. Russian Internet users grew from 43% of the population in 2010 to 55% in 2012 [59]. In Russia, Yandex is a strong commercial competitor of Google.

The publicly available Google Trends data for Russia has several limitations. First, Google does not provide complete results, returning only subregions with the highest search volume. Google data were available for only 16 of Russia's 89 subregions for the term "HIV" and 29 for the term "AIDS" during 2011. Second, Google does not provide raw search data. This makes direct comparisons between subregions and matching with variables representing Internet access determinants complex. We used WordStat as the primary data source, as Yandex made publicly available a complete raw search dataset for all Russian regions and subregions for the full 12 months of 2011. We used Google Trends as a secondary source of aggregated search results for validation purposes.

Case Study: Why is Search-Based HIV Surveillance Important in Russia?

HIV is a serious health problem in the Russian Federation. Russia has the highest cumulative number of PLHIV of any European country, largely concentrated among people who inject drugs (PWID). On December 31, 2011, there were 650,100 PLHIV registered in Russia [60]. However, the true scale of HIV in Russia has long been the subject of considerable debate [61,62]. Feshbach and colleagues' 2005 study compiled data from official and unofficial Russian sources, as well as international agencies, to assess the quality of Russian HIV statistics [63]. The authors suggested that official Russian HIV data are frequently inconsistent, diverge markedly from alternative sources such as UNAIDS (the Joint United Nations Programme on HIV/AIDS), and present major methodological obstacles. The authors concluded that official Russian estimates of HIV prevalence were understated by a multiple of three to five times. Similar findings emerged from a 2007 UNODC (United Nations Office on Drugs and Crime) report that evaluated national data collection mechanisms related to HIV among PWIDs in nine lower income countries including Russia [64].

HIV Surveillance and Hidden Populations

HIV surveillance is further complicated by Russian drug laws, police, medical, and public attitudes. Most international observers regard Russian drug laws as punitive, unsupported by scientific evidence, and ineffective [65]. A 2010 study into police behavior found widespread reports of extrajudicial policing practices, including extortion, torture, and rape of PWIDs [66]. Attitudes among medical staff too are generally negative towards PLHIV [67,68]. Public opinion is also generally negative towards individuals acquiring HIV sexually or through drug use [69]. As a consequence of professional and social attitudes, many PLHIV avoid contact with medical organizations and avoid testing for HIV. Literature suggests there are disincentives for Russian PLHIV accessing health information directly from health professionals.

International researchers generally regard Russian HIV-positive PWIDs as a population hidden from public health surveillance. Since the early 2000s, researchers have sought to improve population estimates and document the conditions experienced by Russian PWIDs living with HIV [70,71]. Traditional surveys and sampling methods among PWIDs are unreliable, as individuals may not report accurately on stigmatized and illegal behaviors. A 2011 study in Russia found that, among 193 HIV-positive participants, only 36% were aware of their HIV status [72]. Another study of HIV-positive Russian PWIDs found persistent high-risk behaviors associated with HIV transmission [73]. Among study participants, 25% had been refused access to medical care, 18% were refused employment or fired, and 6% were forced from family homes. Researchers found 39% of participants had probable clinical depression, and 37% had anxiety levels comparable to psychiatric inpatients. In summary, there is considerable evidence that Russia has large numbers of PLHIV, many of whom are likely to be alienated from the formal health system and be absent from official statistics. The high rates of Internet searches for health information, combined with stigmatization of HIV, suggest that the Internet may be an important resource for PLHIV in Russia.

Russian injecting drug users have generally avoided contact with the formal health system. Between 2004 and 2011, much of the contact with injecting drug users and other groups at high risk of HIV was conducted by donor-funded Russian non-governmental organizations (NGOs). The behavioral surveillance data collected by these NGOs also contributed to Russian national HIV reporting to UNAIDS. However, as the result of government pressures, the number of donor-funded harm reduction NGO projects in Russia decreased from 70 in 2007 to 20 in 2011 [74]. The decrease in NGOs may also have eroded the capacity for data collection from populations at risk of HIV. In 2012, Russia did not report any HIV behavioral surveillance data associated with injecting drug use and sex work [75]. In summary, the progressive dismantling of harm reduction projects in Russia means only surveillance data from individuals formally diagnosed with HIV in government clinics

are available. Injecting drug workers, sex workers, and others at risk of HIV have disappeared from Russian government reporting.

RQ1 Method: Is Search Surveillance a Valid Method for Monitoring HIV in Russia?

To answer this research question, we examined the relationship between HIV prevalence across the Russian Federation and Internet searches for the terms “HIV” and “AIDS”. First, we obtained HIV prevalence data for each region and subregion from the Russian Federal AIDS Centre [60]. We chose 2011 data as this was latest complete dataset available. The Russian Federal AIDS Centre publishes the most timely and comprehensive HIV dataset available. However, these data are limited to formally diagnosed PLHIV and likely exclude many individuals at risk of HIV, or of uncertain serostatus, who deliberately avoid contact with government health services.

Second, we selected two terms to represent HIV searches. These two search terms were “HIV” (VICh in Russian) and “AIDS” (SPID). We referred to the Google Trends related-terms feature [76] to ensure each term referred to the subject of this study. In the case of “HIV”, all terms were related to HIV, whereas the term “AIDS” revealed several unrelated terms (Table 2). For example, the second most popular term associated with “AIDS” referred to the computer game “need for speed”. Based on these results, we anticipated that the search term “HIV” would have a stronger positive correlation with HIV prevalence than the term “AIDS”.

Third, we aggregated Yandex searches for each month of 2011 to produce a single annual search figure for the terms “HIV” and “AIDS” for each of Russia’s 89 subregions (see the map in Multimedia Appendix 1; [77]) covering the data range January 1 to December 31, 2011. In Russian federal statistical compilations, several smaller subregions are routinely aggregated, producing 83 rather than 89 statistical subregions. In our calculations, we used population prevalence of HIV and searches rather than raw figures. This allowed comparison across regions and subregions. We used this single 2011 annual HIV search in our further calculations.

Fourth, we conducted Spearman correlations of per-capita Yandex monthly searches for the terms “HIV” and “AIDS” against HIV prevalence data for all Russian subregions in 2011. We repeated this process with each of eight Russian regions. This provided us with national and regional correlations between search and prevalence data for 2011.

Fifth, we obtained all available Google data for the terms “HIV” and “AIDS” for 2011 and repeated this analysis for validation purposes. Google search data for the term “HIV” were available for 16 regions and for “AIDS” for 29 regions. We then conducted Spearman correlations between Google and Yandex data for validation purposes.

Table 2. Google Trends—Related terms for HIV and AIDS in the Russian Federation in 2011.

Search related terms	Russian	Value
“HIV”		
symptoms HIV	симптомы вич	100
symptoms	симптомы	100
AIDS	спид	65
AIDS HIV	спид вич	65
HIV infection	вич инфекции	35
HIV signs	вич признаки	35
analysis for HIV	анализ на вич	35
HIV infection	вич инфекция	30
HIV dating	вич знакомства	25
HIV photo	вич фото	20
“AIDS”		
test AIDS	тест спид	100
need for speed	нид фор спид	75
AIDS HIV	спид вич	55
HIV	вич	55
AIDS info	спид инфо	50
AIDS centre	спид центр	45
AIDS symptoms	спид симптомы	25
AIDS photo	спид фото	25
AIDS test	спидтест	25
speed hack	спид хак	20

RQ2: What is the Relationship Between the Determinants of Internet Access and Searches for the Term "HIV" Across Russia?

The relationship between Internet search patterns for specific health problems and the prevalence of these problems in populations is now well established. However, Internet users may not be representative of overall populations. Further, the characteristics of the Internet using population cannot be directly ascertained from search pattern data. We sought to extend understanding of the characteristics of Internet searching populations through data matching the determinants of Internet access (ie, age, income, broadband access price, and urban to rural ratios) with search patterns through multivariate analysis.

Several studies have examined the socioeconomic factors associated with HIV prevalence and injecting drug use in Russia. Moran et al investigated the relative importance of several variables in influencing HIV prevalence in a cross-sectional study based solely on Russian federal government statistics [78]. The authors found urbanization, mobility, crime, and income growth associated with HIV prevalence. In 2011, researchers surveyed 711 PWIDs in two large provincial cities [79]. The researchers concluded PWIDs were typical Russians when compared with a random population. However, investigators drew their random sample from 2004 household survey data. While Russian per capita income grew from US $9800 in 2004 to US $17,000 in 2011 [80], the authors did not comment on this potentially important confounder. These two studies illustrate the logistic difficulties of obtaining timely, valid, and independent data in Russia.

Our Methodology

We examined the relationship between spatial patterns of online searches for the term "HIV" and the determinants of Internet access. We used data from RQ1 in our analysis. In RQ1, we demonstrated the relationship between HIV prevalence and searches for "HIV". While this relationship was generally strong, differences in search patterns across regions may reflect differences in the determinants of Internet access as well as differences in HIV rates.

We selected principal component analysis (PCA) to explore the relationship between the determinants of Internet access and searches for the term "HIV". PCA is a method of multivariate analysis for finding patterns in data rather than hypothesis testing. PCA aids in the interpretation of relationships in the original data by transforming the original variables into a new set of variables, the principal components [81,82]. PCA has been widely used in public health to study relationships of health problems to socioeconomic variables. For example, PCA has been used to investigate European tumor prevalence [83], nutritional epidemiology in Greece [84], and epidemiological analysis in low- and middle-income countries [85]. As a consequence, we considered PCA an appropriate method for this exploratory study.

First, we collated the data sources. We obtained search pattern data for the term "HIV" through PCA. We obtained Russian-language data for five determinants of Internet access (see Table 3) from the Russian federal statistics agency for 83 Russian statistical regions [86]. The determinants of Internet access comprised each a single figure for each subregion for 2011. In compiling our data, we sought to most closely align search pattern, HIV prevalence, and determinants of Internet access data.

Second, we conducted PCA on all Russian subregions to produce a national level analysis. We included the determinants of Internet access and per capita search for "HIV" for all subregions. We used a correlation matrix approach to standardize the variables, as we used different units with differing variances. Based on our review of Internet determinants, we anticipated that the variables we chose to analyze would correlate.

Third, we conducted separate PCAs to examine the relationship between HIV search patterns and the determinants of Internet access separately on each of the eight Russian regions. Previous research suggests that, while there is no minimum of variables and cases in PCA, a larger number is preferable [87]. In designing this study, we purposely selected a smaller number of variables. We did this to permit analysis of both national data, as well as of regions with smaller number of subregions. Through analyzing both national and regional PCA separately, we anticipated we would identify additional spatial relationships not obvious at the national level.

Table 3. Determinants of Internet access—List of variables in PCA.

Variable	Determinant of Internet access (abbreviation)
Variable 1	Higher education students per 100,000 population (age)
Variable 2	Percentage aged 25-64 with higher education (education)
Variable 3	Gross regional product per capita (income)
Variable 4	Broadband price per month (Bband price)
Variable 5	Urban / rural population (urbanization)
Variable 6	Searches for HIV per 100,000 population during 2011 (search)

Results

Correlations

We first investigated search surveillance as a valid method for monitoring HIV in Russia. We found generally strong correlations between HIV prevalence and searches for the terms "HIV" and "AIDS". Yandex searches for "HIV" were very strongly correlated with HIV prevalence (Spearman rank-order coefficient [*rs*]=.881, *P*≤.001), whereas "AIDS" was strongly correlated nationally (*rs*=.714, *P*≤.001) (see Table 4). The strength of correlations varied across Russian regions. Several regions were less strongly correlated in Yandex. For example, HIV prevalence and searches in the central and northwestern regions were moderately correlated as a result of outlier data points. Further, Google national searches for the term "HIV" were moderately correlated (*rs*=.672, *P*=.004) with HIV prevalence and weakly correlated with Yandex searches for "HIV" (*rs*=.584, *P*≤.001) (see Table 5).

Second, we examined the relationship between the determinants of Internet access and search patterns for the term "HIV" across Russia. We found considerable variation in the relationship between these determinants and search patterns. We first analyzed national PCA results (Table 6). We determined the number of components to analyze using the Kaiser, Scree, and cumulative variance methods [82]. Kaiser and scree tests suggested three principal components (PCs), and the cumulative variance method suggested four PCs should be analyzed. In PC1 (the first and most important component), HIV search, age, and educational variables were moderately correlated. In PC2, per capita income was most important. This factor was weakly correlated with searches for HIV and explained 23% of the variance. The subsequent two components, which explain less variance, are more difficult to interpret.

Table 4. HIV and AIDS correlations from Yandex—National and all federal regions of Russian Federation.

Region	HIV prevalence per 100,000 pop'n	Searches for "HIV" per 1000 pop'n	Spearman correlation for HIV prevalence & "HIV" (2-tailed *P* value)	Searches for "AIDS" per 1000 pop'n	Spearman correlation for HIV prevalence & term "AIDS" (2-tailed *P* value)
National	446.513	16.995	.881 (≤.001)	19.312	.714 (≤.001)
Central	279.2	21.832	.377 (.006)[a]	21.215	-.123 (.386)
Northwestern	586.6	23.619	.482 (≤.001)[a]	26.383	.209 (.137)
Southern	144.3	8.397	.486 (≤.001)	12.665	.486 (≤.001)
North Caucuses	58.8	2.758	-.179 (.206)	6.666	-.286 (.040)
Volga	437.8	17.322	.793 (≤.001)	20.366	.380 (.005)
Urals	805	24.037	.657 (≤.001)	19.379	.429 (≤.001)
Siberian	528.1	13.962	.804 (≤.001)	15.561	.503 (≤.001)
Far east	166.4	7.473	.017 (.907)	11.197	.083 (.557)

[a]Outliers.

Table 5. Google Trends search results.

	AIDS searches	HIV searches
Number of regions	29	16
Spearman correlation, HIV prevalence (2-tailed *P* value)	.584 (*P*≤.001)	.672 (*P*=.004)
Spearman correlation, Google with Yandex (2-tailed *P* value)	-.289 (*P*=.129)	.223 (*P*=.406)

Table 6. National PCA results.

Importance of components		PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation		1.386	1.172	0.989	0.854	0.737	0.674
Proportion of variance		0.320	0.229	0.163	0.121	0.091	0.076
Cumulative proportion		0.320	0.549	0.712	0.834	0.924	1.000
Loadings							
	Age	-0.560	–	0.105	0.269	0.762	0.121
	Education	-0.479	-0.376	0.313	0.247	-0.345	-0.593
	Income	0.215	-0.652	–	-0.491	0.460	-0.268
	Bband price	0.329	-0.460	0.523	0.422	–	0.478
	Urbanization	0.131	-0.417	-0.777	0.396	-0.110	0.190
	Search	-0.533	-0.200	–	-0.539	-0.278	0.546

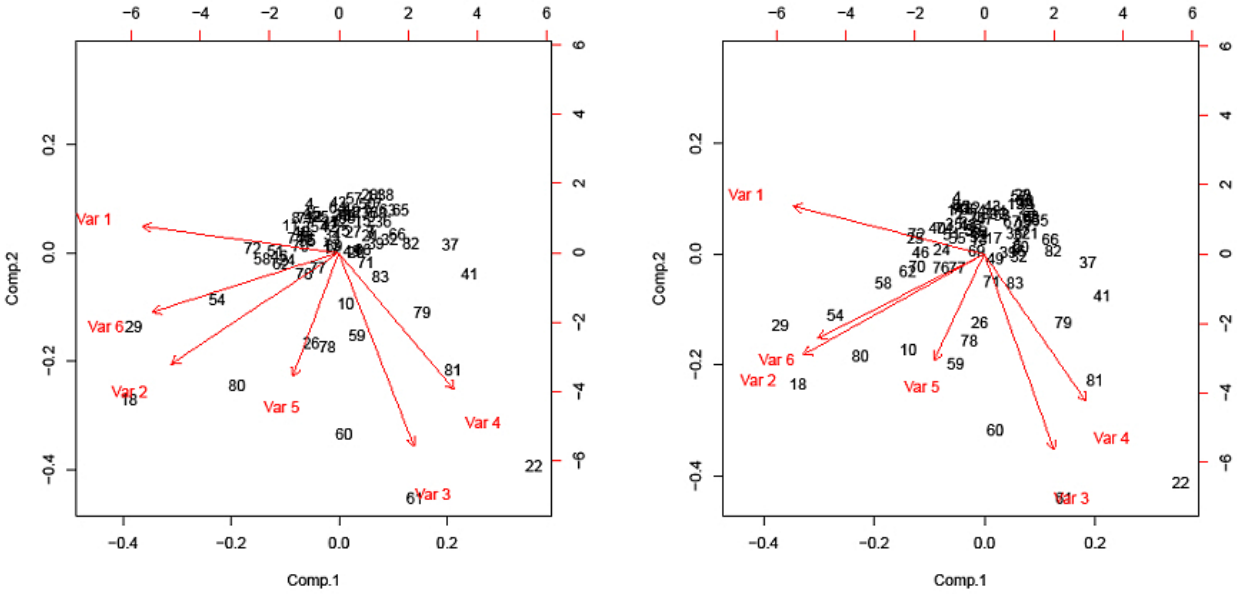
Biplots and Spatial Relationships

We used biplots to explain spatial relationships in our PCA results. Biplots provide a visual representation of PCA data from the first two PCs [88]. Biplots allow identification of clusters of subregions with similar characteristics. Further, the clustering of subregions along vector lines serves to highlight subregions more strongly associated with specific variables. Importantly, the clustering of subregions is subjective and requires additional analysis. On the national HIV search biplot (see Figure 1), PC1 was associated with Vector 3 (income), Vector 4 (broadband price), and Vector 5 (urbanization). PC2 was associated with Vector 1 (age), Vector 2 (education), and Vector 6 (HIV search). We obtained PCA results and biplots for all eight Russian regions. See Multimedia Appendix 2 for a list of subregions referenced in the national PCA. See Multimedia Appendices 3-5 for biplot results for each Russian federal region. Finally, we conducted a separate PCA for HIV prevalence data. We substituted the variable HIV search with HIV prevalence. The results of a PCA incorporating the variable HIV prevalence produced results with a similar form to those incorporating HIV searches at both the national and regional levels.

Table 7. Summary of national biplots.

Region	Relationship	Geographic clusters	Outliers
National	PC1:	Cluster 1	18. Moscow City
	V3, V4, V5	37. Ingushetia	22. Nenets Autonomous Region
	Income, broadband fees, urbanization	41. Chechnya	29. St Petersburg
		79. Amursk	61. Yamalo-Nenets
		81. Sakhalin	
		83. Chukhotka	
		Cluster 2	
		10. Moscow region	
		71. Kemerovo	
		26. Murmansk	
		78. Khabarovsk	
		59. Tyumen	
	PC2:	Cluster 3	
	V1, V2, V6	18. Moscow city	
	Age, education, and HIV prevalence/ search	24. Kaliningrad	
		29. St Petersburg City	
		46. Tatarstan	
		54. Samara	
		58. Sverdlovsk	
		62. Chelyabinsk	
		70. Irkutsk 76. Kamchatka 77. Primorsk 80. Magadan	

Figure 1. National HIV search and HIV prevalence biplots.



Discussion

Principal Findings

Overall, we found search patterns were a valid method of HIV surveillance in the Russian Federation. Furthermore, our research suggests that search patterns for HIV are generally not related to income or broadband price. However across Russian regions, we found considerable variation in the strength of correlations between search and disease prevalence, and the determinants of Internet access. Finally, our analysis suggested that the strong correlations between search and disease prevalence may indicate effective use of the Internet by individuals at risk of HIV and PLHIV.

RQ1: Is Search Surveillance a Valid Method for Monitoring HIV in Russia?

We found online search patterns for HIV were correlated with HIV prevalence in both Google and Yandex at the national level. It is noteworthy that the latest official Russian HIV data available at the time of writing in mid-2013 were for the year 2011. By contrast, Yandex search data were available with a delay of 4 weeks and Google data with a 48-hour delay. This timely availability illustrates the potential contribution of search pattern data to disease surveillance.

Second, we found considerable variation in the strength of correlations among regions in Yandex data. Overall, we found Yandex searches for the term “HIV” and HIV prevalence were most strongly correlated. This suggests PLHIV are more likely to search for “HIV” than “AIDS”. In the North Caucuses and far eastern regions, HIV prevalence was not positively correlated with search. We attribute this to the low HIV prevalence and low search volumes in these regions. By contrast, in the central and northwestern regions, search volumes and HIV prevalence were high, but correlations were moderate. We attributed the weaker correlations to outliers in Yandex data. Removing the central (Moscow subregion) and northwestern region (Leningrad subregion), outliers strengthened correlations from 0.377 to 0.551 and 0.482 to 0.939 respectively. This suggests correlation analysis should routinely account for outlying subregions.

Third, we found Google data are not adequate for subnational HIV surveillance in Russia. We attribute the low correlations to the multiple zero values present in our Google dataset. Of the 15 regions for which Google data were available, many months recorded a zero search value for the term “HIV”. These zero values were consistent with an earlier study [46] of the use of Google search for health policy analysis in Russia that found Google Trends requires an unknown threshold before results are displayed. While national level Google data were correlated with HIV prevalence, our analysis suggests it should not be used for regional analysis.

Finally, our results contribute to understanding of hidden populations of PLHIV in Russia. There is a general consensus that Russian HIV rates are underreported. Previous studies have reported considerable at-risk populations unaware of their HIV status in subregions with high HIV prevalence (eg, [74]). However, we found strong spatial correlations between official HIV rates and searches for HIV. This finding has several interpretations. First, the spatial variation in search results also appears in traditional surveillance. Researchers have found high populations of unknown HIV serostatus in subregions of high HIV prevalence. That is, additional searches for HIV related information by populations at risk of HIV and unknown serostatus may inflate already high search volumes in those subregions with high HIV prevalence. However, these additional searches would not change the overall spatial distribution of search patterns. A second interpretation relates to the search data used. Our analysis relied on annually aggregated search results. Our results are thus a static view of HIV prevalence over a 12-month period. This static view does not capture longitudinal anomalies in search patterns. While this 12-month snapshot was appropriate for the purposes of this study,

monitoring of weekly and monthly search patterns may produce different results and reveal spatial variations in searches.

RQ2: What is the Relationship Between the Determinants of Internet Access and Search Patterns for the Term "HIV" Across Russia?

We analyzed national and regional PCA results separately. First, we examined the two national level biplots. One biplot incorporated HIV prevalence as a variable and the other searches for "HIV". All other variables remained consistent (see Figure 1). We found these two biplots to be near isomorphic.

The separate national biplots containing both HIV prevalence and HIV search produced three logically coherent geographical clusters. National cluster 1 was characterized by low-income, non-ethnic Russian subregions with low HIV prevalence. The exception to this is the Sakhalin subregion, with high per capita income. This clustering occurred along the broadband vector (V4), suggesting high broadband prices and limited access. National cluster 2 was associated with the urbanization vector (V5). It includes urbanized non-metropolitan areas. National cluster 3 included the Russian cities with the highest prevalence of HIV, along the HIV search/prevalence (V6) and education vectors (V2). In addition, it included the Magadan subregion. The Magadan subregion is highly urbanized but has a low HIV prevalence and low population. We attribute the inclusion of Magadan in this cluster as an indicator of potential HIV risk. Conversely, the isolation of Magadan, in northeastern Russia, away from borders and drug routes, suggests a lower risk of HIV transmission through injecting drug use.

Both national PCA biplots featured several outlier subregions (see Figure 1). For example, the Yamalo Nenets subregion (61) was an outlier. This is an oil-producing subregion, with very high per capita incomes and below average HIV rates. It was strongly associated with the income vector (V3). Second, Moscow and St Petersburg were outliers. We attribute these cities' outlier positions to a statistical anomaly. Each city had a rural to urban ratio of zero and very high Internet access rates and incomes. In summary, national level PCA analysis of both the HIV prevalence and HIV search biplots suggested a stronger relationship between broadband access prices in several subregions. See Multimedia Appendix 4 for a discussion of PCA in Russian regions.

In summary, PCA is not a technique that establishes causal relationships. However, based on our preliminary analysis, we suggest that income and broadband prices do not generally appear to be associated with HIV searches, either positively or negatively, in the subregions of highest HIV prevalence. Further research, in the form of confirmatory factor analysis and regression analysis is needed to establish this relationship statistically. Contingent upon the results of this additional analysis, HIV search pattern data may be incorporated into HIV modeling.

Our findings extend beyond an examination of the digital divide in Russia as defined by access to the Internet. There is also a behavioral dimension implicit in our two research questions. Search patterns measure aggregate behavior at the population level, with important issues more frequently searched. Searches for the term "HIV" measure the importance of this disease in a population. Consequently, the generally strong correlations between search patterns and disease prevalence lead us to infer that the Internet is being used effectively by PLHIV. That is, searchers for "HIV" demonstrate the technical competence to search for health information they consider important. However, this is a cautious conclusion, and one that merits further research.

Further Research and Limitations

Our research suggests that further exploratory analysis applying search pattern methods to HIV surveillance in Russia is warranted. First, PWID and sex worker populations may be at increased risk of HIV as the result of the Russian government's censorship of prevention, treatment, and care information [89,90] and decreased behavioral monitoring capacity among internationally funded NGOs. Further, in 2013, concerns emerged about the capacity of independent Russian social research organizations to continue unencumbered data collection [91]. Search methods may present a partial solution to these emerging information constraints. Internet search patterns provide a valid near real time measure of health behaviors in the field at population level.

Second, additional research is required to establish how effectively Russians use the Internet for HIV and health information. Qualitative and survey research among populations at risk of HIV and PLHIV will assist the further development of search surveillance methods and the planning of online interventions. Research in Russia should also examine the quality of health information available to PLHIV, both through domestic and international Russian language websites.

Third, organizations working with at-risk populations and PLHIV may consider initiating studies that establish baseline measures of search patterns for HIV and related diseases. From these baselines, longitudinal studies will be able to rapidly identify unanticipated shifts in spatial and temporal patterns of HIV-related searches and HIV prevalence, well in advance of official incidence and prevalence data.

Fourth, the method described in this paper can be extended to other communicable and non-communicable diseases in Russian-speaking countries. Broader application of this method may require initial disease-by-disease and country-by-country validation. However, even without validation, this method provides a low-cost, rapid, timely initial assessment with which to shape further planning, analysis, and decision making.

Finally, our research had several limitations. First, we were constrained by the absence of time series data. To conduct data matching for the PCA, we used a single aggregate figure to represent total searches nationally and within each subregion. We believe this analysis would be strengthened by a month-to month comparison of HIV prevalence data in each Russian region. Such data are not publicly available. Second, Google is the only data source in most middle income countries. This limits the application of this method. An important exception is China, where Baidu is increasingly being used alongside Google for disease surveillance.

Conclusions

The use of data for disease surveillance has been widely promoted in popular literature. Under the rubric of "big data", journalists have popularized the novel application of Internet search patterns in medicine [92]. Scholars too, have speculated that data availability will lead to the evolution of new models of disease surveillance [93-95]. While the potential application of large scale data analysis in health care has generated considerable popular and scholarly interest, most research has focused on high-income countries with well-functioning public health information systems. We believe it is in middle-income countries that search methods can make the greatest contribution to public health. It is in these countries that traditional surveillance systems are least developed and health data least available.

Clearly, a digital divide between rich and poor countries persists. However, Internet access in middle-income countries is growing rapidly, and online health information is in demonstrably high demand. Based on our preliminary research, we are cautiously optimistic in suggesting that access to the Internet should therefore not be considered a constraint to conducting search studies beyond high-income countries. It is in lower income countries that search pattern surveillance may move beyond a statistical novelty and be incorporated into local health data collection and decision making.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Map of Russian Federation.

[PNG File, 546KB - jmir_v15i11e256_app1.png]

Multimedia Appendix 2

Reference list of Russian regions and subregions for PCA biplots.

[PDF File (Adobe PDF File), 63KB - jmir_v15i11e256_app2.pdf]

Multimedia Appendix 3

Narrative analysis of PCA in Russian regions.

[PDF File (Adobe PDF File), 22KB - jmir_v15i11e256_app3.pdf]

Multimedia Appendix 4

PCA biplots - Russian regional HIV search and HIV prevalence.

[PDF File (Adobe PDF File), 477KB - jmir_v15i11e256_app4.pdf]

Multimedia Appendix 5

Table summarizing PCA biplot results in Russian regions.

[PDF File (Adobe PDF File), 32KB - jmir_v15i11e256_app5.pdf]

References

1. Granka L. Inferring the Public Agenda from Implicit Query Data. Boston: ACM SIGIR Conference; 2009. URL: http://ceur-ws.org/Vol-512/paper07.pdf [accessed 2013-11-05] [WebCite Cache ID 6KtuAa2j2]
2. 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]
3. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
4. 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]
5. Pervaiz F, Pervaiz M, Abdur Rehman N, 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]
6. 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. [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]

7. 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]
8. Wong PW, Fu KW, Yau RS, Ma HH, Law YW, Chang SS, et al. Accessing suicide-related information on the internet: a retrospective observational study of search behavior. J Med Internet Res 2013;15(1):e3 [FREE Full text] [doi: 10.2196/jmir.2181] [Medline: 23305632]
9. Jena AB, Karaca-Mandic P, Weaver L, Seabury SA. Predicting new diagnoses of HIV infection using internet search engine data. Clin Infect Dis 2013 May;56(9):1352-1353. [doi: 10.1093/cid/cit022] [Medline: 23334812]
10. Fox S, Duggan M. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2013. Health Online 2013 URL: http://www.pewinternet.org/~/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2013-10-17] [WebCite Cache ID 6KQX6Gwdl]
11. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
12. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. BMC Cancer 2011;11:442 [FREE Full text] [doi: 10.1186/1471-2407-11-442] [Medline: 21993136]
13. Peslak A. First Monday. 2004 Mar 01. An analysis of regional and demographic differences in United States Internet usage URL: http://firstmonday.org/ojs/index.php/fm/article/view/1124 [accessed 2013-11-05] [WebCite Cache ID 6KtlONfCd]
14. Vicente MR, López AJ. Some empirical evidence on Internet diffusion in the New Member States and Candidate Countries of the European Union. Applied Economics Letters 2008 Oct 24;15(13):1015-1018. [doi: 10.1080/13504850600972352]
15. Chinn MD, Fairlie RW. ICT use in the developing world: an analysis of differences in computer and internet penetration. Review of International Economics 2010;18(1):153-167. [doi: 10.1111/j.1467-9396.2009.00861.x]
16. Li N, Kirkup G. Gender and cultural differences in Internet use: A study of China and the UK. Computers & Education 2007 Feb;48(2):301-317. [doi: 10.1016/j.compedu.2005.01.007]
17. Gil-Garcia JR, Helbig NC, Ferro E. Is it only about Internet access? An empirical test of a multi-dimensional Digital Divide. In: EGOV 2006 proceedings. Berlin: Springer; 2006 Presented at: Electronic Government 5th International Conference; Sept. 4-8, 2006; Krakow, Poland. [doi: 10.1007/11823100_13]
18. Chaudhuri A, Flamm KS, Horrigan J. An analysis of the determinants of internet access. Telecommunications Policy 2005 Oct;29(9-10):731-755. [doi: 10.1016/j.telpol.2005.07.001]
19. Demoussis M, Giannakopoulos N. Facets of the digital divide in Europe: Determination and extent of internet use. Economics of Innovation and New Technology 2006 Apr;15(3):235-246. [doi: 10.1080/10438590500216016]
20. Horrigan JB. Federal Communications Commission. 2010. Broadband adoption and use in America URL: http://transition.fcc.gov/DiversityFAC/032410/consumer-survey-horrigan.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGvdN15]
21. Lera-López F, Billon M, Gil M. Determinants of Internet use in Spain. Economics of Innovation and New Technology 2011 Mar;20(2):127-152. [doi: 10.1080/10438590903378017]
22. Qiang CZW, Rossotto CM, Kimura K. Economic impacts of broadband. In: Information and Communications for Development. Washington, DC: World Bank; 2009:35-50.
23. OECD. 2001. Understanding the Digital Divide URL: http://www.oecd.org/internet/ieconomy/1888451.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGpjtkV]
24. Hilbert M. The end justifies the definition: The manifold outlooks on the digital divide and their practical usefulness for policy-making. Telecommunications Policy 2011 Sep;35(8):715-736. [doi: 10.1016/j.telpol.2011.06.012]
25. Hamel JY. UNDP Human Development Research Paper. New York: UNDP; 2010. ICT4D and the human development and capabilities approach: the potentials of information and communication technology URL: http://hdr.undp.org/es/informes/mundial/idh2010/trabajos/HDRP_2010_37.pdf [accessed 2013-08-30] [WebCite Cache ID 6JFiv5EWO]
26. Fox S, Duggan M. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2013. Health Online 2013 URL: http://www.pewinternet.org/Reports/2013/Health-online.aspx [accessed 2013-08-14] [WebCite Cache ID 6IrGmeWvy]
27. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
28. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
29. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
30. Fox S, Purcell K. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2010. Chronic disease and the Internet URL: http://web.pewinternet.org/~/media/Files/Reports/2010/PIP_Chronic_Disease_with_topline.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGj0jbz]
31. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. Urology 2013 Jan;81(1):37-42. [doi: 10.1016/j.urology.2011.04.024] [Medline: 21676450]

32. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
33. Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information seeking. J Health Commun 2008 Mar;13(2):181-199 [FREE Full text] [doi: 10.1080/10810730701854110] [Medline: 18300068]
34. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
35. Fox S. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2008. The engaged e-patient population URL: http://nicolaziady.com/wp-content/uploads/2011/07/PIP_Health_Aug08.pdf.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGdiDcm]
36. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
37. Kalichman SC, Cherry C, Cain D, Weinhardt LS, Benotsch E, Pope H, et al. Health information on the Internet and people living with HIV/AIDS: information evaluation and coping styles. Health Psychol 2006 Mar;25(2):205-210. [doi: 10.1037/0278-6133.25.2.205] [Medline: 16569112]
38. Kalichman SC, Benotsch EG, Weinhardt LS, Austin J, Luke W. Internet use among people living with HIV/AIDS: association of health information, health behaviors, and health status. AIDS Educ Prev 2002 Feb;14(1):51-61. [Medline: 11900110]
39. Samal L, Saha S, Chander G, Korthuis PT, Sharma RK, Sharp V, et al. Internet health information seeking behavior and antiretroviral adherence in persons living with HIV/AIDS. AIDS Patient Care STDS 2011 Jul;25(7):445-449 [FREE Full text] [doi: 10.1089/apc.2011.0027] [Medline: 21682586]
40. BUPA. Bupa health pulse 2011: Global trends, attitudes and influences. London: BUPA; 2011. URL: http://www.bupa.com/media/288798/bupa_health_pulse_report_2011.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGWUag4]
41. McDaid D, Park AL. Online health: untangling the web. London: BUPA; 2010. URL: http://www.bupa.com/media/44806/online_20health_20-_20untangling_20the_20web.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGTnXXS]
42. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
43. Aguiar M, Boutenko V, Michael D, Rastogi V, Subramanian A, Zhou Y. The Internet's New Billion: Digital Consumers in Brazil, Russia, India, China, and Indonesia. Boston: Boston Consulting Group; 2010. URL: http://www.bcg.com/documents/file58645.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGR8Yom]
44. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011 May;5(5):e1206 [FREE Full text] [doi: 10.1371/journal.pntd.0001206] [Medline: 21647308]
45. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]
46. Zheluk A, Gillespie JA, Quinn C. Searching for truth: internet search patterns as a method of investigating online responses to a Russian illicit drug policy debate. J Med Internet Res 2012;14(6):e165 [FREE Full text] [doi: 10.2196/jmir.2270] [Medline: 23238600]
47. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
48. Choi H, Varian H. Predicting the present with google trends. Economic Record 2012 Jun;88(s1):2-9. [doi: 10.1111/j.1475-4932.2012.00809.x]
49. Chunara R, Freifeld CC, Brownstein JS. New technologies for reporting real-time emergent infections. Parasitology 2012 Dec;139(14):1843-1851 [FREE Full text] [doi: 10.1017/S0031182012000923] [Medline: 22894823]
50. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
51. Khan K, McNabb SJ, Memish ZA, Eckhardt R, Hu W, Kossowsky D, et al. Infectious disease surveillance and modelling across geographic frontiers and scientific specialties. Lancet Infect Dis 2012 Mar;12(3):222-230. [doi: 10.1016/S1473-3099(11)70313-9] [Medline: 22252149]
52. Madoff LC, Fisman DN, Kass-Hout T. A new approach to monitoring dengue activity. PLoS Negl Trop Dis 2011 May;5(5):e1215 [FREE Full text] [doi: 10.1371/journal.pntd.0001215] [Medline: 21647309]
53. Curioso WH, Kurth AE. Access, use and perceptions regarding Internet, cell phones and PDAs as a means for health promotion for people living with HIV in Peru. BMC Med Inform Decis Mak 2007;7:24 [FREE Full text] [doi: 10.1186/1472-6947-7-24] [Medline: 17850656]
54. Stephens-Davidowitz S. The cost of racial animus on a black presidential candidate: using Google search data to find what surveys miss. SSRN Journal 2012:1-55. [doi: 10.2139/ssrn.2238851]
55. 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]

56. Netmarketshare.com. Aliso Viejo, CA: Net Applications; 2012 Jun 15. Search Engine Market Share URL: http://www.netmarketshare.com/report.aspx?qprid=4&qptimeframe=M&qpsp=148 [accessed 2012-06-16]
57. Crisp T. ComScore Voices. Reston, VA: comscore.com; 2011. Yandex: From Russia with Love URL: http://blog.comscore.com/2011/06/yandex_russia_with_love.html [accessed 2013-08-14] [WebCite Cache ID 6IrGCavjr]
58. Comscore Press Releases. London, UK; 2011 Nov 14. comScore Releases Overview of European Internet Usage in September 2011 URL: http://www.comscore.com/Insights/Press_Releases/2011/11/comScore_Releases_Overview_of_European_Internet_Usage_in_September_2011 [accessed 2013-11-05] [WebCite Cache ID 6KtrskSAe]
59. WCIOM. Russ Public Opin Res Cent WCIOM. Moscow; 2012 Feb 13. Russians networking: rating of social media URL: http://wciom.com/index.php?id=61&uid=63 [accessed 2012-04-05] [WebCite Cache ID 6BTiwJdNa]
60. Pokrovsky V, Ladnaya N, Sokolova E, Buravtsova E. HIV infection - Information Bulletin No 36. Moscow: Federal Scientific-methodological center for the prevention of AIDS; 2012. URL: http://www.hivrussia.ru/files/bul_36.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrOXIJ7V]
61. Naskhoev M, Sergeyev B. AIDS in the Commonwealth of Independent States. Chisinau: UNAIDS; 2008. URL: http://aids.md/aids/files/1012/CIS%20MAP%20Report_English%20Final%204%20print.doc [accessed 2013-11-06] [WebCite Cache ID 6KtneBHLq]
62. Maier CB, Martin-Moreno JM. Quo vadis SANEPID? A cross-country analysis of public health reforms in 10 post-Soviet states. Health Policy 2011 Sep;102(1):18-25. [doi: 10.1016/j.healthpol.2010.08.025] [Medline: 20864203]
63. Feshbach M, Galvin CM. Wash DC Woodrow Wilson Int Cent Sch. 2005. HIV/AIDS in Russia an analysis of statistics URL: https://docs.google.com/file/d/0ByO7HajOC-FJOGM4NTRlM2MtNWMwOC00ZmY1LWJhOGQtOGEyMzU5OWRiYTk5/edit?usp=sharing [accessed 2013-09-14] [WebCite Cache ID 6JcUjo7lr]
64. Burrows D, Birgin R, Burns K, Zheluk A. Measuring Coverage of HIV Prevention and Care Services for Injecting Drug Users. Vienna: UNODC; 2009.
65. Elovich R, Drucker E. On drug treatment and social control: Russian narcology's great leap backwards. Harm Reduct J 2008;5:23 [FREE Full text] [doi: 10.1186/1477-7517-5-23] [Medline: 18577225]
66. Sarang A, Rhodes T, Sheon N, Page K. Policing drug users in Russia: risk, fear, and structural violence. Subst Use Misuse 2010 May;45(6):813-864 [FREE Full text] [doi: 10.3109/10826081003590938] [Medline: 20397872]
67. Gerber T, Mendelson S. csis.org. Madison: University of Wisconsin-Madison; 2006 Jan. A Survey of Russian Doctors on HIV/AIDS URL: http://csis.org/files/media/csis/pubs/060707_ruseura_mendelsondocrepjan06.pdf [accessed 2013-08-29] [WebCite Cache ID 6JDyt6pqu]
68. Bikmukhametov DA, Anokhin VA, Vinogradova AN, Triner WR, McNutt LA. Bias in medicine: a survey of medical student attitudes towards HIV-positive and marginalized patients in Russia, 2010. J Int AIDS Soc 2012;15(2):17372 [FREE Full text] [Medline: 23031336]
69. Balabanova Y, Coker R, Atun RA, Drobniewski F. Stigma and HIV infection in Russia. AIDS Care 2006 Oct;18(7):846-852. [doi: 10.1080/09540120600643641] [Medline: 16971297]
70. Kozlov AP, Shaboltas AV, Toussova OV, Verevochkin SV, Masse BR, Perdue T, et al. HIV incidence and factors associated with HIV acquisition among injection drug users in St Petersburg, Russia. AIDS 2006 Apr 4;20(6):901-906. [doi: 10.1097/01.aids.0000218555.36661.9c] [Medline: 16549975]
71. Heimer R, White E. Estimation of the number of injection drug users in St. Petersburg, Russia. Drug Alcohol Depend 2010 Jun 1;109(1-3):79-83 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.12.010] [Medline: 20060238]
72. Niccolai LM, Verevochkin SV, Toussova OV, White E, Barbour R, Kozlov AP, et al. Estimates of HIV incidence among drug users in St. Petersburg, Russia: continued growth of a rapidly expanding epidemic. Eur J Public Health 2011 Oct;21(5):613-619 [FREE Full text] [doi: 10.1093/eurpub/ckq115] [Medline: 20798184]
73. Amirkhanian YA, Kelly JA, Kuznetsova AV, DiFranceisco WJ, Musatov VB, Pirogov DG. People with HIV in HAART-era Russia: transmission risk behavior prevalence, antiretroviral medication-taking, and psychosocial distress. AIDS Behav 2011 May;15(4):767-777. [doi: 10.1007/s10461-010-9793-x] [Medline: 20803063]
74. esvero.ru. Global Fund Transitional Funding Mechanism single country application: Sections 1-2. Geneva: theglobalfund.org; 2012. URL: http://www.theglobalfund.org/grantdocuments/RUS(ESVERO)-TFM-HA_Proposal_0_en/ [accessed 2013-08-14] [WebCite Cache ID 6IrFRcmKU]
75. UNAIDS. Report on the Global AIDS Epidemic. Geneva; 2012. URL: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrFNuEBS]
76. Google.com. What are top searches?. Mountain View, CA; 2013. URL: https://support.google.com/trends/answer/99742?hl=en [accessed 2013-08-30] [WebCite Cache ID 6JFi8oHwv]
77. Russian Regions.: Wikipedia.org; 2010 Mar 09. URL: http://en.wikipedia.org/wiki/File:Russian_Regions-EN.svg [accessed 2013-09-03] [WebCite Cache ID 6JLYxbbIc]
78. Moran D, Jordaan JA. HIV/AIDS in Russia: determinants of regional prevalence. Int J Health Geogr 2007;6:22 [FREE Full text] [doi: 10.1186/1476-072X-6-22] [Medline: 17553136]

79. Wall M, Schmidt E, Sarang A, Atun R, Renton A. Sex, drugs and economic behaviour in Russia: a study of socio-economic characteristics of high risk populations. Int J Drug Policy 2011 Mar;22(2):133-139. [doi: 10.1016/j.drugpo.2010.10.001] [Medline: 21055913]
80. CIA World Fact Book - Russia. Washington, DC: Central Intelligence Agency; 2013. URL: https://www.cia.gov/library/publications/the-world-factbook/geos/rs.html [accessed 2013-08-14] [WebCite Cache ID 6IrG6HL0c]
81. Joliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. Stat Methods Med Res 1992;1(1):69-95. [Medline: 1341653]
82. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pr Assess Res Eval 2005;10(7):1-9 [FREE Full text]
83. Decarli A, La Vecchia C. Environmental factors and cancer mortality in Italy: correlational exercise. Oncology 1986;43(2):116-126. [Medline: 3951786]
84. Kourlaba G, Panagiotakos DB, Mihas K, Alevizos A, Marayiannis K, Mariolis A, et al. Dietary patterns in relation to socio-economic and lifestyle characteristics among Greek adolescents: a multivariate analysis. Public Health Nutr 2009 Sep;12(9):1366-1372. [doi: 10.1017/S1368980008004060] [Medline: 19063765]
85. Howe LD, Galobardes B, Matijasevich A, Gordon D, Johnston D, Onwujekwe O, et al. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: a methods of measurement in epidemiology paper. Int J Epidemiol 2012 Jun;41(3):871-886 [FREE Full text] [doi: 10.1093/ije/dys037] [Medline: 22438428]
86. Russian Federal State Statistics Service. Moscow: Rosstat; 2013. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/ [accessed 2013-08-14] [WebCite Cache ID 6IrF5tUEs]
87. Jolliffe IT. Principal component analysis. New York: Springer; 2002.
88. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. Biometrika 1971 Dec;58(3):453. [doi: 10.2307/2334381]
89. HRW. Russia: Government Shuts HIV-Prevention Group's Website. 2012. URL: http://www.hrw.org/news/2012/02/08/russia-government-shuts-hiv-prevention-group-s-website [WebCite Cache ID 6KxrHaruJ]
90. Golichenko M, Sarang A. Atmospheric pressure - Russian Drug Policy as a Driver for Violations of the UN Convention against Torture. Moscow: Andrey Rylkov Foundation; 2011. URL: http://www.aidslaw.ca/publications/interfaces/downloadFile.php?ref=1949 [accessed 2013-08-14] [WebCite Cache ID 6IrFBV2ZQ]
91. Shlapentokh V. ODRussia. London, UK: opendemocracy.net; 2013 Jun 05. Sticks and stones URL: http://www.opendemocracy.net/od-russia/vladimir-shlapentokh/sticks-and-stones [accessed 2013-08-14] [WebCite Cache ID 6IrFE80lT]
92. Mayer-Schönberger V, Cukier K. Big Data: A Revolution that Will Transform how We Live, Work, and Think. London, UK: Eamon Dolan/Houghton Mifflin Harcourt; 2013.
93. 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]
94. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
95. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013;10(4):e1001413 [FREE Full text] [doi: 10.1371/journal.pmed.1001413] [Medline: 23565065]

Abbreviations

AIDS: Acquired Immune Deficiency Syndrome
HIV: Human Immunodeficiency Virus
NGO: non-governmental organization
PC: principal component
PCA: principal components analysis
PLHIV: people living with HIV
PWID: people who inject drugs
UNAIDS: Joint United Nations Programme on HIV/AIDS
UNODC: United Nations Office on Drugs and Crime

Edited by G Eysenbach; submitted 03.09.13; peer-reviewed by A Jena; comments to author 27.09.13; revised version received 18.10.13; accepted 22.10.13; published 12.11.13.

Please cite as:
Zheluk A, Quinn C, Hercz D, Gillespie JA
Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study
J Med Internet Res 2013;15(11):e256
URL: http://www.jmir.org/2013/11/e256/
doi:10.2196/jmir.2936
PMID:24220250

Original Paper

Electronic Word of Mouth on Twitter About Physical Activity in the United States: Exploratory Infodemiology Study

Ni Zhang[1], MA, MPH, PhD; Shelly Campo[2], PhD; Kathleen F Janz[3], EdD; Petya Eckler[4], PhD; Jingzhen Yang[5], MPH, PhD; Linda G Snetselaar[6], PhD; Alessio Signorini[7], M Comp Sc

[1]The University of Iowa alumnus, Iowa City, IA, United States

[2]The University of Iowa, Department of Community and Behavioral Health, Iowa City, IA, United States

[3]The University of Iowa, Department of Health and Human Physiology, Iowa City, IA, United States

[4]University of Strathclyde, School of Humanities, Glasgow, United Kingdom

[5]Kent State University, Department of Social and Behavioral Sciences, Kent, OH, United States

[6]The University of Iowa, Department of Epidemiology, Iowa City, IA, United States

[7]The University of Iowa, Department of Computer Science, Iowa City, IA, United States

Corresponding Author:
Ni Zhang, MA, MPH, PhD
The University of Iowa alumnus
N400 CPHB, 105 River St
The University of Iowa
Iowa City, IA, 52242
United States
Phone: 1 319 541 4631
Fax: 1 319 384 1474
Email: nizhang515@gmail.com

Abstract

Background: Twitter is a widely used social medium. However, its application in promoting health behaviors is understudied.

Objective: In order to provide insights into designing health marketing interventions to promote physical activity on Twitter, this exploratory infodemiology study applied both social cognitive theory and the path model of online word of mouth to examine the distribution of different electronic word of mouth (eWOM) characteristics among personal tweets about physical activity in the United States.

Methods: This study used 113 keywords to retrieve 1 million public tweets about physical activity in the United States posted between January 1 and March 31, 2011. A total of 30,000 tweets were randomly selected and sorted based on numbers generated by a random number generator. Two coders scanned the first 16,100 tweets and yielded 4672 (29.02%) tweets that they both agreed to be about physical activity and were from personal accounts. Finally, 1500 tweets were randomly selected from the 4672 tweets (32.11%) for further coding. After intercoder reliability scores reached satisfactory levels in the pilot coding (100 tweets separate from the final 1500 tweets), 2 coders coded 750 tweets each. Descriptive analyses, Mann-Whitney *U* tests, and Fisher exact tests were performed.

Results: Tweets about physical activity were dominated by neutral sentiments (1270/1500, 84.67%). Providing opinions or information regarding physical activity (1464/1500, 97.60%) and chatting about physical activity (1354/1500, 90.27%) were found to be popular on Twitter. Approximately 60% (905/1500, 60.33%) of the tweets demonstrated users' past or current participation in physical activity or intentions to participate in physical activity. However, social support about physical activity was provided in less than 10% of the tweets (135/1500, 9.00%). Users with fewer people following their tweets (followers) (*P*=.02) and with fewer accounts that they followed (followings) (*P*=.04) were more likely to talk positively about physical activity on Twitter. People with more followers were more likely to post neutral tweets about physical activity (*P*=.04). People with more followings were more likely to forward tweets (*P*=.04). People with larger differences between number of followers and followings were more likely to mention companionship support for physical activity on Twitter (*P*=.04).

Conclusions: Future health marketing interventions promoting physical activity should segment Twitter users based on their number of followers, followings, and gaps between the number of followers and followings. The innovative application of both

marketing and public health theory to examine tweets about physical activity could be extended to other infodemiology or infoveillance studies on other health behaviors (eg, vaccinations).

(J Med Internet Res 2013;15(11):e261) doi:10.2196/jmir.2870

KEYWORDS

Twitter messaging; social marketing; motor activity

Introduction

Background

Twitter, a microblogging service and social networking site [1,2], provides a public platform to study the distribution and determinants of information with the ultimate aim to inform public health and public policy. This has been referred to as infodemiology or infoveillance with the primary aim of surveillance [3,4]. A few pioneering infoveillance scholars have successfully used Twitter to monitor people's status updates to track illness over time, often referred to as syndromic surveillance [3-6], such as during the H1N1 outbreak [7,8]. Other infoveillance studies analyzed how people share health information on Twitter and have monitored their health-related behaviors [3,4], such as antibiotic use [9], drug abuse [10], dietary behavior [11], and smoking behavior [12]. However, to date little larger-scale research has addressed the distribution of information about health behaviors among personal Twitter users rather than organizational Twitter accounts in the United States.

The current exploratory study aims to fill this gap and to inform the development of future health marketing interventions aiming to promote physical activity in the United States on this far-reaching communication platform. The Pew Internet & American Life Project found that 13% of online adults [13] and 8% of teenagers (aged 12-17 years) use Twitter [14] in the United States. In 2009, 12% of the people who looked online for health information also used Twitter to share health updates about themselves or to see updates about others [15].

The current study focuses on the health behavior of physical activity, in part to address the epidemic of inactivity in the United States. According to the 2009 US Youth Risk Behavior Survey, only 18% of students in grades 9 to 12 participated in at least 60 minutes of physical activity per day, and only 33% attended physical education class daily [16]. An examination of physical activity prevalence in the United States, derived from the results of the 2008 National Health Interview Survey, found that fewer than half (43.5%) of adults were aerobically active, a little over one-fifth (21.9%) met the muscle-strengthening guideline, and only 18.2% met both the muscle-strengthening guideline and were aerobically active throughout the year [17].

The current study examines the dissemination and sharing of information about physical activity among personal Twitter users rather than organizational Twitter accounts, physical activity facility, or physical activity equipment company accounts. In marketing research, digital sharing among individuals is called electronic word of mouth (eWOM), which refers to the online information exchange between and among a large number of consumers about a product [18]. In this study, the product is physical activity and the consumers are Twitter users. We focused on eWOM rather than information sent from organizational accounts because eWOM on social networking sites features higher response rates and can be archived in a manner that extends influence to more receivers over longer periods of time compared to other marketing techniques [19]. Additionally, we were interested in the organic communication among users about physical activity, which occurs outside organizational influences and is typically the target of interventions.

To suggest guidelines for designing future health marketing interventions aiming to promote physical activity on Twitter, the current study examines both the format and the content of eWOM about physical activity on Twitter. We innovatively applied marketing and health behavior principles simultaneously. The marketing principles we examined included valence, eWOM components (ie, opinion leadership and opinion seeking), and eWOM consequences (ie, forwarding and chatting). In the context of health behavior theory, we examined physical activity modeling, social support, and negativity. Furthermore, this study examined how the characteristics of eWOM varied among tweets from users with different networking characteristics, including number of followers, number of followings (Twitter accounts that a user is following), and ratios of number of followers to followings.

Marketing Aspects

Valence

Marketers are especially interested in whether their products are talked about positively, negatively, or neutrally [20]. Positive word of mouth (WOM) includes "relating pleasant, vivid, and novel experiences, recommendations to others, and even conspicuous display" whereas negative WOM includes "behaviors such as product denigration, relating unpleasant experiences, rumor, and private complaining" [21]. Exposure to negative WOM is associated with low probability of purchasing a product, whereas positive WOM is associated with high probability of purchasing [22]. For example, valence of eWOM has been found to influence box office revenue [23] and book sales [24]. Infodemiology studies have also analyzed the sentiment in surveillance of health beliefs [25] and tobacco-related tweets [12].

Components of Electronic Word of Mouth

In addition to valence, the current study addresses the mechanism of interactions, referred to as components of WOM [26] or eWOM [27], between Twitter users regarding physical activity. According to the path model of antecedents and consequences of online WOM, eWOM is composed of 2 forms

of interactions. Online opinion leadership is the process by which people attempt to influence others' purchasing behavior for a certain product. Online opinion seeking refers to the process by which people seek advice when purchasing a certain product [27].

A few infodemiology studies have explored how the information about different health conditions is presented on Twitter. For example, in the context of H1N1, approximately 10% of the tweets were in the form of questions [7]. In contrast, for concussions, approximately 1.4% of the tweets sought explicit advice [28]. However, little research has explored the mechanisms of interactions between Twitter users on Twitter about "purchasing" a specific health behavior (ie, physical activity participation).

Consequences of Electronic Word of Mouth

Based on the path model of antecedents and consequences of online word of mouth, eWOM has 2 consequences: forwarding and chatting [27]. With a limit of 140 characters (including all punctuation and spaces), Twitter is a rapid mode of communication that permits frequent updates [2]. Thus, Twitter is largely used for daily chatter, conversations, sharing information, and reporting news [2]. In addition, tweets can be archived and retrieved later by followers [29,30], extending their possible influence on others. On the other hand, the forwarding function of Twitter enables viral advertising, which is "a widely used form of unpaid communication through persuasive messages created by identifiable sponsors and distributed among peers on interactive digital platforms" [31]. Viral advertising can exponentially increase the number of people who receive a particular message and can work in conjunction with eWOM to drive communication about a topic or message.

Health Behavior Aspects

Overview

Physical activity as a health behavior is a unique "product." Purchasing in this case refers to participation. Moreover, the purchasing can be influenced by different social factors. Thus, in addition to the traditional eWOM characteristics for commercial products, this study also examined the health behavior aspects of eWOM about physical activity participation.

Physical Activity Modeling

Based on social cognitive theory (SCT), health behaviors can be acquired through observational learning or modeling, which is to watch and mimic the actions and outcomes of others' behavior [32]. Observational learning can occur through many channels: face-to-face observation [32], mass media [33], and online interactions [34]. The current study examined how Twitter users provide opportunities for others to engage in observational learning about physical activity, a behavior that we refer to as physical activity modeling. In addition to actual past and current participation in physical activity, this study investigates eWOM related to the intention to participate in physical activity. Research suggests that observational learning is acquired not only from viewing others' actions, but also through perceiving the models' intention and then imitating their goal [35,36].

Social Support and Social Negativity

Apart from observational learning, SCT posits that one's behavior is influenced by their social environment, which includes family members, friends, and acquaintances [32]. Influence exerted in a social environment can include both support and negativity [37], which have been found to play an important role in predicting physical activity participation [38,39]. Examining whether individuals exert social support and/or negativity on Twitter could provide insight into whether and to what extent Twitter might be used as a channel for social influence in future physical activity interventions. Chogahara [37] further categorized social support into 3 dimensions: companionship support, informational support, and esteem support. Chogahara also classified negativity into 3 dimensions: inhibitive, justifying, and criticizing behavior [37]. Examining these dimensions could help public health interventions target certain dimensions of social influence.

Therefore, regarding the information exchange among individuals about physical activity on Twitter, what is the distribution of (1) valence (positive, neutral, and negative); (2) eWOM components (opinion leadership and online opinion seeking); (3) eWOM consequences (chatting and forwarding); (4) physical activity modeling (communicating intention, past behavior, and current behavior); (5) social support (companionship, informational and esteem support); and (6) social negativity (inhibitive, justifying, and criticizing behavior)?

Networking Characteristics

Overview

Internet social connection is an antecedent of eWOM [27]. Twitter has a unique social networking function. It enables users to choose whom to receive information from, called followings, and who can receive their information, called followers [1,2]. Both the number of followers and followings for a user are shown on the Twitter profile and can be obtained by Twitter's application programming interface if users set their profile as public [1,2]. Thus, this study focused on 3 aspects of networking characteristics visible or easy to be estimated by other users: number of followers, number of followings, and the ratio of number of followers to followings.

Number of Followers and Followings

The number of contacts is an important aspect of traditional WOM [20]. The number of people following an individual is an indication of that person's popularity on Twitter [40,41]. Popularity, in turn, is an indicator of potential influence [41]. On the other hand, the number of followings can be seen as an indicator of inquisitiveness or how much of an expert one is [42]. To guide future physical activity interventions on Twitter, the current study also explores how the number of followers and followings is associated with the way a user talks about physical activity on Twitter.

Therefore, for tweets about physical activity, how does the number of followers and followings relate to differences in (1) valence, (2) eWOM components, (3) eWOM consequences, (4)

physical activity modeling, (5) social support, and (6) social negativity?

Followers Versus Followings

The difference between the number of followers and followings can also provide useful networking information [1,43]. Twitter users perceive other users with a narrower gap between the number of followers and followings as more credible [42]. Thus, to provide insights into designing future physical activity marketing interventions on Twitter, this study examined how the people with wider and narrower gaps between the number of followers and followings talked about physical activity differently on Twitter.

Specifically, in tweets about physical activity, how does the gap between number of followers and followings relate to differences in (1) valence, (2) eWOM components, (3) eWOM consequences, (4) physical activity modeling, (5) social support, and (6) social negativity?

Methods

Data Retrieval

Using a Twitter-streaming application programming interface, 1 million tweets posted between January 1 and March 31, 2011, in the United States containing 1 of 113 key physical activity words (see Multimedia Appendix 1) in either hashtags or the text body were retrieved. The first tweet with a physical activity keyword was posted at 03:33:39 Coordinated Universal Time (UTC) on Tuesday, January 4, 2011. The millionth tweet with a physical activity keyword was posted at 00:29:34 UTC on Thursday, March 31, 2011. The keywords included all activities from lists of published physical activity measures (eg, the Physical Activity Questionnaire for adults [44], the compendium of physical activities [45], and lists of fitness programs available at a Midwestern university [46]). Synonyms were grouped after consulting a standard thesaurus and dictionaries of American slang. We also pilot-tested the keywords to ensure the list adequately addressed the physical activity content and word usage among Twitter users. To ascertain the inclusion of tweets about similar types of physical activity, different tenses, word forms (eg, walk, walking, walked), and popular Internet expressions (eg, bball and B-ball for basketball) were also used as keywords. The keywords used to search included, but were not limited to, biking, climbing, golf, hockey, jogging, pull-up, sit-up, swimming, tennis, treadmill, walking, yoga, and Zumba (see Multimedia Appendix 1).

Scanning and Sampling

Two coders (native English speakers) were trained to scan tweets to exclude those that were not about physical activity and not from personal accounts. The exclusion criteria covered (1) tweets posted by an organization (discerned by username) and (2) tweets that included 1 of the keywords but were not about physical activity (eg, some advertisement about physical activity equipment). For example, a tweet including the word “pump” in reference to filling one’s gas tank was excluded. Non-English tweets were also excluded.

First, 30,000 tweets were randomly selected from the pool of 1 million tweets containing physical activity keywords. Second, the 30,000 tweets were sorted based on numbers generated by a random number generator. Third, the 2 coders scanned the first 16,100 tweets and yielded 4672 (29.02%) tweets that they both agreed to be about physical activity and were from personal accounts. Finally, 1500 tweets were randomly selected from the 4672 tweets (32.11%) for further coding. All 1500 selected tweets were from unique users as determined by user and Twitter account names.

Coding

The unit of analysis was a single tweet. The main concepts coded included (1) valence of eWOM, (2) components of eWOM,(3) consequences of eWOM, (4) physical activity modeling, (5) social support, and (6) social negativity. Coders could select all values that applied for most of the concepts except for physical activity modeling. See Table 1 for the coding scheme.

Intercoder reliability was calculated using 100 tweets. Two graduate students—a master’s and a doctoral student in public health—were trained and then completed a preliminary round of coding 100 tweets separate from the 1500 tweets. After the first round of reliability calculation, disagreements between the coders were discussed and the coding scheme was revised based on these discussions. Because some important variables were skewed (eg, showing in very few instances), this study used Holsti’s method to determine intercoder reliability [47]. The intercoder reliability scores ranged from 0.83 to 0.98 and were all acceptable (see Table 1). After the intercoder reliability was estimated, the 2 coders each coded 750 tweets.

Statistical Analysis

Descriptive analyses were conducted for the timing of posting the tweets, the number of followers, and people the users were following. Descriptive analyses for all eWOM characteristics were performed for (1) valence, (2) eWOM components, (3) eWOM consequences, (4) physical activity modeling, (5) social support, and (6) social negativity. Because the distribution of the number of followers and followings was quite skewed, we used the nonparametric Mann-Whitney *U* test (which does not require the normal distribution assumption) [48] to investigate if the distribution of the number of followers and followings differed across different eWOM characteristics.

For the gap between the number of followings and followers, a narrow gap was defined as a ratio of 0.9-1.1 between the number of followers and followings, whereas a wide gap group was defined as a ratio less than 0.9 or higher than 1.1 [42]. Fisher exact test was used to test if eWOM characteristics differed between the 2 groups. Fisher exact test was chosen because the data were skewed, and in most tables, 1 or more cells had expected counts less than 5 [49]. Twenty cases with zero followings were deleted, because their ratios between the number of followers and followings were not able to be calculated. Thus, the total number (N) in the analysis was 1480.

All analyses were conducted in SPSS 20 (IBM Corp, Armonk, NY, USA).

Table 1. Coding scheme and intercoder reliability scores of tweets about physical activity (PA).

Variables	Descriptions	Real tweet examples[a]	Reliability scores[b]
Valence (select all that apply)			
	Explicit sentiments associated with either a kind of PA itself or participating in PA or the environment where PA takes place		
Positive	Complimenting and relating pleasant and vivid experiences; praise/favorableness	A best day of personal fitness EVER=running around kinnick. :) football season get here faster-rrr	0.91
Negative	Complaining and the relating of unpleasant experiences	No one wants to go to the rec with me... #wah	0.96
Neutral	No sentiment	fooood, cleaning, gym...	0.88
Components of eWOM about PA (select all that apply)			
Online opinion leadership	Giving out information or opinions about PA (including PA itself or participating in PA or the environment where PA takes place)	Morning jog. Tennis later. We love sports!	0.97
Online opinion seeking	Asking for information and opinions of PA (including PA itself or participating in PA or the environment where PA takes place)	Shall I go swimming or take a bangin nap when i get off work in 30mins?	0.97
Communication consequences of eWOM (select all that apply)			
Online chatting	Provide plain text	Finally gym time. If anyone see's JLove on Michigan Ave. Tell her I said no to the pretty red sole shoes in the window!!	0.94
Online forwarding	Forward what other people said or the content of a Web page about PA. It could include sharing the Twitter messages from other Twitter users. When the Twitter message contains RT," code all the contents before and/or following RT. Forwarding the Twitter messages from Web, when the Twitter message contains a website link (URL). If a tweet includes a link, do not need to analyze the content in the link. But you can follow the link to help understanding	RT @* Im bout to go swimming...	0.94
PA modeling (single choice)			
	Either ones' own experience or others experience including previous experience and intention to future experience		
Intention to participate in a PA	A statement showing the participant is going to or needs to participate in PA (including intent to and nonintent to)	Feels like going to the gym	0.83
Past experience of participating in a PA	A statement that they have done a PA but did not give any advice or support to others	Went to the gym. Tired. Hanging out at older sis house now. Nap is probably needed	0.86
Current status of participating in a PA	A statement indicating the participation in PA right at the moment when posting the messages	@*still running :)) 2 more km. .and you? Are you still hungry? :D	0.93
Social support and social negativity (select all that apply)			
	Not about oneself; need another person involved		
Companionship support	Partnership assistance of a PA that suggests "we participate together"(components: coplanning, cooperation, coparticipation, reminding, rescheduling, offering, willingness)	Went on an enjoyable run with the lovely *. Now to do homework for the rest of the night...	0.86
Informational support	Knowledge assistance of a PA that suggests "you should know" (components: enlightenment, rationalization, clarification, program referral, intensity suggestion, activity recommendation, supporter referral, problem solving, and goal direction)	@* I recommend it for a beginners workout. . DM and I can give you more information.	0.97
Esteem support	Esteem information provision of a PA that suggests "you are good" (components: mastery recognition, social comparison, affirmation, respect, reinforcement, interest, and reassurance)	@LAEasyMeals Congrats! Can't imagine running 26.2 in that heat but well done!	0.97

Variables	Descriptions	Real tweet examples[a]	Reliability scores[b]
Inhibitive behavior	Disapproval and discouraging behavior that suggests "you should not participate in PA" (components: warning, delimitation, worrying, forbidding, threatening, disapproving, and rejection)	@* hott as hell in southeast gym	0.98
Justifying behavior	Excusing and overprotective behavior that suggests "you don't need to participate in PA" (components: excuse-giving, compromising, exempting, pardoning, and ignoring)	"Girl! YOUR body is a solid A++++++++ you don't need to work out!!" - @*	0.95
Criticizing behavior	Demanding and blaming behavior that suggests "you are not good at doing PA" (components: exclusion, demanding, nagging, contempt, bothering, depressing, and ridicule)	I've been told I play basketball like a girl haha...	0.97

[a]Personal names replaced with * to maintain confidentiality.

[b]Using Holsti's method (n=100).

Results

Overview

All 1500 tweets were posted by 1500 distinct users in the first quarter (January 1 to March 31) of 2011 in the United States. Tuesday was the most popular day for posting (245/1500, 16.33% of posts) and followed closely by Monday (234/1500, 15.60%), Wednesday (228/1500, 15.20%), and Thursday (227/1500, 15.13%). Friday was least popular (182/1500, 12.13%), followed by Sunday (184/1500, 12.27%) and Saturday (200/1500, 13.33%). Figure 1 presents a snapshot of the hourly distribution of all 1500 tweets from 0 (00:00-00:59) to 23 (23:00-23:59) UTC.

Users had an average of 576 followers (SD 3183, range 0-82,874, median 122). Users followed an average of 368 people (SD 976, range 0-25,069, median 148). Users posted 6630 total tweets on average, ranging from 1 to 167,517 (SD 14,158). Figure 2 presents the relationship between number of followers and followings with the axes in logarithmic scales among the 1472 Twitter users. We excluded 28 users who did not have any followers or followings from the 1500 Twitter users.

Descriptive Distribution

Table 2 presents the numbers of tweets in each category (eg, positive) of eWOM characteristics (eg, valence). Because 1 tweet can present in more than 1 category of each characteristic (except for physical activity modeling), the numbers in the categories of each eWOM characteristic (except for physical activity modeling) are not mutually exclusive.

Regarding the distribution of eWOM valence (ie, positive, neutral, or negative), approximately 85% of the tweets (1270/1500, 84.67%) had neutral valence only. Tweets with only negative valence comprised less than 3% (41/1500, 2.73%) of the total. In addition, there was 1 tweet (1/1500, 0.07%) that had both positive and neutral value, 1 (1/1500, 0.07%) that had both negative and neutral value, and 4 (4/1500, 0.27%) that had both positive and negative values.

Regarding the distribution of online opinion leadership and online opinion seeking, nearly all tweets (1464/1500, 97.60%) illustrated online opinion leadership only. Online opinion seeking alone was rare (26/1500, 1.73%). In addition, 10 tweets (10/1500, 0.67%) performed both online opinion leadership and opinion seeking.

For the 2 consequences of eWOM (online chatting and online forwarding), approximately 9 in 10 tweets (1354/1500, 90.27%) were in the form of online chatting only. Online forwarding alone occurred in about 1 in 13 tweets (108/1500, 7.20%). Another 36 tweets were in the both forms of online chatting and forwarding.

For the distribution of physical activity modeling, including communicating intention to participate in physical activity, past experience participating in physical activity, and current participation (and intention to participate) in physical activity, approximately 60% (905/1500, 60.33%) of tweets were related to 1 of these 3 areas. Of the tweets that mentioned intention or behavior, more than half (469/905, 51.82%) were about past behavior.

Regarding the distribution of different dimensions of social support, more than 90% of the tweets did not mention any social support (1364/1500, 90.93%). Among the tweets that offered social support (135/1500, 9.00%), informational support was the most frequent (63/135, 46.67%). Social negativity occurred in less than 2% of the tweets (18/1500, 1.20%).

Association Between Number of Followers and Electronic Word of Mouth

The number of followers differed between tweets with positive valence (n=186, mean 508, SD 2919) and others (n=1314, mean 586, SD 3219; P=.02). The number of followers also differed between tweets with neutral value (n=1272, mean 595, SD 3270) and others (n=228, mean 473, SD 2645; P=.04) in response to valence. There were no significant associations between number of followers and eWOM components, eWOM consequences, physical activity modeling, social support, and social negativity.

Association Between Number of Followings and Electronic Word of Mouth Characteristics

We explored the association between the number of followings and the different aspects of eWOM (see Table 2). The number of followings differed between tweets with positive value (n=186, mean 283, SD 506) and others (n=1314, mean 380, SD 1025; P=.04).

There were no significant associations between number of followings and eWOM components, physical activity modeling, social support, and social negativity. The number of followings differed between users who forwarded tweets about physical activity (n=139, mean 375, SD 814) and others (n=1361, mean 367, SD 992; *P*=.04). In summary, people who talked positively about physical activity were likely to follow fewer people, whereas people who forwarded information about physical activity were likely to follow more people.

Gaps Between Number of Followers and Followings and Electronic Word of Mouth Characteristics

There were no significant associations between the gap between numbers of followers and followings and valence, eWOM components, eWOM consequences, physical activity modeling, and social negativity. For social support, it was found that people with a wider gap between the number of followers and followings were more likely to provide companionship (*P*=.04). Table 3 presents the distribution of companionship support based on the gaps between number of followers and followings (OR 0.31, 95% CI 0.10-1.0).

Figure 1. Hourly distribution of all 1500 tweets during the day from hour 0 (00:00-00:59) to hour 23 (23:00-23:59) Coordinated Universal Time (UTC).

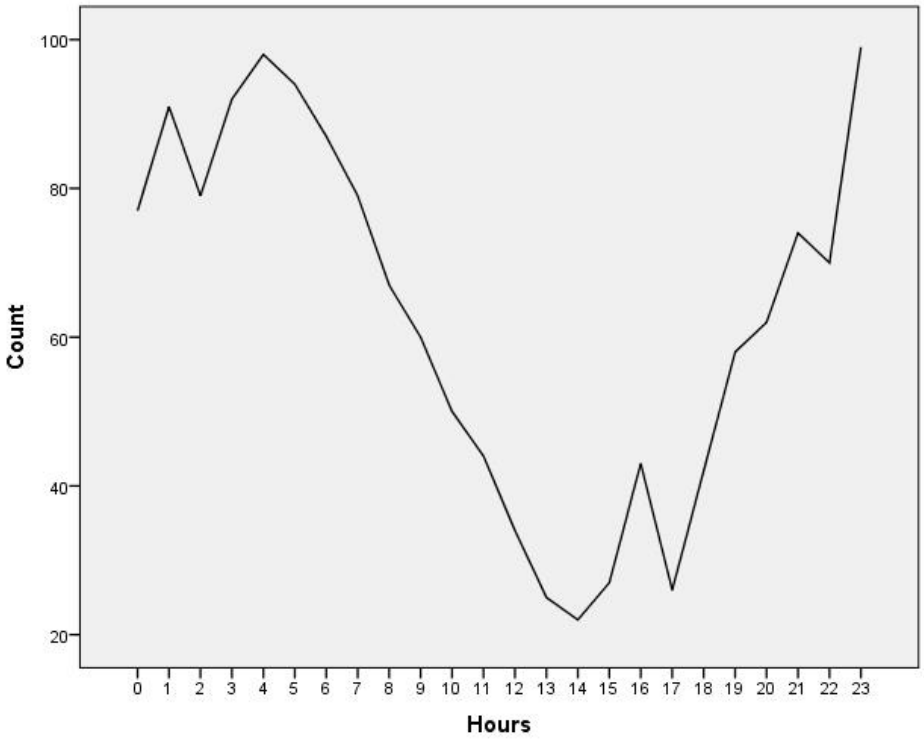


Figure 2. Scatterplot of the number of followers and followings of users (n=1472).

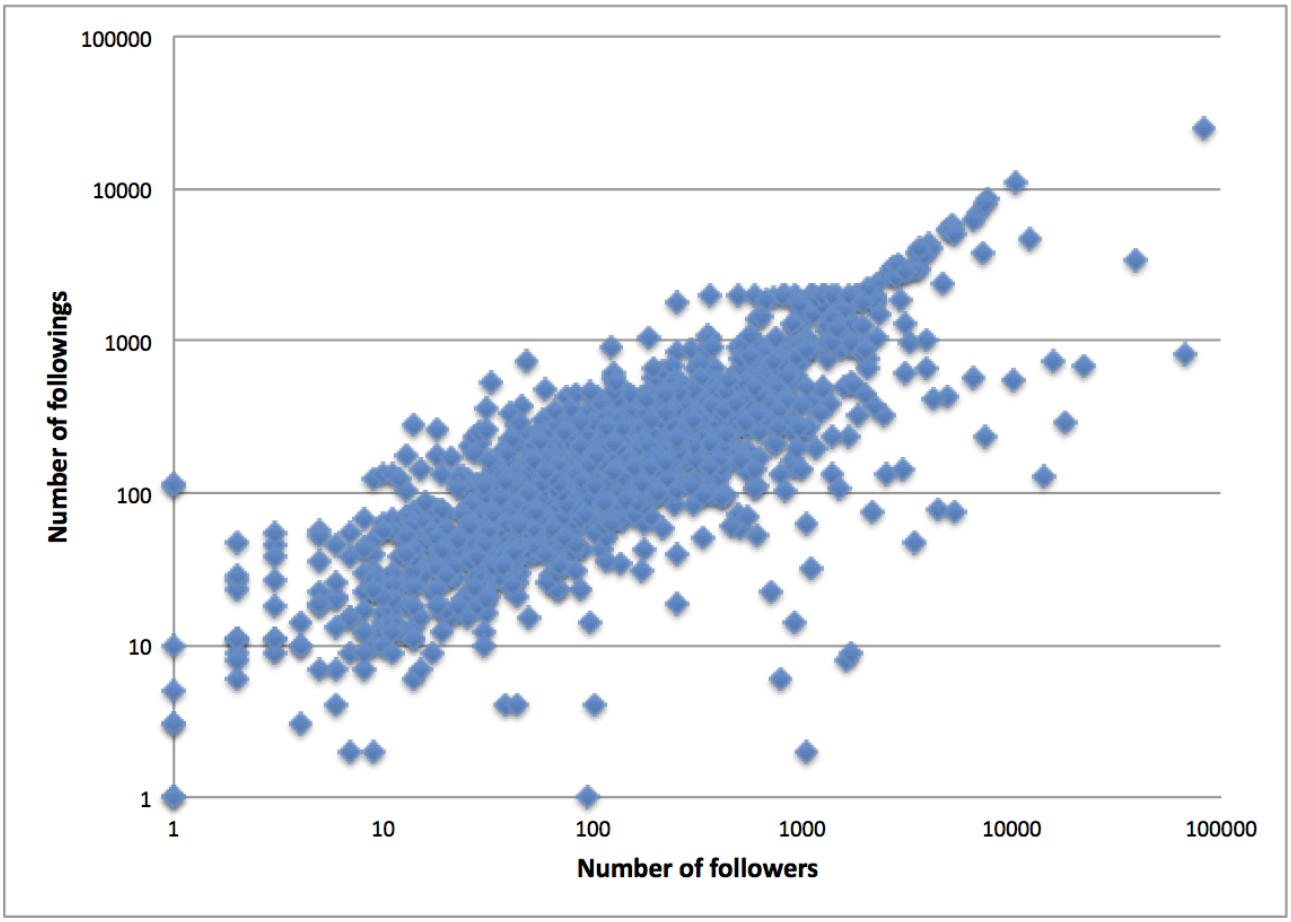


Table 2. Descriptive electronic word of mouth (eWOM) characteristics of tweets about physical activity (PA).

Characteristics of eWOM	n[a]	Number of followers Mean (SD)	Number of followings Mean (SD)
Valence			
Positive	188	508 (2919)[b]	283 (506)[b]
Negative	46	302 (517)	308 (521)
Neutral	1272	595 (3273)[b]	382 (1039)
Components of eWOM			
Online opinion seeking	36	383 (657)	332 (456)
Online opinion leadership	1474	583 (3220)	370 (986)
Consequences of eWOM			
Chatting	1390	582 (3316)	367 (993)
Forwarding	144	535 (1478)	375 (814)[b]
PA modeling	905		
Intention	336	397 (1381)	278 (469)
Past behavior	469	600 (2474)	325 (604)
Current behavior	100	336 (636)	326 (566)
Social support and negativity			
Social support			
Companionship support	51	440 (827)	310 (560)
Informational support	63	1899 (8504)	900 (1952)
Esteem support	23	420 (971)	193 (168)
None	1364	523 (2780)	348 (922)
Social negativity			
Inhibitive behavior	10	1354 (1773)	377 (384)
Criticizing behavior	13	128 (177)	141 (166)
None	1481	577 (3201)	368 (981)

[a]The n in categories of each eWOM characteristics (except for PA modeling) were not mutually exclusive.
[b]$P<.05$.

Table 3. Companionship support and gaps between number of followers and followings.

Gaps between number of followers and followings	Companionship support, n (%)		Total, n
	Yes	No	
Narrow gap (0.9-1.1)	3 (1.2)	240 (98.8)	243
Wide gap (<0.9 or >1.1)	48 (3.9)	1189 (96.1)	1237
Total	51 (3.4)	1429 (96.6)	1480

Discussion

Principal Findings

This exploratory study examined whether and how people talk about physical activity on Twitter. First, this study examined the valence of physical activity eWOM, an important marketing concern. Second, it explored the components of the path model of antecedents and consequences of online word of mouth [27]. Third, it addressed 3 important constructs of SCT [32]: observational learning (physical activity modeling in this study), social support, and social negativity. Finally, this study tested for differences in eWOM characteristics associated with the number of followers, followings, and the ratio of number of followers to followings.

The distribution of valence across physical activity tweets was different from other commercial products. In a study examining tweets for a variety of products (eg, automotive, computer hardware, consumer electronics, energy, fast food, Internet

services, personal care, sporting goods, and transportation), approximately 60% of the tweets were positive, 12% were neutral, and 23% were negative [1]. One recent content analysis on tobacco-related tweets found more positive than negative or neutral tweets [12]. However, in the current study, most tweets (85%) were neutral. This finding reflects a possible difference in eWOM between tangible commercial products and health behavior. For instance, commercial products involve tangible costs and benefits, and the transaction can be completed relatively easily in a short time. Physical activity involves more intangible costs, including time and energy, and usually takes longer to "consume." Moreover, the potential benefits of engaging in physical activity can take even longer to observe. This may preclude users from commenting positively or negatively about physical activity. An alternative explanation is that people might be less willing to comment or have more difficulty commenting on their own behaviors than on commercial products. When people comment on a product or service, they evaluate third-party providers, which is a relatively easy task. When discussing physical activity, however, they have to evaluate their own behaviors and their own selves, which may be more difficult cognitively.

The number of positive physical activity tweets was 4 times higher than negative physical activity tweets. This finding is consistent with past literature, which has shown that positive WOM was more common than negative WOM in 15 studies, with an average incidence ratio of 3:1 [50].

For eWOM components, the results of this study indicate that Twitter is currently used more often to provide opinions or information than to seek opinions or information about physical activity. This finding is consistent with a content analysis of tweets about concussions, in which researchers found that only approximately 1.4% of tweets sought explicit advice [28]. Our finding also suggests that posting public messages on Twitter is not yet a popular method for seeking physical activity information or opinions. The low percentage of tweets seeking opinion or information might indicate that people are using more traditional WOM communication or other kinds of eWOM channels to seek information. People could also be sending direct tweets, which are private between 2 users, to seek opinions and/or information about physical activity.

Regarding eWOM consequences, chatting was more common than forwarding among the physical activity tweets, a finding consistent with a previous study that found that one of Twitter's main functions was daily chatter [2]. It is also consistent with the primary usage of social network sites for health information: health updates and queries [13].

Physical activity modeling was represented in more than half of the tweets (60%). This finding is not surprising because the Pew Internet & American Life Project found that among 27% of Internet users the most common use of online health communication was to track weight changes, manage diets, record exercise routines (which could qualify as physical activity modeling), or follow some other health indicators or symptoms [13]. This finding is also consistent with a previous study regarding concussion reporting on Facebook: most of the posts (65%) shared a personal experience [51].

In addition to examining how people might model physical activity in their tweets, this study is the first to examine both social support and negativity via eWOM on Twitter. Given that eWOM about daily routine is the most common use of Twitter [2], our finding that only 10% of tweets provided any kind of social support or negativity is not surprising. The results of this study suggest that Twitter is currently not a popular platform for social influence attempts regarding physical activity. However, people could be using direct tweets to ask for or to provide their followers with social support. These direct, private tweets were not available for this study. Future research might incorporate those tweets to examine how people perceive social support or social negativity and these messages are influential. On the other hand, another popular social networking site, Facebook, has been linked to social support among users [52]. This may indicate that certain characteristics of Twitter make it an unlikely place to seek and obtain social support, unlike other online platforms, such as Facebook or discussion groups. Such features could be Twitter's immediacy and the forced brevity of the updates (only 140 characters).

Regarding the association between the number of followers and followings and the eWOM characteristics, the results suggested that people with fewer followers and followings were more likely to talk positively about physical activity on Twitter. People with more followers were more likely to post neutral tweets about physical activity. People with more followings were more likely to forward tweets. These findings suggest that people with different number of followers and followings may have different motivations for using Twitter regarding physical activity. People with fewer followers and followings might be more likely to connect with a close social network on Twitter and talk about physical activity positively for fun, whereas people with more followers and followings might be more likely to use Twitter primarily for information sharing about physical activity. However, future research is needed to further examine the reasons and confirm these suggestions.

Finally, a surprising finding is that people who had a wider gap between the number of followers and followings were more likely to mention companionship support on Twitter. This contradicted the intuition that a narrower gap between the number of followers and followings might indicate higher reciprocity between actual friends, which could result in more mentioning of social support on Twitter. This result can be explained by the finding from another study that Twitter is a sparse network for actual friends rather than a dense network between followers and followings [53]. Only approximately one-third of the users on Twitter are followed by their followings [1]. So the difference between number of followers and followings might not reflect the number of actual friends. It could be possible that people who have a wider gap between number of followers and followings might have more actual friends on Twitter to whom they provide companionship support.

There could be other alternative explanations. For example, because companionship support of physical activity requires the geographic accessibility and proximity of 2 people or more, the offline relationship between the users is indispensable. Thus, people with a narrower gap between number of followers and followings might be receiving and offering companionship

support through other offline channels, such as face-to-face, telephone, text messaging, or even through direct tweets between one another that are private and, thus, could not be retrieved in our study.

Practical Implications

Considering the low prevalence of positive tweets in contrast to the high proportion of physical activity modeling, future interventions should encourage people not only to chat about their physical activity intention or participation, but also to express the benefits of physical activity and their positive experiences with it.

Examining tweets for the components of SCT suggests that Twitter is currently mostly used for general observational learning of physical activity instead of exerting social support or social negativity. More innovative methods, such as infovigil robot, can be used to direct Twitter users to social support interventions after they post any tweets about physical activity [3]. In addition, examining tweets based on the path model of antecedents and consequences of online WOM [27] can inform public health practitioners about specific communication strategies that can be used to promote physical activity on Twitter. Future interventions could encourage Twitter users to provide opinion or information about physical activity through chatting because this study found that most tweets were examples of opinion leadership rather than opinion seeking.

Findings about how eWOM characteristics differed among Twitter users with different networking characteristics can provide insights into segmentation of audiences in future physical activity marketing interventions on Twitter. The association between the number of followers and followings and the valence of eWOM about physical activity indicates that interventions encouraging positive discussion of physical activity could start by enrolling individuals with fewer followers and followings and observing and learning how they talk positively about physical activity.

Because people with more followings tended to forward opinions or information about physical activity on Twitter suggests that public health practitioners could target people with more followings in future physical activity marketing interventions. Public health practitioners could develop Twitter accounts to promote physical activity and encourage Twitter users to follow the accounts and retweet tweets about physical activity to their followers.

Limitations and Future Research

The first limitation of this study is associated with the sampling method. This study used a list of keywords to retrieve tweets about physical activity. Although a comprehensive list of keywords was generated based on the physical activity checklist of the Physical Activity Questionnaire (Adults) [44], the compendium of physical activities developed by Ainsworth et al [45], and the list of fitness programs and intramural sports at the Midwestern university where this research was conducted [46], the possibility that some of the keywords were not captured (eg, the words with hierarchical relationships) cannot be overlooked. Future study should explore more vocabulary for keywords and use query expansion techniques to group similar keywords.

A second limitation relates to the lack of demographic information about the Twitter users. This limitation has also been observed in other studies [7]. Nevertheless, determining differences in talking about physical activity on Twitter based on different demographic characteristics, such as age and race, was not possible in this study. Future research might obtain permission from Twitter users to collect their demographic information. Moreover, future study could retrieve user-aggregated data and perform studies on the scale of individual users [6].

A third limitation is the lack of information about other social network characteristics of the Twitter users. Future studies would benefit from collecting information about reciprocal followings, which indicates the potential for interactive communication between users and their followers [54]. Moreover, this study used only the number of followers as an indicator of influence. Other indicators of influence were beyond the scope of the study; for example, neither message value (measured by the frequency of tweets passed between users) nor name value (measured by the frequency with which a name is mentioned in tweets from other users) was measureable in this study [41]. Future research can retrieve more information about message value and name value to help identify opinion leaders of physical activity and to investigate how they talk about physical activity on Twitter. Future study can also examine the social circles about physical activity on Twitter, adapting the method of a recent study on social circles about prescription drug abuse on Twitter [10].

Fourth, considering the effects of social support on individuals' physical activity based on SCT [32], future formative research using qualitative methods, such as interviews or focus groups, is needed to explore the predictors and barriers of using Twitter as a social support platform. The results could guide public health practitioners to develop interventions that encourage people to provide more social support via eWOM about physical activity on Twitter.

Although the tweets selected represented a random sample, the number of tweets about physical activity retrieved and coded was limited. First, because of the huge number of tweets posted every day, this study could not retrieve tweets in the United States for an entire year to control for physical activity variation in different seasons. We only included the tweets in the first quarter of the year. Second, although human coding enables the accurate categorization for different characteristics of eWOM at the same time, because of the labor intensity in manual coding, 2 coders only coded 1500 tweets in this study. Future study could use crowdsourcing experiments to conduct large-scale studies to provide a broader picture of the eWOM about physical activity on Twitter in the United States. Guided by the coding scheme invented and tested in this study, future studies could also explore a machine learning application and compare human coding and computer coding. With a larger number of tweets, future study could also localize different physical activities and physical activity eWOM by geographic region [5,6].

Conclusions

This study is the first to examine the content of eWOM about physical activity on Twitter. Twitter demonstrated potential for chatting and physical activity modeling (ie, intention, past behavior, and current behavior), as well as talking neutrally about physical activity. To guide future physical activity marketing interventions on Twitter, this study also provides insights into segmenting audiences based on user profile information about number of followers and followings. Having more followings was associated with forwarding information. Having fewer followers and fewer followings was associated with talking positively about physical activity. Having more followers was associated with talking neutrally about physical activity. Having a wider gap between the number of followers and followings was associated with mentioning companionship social support about physical activity on Twitter. Future studies could apply the innovative perspectives from marketing and public health used in this exploratory study for larger-scale infodemiology studies that could also examine tweets about physical activity as well as other health behaviors.

Acknowledgments

Laura Schwab-Reese and Laurel Whitis, students at the Department of Community and Behavioral Health, The University of Iowa, assisted in the coding. Julie Andsager, Professor at the School of Journalism & Communication, The University of Iowa, provided suggestions in the content analysis support group.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Physical activity keywords.

[PDF File (Adobe PDF File), 21KB - jmir_v15i11e261_app1.pdf]

References

1. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? 2010 Presented at: The 19th International Conference on World Wide Web; April 26-30, 2010; Raleigh, NC.
2. Java A, Song X, Finin T, Tseng B. Why we twitter: Understanding microblogging usage and communities. 2007 Presented at: The 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis; August 12-15, 2007; San Jose, CA p. 56-65. [doi: 10.1145/1348549.1348556]
3. 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]
4. Eysenbach G. Infodemiology and Infoveillance. Am J Prev Med 2011;40(5):S154. [doi: 10.1016/j.amepre.2011.02.006]
5. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
6. Paul MJ, Dredze M. You are what you tweet: Analyzing Twitter for public health. 2011 Presented at: The Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Catalonia, Spain.
7. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
8. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
9. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. American Journal of Infection Control 2010;38(3):182-188. [doi: 10.1016/j.ajic.2009.11.004]
10. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
11. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]
12. Myslín M, Zhu SH, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
13. Fox S. The social life of health information, 2011. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/Reports/2011/Social-Life-of-Health-Info/Summary-of-Findings.aspx [accessed 2013-11-12] [WebCite Cache ID 6L4qhJtDM]

14. Lenhart A, Purcell K, Smith A, Zickuhr K. Social media and mobile Internet use among teens and young adults. Washington, DC: Pew Internet & American Life Project; 2010 Feb 03. URL: http://web.pewinternet.org/~/media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed 2013-11-12] [WebCite Cache ID 6L4qiYzr4]
15. Lenhart A, Fox A. Twitter and status updating. Washington, DC: Pew Internet & American Life Project; 2009 Feb 12. URL: http://www.pewinternet.org/Reports/2009/Twitter-and-status-updating.aspx [accessed 2013-11-13] [WebCite Cache ID 6L6OMimdO]
16. Centers for Disease Control and Prevention. 2012. An explanation of United States physical activity surveys 2010 URL: http://www.cdc.gov/physicalactivity/professionals/data/explanation.html [accessed 2013-11-12] [WebCite Cache ID 6L4qqYPyL]
17. Carlson SA, Fulton JE, Schoenborn CA, Loustalot F. Trend and prevalence estimates based on the 2008 Physical Activity Guidelines for Americans. Am J Prev Med 2010 Oct;39(4):305-313. [doi: 10.1016/j.amepre.2010.06.006] [Medline: 20837280]
18. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing 2004 Jan;18(1):38-52. [doi: 10.1002/dir.10073]
19. Trusov M, Bucklin RE, Pauwels K. Effects of word-of-mouth versus traditional marketing: Findings from an Internet social network site. Journal of Marketing 2009;73:90-102 [FREE Full text]
20. Anderson EW. Customer satisfaction and word of mouth. Journal of Service Research 1998 Aug 01;1(1):5-17. [doi: 10.1177/109467059800100102]
21. Arndt J. Role of product-related conversations in the diffusion of a new product. Journal of Marketing Research 1967 Aug;4(3):291-295. [doi: 10.2307/3149462]
22. Duan W, Gu B, Whinston AB. Do online reviews matter? — An empirical investigation of panel data. Decision Support Systems 2008 Nov;45(4):1007-1016. [doi: 10.1016/j.dss.2008.04.001]
23. Liu Y. Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of Marketing 2006 Jul;70(3):74-89. [doi: 10.1509/jmkg.70.3.74]
24. Chevalier J, Mayzlin D. The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research 2006 Aug;43(3):345-354. [doi: 10.1509/jmkr.43.3.345]
25. Bhattacharya S, Tran H, Srinivasan P, Suls J. Belief surveillance with twitter. 2012 Presented at: The 3rd Annual ACM Web Science Conference; June 22-24, 2012; Evanston, IL.
26. Flynn LR, Goldsmith RE, Eastman JK. Opinion leaders and opinion seekers: Two new measurement scales. Journal of the Academy of Marketing Science 1996 Mar 01;24(2):137-147. [doi: 10.1177/0092070396242004]
27. Sun T, Youn S, Wu G, Kuntaraporn M. Online word-of-mouth (or mouse): An exploration of its antecedents and consequences. J Comp Mediated Comm 2006 Jul;11(4):1104-1127. [doi: 10.1111/j.1083-6101.2006.00310.x]
28. Sullivan SJ, Schneiders AG, Cheang CW, Kitto E, Lee H, Redhead J, et al. 'What's happening?' A content analysis of concussion-related traffic on Twitter. Br J Sports Med 2012 Mar;46(4):258-263. [doi: 10.1136/bjsm.2010.080341] [Medline: 21406451]
29. Gelb BD, Sundaram S. Adapting to “word of mouse”. Business Horizons 2002 Jul;45(4):21-25. [doi: 10.1016/S0007-6813(02)00222-7]
30. Kiecker P, Cowles D. Interpersonal communication and personal influence on the Internet: A framework for examining online word-of-mouth. Journal of Euromarketing 2002 Nov 20;11(2):71-88. [doi: 10.1300/J037v11n02_04]
31. Eckler P, Rodgers S. Viral advertising: a conceptualization. In: Cheng H, editor. The Handbook of International Advertising Research. Hoboken, NJ: Wiley-Blackwell; 2014.
32. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
33. Bandura A. Social Cognitive Theory of mass communication. Media Psychology 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]
34. Chen Y, Wang Q, Xie J. Online social interactions: A natural experiment on word of mouth versus observational learning. Journal of Marketing Research 2011;48:238-254. [doi: 10.2139/ssrn.1501843]
35. Bekkering H, Wohlschläger A, Gattis M. Imitation of gestures in children is goal-directed. Quarterly Journal of Experimental Psychology 2000;53:153-164. [doi: 10.1080/027249800390718]
36. Erlhagen W, Mukovskiy A, Bicho E. A dynamic model for action understanding and goal-directed imitation. Brain Res 2006 Apr 14;1083(1):174-188. [doi: 10.1016/j.brainres.2006.01.114] [Medline: 16616516]
37. Chogahara M. A multidimensional scale for assessing positive and negative social influences on physical activity in older adults. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 1999 Nov 01;54B(6):S356-S367. [doi: 10.1093/geronb/54B.6.S356]
38. Courneya KS, Plotnikoff RC, Hotz SB, Birkett NJ. Social support and the theory of planned behavior in the exercise domain. American Journal of Health Behavior 2000 Jul 01;24(4):300-308. [doi: 10.5993/AJHB.24.4.6]

39. Okun MA, Ruehlman L, Karoly P, Lutz R, Fairholme C, Schaub R. Social support and social norms: Do both contribute to predicting leisure-time exercise? American Journal of Health Behavior 2003 Sep 01;27(5):493-507. [doi: 10.5993/AJHB.27.5.2]
40. Weng J, Lim EP, Jiang J, He Q. TwitterRank: Finding topic-sensitive influential twitterers. 2010 Presented at: The third ACM international conference on Web search and data mining; February 4-6, 2010; New York, NY URL: http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1503&context=sis_research
41. Cha M, Haddadi H, Benevenuto F, Gummadi KP. Measuring user influence in Twitter: The million follower fallacy. 2010 Presented at: The Fourth International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
42. Westerman D, Spence PR, Van Der Heide B. A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. Computers in Human Behavior 2012 Jan;28(1):199-206. [doi: 10.1016/j.chb.2011.09.001]
43. Krishnamurthy B, Gill P, Arlitt M. A few chirps about Twitter. 2008 Presented at: Annual conference of the Association for Computing Machinery's Special Interest Group on Data Communication; August 17-22, 2008; Seattle, WA.
44. Copeland JL, Kowalski KC, Donen RM, Tremblay MS. Convergent validity of the Physical Activity Questionnaire for Adults: the new member of the PAQ Family. Journal of Physical Activity & Health 2005;2(2):216-229 [FREE Full text]
45. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine & Science in Sports & Exercise 2000;32(Supplement):S498-S516. [doi: 10.1097/00005768-200009001-00009]
46. Recreational services. The University of Iowa URL: http://recserv.uiowa.edu/Apps/Default.aspx?AspxAutoDetectCookieSupport=1 [accessed 2013-11-12] [WebCite Cache ID 6L4v38Kia]
47. Holsti OR. Content Analysis for the Social Sciences and Humanities. Don Mills, ON: Addison-Wesley; 1969.
48. Siegel S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill College; 1956.
49. Woodward M. Epidemiology: Study Design and Data Analysis (2nd ed). Boca Roton, FL: Chapman & Hall/CRC Texts in Statistical Science; 2008.
50. East R, Hammond K, Wright M. The relative incidence of positive and negative word of mouth: A multi-category study. International Journal of Research in Marketing 2007 Jun;24(2):175-184. [doi: 10.1016/j.ijresmar.2006.12.004]
51. Ahmed OH, Sullivan SJ, Schneiders AG, McCrory P. iSupport: do social networking sites have a role to play in concussion awareness? Disabil Rehabil 2010 Jan;32(22):1877-1883. [doi: 10.3109/09638281003734409] [Medline: 20367328]
52. Manago AM, Taylor T, Greenfield PM. Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. Dev Psychol 2012 Mar;48(2):369-380. [doi: 10.1037/a0026338] [Medline: 22288367]
53. Huberman BA, Romero DM, Wu F. Social networks that matter: Twitter under the microscope. SSRN Journal 2008 Dec 05:1-9. [doi: 10.2139/ssrn.1313405]
54. Cheng J, Romero DM, Meeder B, Kleinberg J. Predicting reciprocity in social networks. 2011 Presented at: IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing; Oct 9-11, 2011; Boston, MA URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6113094 [doi: 10.1109/PASSAT/SocialCom.2011.110]

Abbreviations

eWOM: electronic word of mouth
PA: physical activity
SCT: social cognitive theory
WOM: word of mouth

Edited by G Eysenbach; submitted 08.08.13; peer-reviewed by S Sullivan, WY Wang, S Bhattacharya; comments to author 26.08.13; revised version received 24.09.13; accepted 06.11.13; published 20.11.13.

Please cite as:
Zhang N, Campo S, Janz KF, Eckler P, Yang J, Snetselaar LG, Signorini A
Electronic Word of Mouth on Twitter About Physical Activity in the United States: Exploratory Infodemiology Study
J Med Internet Res 2013;15(11):e261
URL: http://www.jmir.org/2013/11/e261/
doi:10.2196/jmir.2870
PMID:24257325

Original Paper

The Good, Bad, and Ugly of Online Recruitment of Parents for Health-Related Focus Groups: Lessons Learned

Susan Quach[1], MSc; Jennifer A Pereira[1], PhD; Margaret L Russell[2], MD, PhD; Anne E Wormsbecker[1,3], MD, MPH; Hilary Ramsay[4], BA; Lois Crowe[4], BA; Sherman D Quan[5], MSc; Jeff Kwong[1,6,7,8,9], MD, MSc, CCFP, FRCPC

[1]Public Health Ontario, Toronto, ON, Canada
[2]University of Calgary, Calgary, AB, Canada
[3]Division of Paediatric Medicine, Hospital for Sick Children, Toronto, ON, Canada
[4]Bruyere Research Institute, Ottawa, ON, Canada
[5]Trillium Health Partners, Mississauga, ON, Canada
[6]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
[7]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[8]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[9]University Health Network, Toronto, ON, Canada

Corresponding Author:
Jeff Kwong, MD, MSc, CCFP, FRCPC
Institute for Clinical Evaluative Sciences
G1 06, 2075 Bayview Avenue
Toronto, ON,
Canada
Phone: 1 4164804055 ext 7665
Fax: 1 4164806048
Email: jeff.kwong@utoronto.ca

Abstract

Background: We describe our experiences with identifying and recruiting Ontario parents through the Internet, primarily, as well as other modes, for participation in focus groups about adding the influenza vaccine to school-based immunization programs.

Objective: Our objectives were to assess participation rates with and without incentives and software restrictions. We also plan to examine study response patterns of unique and multiple submissions and assess efficiency of each online advertising mode.

Methods: We used social media, deal forum websites, online classified ads, conventional mass media, and email lists to invite parents of school-aged children from Ontario, Canada to complete an online questionnaire to determine eligibility for focus groups. We compared responses and paradata when an incentive was provided and there were no software restrictions to the questionnaire (Period 1) to a period when only a single submission per Internet protocol (IP) address (ie, software restrictions invoked) was permitted and no incentive was provided (Period 2). We also compared the median time to complete a questionnaire, response patterns, and percentage of missing data between questionnaires classified as multiple submissions from the same Internet protocol (IP) address or email versus unique submissions. Efficiency was calculated as the total number of hours study personnel devoted to an advertising mode divided by the resultant number of unique eligible completed questionnaires .

Results: Of 1346 submitted questionnaires, 223 (16.6%) were incomplete and 34 (2.52%) did not meet the initial eligibility criteria. Of the remaining 1089 questionnaires, 246 (22.6%) were not from Ontario based on IP address and postal code, and 469 (43.1%) were submitted from the same IP address or email address (multiple submissions). In Period 2 vs Period 1, a larger proportion of questionnaires were submitted from Ontario (92.8%, 141/152 vs 75.1%, 702/937, P<.001), and a smaller proportion of same IP addresses (7.9%, 12/152 vs 47.1%, 441/937, P<.001) were received. Compared to those who made unique submissions, those who made multiple submissions spent less time per questionnaire (166 vs 215 seconds, P<.001), and had a higher percentage of missing data among their responses (15.0% vs 7.6%, P=.004). Advertisements posted on RedFlagDeals were the most efficient for recruitment (0.03 hours of staff time per questionnaire), whereas those placed on Twitter were the least efficient (3.64 hours of staff time per questionnaire).

Conclusions: Using multiple online advertising strategies was effective for recruiting a large sample of participants in a relatively short period time with minimal resources. However, risks such as multiple submissions and potentially fraudulent information

need to be considered. In our study, these problems were associated with providing an incentive for responding, and could have been partially avoided by activating restrictive software features for online questionnaires.

(J Med Internet Res 2013;15(11):e250) doi:10.2196/jmir.2829

KEYWORDS

parents; data collection; communication; social media; Internet

Introduction

Study recruitment is a challenging aspect of health research. Strategic communication, planning, and marketing are essential for reaching potential participants. By identifying a target population's demographic characteristics and interests, researchers can select methods that represent the best investment for study recruitment given the study objectives and budgetary, human resource, and time constraints [1,2].

Historically, researchers have relied on traditional methods, such as random-digit dialing, school or workplace recruitment, mass postal mailing, poster displays, and mass media campaigns, to recruit participants for research studies. However, these methods are often resource intensive and costly, sometimes with limited reach [3]. More recently, researchers have used the Internet to recruit younger participants for studies, by using paid advertising (eg, Google AdWords, Facebook Ads), links on websites, email lists, and online classified advertisement sites [4-6]. Compared with traditional strategies, Internet-based methods may achieve desired sample sizes with lower costs for distribution of materials [7]. They allow for rapid, convenient participation for study participants, and provide an easier means to share information across social networks, as well as an efficient process to respond to inquiries. Internet-based recruitment methods can be used to target difficult-to-reach groups who cannot be accessed by traditional means (eg, highly mobile individuals, those without a landline telephone) provided those groups have Internet access [8]. Further, respondent anonymity may decrease bias regarding sensitive health issues [5,9,10]. In contrast, anonymity can also reduce an individual's fear of behavioral consequences, resulting in decreased honesty and increased inappropriate activity from the perspective of the researcher, such as creating multiple identities and/or falsifying information for personal benefit [11]. This type of activity, termed "gaming the system," may be related to offering an incentive [12].

In 2012, we began recruitment for a qualitative study to examine and understand parental perceptions of the advantages and disadvantages of expanding influenza immunization programs to include delivery to children in elementary schools in Ontario, Canada, as well as the programmatic characteristics that would facilitate its adoption in this province. Given that eligible participants were parents of school-aged children, and likely busy individuals within the age range from early 20s to middle age, we used multiple advertising modes and conventional mass media to identify and recruit participants for focus groups.

In this paper, we evaluate the impact of the recruitment methods on study recruitment response, and compare the relative efficiencies of different advertising modes. The aims of this paper were to (1) compare the demographic attributes of eligible participants who completed the questionnaire during an incentive versus a nonincentive recruitment period; (2) compare questionnaire response patterns between multiple and unique submissions; (3) assess the uptake, defined as the number of times advertisements were viewed among eligible participants who made unique submissions; and (4) measure efficiency, defined as the average time invested by study personnel to recruit a unique eligible participant by advertising mode.

Methods

Participants

This is a descriptive study of data collected from an online questionnaire used to screen participants for a qualitative study on parental perceptions of school-based influenza immunization. Participants satisfied eligibility criteria if they were (1) English language proficient; (2) resided in Ontario, Canada; (3) parents or guardians of at least 1 child between the ages of 4 and 18 years in an Ontario school; and (4) involved in decisions about their child's immunizations.

Procedures

We invited participants to complete a 5-minute online questionnaire that had been pretested for usability and technical functionality. From October 9 to 11, 2012 (period 1), we offered a CAD $5 Amazon electronic gift card as an incentive to those participants who met the self-reported eligibility criteria and completed the questionnaire. Because of the unexpectedly high volume of submissions received, we temporarily closed the questionnaire on October 11, 2012. The questionnaire was reopened from November 16, 2012 to February 17, 2013 (period 2) to recruit additional participants. In mid-January 2013, we focused on recruiting participants from rural areas. During period 2, no incentive was provided and software features were activated to permit submission of only 1 questionnaire per Internet protocol (IP) address.

Ethics approval was obtained from the University of Toronto Health Sciences Research Ethics Board and the Bruyère Continuing Care Research Ethics Board.

Online Questionnaire

When participants entered the study website, they were asked to review information about the study. Participants were directed to an initial 5-question eligibility screen that asked whether they read English, wrote in English, resided in Ontario, were parents or guardians of at least 1 child between the ages of 4 and 18 years in an Ontario school, and were involved in decisions about their child's immunizations. Another question asked participants how they heard about the study. The software automatically directed eligible individuals to the rest of the questionnaire,

whereas those who were ineligible were directed to a page that prevented them from accessing the questionnaire. Participants could quit the questionnaire at any time. The full questionnaire included questions about number of children, the grade level of each child, whether participants or their children had previously received influenza immunization, and questions about participant's age, sex, postal code, marital status, education, and ethnicity. Participants were asked if they were willing to engage in a focus group. If they agreed, they were prompted to indicate the type of focus group they preferred (teleconference or Internet forum) and to provide their contact information (name, email, and phone number). Only those who agreed to participate in the focus groups were required to provide their names. Email addresses were required when the incentive was offered or when participants agreed to engage in the Internet forum focus groups. Each page of the questionnaire had 1 to 3 questions, and there were a maximum of 17 pages for those who agreed to participate in the focus groups. FluidSurveys Version 4.0 (Ottawa, ON, Canada) was used for data collection and data were subsequently exported into Microsoft Excel 2010 and Stata SE 10.0 (StataCorp LP, College Station, TX, USA) for analysis.

Recruitment

Our primary recruitment strategy involved several free online advertising modes. These were (1) social media (Twitter and Facebook), (2) online classified advertisement websites (Craigslist and Kijiji), (3) advertisements on deal forum websites (RedFlagDeals, SmartCanucks), and (4) postings to public health email lists and websites. As a result of these efforts, we were contacted by conventional mass media outlets who further publicized the study. In the last month of period 2, we also issued media releases to rural newspapers. Table 1 contains a list of the advertising modes and number of advertisements. Multimedia Appendix 1 includes a sample of the text used for the online advertisements.

Table 1. Advertising activities to recruit participants.

Advertising mode and source	Number of advertisements
Social media	
Twitter	57
Facebook	339
Online classified advertisement websites	
Kijiji	71
Craigslist	57
Deal forum websites	
SmartCanucks	7
RedFlagDeals	2
Press releases to conventional mass media	
Rural newspapers	3 news articles in rural papers

Online Advertising Modes

Social Media

We created and used a list of public health organizations, popular parent blogging groups or individuals (eg, @Canadian_moms, @ParentSource), parent magazines or websites (eg, Today's Parent), local and national news agencies, physicians, and health reporters to follow and connect with on our social media channels.

We created a Facebook page (School Flu Shots-Ontario Study, created in September 2012) because Facebook is the most popular social media website in Canada [13]. We posted information about the study, such as the purpose, our research team, updates, and links to the study's website and questionnaire [14]. Messages were posted between 1 pm and 4 pm on weekdays, the peak Facebook activity hours, on a weekly or biweekly basis during periods 1 and 2 [15].

Twitter is a social networking service that allows users to send and resend text messages within a 140-character limit. We created a Twitter account (@SchoolFluShots) and posted messages about the study and questionnaire links to generate interest. We tweeted twice per week during periods 1 and 2, and every weekday during the rural recruitment period (the last month of period 2). We also spent time monitoring conversations and identifying new groups and individuals with whom to connect.

Online Classified Advertisement Websites

Ads were posted on Craigslist and Kijiji, free classified advertisement websites that are used widely in Canada to advertise goods and services. We posted messages and links about the study in the "Jobs" and "Volunteers" categories for each of the 20 and 30 Ontario cities with listings pages on Craigslist and Kijiji, respectively. The ads were monitored daily over periods 1 and 2, and were reposted once they no longer appeared on the first page of results for the category, which was every 7 to 10 days for most cities.

Deal Forum Websites

SmartCanucks and RedFlagDeals are popular deal forum websites that advertise coupons, promotions, events, and freebies. Users can post messages on different threads that can be viewed by all website visitors. At the beginning of period 1,

we posted 2 messages on RedFlagDeals under the "Freebies" and "Parenting" categories. We posted 28 messages on SmartCanucks under the "Paid Surveys and Mystery Shopping," "Canadian Parents," and 21 individual Ontario city categories. The ads in Smart Canucks were later merged into 7 ads and remained active during periods 1 and 2. On a daily basis, we monitored both RedFlagDeals and SmartCanucks for new replies from community members.

Public Health Email Lists and Websites

We sent emails to 148 Ontario health care organizations requesting they assist with study recruitment by sending a generic study email to their email list or posting a message on their own websites. We did not follow up to identify the number of organizations that complied with this request.

Press Releases to Conventional Mass Media

In the last month of period 2, we issued media releases to 35 rural newspapers, which resulted in 3 news articles (2 online and 1 paper) about the study.

Unsolicited Mass Media Contacts Resulting From Online Classified Advertisements

During period 2, we were contacted by the Canadian Broadcasting Corporation (CBC) requesting an interview. Our press releases had not targeted this news agency; the contacts were initiated as a result of our advertising on Kijiji. This resulted in 2 interviews that, in turn, led to 4 CBC news items (eg, online articles and videos, radio coverage, and television coverage) and 1 additional interview by a local Ottawa, Ontario, newspaper. The newspaper published an online video and a news article.

Analysis

To examine the effects of the incentive and software restrictions, we compared paradata (ie, data regarding the survey process), such as survey completion time, IP address, country of the IP address, and missing data, and questionnaire responses for the period when we provided an incentive with no software restrictions (period 1) to the period without an incentive and software restrictions activated to limit access to the questionnaire (ie, single questionnaire per IP address) (period 2). We validated the self-reported Ontario residential status from the eligibility screen by examining the country of origin of the IP address (Canada vs elsewhere), and the self-reported postal code. FluidSurveys automatically checks an online database of IP addresses from different countries to determine the location of the IP address (email communication, June 2013). If the country for the IP address and the postal code were missing, we did a manual reverse lookup of IP addresses to determine if they were from Ontario by using an online IP address reverse lookup website [16]. We estimated the proportion of submissions originating from validated Ontario residents (ie, both postal code and IP address from Ontario).

We compared the distribution of questionnaires by advertising mode, median time for questionnaire completion, and the proportion of missing responses for demographic questions between multiple versus unique submissions. We classified the questionnaires as multiple submissions if more than 1 submission was received from the same IP address or if the email address was classified as nonunique. We used automated software procedures to check for multiple submissions from the same IP addresses and for nonunique email addresses. Email addresses were classified as nonunique if the same email address was (1) included on more than 1 submission even if originating from a different IP address, or (2) the email address was a close variant of another submission (eg, jim.doe@gmail.com and jim.doe@yahoo.com) based on a manual scan of each email address by 1 reviewer.

Bivariate analysis of statistical significance of time was performed using the Wilcoxon rank sum test. Proportions were compared using chi-square tests or Fisher exact tests. *P* values are presented for descriptive purposes. Stepwise logistic regression was used to identify demographic variables that differentiated unique eligible participants that were offered (period 1) or not offered the incentive (period 2). The independent demographic variables included age group (categorical), sex, marital status (single or married), self-identified ethnicity (categorical), and educational attainment (categorical). We also examined an adjusted model that included a variable for rural residence within the province of Ontario. *Rural* was defined as having a postal code with zero in the second position of the 6-digit postal code [17].

To evaluate uptake, we tracked the number of times advertisements posted on Kijiji, RedFlagDeals, and SmartCanucks were viewed. As indicators of social media popularity, we tracked the number of likes and friends on Facebook. We also tracked the number of retweets, mentions, and followers on Twitter. We could not track the number of views for the media-related activities, email lists, or websites. Respondents self-reported the source from which they heard about the study based on a close-ended checklist with the following categories: Facebook, Twitter, Kijiji, Craigslist, RedFlagDeals, SmartCanucks, email lists, word of mouth, friends or family, websites, and other. Those who checked other or website were asked to provide more details. From this, we calculated the proportion of completed questionnaires from unique eligible participants by advertising mode.

To assess the efficiency of each advertising mode, we summed the number of personnel hours required to advertise and monitor responses for each source, and divided this by the number of unique eligible participants identified using that source.

Results

Overview

During the 5-month recruitment period, we received 1346 questionnaires. During period 1 (incentive offered, software restrictions inactive) 1124/1346 (83.51%) were received. When recruitment reopened (period 2), 222/1346 (16.49%) were received. This increased minimally when several study-related media stories were published online (Figure 1).

Of the 1346 questionnaires, we excluded 257 (19.09%) because they were incomplete (n=223) or because the potential participants did not meet the initial eligibility criteria (n=34), leaving 1089 for further analysis. We also excluded 246 that

were not from Ontario based on the IP address and postal code, 65 based on similar or identical email addresses, and 298 based on nonunique IP addresses. This left 480 questionnaires from unique participants who were eligible for the qualitative study of school-based influenza immunization for which we had been recruiting participants (Figure 2). Of the 246 (22.6%) questionnaires that were not from Ontario, 26 (10.6%) were from other Canadian provinces, 182 (74.0%) were from the United States, and the remaining 38 (15.4%) were from outside North America (eg, India, Japan, United Kingdom, China).

Of the 1089 questionnaires retained for further analysis, 636 (58.40%) were from unique IP addresses and 453 (41.60%) originated from 118 nonunique IP addresses (ie, were classified as multiple submissions). Most of the nonunique IP addresses (105/118, 89.0%) occurred 2 to 5 times. Five nonunique IP addresses occurred 17 to 30 times. Of the 1089 questionnaires, 932 (85.58%) listed unique email addresses, 62 (5.69%) were missing email addresses, and 94 (8.63%) listed 39 nonunique email addresses. Most of the nonunique email addresses (32/39, 82%) occurred twice, whereas 7 occurred 3 times, and 1 email address occurred 5 times. Approximately one-third of the nonunique email addresses were variations of other email addresses and were only detected by manual review.

Figure 1. Study timeline showing questionnaire submissions in relation to the use of incentives and survey restrictions throughout the study (October 2012-February 2013).

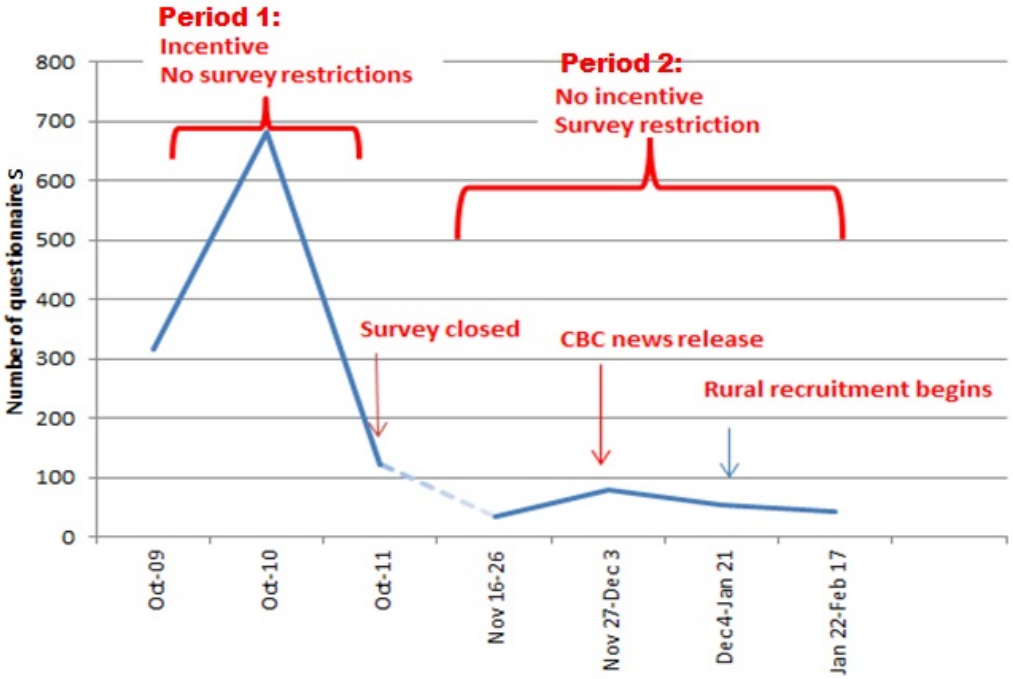
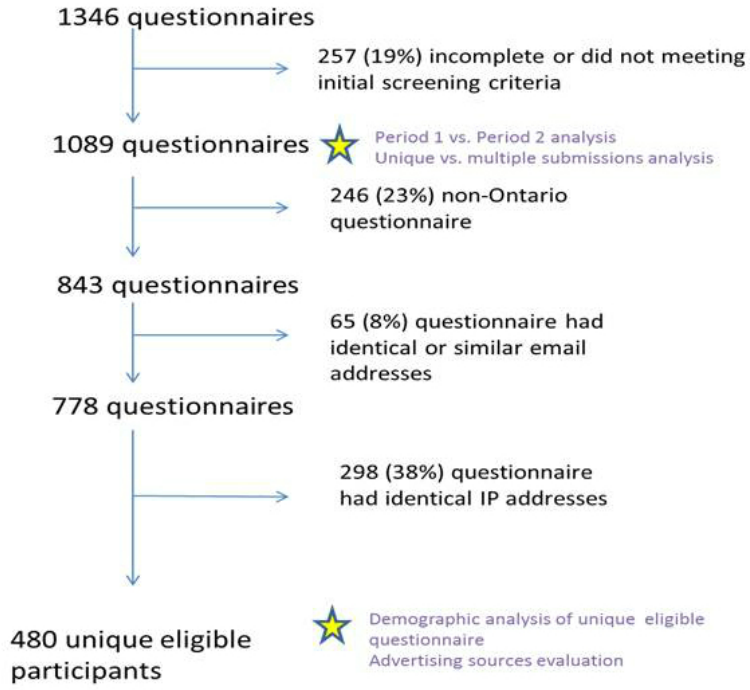


Figure 2. Participant flow diagram.



Effects of Incentives and Software Restrictions (Period 1 vs Period 2)

Compared to period 1, period 2 was associated with a larger proportion of questionnaires being submitted from Ontario (92.8%, 141/152 vs 75.1%, 702/937, *P*<.001), and a smaller proportion of same IP address submissions (7.9%, 12/152 vs 47.1%, 441/937, *P*<.001). Of the 12 questionnaires submitted during period 2 that originated from nonunique IP addresses, 10 (7%) were from IP addresses from which submissions had also been received during period 1. Similar proportions of submissions from unique email addresses were received during both period 1 and period 2 (86%, 806/937 vs 87%, 132/152, *P*=.74). The median time for questionnaire completion was significantly longer for responses submitted during period 2 compared to period 1 (266 seconds vs 181 seconds, *P*<.001).

Multiple vs Unique Submissions

The median time to complete a questionnaire for multiple submissions was less than that for questionnaires that were unique submissions (166 seconds vs 215 seconds, *P*<.001). We examined the time interval between submissions that were classified as multiple submissions (n=469) on the basis of IP address or email address: 272 (58.0%) were submitted within 5 minutes of the previous submission, 375 (80.0%) were submitted within an hour, and 450 (95.9%) were submitted within 24 hours.

Questionnaires from multiple submissions had a larger proportion of missing data than those from unique submissions (15% vs 8%, *P*=.004) (Table 2).

Among unique eligible participants, higher proportions of persons aged 40 years or older (39.2%, 49/125 vs 27.0%, 96/355), of female sex (83.2%, 104/125 vs 41.1%, 146/355), and of self-reported white ethnicity (70.4%, 88/125 vs 43.9%, 156/355) completed the questionnaire during period 2 compared to period 1 (Table 3).

In the adjusted model, sex and ethnicity were significant; females were more likely to participate during period 2 compared to period 1 (Table 4), whereas persons who self-identified as being of Chinese ethnicity had lower odds (OR 0.05, 95% CI 0.01-0.20) of participating during period 2 compared to those of white ethnicity. No other demographic variables were significant at the .05 level in the multivariate model. The adjusted model that also included rural versus urban residence within the province of Ontario showed similar results (Table 3). Participants living in urban locations had lower odds of participating during period 2, which was consistent with the study's recruitment goal to target individuals from rural areas during the last month of the study.

Uptake of Advertising Modes

During period 1, more than half of participants who completed questionnaires (192/355, 54.1%) were recruited through RedFlagDeals, whereas Craigslist or Kijiji, conventional mass media, and public health email lists or websites comprised the smallest contributions (<2% combined) (Table 5). However, during period 2, 28.0% (35/125) of participants who completed questionnaires were recruited through conventional mass media, 20.0% (25/125) through Craigslist or Kijiji, and 6.4% (8/125) through RedFlagDeals.

Efficiency of Advertising Modes

Recruitment from RedFlagDeals was the most efficient, requiring 0.03 hours of staff time per questionnaire (Table 6). Two ads were posted on RedFlagDeals at the beginning of the study that resulted in 5077 views. In contrast, recruitment via Twitter was the least efficient, requiring 3.64 hours per questionnaire.

Table 2. Proportion of missing data for each demographic variable.

Variable	Multiple submissions, n (%) n=469	Unique submissions, n (%) n=620
Postal code	176 (37.5)	80 (12.9)
Education	48 (10.2)	25 (4.0)
Age	25 (5.3)	32 (5.2)
Single status	28 (6.0)	49 (7.9)
Ethnicity	75 (16.0)	49 (7.9)
Total	352 (15.0)	235 (7.6)

Table 3. Demographic characteristics of unique eligible participants[a] (n=480).

Characteristic	Period 1, n (%) n=355	Period 2, n (%) n=125
Sex (female)	146 (41.1)	104 (83.2)
Age (years)		
<20	2 (0.6)	1 (0.8)
20-29	81 (22.8)	11 (8.8)
30-39	153 (43.1)	60 (48.0)
≥40	96 (27.0)	49 (39.2)
No answer	23 (6.5)	4 (3.2)
Marital status		
Single parent	50 (14.1)	19 (15.2)
Married	262 (73.8)	106 (84.8)
No answer	43 (12.1)	0 (0)
Ethnicity		
White	156 (43.9)	88 (70.4)
Chinese	84 (23.7)	3 (2.4)
South Asian	36 (10.1)	8 (6.4)
Southeast Asian/Filipino	14 (3.9)	2 (1.6)
Korean/Japanese	6 (1.7)	1 (0.8)
Mixed	7 (2.0)	7 (5.6)
Other (Arab, Black, Latin American, West Asian, Aboriginal)	16 (4.5)	12 (9.6)
No answer	36 (10.1)	4 (3.2)
Education		
High school or less	29 (8.2)	6 (4.8)
At least some postsecondary education (eg, college, university)	307 (86.5)	117(93.6)
No answer or other	19 (5.3)	2 (1.6)
Location		
Rural[b]	8 (2.3)	22 (17.6)
Urban	301 (84.8)	101 (80.8)
No answer	46 (13.0)	2 (1.6)

[a]Unique eligible participants exclude those identified as multiple submissions or those who did not meet the geographical criterion (Ontario residence).
[b]Rural location was based on the second digit of the 6-digit postal code being zero (eg, N0P 1L0) [17].

Table 4. Adjusted odds ratios[a] comparing demographic characteristics of unique eligible participants[b] from period 2 to period 1.

Characteristic	Adjusted OR (95% CI)	*P* value
Sex (female)	7.67 (4.12-14.27)	<.001
Age (years)		
<29[c]	1.00	
30-39	1.56 (0.67-3.56)	.29
≥40	2.26 (0.96-5.29)	.06
Marital status		
Married	1.00	
Single	0.70 (0.34-1.41)	.32
Ethnicity		
White	1.00	
Chinese	0.05 (0.01-0.20)	<.001
South Asian	0.65 (0.25-1.67)	.37
Southeast Asian/Filipino	0.37 (0.07-1.92)	.24
Korean/Japanese	0.46 (0.04-5.13)	.53
Mixed	2.28 (0.65-7.96)	.20
Other (Arab, Black, Latin American, West Asian, Aboriginal)	2.26 (0.96-5.29)	.06
Education		
High school or less	1.00	
At least some postsecondary education (eg, college, university)	2.22 (0.78-6.29)	0.133

[a]Adjusted odds ratio simultaneously adjusted for all variables listed in the table.

[b]Unique eligible participants excludes those identified as multiple submissions or those who did not meet the geographical criterion (Ontario residence) (n=390).

[c]<20 years was combined with 20-29 years because of the small sample size.

Table 5. How unique eligible participants heard about the study (n=480).

Advertising mode	Period 1, n (%) n=355	Period 2, n (%) n=125
RedFlagDeals	192 (54.1)	8 (6.4)
Friend or family	50 (14.1)	6 (4.8)
Facebook	47 (13.2)	21 (16.8)
SmartCanucks	17 (4.8)	10 (8.0)
Word of mouth	16 (4.5)	5 (4.0)
Twitter	11 (3.1)	2 (1.6)
Prefer not to answer	10 (2.8)	2 (1.6)
Public health email list or website	5 (1.4)	11 (8.8)
Craigslist/Kijiji	4 (1.1)	25 (20.0)
Press releases to conventional mass media	0 (0)	35 (28.0)

Table 6. Hours of staff time required for submissions from each advertising mode, the number of unique eligible participants[a] recruited from each mode, and the efficiency (number of staff hours per questionnaire) and uptake of each mode.

Advertising mode and source	Hours, n	Unique eligible participants, n	Efficiency	Uptake
Social media				
Twitter	47	13	3.64	13 retweets, 15 mentions, 112 followers, 469 following
Facebook	37	68	0.54	16 likes
Online classified advertisement websites				
Kijiji/Craiglist[b]	22	29	0.77	1193 views (Kijiji only)
Conventional mass media	4	16	0.25	Unable to assess
Email lists or websites	3	13	0.23	Unable to assess
Deal forum websites				
SmartCanucks	6	26	0.22	3579 views
RedFlagDeals	6	202	0.03	5077 views

[a]Total number of unique eligible participants reported is less than the number of unique eligible participants(367/480) because some participants provided no response or provided a response that could not be linked with a specific source(eg, friends or family, word of mouth).

[b]Craigslist did not provide any information on the number of views.

Discussion

Principal Results

By using multiple online advertising strategies, we recruited a large sample of participants in a relatively short time period with minimal resources. Although the Internet was an effective recruitment medium, we also observed a significant amount of suspected gaming during the recruitment process because some participants submitted multiple questionnaires. In contrast to those who made unique submissions, people who made multiple submissions from the same IP address or email address spent less time per questionnaire and their responses had a greater percentage of missing data. Further analysis revealed that approximately one-quarter of those who completed the questionnaire were not from Ontario, despite passing the initial eligibility screen of self-reported Ontario residence. Females were more likely to participate in period 2 (no incentive) compared to period 1 (with incentive), whereas persons who self-identified as Chinese ethnicity were less likely to participate in period 2 compared to those of self-identified white ethnicity. Recruiting participants via RedFlagDeals was the most efficient in terms of research staff time and had the highest uptake, whereas Twitter was the least efficient. The anonymity feature of online questionnaires may allow individuals the freedom to engage in behavior undesirable to researchers, such as completing questionnaires multiple times or providing false information. These behaviors are probably based on a desire to receive the maximum amount of an incentive possible, rather than a primary desire to contribute to the actual research. Failure to identify this type of behavior can compromise data integrity. Many of our findings and recommendations (summarized in Textbox 1) are consistent with those included in the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) statement as ways to prevent and detect multiple entries from the same individual [18]. We have identified several additional methods to help ensure authentic data are obtained from unique persons beyond enabling cookies and conducting IP checks.

Textbox 1. Recommendations for researchers and designers for Internet-based questionnaires.

- Include a disclaimer stating that any activity such as completing the questionnaire multiple times will be considered to be suspicious, the entries classified as invalid, and that no incentive will then be provided.
- Activate restrictions such as cookies or IP address restrictions to prevent participants from submitting multiple questionnaires.
- Include logic checks and repeated questions with definite answers (eg, username, birth date) throughout the questionnaire.
- Choose an incentive that is accessible and appealing to the geographical target. If there are sufficient funds in the budget, use post office mail instead of email to deliver the incentive. This provides another opportunity to validate a geographical criterion and detect if multiple submissions are present.
- Examine the paradata (eg, questionnaire completion time, IP address) to identify inappropriate activity or gaming.
- Deploy questionnaire-specific links for each advertising mode (multiple site entry linkage) to help identify the questionnaire participant's source, rather than relying on self-report.

Gaming

Preventing multiple submissions is a significant challenge with online recruitment. Users can easily create multiple email accounts or use different computers (thus different IP addresses) to complete questionnaires several times to maximize receipt of an incentive. In this study, approximately 43.07% (469/1089) of our participants made multiple submissions, which is significantly higher than reported in other studies. In an online questionnaire of men who have sex with men (MSM), Bowen et al [12] detected that 33% of their sample made multiple submissions, with the largest contributions coming from participants classified as "infrequent" (2-5 submissions; 36% of all multiple submissions) and "hackers" (>30 submissions; 34% of all multiple submissions). In another study of Latino MSM, 10% of the respondents made multiple submissions and 55% of these came from 1 person [10]. Differences in these rates across studies may be explained by how multiple submissions were defined and identified, where the study was advertised, the type of survey restrictions implemented to control or limit multiple submissions, the value of the incentive, and the type of participants targeted. For example, to gain more exposure, we posted on the deal forum websites under the "freebie" and "paid survey" categories because these were popular. In retrospect, we suspect that posting under these categories may particularly attract people who are highly motivated to obtain incentives and, thus, are more likely to make multiple submissions. We found that incentives seemed to have the least impact on the response rate of women and the greatest impact on the response rate by individuals who self-reported as being of Chinese ethnicity. This may be of value to researchers who are recruiting individuals of a specific gender, race, or ethnicity, and would be worthwhile to explore further.

Providing an incentive has been associated with increasing the frequency of multiple submissions. Bowen et al [12] found that men who were designated as eligible for compensation were 6 times more likely to have multiple submissions than the unpaid group [12]. Similarly, we found that the proportion of questionnaires submitted from nonunique IP addresses during period 1 (incentive offered) was 6 times larger than during period 2, the no-incentive period (47% vs 8%, respectively). Although we also activated software restrictions during period 2 to prevent multiple submissions from nonunique IP addresses, we suspect the incentive was the main driver of multiple submissions because the proportion of questionnaires received from non-Ontario residents was significantly higher when there was an incentive and the use of software restrictions should not have impacted that. Participants who made multiple submissions from the same IP address or email address spent less time per questionnaire, generally completed all the questionnaires consecutively within a short period of time, and had a higher percentage of missing data in their responses. This is not surprising because participants who have completed the questionnaire multiple times are likely to be familiar with the questions and therefore take less time to complete subsequent submissions. Also, by consecutively filling out the questionnaires and skipping questions, participants can maximize the total value of incentives received in a relatively short period of time. This type of behavior can have detrimental effects on the research budget; we estimate that a total of CAD $1485 (297 CAD $5 giftcards) was spent on individuals who were likely gaming the system.

Several strategies have been recommended to prevent or minimize the occurrence of multiple submissions: requiring participants to provide a unique identifier at the start of the questionnaire; using questions with definite answers (eg, birth date, name, phone numbers) repeated in the questionnaire, including questions specifically designed to catch dishonest participants; activating software restrictions to prevent participants from entering the questionnaire from the same IP address; and enabling cookies to prevent participants from entering the questionnaire from the same computer station [7,19-21]. Although none of these strategies are perfect, they provide extra barriers to detect and prevent multiple submissions. Prior studies using social media and other online resources as recruitment tools have advised implementing a protocol which includes checking IP and email addresses and following up with telephone confirmations as needed to guard against this behavior [22,23]. Conducting a reverse lookup of IP address location or postal code (as we did) can also identify participants who are outside the geographical target.

Incentive Type

We suspect that the amount and type of incentive in this study may have contributed to the high rate of fraudulent data (ie, submissions from outside Ontario although the respondent indicated Ontario residence) in our sample. The value of the incentive (CAD $5 gift card) may have been too high for simply completing a 5-minute questionnaire. We used electronic Amazon gift cards that can be used in a wide variety of jurisdictions because they were easy to distribute by email, which can be used to purchase a wide variety of items. It is possible that a small percentage of parents truly resident in Ontario might have been visiting outside of Canada when the questionnaire was completed, but we doubt that this was the case for most persons classified as not being Ontario residents. For studies that are attempting to recruit participants from a specific geographic area, we recommend the use of an incentive that is only locally accessible (eg, local grocery chain gift card). Alternately, incentives could be sent to recipients by first-class mail only because the post office will return them if the mailing address is not valid, providing another opportunity to validate geographic residence.

Advertising Mode Evaluation

In this study, we found that the deal forum websites (RedFlagDeals and SmartCanucks) had the largest uptake while requiring the least amount of time to maintain or update. Online classified advertisement websites (Craigslist and Kijiji) were also helpful with recruiting participants, especially during period 2. A reporter from the CBC national news corporation contacted us after seeing the advertisement in Kijiji, which resulted in our first media exposure and generated subsequent interviews and media stories. Without this exposure, we might have had more difficulty recruiting during period 2, given there was no incentive and most participants only heard about the study through conventional media during this period.

Using social media (Facebook and Twitter) was not as efficient for study recruitment as we had anticipated. Previous studies have examined using Facebook for recruitment, but these studies relied on paid advertisements [5,6]. In our case, we disseminated recruitment information by social networks, of which most groups were professional health care organizations. We likely would have had better success with these recruitment methods if we were more widely known before our online presence or had a larger network of followers on social media. Given the novelty of using social media for study recruitment and the limited resources available, we were hesitant to engage with our audience on topics unrelated to study recruitment, which may have affected the interest level. Further, most of our communication was one-way and focused entirely on study recruitment, which may not have been appealing to our audience. Instead, it would have been helpful to actively engage with our audience about related topics of interest (eg, immunization, cold and flu tips). When developing a social media strategy for study recruitment, researchers need to consider how to balance the needs and interests of their social media audience with the study objectives.

Strengths and Limitations

We explored a range of relatively novel advertising approaches for the purpose of recruiting parents for research. We also conducted automated and manual reviews to scan email addresses to identify multiple submissions. Although it was more resource intensive, manual methods allowed us to identify additional email addresses leading to identification of multiple submissions that would not have been detected by automated methods alone. To validate one of the eligibility criteria, we reviewed all the postal codes to ensure they were from Ontario and did a reverse lookup of IP addresses to verify an Ontario address when the postal code and country of IP address was absent. The limitation of relying on an IP address is that it can easily be manipulated or hidden. Ultimately, we successfully recruited all participants (n=55) needed for the focus groups with acceptable cost. Even so, a participant who self-reported being a parent during screening revealed that this was not true while participating in a focus group. This person left the focus group claiming she had not clearly understood this criterion for eligibility. This highlights the challenges in ensuring the validity of self-reported information, particularly for online responses.

We may have eliminated some legitimate respondents by removing questionnaires deemed to have been multiple submissions; it is possible that different people can share the same IP address. We may also have eliminated legitimate Ontario residents by removing questionnaires from non-Ontario IP addresses; some Ontario residents may have been traveling when responding or, because of the use of Web proxies, virtual private networks, and mobile networks, the IP address may not have been representative of the participant's physical location. However, given that our analysis suggested gaming, we elected to err on the side of caution in retaining participants for focus group selection. The recruitment methods used in this study tended to attract a younger and more educated population, and should be used with caution if the main objective is to obtain a representative sample of parents. We defined a rural participant as someone who had a postal code with a zero in the second position of the 6-digit postal code. This indicates residence in an area that is not accessible by letter carriers. This may differ from definitions used outside of Canada; therefore, recommendations from rural participants in this study may be not be generalizable elsewhere.

A key limitation of online recruitment methods includes the inability to reach socioeconomically or educationally disadvantaged groups who may lack the skills to use, or have adequate access to, the Internet [24]. Further, although we asked participants how they heard about the study, the validity of this response may also be an issue because this was based on self-report. Such self-reports can be validated if the study design uses unique study URLs for each advertising mode (multiple site entry technique) [21]. We did not track the number of advertisements that public health posted about the study; therefore, we cannot fully assess the contribution from this mode. Additionally, because this stage of our study involved only a single brief questionnaire, we were not able to evaluate the effect of the advertising modes on retention rates. Such data would be very valuable to researchers. Finally, we conducted a series of post hoc analyses to examine differences between the patterns of response for different scenarios (unique vs multiple submissions, period 1 vs period 2), increasing the risk of type 1 error.

Conclusions

Our study has identified that the Internet can be a useful way to recruit parents for research in a relatively short time with limited resources, but it also identified important elements that researchers need to consider in order to fully utilize the Internet for study recruitment. When studies are conducted face-to-face, participants may assume more accountability for their behavior than with research conducted online. The anonymity of online research could lead some individuals to engage in dishonest behavior from which they would otherwise refrain if they knew they could be easily identified. When incentives are linked to email addresses, it is easy for individuals to create new accounts to collect more than 1 incentive.

It is imperative for researchers to implement control measures in the study design and questionnaire programming stages to limit and detect gaming. As the Internet evolves and more individuals are accessible online, it will be critical for researchers and questionnaire designers to understand how best to address these challenges to ensure and preserve data integrity.

Acknowledgments

The Canadian Association for Immunization Research and Evaluation provided networking assistance. The Public Health Agency of Canada/Canadian Institutes of Health Research Influenza Research Network provided funding. The PCIRN Program Delivery and Evaluation Group members are: Julie Bettinger, Nicole Boulianne, Stephanie Brien, David Buckeridge, Larry Chambers,

Natasha Crowcroft, Lois Crowe, Shelley Deeks, Michael Finkelstein, Maryse Guay, Jemila Hamid, Faron Kolbe, Jeff Kwong, Allison McGeer, Jennifer Pereira, Susan Quach, Sherman Quan, Margaret Russell, Beate Sander, Doug Sider, Chris Sikora, Anne Wormsbecker, and Anne-Luise Winter.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample of text used for advertisements.

[PDF File (Adobe PDF File), 12KB - jmir_v15i11e250_app1.pdf]

References

1. Sherr L, Davé S, Lucas P, Senior R, Nazareth I. A feasibility study on recruiting fathers of young children to examine the impact of paternal depression on child development. Child Psychiatry Hum Dev 2006;36(3):295-309. [doi: 10.1007/s10578-005-0004-3] [Medline: 16362240]
2. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. Advances in Psychiatric Treatment 2003;9:229-238. [doi: 10.1192/apt.9.3.229]
3. 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]
4. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
5. 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]
6. 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]
7. Matsuo H, McIntyre KP, Tomazic T, Katz B. The online survey: its contributions and potential problems. In: ASA Section on Survey Research Methods. 2004 Presented at: American Statistical Association; Aug 8-12, 2004; Toronto, Canada p. 8-12.
8. Mustanski BS. Getting wired: exploiting the Internet for the collection of valid sexuality data. J Sex Res 2001;38(4):292-301.
9. McCabe SE, Boyd CJ, Couper MP, Crawford S, D'Arcy H. Mode effects for collecting alcohol and other drug use data: Web and U.S. mail. J Stud Alcohol 2002 Nov;63(6):755-761. [Medline: 12529076]
10. Konstan JA, Rosser BRS, Ross MW, Stanton J, Edwards WM. The story of subject naught: a cautionary but optimistic tale of Internet survey research. J Comput-Mediat Comm 2005;10(2):11. [doi: 10.1111/j.1083-6101.2005.tb00248.x]
11. Joinson AN. Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications. San Diego, CA: Academic Press; 1998.
12. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in Internet research: preserving data integrity. AIDS Behav 2008 Nov;12(6):964-973 [FREE Full text] [doi: 10.1007/s10461-007-9352-2] [Medline: 18240015]
13. Buck L. The Cord. 2013. Social media use climbs in Canada URL: http://www.thecord.ca/social-media-use-climbs-in-canada/ [accessed 2013-06-21] [WebCite Cache ID 6HXzMpDlj]
14. Public Health Ontario, PHAC/CIHR Influenza Research Network. School flu shots. 2013. URL: http://www.schoolflushots.ca/ [accessed 2013-05-22] [WebCite Cache ID 6GoSViIQG]
15. bitly blog. 2012 May 08. Time is on your side URL: http://blog.bitly.com/post/22663850994/time-is-on-your-side [accessed 2013-05-22] [WebCite Cache ID 6GoShUa8E]
16. WhatIsMyIP. 2013. IP address lookup URL: http://www.whatismyip.com/ip-tools/ip-address-lookup/ [accessed 2013-06-21] [WebCite Cache ID 6HY0x0t5n]
17. Canada Post. 2013. Addressing guidelines URL: http://www.canadapost.ca/tools/pg/manual/PGaddress-e.asp [accessed 2013-06-21] [WebCite Cache ID 6HXzl4rUl]
18. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
19. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004;55:803-832. [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]
20. Nosek BA, Banaji MR, Greenwald AG. E-Research: Ethics, security, design, and control in psychological research on the Internet. J Social Isssues 2002 Jan;58(1):161-176. [doi: 10.1111/1540-4560.00254]
21. Reips UD. The Web experiment method: advantages, disadvantages, solutions. In: Psychological Experiments on the Internet. San Diego, CA: Academic Press; 2000:89-92.

22. 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]
23. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
24. Suarez-Balcazar Y, Balcazar FE, Taylor-Ritzler T. Using the Internet to conduct research with culturally diverse populations: challenges and opportunities. Cultur Divers Ethnic Minor Psychol 2009 Jan;15(1):96-104. [doi: 10.1037/a0013179] [Medline: 19209984]

Abbreviations

CBC: Canadian Broadcasting Corporation
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IP: Internet protocol
MSM: men who have sex with men

Edited by G Eysenbach; submitted 15.07.13; peer-reviewed by S Close, L Taylor; comments to author 15.08.13; revised version received 03.09.13; accepted 15.09.13; published 11.11.13.

Please cite as:
Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, Quan SD, Kwong J
The Good, Bad, and Ugly of Online Recruitment of Parents for Health-Related Focus Groups: Lessons Learned
J Med Internet Res 2013;15(11):e250
URL: http://www.jmir.org/2013/11/e250/
doi:10.2196/jmir.2829
PMID:24231040

Original Paper

Comparison of US Panel Vendors for Online Surveys

Benjamin M Craig[1,2], PhD; Ron D Hays[3,4], PhD; A Simon Pickard[5], PhD; David Cella[6], PhD; Dennis A Revicki[7], PhD; Bryce B Reeve[8], PhD

[1]Moffitt Cancer Center, Tampa, FL, United States
[2]University of South Florida, Tampa, FL, United States
[3]University of California, Los Angeles, Los Angeles, CA, United States
[4]RAND, Santa Monica, CA, United States
[5]Department of Pharmacy Systems, Outcomes, and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, United States
[6]Northwestern University, Chicago, IL, United States
[7]Outcomes Research, Evidera, Bethesda, MD, United States
[8]University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Benjamin M Craig, PhD
Moffitt Cancer Center
12902 USF Magnolia Drive
MRC-CANCONT
Tampa, FL, 33612-9416
United States
Phone: 1 813 745 6710
Fax: 1 813 745 6525
Email: Benjamin.Craig@moffitt.org

Abstract

Background: Despite the increasing use of panel surveys, little is known about the differences in data quality across panels.

Objective: The aim of this study was to characterize panel survey companies and their respondents based on (1) the timeliness of response by panelists, (2) the reliability of the demographic information they self-report, and (3) the generalizability of the characteristics of panelists to the US general population. A secondary objective was to highlight several issues to consider when selecting a panel vendor.

Methods: We recruited a sample of US adults from 7 panel vendors using identical quotas and online surveys. All vendors met prespecified inclusion criteria. Panels were compared on the basis of how long the respondents took to complete the survey from time of initial invitation. To validate respondent identity, this study examined the proportion of consented respondents who failed to meet the technical criteria, failed to complete the screener questions, and provided discordant responses. Finally, characteristics of the respondents were compared to US census data and to the characteristics of other panels.

Results: Across the 7 panel vendors, 2% to 9% of panelists responded within 2 days of invitation; however, approximately 20% of the respondents failed the screener, largely because of the discordance between self-reported birth date and the birth date in panel entry data. Although geographic characteristics largely agreed with US Census estimates, each sample underrepresented adults who did not graduate from high school and/or had annual incomes less than US $15,000. Except for 1 vendor, panel vendor samples overlapped one another by approximately 20% (ie, 1 in 5 respondents participated through 2 or more panel vendors).

Conclusions: The results of this head-to-head comparison provide potential benchmarks in panel quality. The issues to consider when selecting panel vendors include responsiveness, failure to maintain sociodemographic diversity and validated data, and potential overlap between panels.

(J Med Internet Res 2013;15(11):e260) doi:10.2196/jmir.2903

KEYWORDS

survey methods; community surveys; sampling bias; selection bias; Internet; data sources

Introduction

The dramatic growth of the use of panel vendors for online survey research has been described as "one of the most compelling stories of the last decade" [1]. A panel vendor is an organization that recruits and matches participants to a specified target audience of a survey to collect reliable quantitative information about the participants' preferences and behaviors. They provide a wide range of services that allow researchers to expeditiously accrue survey respondents while protecting their anonymity, including maintaining the panel (recruitment and database), verifying identities (quality control), selecting and inviting panelists, compensating respondents for participation, and delivering panel entry and survey data for all invited panelists. In the field of survey research, the use of a panel vendor is attractive because of the widespread availability of Internet-linked devices (eg, tablets, smartphones, and laptops), enabling panel vendors to target difficult-to-reach populations. Online surveys through panel vendors allow researchers to collect data efficiently and inexpensively (eg, less than US $10/completed survey), but this approach has its shortcomings. A preliminary literature search revealed multiple studies comparing survey mode (eg, online, face-to-face, telephone, postal) [2-10]; however, systematic comparisons of panel vendors for online surveys were not found, which is particularly troubling because of their almost exclusive use of nonprobability recruitment methods.

Because most panel vendors rely on nonprobability-based recruitment, their samples have unknown representativeness of target populations. In an attempt to address this uncertainty, panel vendor users typically specify a target number of respondents along with selected demographic characteristics (quota sampling) and rely on poststratification adjustments (analytic weights) to compensate for nonresponse and noncoverage. That is, panel respondents are weighted so that the marginal distributions of gender, age, race/ethnicity, education, and other demographic characteristics match a target distribution (eg, US Census). Quota sampling is necessary because of concerns that standing panels differ from the general population in that they are more educated, have a higher income, and are more likely to be younger, white, and non-Hispanic [1].

Panel vendors often advertise that they have large and nationally representative panels for timely and accurate online surveys of the United States. In addition to favoring nationally representative samples, most survey researchers prefer that invited panelists respond quickly and accurately. This study conducted a head-to-head comparison of 7 US panel vendors by giving each the same task and comparing them along 3 basic quality criteria:

1. Data efficiency: Do their invited panelists respond quickly?
2. Data validity: Are their respondents who they say they are?
3. Panel representativeness: Do their respondents have similar characteristics to the US general population?

This paper excludes potentially relevant details beyond the survey screener (eg, dropout rates and survey completion), which may be examined in future research and will likely vary by survey instrument. The purpose of this study was to typify panel vendors and their respondents and to provide helpful information, while acting in accordance to contract restrictions. Therefore, panel vendor names were withheld to protect their anonymity and were assigned a rank-based identifier based on the proportion of invited respondents who consented (panel vendor [PV]; PV1 to PV7). All study procedures were approved by the University of South Florida Institutional Review Board (IRB # Pro00000076).

Methods

Panel Vendors

A standardized scope of work was described in the request for quotes (RFQ), including the following 7 requirements:

1. 1000 US adult respondents at less than US $10 per completed survey;
2. 18 demographic quotas filled using their own panel (no partners or brokers);
3. Third-party survey hosting with email invitation only (no river sampling, routing, or banners);
4. Panel entry data on each invited panelist (ie, date of birth, gender, race/ethnicity, socioeconomic status, health status, and geographic location);
5. Date and time of invitation and panel enrollment for all invited panelists;
6. Vendor affiliation with either the European Society for Opinion and Market Research (ESOMAR), the Council of American Survey Research Organizations (CASRO), and/or the American Association for Public Opinion Research (AAPOR); and
7. Vendor responses to the ESOMAR 26, an industry standard and a survey instrument designed to help research buyers of online samples [11].

All possible panel vendors were contacted and those who met the requirements were included. Some (if not most) panel vendors are panel or list "brokers" (ie, they outsource surveys to partners who then administer it to their own panels) and did not meet the requirements of this study. Requirements specified that all respondents be recruited from a single proprietary source of standing panelists (no partners) and be invited using a generic email invitation. Sampling from multiple panels or without invitation would have complicated the delivery of panel entry data on all invitees (requirement #4).

First, to identify all possible panel vendors, the third-party survey host (ie, website provider) was asked to recommend panel vendors based on their prior experience. Second, a review was performed of all panel vendor advertisements and membership to standard-setting organizations in online research, such as AAPOR, ESOMAR, CASRO, and Quirk's Marketing Research Review [11-14]. Third, referrals were solicited from experts in the field based on their experiences with various panel vendors. Each time a potential panel vendor was identified, their website was evaluated to ascertain whether they met the study requirements.

As of March 2012, 134 panel vendors were identified and reviewed [15]. After removing panel brokers and those who could not meet the 7 requirements, RFQs were sent to 23 panel

vendors; however, only 12 panel vendors provided a quote that met the 7 requirements. Once the project's scope of work was agreed upon by the panel vendor representative, a contract was sent to the panel vendor legal team for review and approval, with a turnaround time of as little as 1 week or as long as 5 months. Among the 12 panel vendors, 5 declined, largely because of disagreements between the panel vendor legal team and account representatives about the scope of work. The time from start to finish—RFQ to signing—ranged from 3 weeks to 6 months for the successful 7 panel vendor collaborations.

Survey Design

Overview

At launch, panel vendors sent generic email invitations directly to their respective panelists in a series of waves. For this study, response time is defined by the number of days from the time of invitation to the time of consent (ie, do their invited panelists respond quickly?). This measure is analogous to response time in postal surveys, in which response is intrinsically linked to consent because nonconsenting respondents rarely return postal surveys. In this study, some respondents closed their browser on the consent page or stated nonconsent, but returned later to consent and proceed with the online survey (ie, response time).

After consenting, respondents answered questions about their current US state of residence, ZIP code, age, birth date, gender, race, and ethnicity, and were assessed for 4 technical requirements that enabled participation in the survey [15].

Complete Pass-Through Information

Upon clicking the invitation link, a respondent identification number was sent (ie, passed through) to the survey website to identify the panelist for payment purposes.

Complete Panel Entry Information

Each panel vendor was required to deliver the time and date of invitation for each invited panelist, which was linked to survey responses by using the respondent identification number.

Complete or Valid Geolocation

The Internet Protocol (IP) address was invalid if it was for a proxy server or if its geolocation was unknown or outside the United States. Respondents were required to be within the 50 US states or the District of Columbia for this study.

JavaScript Enabled

JavaScript was required for the survey software to function as designed.

To assess discordance, survey responses were compared with each other (ie, age and birth date; state and ZIP code), with the panel entry data (eg, gender), and with IP address data (eg, state). Discordance in race and ethnicity was not assessed because of differences in questions used by panel vendors at panel entry. Proof of valid identity was defined as responses that meet the technical requirements of the survey and were concordant (ie, are their respondents who they say they are?).

The last page of the screener asked respondents about their annual income and educational attainment. Panel vendors were required to fill 18 demographic quotas. Taking into account the demographic quotas, the validated respondents were assessed as to their representativeness of the US population in terms of income, education, and current US state of residence based on US Census data [16,17].

Statistical Analysis

To assess response time, the proportion of invitees who responded on the same day, next day, and second day by panel vendor were estimated. Although some panel vendors provided invitation times, no panel vendor listed time zones that would allow hourly analysis. In violation of scope of work, PV7 did not provide invitation dates and was excluded from the response time analysis.

To validate respondent identity, the proportion of consented respondents who failed to meet the technical criteria, failed to complete the screener questions, and provided discordant responses were estimated. Discordance in self-reported responses was based on 4 indicators: (1) gender differed from panel entry data, (2) reported age differed from reported birth date, (3) birth month and year differed from panel entry data, and (4) ZIP code differed from current US state. The proportion of consented respondents with each discordant indicator is reported along with 2 further indicators: current US state disagreed with IP state, and birth date disagreed with panel entry data. These indicators were excluded from the definition of discordance because of panel-specific issues. Specifically, the bulk of IP state data were lost for PV2 because of an error in the Web-based survey software. In addition, PV4 reported that the day of birth in their panel entry data defaulted to the first of the month for a large portion of their panelists.

Before examining representativeness, respondents across panels were compared on self-reported birth date and ZIP code to assess the potential for overlap within and between panel vendors (ie, respondents completed the survey multiple times because they were panelists in more than 1 panel or were enrolled multiple times in the same panel). Under independence, the probability of finding a specific ZIP–birth combination from 1 panel in another panel is the product of 3 factors: (1) combined sample size of all other panels (N), (2) proportion of respondents in all panels with that birthdate (Sbirth), and (3) the proportion of respondents in all panels with that ZIP code (SZIP) This probability increases with sample size, the proportion with that birth date, and the proportion with that ZIP code (S=N × Sbirth × SZIP). Potential overlap between a panel and all other panels due to chance was estimated by the mean of these probabilities and compared to the actual overlap between panel vendors. This is a conservative estimate because birth dates may naturally cluster within ZIP codes (eg, universities, retirement communities).

To assess representativeness among the respondents who passed the screener, sampling weights were applied to 18 demographic quotas by panel vendor. Each quota was defined by age (18-34 years; 35-54 years; >55 years), gender (male; female), and race/ethnicity (Hispanic; black, non-Hispanic; white or other, non-Hispanic). The proportion of the sample in 6 education categories, 8 annual household income categories, and 9 US Census Bureau divisions were estimated. For comparison, these

estimates were presented alongside national estimates from the 2010 American Community Survey (ACS) [18,19].

Results

Figure 1 shows response rates by day among 6 of the 7 panel vendors, demonstrating that PV1 and PV2 had approximately double the response rates of the other panel vendors. The majority of invited panelists who consented did so on the day of invitation with further accrual on the next day, particularly when the invitations were sent in the evenings. Few consented on the second day after invitation.

Figure 2 describes the proportion of consented respondents who failed the screener. PV1 had a higher proportion largely because it was unable to find panel entry data for more than 1000 of those recruited, which was a technical requirement for this study. Only a few panelists failed to complete the screener after consent, which may be attributable to drop out or Internet disruptions. Aside from PV1's panel entry data issue, approximately 20% of the consented respondents failed the screener, with the majority of loss because of discordant responses.

Figure 3 indicates the proportion of discordant responses among the respondents who completed the screener. Self-reported gender discordance ranged from 1% (PV5) to 2% (PV2), suggesting mostly agreement with panel vendor data. Self-reported age and ZIP code largely agreed with self-reported birth date and current US state with discordance ranging from 4% (PV7) to 6% (PV2) and from 3% (PV6) to 5% (PV2), respectively. Self-reported current US state using the IP state for 7% (PV7) to 9% (PV1) of respondents was unable to be verified, which was largely attributable to missing data. IP state was not captured for PV2 because of an error in the survey software.

The largest source of variability in discordance between panel vendors was birth date. Discordance in birth month and year for PV4 was twice that of PV3 (4% vs 9% and 3% vs 8%, respectively). Discordance in date of birth was greatest for PV1, PV2, and PV4. Because of 30% discordance in date of birth, PV4 acknowledged their use of the first of the month as their default, which largely invalidated the birth dates in their panel entry data. PV2 reported that some of their panelists intentionally report inaccurate date of birth to protect their identities.

To understand better the relationship among panel vendor samples, overlap of panel members was measured by the proportion of respondents in a panel vendor sample who reported a birth date and ZIP code identical to 1 or more respondents in another panel vendor sample. Although some repetition may be due to chance alone, Table 1 depicts systematic relationships between and within panel vendor–specific samples. Due to within-sample repetition, PV1 and PV6 may have allowed a small number of respondents to participate more than once (2% and 0.1%, respectively).

Figure 1. Proportion of invited panels who consented by day and panel vendor.

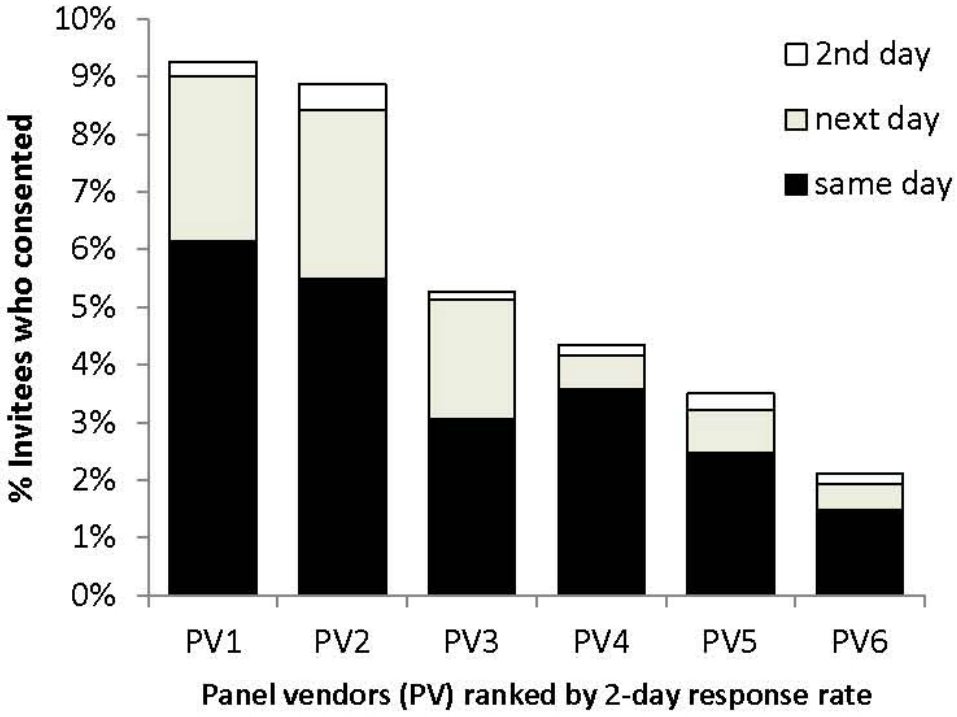


Figure 2. Proportion of consented respondents who failed the screener by reason and panel vendor.

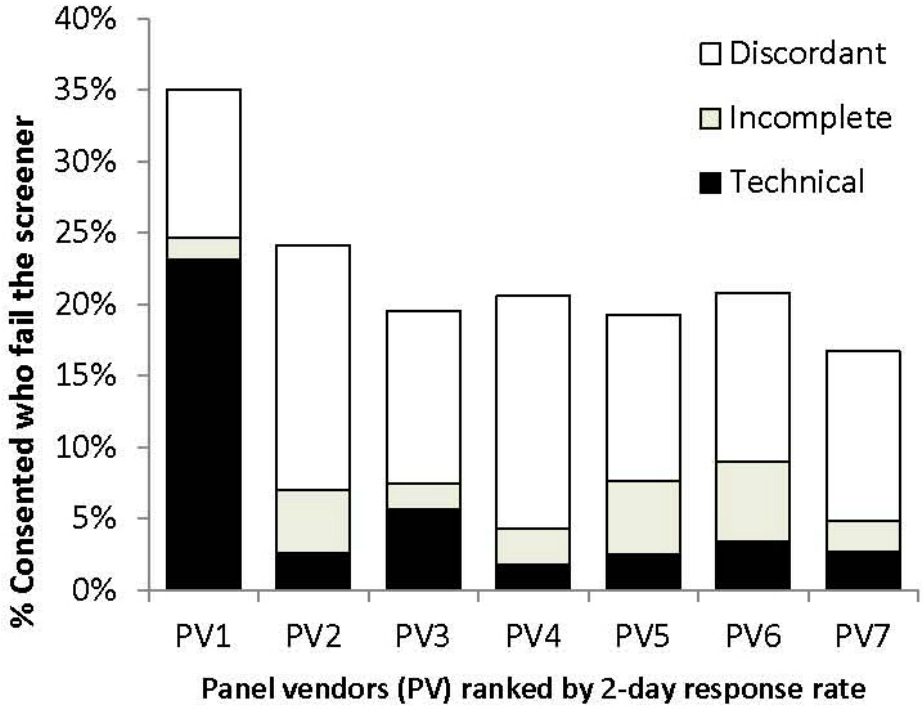


Figure 3. Proportion of discordant responses by self-reported attribute and panel vendor.

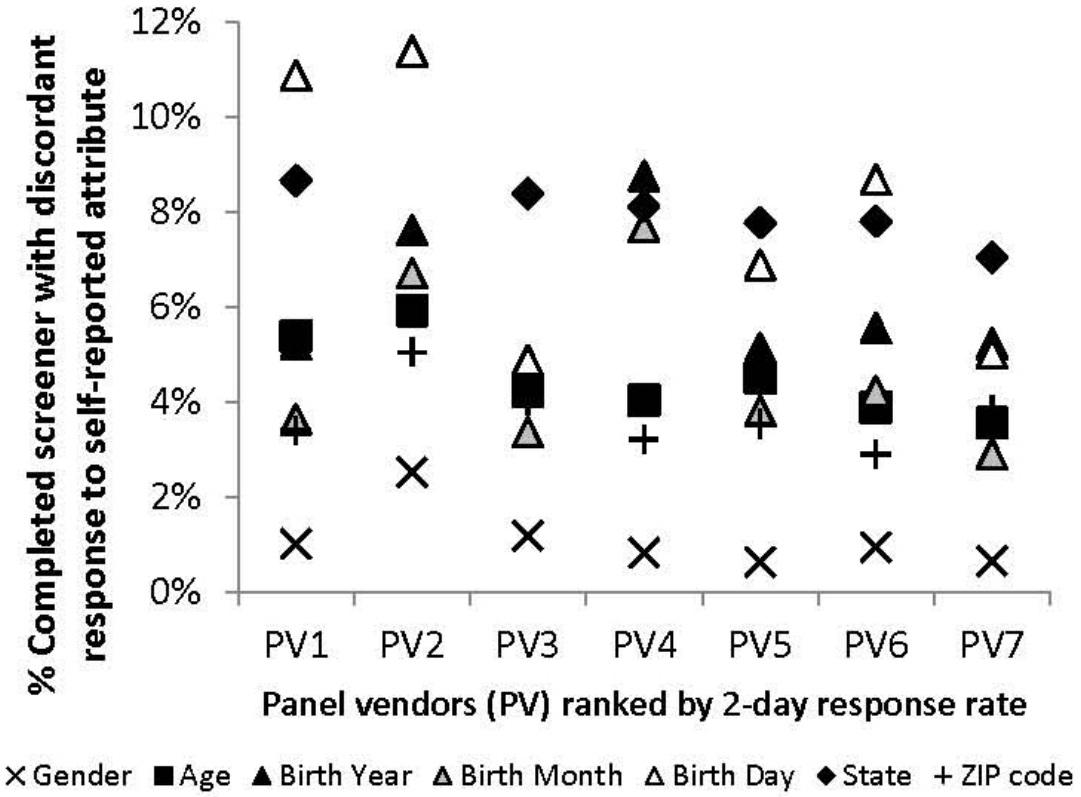


Table 1. Overlap[a] within and between panel vendors (PV).

PV	PV1	PV2	PV3	PV4	PV5	PV6	PV7
Sample size	4183	3874	3158	2190	3644	2704	2154
PV1	1.7%	0.3%	6.2%	8.6%	8.1%	12.2%	5.6%
PV2	2.0%	0.0%	0.3%	0.5%	0.4%	0.6%	0.2%
PV3	6.3%	0.3%	0.0%	7.4%	4.8%	5.8%	3.2%
PV4	6.1%	0.3%	5.1%	0.0%	4.6%	4.7%	3.1%
PV5	8.6%	0.4%	5.5%	7.7%	0.0%	5.6%	4.2%
PV6	9.5%	0.3%	4.9%	5.7%	4.1%	0.1%	3.5%
PV7	4.5%	0.1%	2.2%	3.0%	2.5%	2.9%	0.0%
Any	23.2%	1.3%	18.6%	25.2%	19.7%	23.9%	14.6%

[a] Overlap is measured by the proportion of screened respondents in the column PV sample who reported a birth date (ie, day, month, year) and 5-digit ZIP code identical to 1 or more screened respondents in the row PV sample.

Aside from PV2, all pairs of panel vendors had more than 2% overlap, suggesting recruitment from a common source. PV2 had the least overlap with any other panel vendor (1%). The greatest overlap was PV6 respondents who reported identically to 1 or more PV1 respondents (12%). Aside from PV2, the other panel vendors appeared to draw 15% to 25% of their sample from a common pool, likely because of panelists enrolling with multiple panel vendors. Assuming birth date and ZIP codes are unrelated, the predicted overlap because of chance ranged from 0.027% to 0.039%, which is less than the overlap observed between panels.

Figures 4-6 illustrate the representativeness of the panel vendor samples after applying demographic weights to the respondents who passed the screener. PV2 was distinct from the other 6 panel vendors (PV1 and PV3-PV7), favoring higher income and educational attainment as well as respondents in Western states. This relative skewness in PV2 improved representativeness in graduate education, incomes over US $150,000, and in the Pacific region, although sacrificing representativeness at high school education or less, incomes less than US $25,000, and along the Atlantic coast states.

Compared to the 2010 ACS estimates, the geographic differences seemed minor (all within a band of 4%). All panel vendors underrepresented adults who did not graduate from high school or had annual incomes less than US $15,000, with PV2 having the largest deficiency (–14% and –9%, respectively). However, 2010 ACS estimates included persons in institutionalized settings (eg, skilled nursing facilities, adult correctional facilities, and psychiatric hospitals) where Internet access may be restricted.

Figure 4. Weighted proportion of respondents who passed screener by educational attainment and panel vendor (PV).

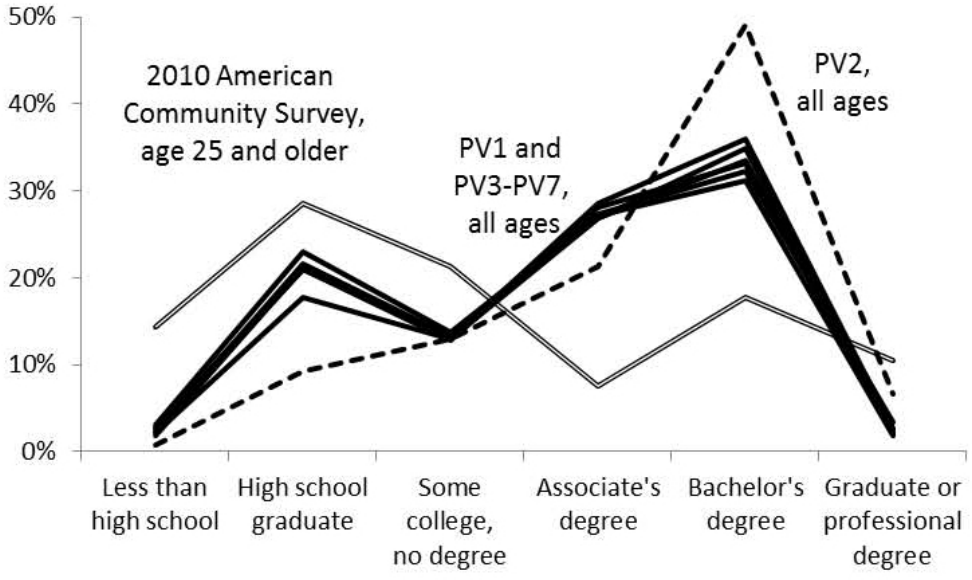


Figure 5. Weighted proportion of respondents who passed screener by annual household income in 2011 and panel vendor (PV).

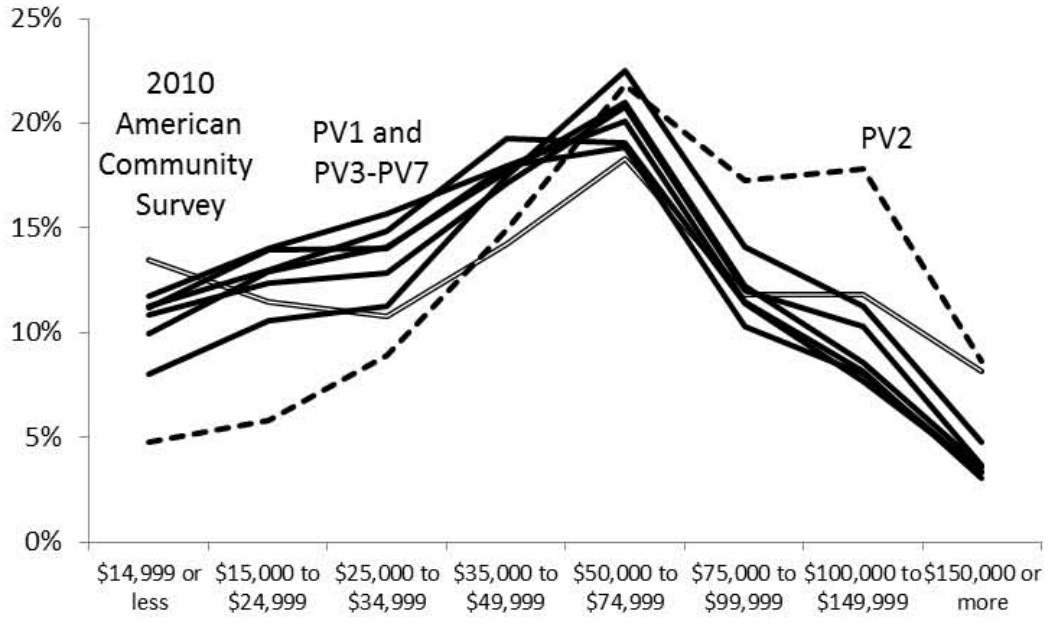
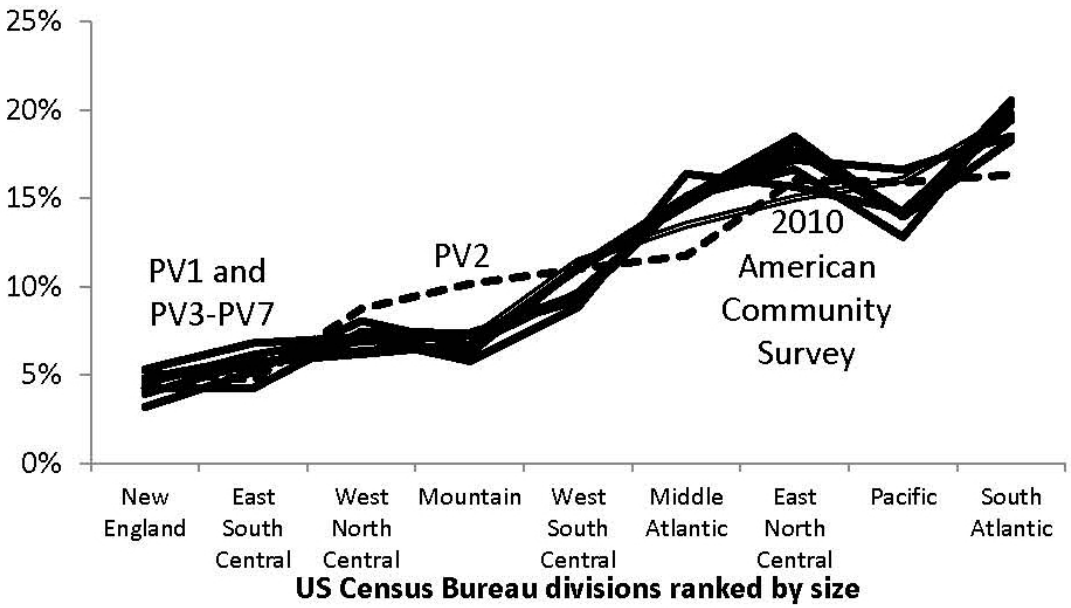


Figure 6. Weighted proportion of respondents who passed screener by US Census division and panel vendor (PV).



Discussion

The increased use of panel vendors for online survey research makes it essential to understand the variability, differences, and shortcomings of panel vendors. The results of the study show variability in panel quality between the vendors and provide practical benchmarks for survey researchers. Points to consider when selecting panel vendors include responsiveness, failure to maintain sociodemographic diversity and validated data, and potential overlap between panels. Additionally, choosing to use an online survey panel offers advantages and disadvantages; therefore, the survey task itself should determine the mode used. In a matter of weeks and with a modest budget (<US $10 per complete), a general sample of US adults from 7 panel vendors was recruited. In addition to the 7 technical requirements for panel vendors, the project results may serve as benchmarks for panel vendor users who expect a 2-day 10% response rate with less than 10% of the consenting respondents failing the screener.

Discordance in birth dates is more of an issue compared with ZIP codes and other demographic characteristics. Six of the 7 panel vendors (PV1 and PV3-PV7) had a 25% overlap in respondents, which may explain why respondents in these 6 panel vendor samples provided similar educational attainment, geographic, and income responses. This study found that all panel vendor samples underrepresent low socioeconomic respondents, particularly PV2, which may motivate the greater use of socioeconomic status quotas in future work.

It has been argued that the representativeness of US panels is approximately that of random digit dialing surveys. In the 1970s, 89% of the US population had landline phones in contrast to less than 80% now [1]. Approximately 72% of the US population uses the Internet on a regular basis [1]. Even with recent increases in access to the Internet, online surveys have shown biases toward younger age, college education, higher socioeconomic status, English speaking, white, non-Hispanic ethnicity, literate, nonvisually impaired, and persons with low time costs [4,5,8,9,20,21]. A number of other studies have noted the nonrepresentativeness of panel entry data, perhaps attributable to self-selection or demographic and other unmeasured characteristic differences between panels and the general population [20,22-25]. These observable differences may affect the generalizability and self-selection for participation in panels may compromise external validity. For example, in the National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System (PROMIS) project [26], despite a sample-matching methodology to ensure a sufficient representation in important subgroups, the resulting sample remained different from the 2010 US Census. Compared to the census data, the PROMIS unweighted general population sample was 5 years older, had a higher percentage of males, and had a higher percentage of those having a college education [21,27].

Most panels are recruited using nonprobability-based methods (eg, self-selection). Panel vendor users may apply sample-matching or quota-sampling approaches so that observable characteristics of respondents are similar to a target population. Despite potential similarity on measured variables, it is uncertain how these respondents compare to the desired population on unmeasured variables. For example, interest in participating in different types of surveys may further compromise generalizability. Interested participants may be different from uninterested participants in unobservable ways, as previous research has shown that women are more interested in health topics compared to men [28]. Finally, recruitment rates for participation in panels are only 33% or less [29], which may result in greater nonrepresentativeness of the potential panel-sampling frame. This limitation may be acceptable for research that is not intended to produce precise estimates of a target population.

Aside from panel attributes, response and completion rates vary between online and other modes of administration [6,20]. Advantages of online panels are similar to mailed surveys in that they can be completed at the convenience of the respondent, and the absence of an interviewer may reduce the possibility of social desirability bias. Disadvantages include the inability of respondents to obtain clarification for confusing questions or the inability of interviewers to know when a respondent does not understand the survey questions/tasks. Like postal surveys, respondents may complete the survey with minimal attention to the questions because of distractions and multitasking or they may not complete it at all. Unlike postal surveys, respondents may complete the same online survey multiple times, particularly if the study uses multiple panel vendors or a panel vendor with insufficient control. In addition, some respondents deliberately speed through online surveys just to obtain the incentive. For these reasons, it is important to consider the nature of the survey task before deciding whether an online survey is the best approach to a study.

There are many merits to population-based samples and modes for a variety of research questions and settings. Where population-based sampling is not attainable, a variety of weighting schemes can be applied to results of a panel survey in such a way as to minimize the bias introduced by lack of representation on key demographic characteristics. It is also noteworthy that some studies do not require representativeness to retain their internal validity. Experimental designs that test theory-driven hypotheses within a defined sample can legitimately test those hypotheses, even when the sample is not representative of a larger population (eg, oversampling). Similarly, testing the psychometric performance of questionnaires, particularly the item-level statistics associated with banks of questions that are used to define and measure an underlying concept, does not require demographic representativeness of the sample. Within this context, it is far more important that the sample be sufficiently heterogeneous in regards to covering the full range of the subject that is being measured. In other words, sample representativeness on socioeconomic variables is not always essential for good survey science. It depends on the nature of the study.

To our knowledge, this is the first study to perform a head-to-head comparison of panel vendors on 3 key criteria relevant to researchers: data efficiency, data validity, and panel representativeness. This study demonstrates the variability of panel quality and that no tested panel vendor performed well in meeting their claims of nationally representative samples and

high data quality. In addition to the 3 basic criteria examined in this study, other criteria, such as cost, service, mode (eg, telephone), other countries, and technical capabilities, may be examined in future work. Online survey researchers can objectively apply these benchmarks to assess their own panel experiences and to improve the field of online survey research.

Acknowledgments

The authors thank Carol Templeton and Michelle Owens at Moffitt Cancer for their contributions to the research and creation of this paper. We would also like to thank Joseph Lipscomb and Paul Brown, members of the Expert Advisory Panel, for their continued collaboration and contribution to the PROMIS Valuation Study. Funding support for this research was provided by an NCI R01 grant (1R01CA160104). Ron D Hays was supported in part by grants from the NIA (P30-AG021684) and the NIMHD (P20MD000182).

Conflicts of Interest

None declared.

References

1. Baker R, Blumberg S, Brick JM, Couper MP, Courtright M, Dennis M, et al. AAPOR Report on Online Panels. 2010 Mar. URL: http://www.aapor.org/AM/Template.cfm?Section=AAPOR_Committee_and_Task_Force_Reports&Template=/CM/ContentDisplay.cfm&ContentID=2223 [accessed 2013-11-12] [WebCite Cache ID 6L4jObT1S]
2. Berrens RP, Bohara AK, Jenkins-Smith H, Silva C, Weimer DL. The advent of Internet surveys for political research: A comparison of telephone and Internet samples. Polit Anal 2003;11:1-22. [doi: 10.1093/pan/11.1.1]
3. Duffy B, Smith K, Terhanian G, Bremer J. Comparing data from online and face-to-face surveys. Int J Market Res 2005;47:615-639.
4. Grandjean BD, Taylor PA, Nelson NM. Comparing an Internet panel survey to mail and phone surveys on “willingness to pay” for environmental quality: a national mode test. In: The American Association for Public Opinion Research (AAPOR) 64th Annual Conference, May. Hollywood, FL; 2009 Presented at: The American Association for Public Opinion Research (AAPOR) 64th Annual Conference; May 14-17, 2009; Hollywood, FL p. 5779-5793.
5. Hirsch O, Hauschild F, Schmidt MH, Baum E, Christiansen H. Comparison of Web-based and paper-based administration of ADHD questionnaires for adults. J Med Internet Res 2013;15(3):e47 [FREE Full text] [doi: 10.2196/jmir.2225] [Medline: 23518816]
6. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a randomized comparison (2). J Med Internet Res 2004 Sep 24;6(3):e30 [FREE Full text] [doi: 10.2196/jmir.6.3.e30] [Medline: 15471756]
7. Rao K, Kaminska O, McCutcheon AL. Recruiting probability samples for a multi-mode research panel with Internet and mail components. Public Opinion Quarterly 2010 Feb 16;74(1):68-84. [doi: 10.1093/poq/nfp091]
8. Roster CA, Rogers RD, Albaurn G, Klein D. A comparison of response characteristics from web and telephone surveys. Int J Market Res 2004;46:359-373.
9. Whitehead L. Methodological issues in Internet-mediated research: a randomized comparison of internet versus mailed questionnaires. J Med Internet Res 2011;13(4):e109 [FREE Full text] [doi: 10.2196/jmir.1593] [Medline: 22155721]
10. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing theaccuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. Public Opinion Quarterly 2011 Oct 05;75(4):709-747. [doi: 10.1093/poq/nfr020]
11. ESOMAR. 26 Questions To Help Research Buyers Of Online Samples. Amsterdam: ESOMAR; 2008. Knowledge and Standards URL: http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_26-Questions-To-Help-Research-Buyers-Of-Online-Samples.pdf [accessed 2013-04-09] [WebCite Cache ID 6KuZxVobQ]
12. American Association For Public Opinion Research. 2013. URL: http://www.aapor.org/Home.htm [accessed 2013-05-16] [WebCite Cache ID 6JsInebXC]
13. Quirk's Marketing Research Review. Saint Paul, MN: Quirk Enterprises; 2011 Nov. 2011 Research Panels Directory URL: http://www.quirks.com/pdf/201111_quirks.pdf [accessed 2011-11-01] [WebCite Cache ID 6KucHafFY]
14. Council of American Survey Research Organizations. 2013. URL: http://www.casro.org/ [accessed 2013-05-16] [WebCite Cache ID 6JsJ0CRsX]
15. Craig BM, Reeve BB. Methods Report on the PROMIS Valuation Study: Year 1. 2012 Sep 27. URL: http://labpages.moffitt.org/craigb/Publications/Report120928.pdf [accessed 2013-09-24] [WebCite Cache ID 6JsJHcdnH]
16. United States Census Bureau. United States Census 2010. URL: http://www.census.gov/2010census/pdf/2010_Questionnaire_Info.pdf [accessed 2013-11-05] [WebCite Cache ID 6KudgfQrS]

17. Zip Code Tabulation Areas (ZCTAs). 2012. United States Census Bureau URL: http://www.census.gov/geo/reference/zctas.html [accessed 2013-03-13] [WebCite Cache ID 6Kucl3NLD]
18. Selected Social Characteristics in the United States: 2010 American Community Survey 1-Year Estimates. 2010. United States Census Bureau URL: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_1YR_DP03&prodType=table [accessed 2013-05-16]
19. Selected Social Characteristics in the United States: 2010 American Community Survey 1-Year Estimates. 2010. United States Census Bureau URL: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_1YR_DP02&prodType=table [accessed 2013-05-16]
20. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002;4:e13. [doi: 10.2196/jmir.4.2.e13]
21. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the patient-reported outcomes measurement information system Internet panel. J Clin Epidemiol 2010 Nov;63(11):1169-1178 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.11.021] [Medline: 20688473]
22. Brick JM. The future of survey sampling. Public Opinion Quarterly 2011 Dec 15;75(5):872-888. [doi: 10.1093/poq/nfr045]
23. Sparrow N. Developing reliable online polls. Int J Market Res 2006;48:659-680.
24. Sparrow N. Quality issues in online research. J Advert Res 2007;47:179-182.
25. Wood FB, Benson D, LaCroix EM, Siegel ER, Fariss S. Use of Internet audience measurement data to gauge market share for online health information services. J Med Internet Res 2005 Jul 1;7(3):e31 [FREE Full text] [doi: 10.2196/jmir.7.3.e31] [Medline: 15998622]
26. PROMIS Health Organization, PROMIS Cooperative Group. PROMIS-29 Profile v1.0. Investigator Format 2008-2012. URL: https://www.assessmentcenter.net/ac1//files/pdf/44b7636201a34267a9213db7f69f2c6d.pdf [accessed 2012-08-28] [WebCite Cache ID 6KubEo6qZ]
27. Sugiura H, Ohkusa Y, Akahane M, Sano T, Okabe N, Imamura T. Development of a web-based survey for monitoring daily health and its application in an epidemiological survey. J Med Internet Res 2011;13(3):e66 [FREE Full text] [doi: 10.2196/jmir.1872] [Medline: 21946004]
28. Fox S, Rainie L. The online health care revolution: How the Web helps Americans take better care of themselves. Washington, DC: Pew Internet & American Life Project; 2000 Nov. URL: http://www.pewinternet.org/ [accessed 2013-11-05] [WebCite Cache ID 6KueySfNc]
29. Callegaro M, DiSogra C. Computing response metrics for online panels. Public Opinion Quarterly 2009 Jan 29;72(5):1008-1032. [doi: 10.1093/poq/nfn065]

Abbreviations

AAPOR: American Association for Public Opinion Research
CASRO: Council of American Survey Research Organizations
ESOMAR: European Society for Opinion and Market Research
IP: Internet Protocol
NIH: National Institutes of Health
PROMIS: Patient-Reported Outcomes Measurement Information System
PV: panel vendor
RFQ: request for quotes

Edited by G Eysenbach; submitted 27.08.13; peer-reviewed by L Whitehead, Y Zhang, S Shiffman; comments to author 19.09.13; revised version received 23.10.13; accepted 26.10.13; published 29.11.13.

Please cite as:
Craig BM, Hays RD, Pickard AS, Cella D, Revicki DA, Reeve BB
Comparison of US Panel Vendors for Online Surveys
J Med Internet Res 2013;15(11):e260
URL: http://www.jmir.org/2013/11/e260/
doi:10.2196/jmir.2903
PMID:24292159

Original Paper

Factors Associated With Intention to Use Internet-Based Testing for Sexually Transmitted Infections Among Men Who Have Sex With Men

Mark Gilbert[1,2], MD, MHSc, FRCPC; Travis Salway Hottes[1], MSc; Thomas Kerr[3,4], MA, PhD; Darlene Taylor[1,2], RN, MSc; Christopher K Fairley[5], MBBS, PhD; Richard Lester[1,3], MD, FRCPC; Tom Wong[6], MD, MPH, FRCPC; Terry Trussler[7], EdD; Rick Marchand[7], PhD; Jean Shoveller[2], PhD; Gina Ogilvie[1,3], MD, MSc, FCFP, DrPH

[1]Clinical Prevention Services, BC Centre for Disease Control, Vancouver, BC, Canada

[2]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[3]Department of Medicine, University of British Columbia, Vancouver, BC, Canada

[4]BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada

[5]Melbourne School of Population and Global Health, University of Melbourne, Victoria, Australia

[6]Public Health Agency of Canada, Ottawa, ON, Canada

[7]Community Based Research Centre for Gay Men's Health, Vancouver, BC, Canada

Corresponding Author:
Mark Gilbert, MD, MHSc, FRCPC
Clinical Prevention Services
BC Centre for Disease Control
655 West 12th Avenue
Vancouver, BC, V5Z4R4
Canada
Phone: 1 6047075615
Fax: 1 6047075604
Email: mark.gilbert@bccdc.ca

Abstract

Background: Internet-based testing programs are being increasingly used to reduce testing barriers for individuals at higher risk of infection, yet the population impact and potential for exacerbation of existing health inequities of these programs are not well understood.

Objective: We used a large online sample of men who have sex with men (MSM) in Canada to measure acceptability of Internet-based testing and perceived advantages and disadvantages of this testing approach.

Methods: We asked participants of the 2011/2012 Sex Now Survey (a serial online survey of gay and bisexual men in Canada) whether they intended to use Internet-based testing and their perceived benefits and disadvantages of use. We examined whether intention to use was associated with explanatory variables spanning (A) sociodemographics, (B) Internet and technology usage, (C) sexually transmitted infections (STI)/ human immunodeficiency virus (HIV) and risk, and (D) health care access and testing, using multivariable logistic regression (variable selection using Bayesian information criterion).

Results: Overall, intention to use was high (5678/7938, 71.53%) among participants with little variation by participant characteristics. In our final model, we retained the variables related to (B) Internet and technology usage: use of Internet to cruise for sex partners (adjusted odds ratio [AOR] 1.46, 95% CI 1.25-1.70), use of Internet to search for sexual health information (AOR 1.36, 95% CI 1.23-1.51), and mobile phone usage (AOR 1.19, 95% 1.13-1.24). We also retained the variables for (D) health care access and testing: not “out” to primary care provider (AOR 1.24, 95% CI 1.10-1.41), delayed/avoided testing due to privacy concerns (AOR 1.77, 95% CI 1.49-2.11), and delayed/avoided testing due to access issues (AOR 1.65, 95% CI 1.40-1.95). Finally, we retained the variable being HIV positive (AOR 0.56, 95% CI 0.46-0.68) or HIV status unknown (AOR 0.89, 95% CI 0.77-1.01), age <30 years (AOR 1.41, 95% CI 1.22-1.62), and identifying as bisexual (AOR 1.18, 95% CI 1.04-1.34) or straight/other (AOR 0.67, 95% CI 0.50-0.90). The greatest perceived benefits of Internet-based testing were privacy (2249/8388, 26.81%), general convenience (1701/8388, 20.28%), and being able to test at any time (1048/8388, 12.49%). The greatest perceived drawbacks

were the inability to see a doctor or nurse (1507/8388, 17.97%), wanting to talk to someone about results (1430/8388, 17.97%), not wanting online results (1084/8388, 12.92%), and low trust (973/8388, 11.60%).

Conclusions: The high and wide-ranging intention to use that we observed suggests Internet-based testing has the potential to reach into all subgroups of MSM and may be particularly appealing to those facing current barriers to accessing STI/HIV testing and who are more comfortable with technology. These findings will be used to inform the promotion and further evaluation of an Internet-based testing program currently under development in British Columbia, Canada.

(J Med Internet Res 2013;15(11):e254) doi:10.2196/jmir.2888

KEYWORDS

homosexuality; male; Internet; testing; human immunodeficiency virus; sexually transmitted infection; health equity; patient acceptance of health care

Introduction

Public health agencies are increasingly turning to the Internet for the delivery of sexual health services in order to reduce barriers to access and reach people at heightened risk of sexually transmitted infections (STI) [1-3]. Internet-based testing reflects a recent and fundamental shift in delivery of testing services, from provider-mediated to patient-centered testing [4-6], with the aim of reducing barriers to accessing traditional testing services (such as needing to travel to a clinic or waiting for an appointment, or privacy and confidentiality concerns). Internet-based testing programs typically involve requesting a home self-collection kit or downloading a test requisition and presenting it to a specimen collection site, then receiving results online or by phone. The scope of public programs varies widely, from state or country-wide programs for chlamydia screening [7-9], to more local programs that offer one or more STI/ human immunodeficiency virus (HIV) tests and may be targeted to a specific population such as gay, bisexual, and other men who have sex with men (MSM) [10,11].

The evidence of the impact of Internet-based testing programs is beginning to accumulate, primarily for population-based chlamydia screening programs [8,12,13]. However, substantial knowledge gaps remain regarding the impact of these services at a population level, such as the reach and diffusion of programs within populations at higher risk for STI/HIV [14]. An important concern with the introduction of new health technologies is that uptake may be concentrated among individuals who already have good access to health services (often correlated with socioeconomic status) and not among individuals who need it most [15,16]. For example, if the uptake of Internet testing programs is highest among individuals who already have adequate access to testing, these programs may run the risk of reinforcing rather than reducing health inequities if not accessed by individuals currently facing barriers to testing. The alluring promise of, but widespread lack of delivery by, online technologies to expand health and health access to larger portions of the population, particularly the more marginalized, is a topic of concern [17-21]. This concept applies not just to, but within, marginalized populations such as MSM where both sexual risk, disease prevalence, and appropriate access to health care are unevenly distributed [22].

Internet-based interventions are widely recognized as a valuable tool for promoting sexual health among MSM [23]. MSM have a high burden of STIs and HIV in Canada, as well as in most industrialized countries. MSM comprise approximately 50% of all incident HIV and prevalent infections in Canada [24] with high rates of other STIs including ongoing syphilis outbreaks in several provinces and countries with comparable STI epidemiology [25]. MSM are also likely to turn to the Internet to look for sexual health information or support, and finding sex partners through sex-seeking websites is widespread, including by men at greater sexual risk of infection [23,26]. Studies of Internet-based testing programs have demonstrated that significant numbers of MSM use these services [16,18]. However, few studies assessing the reach or broader acceptability of Internet-based testing among MSM have been published [4]. Studies have examined the willingness of MSM to access anonymous home HIV testing as part of online research [27] and the acceptability of other online interventions such as partner notification [28-31]; overall, acceptability is high with small but significant differences across subgroups. Factors associated with acceptability or uptake of Internet-based sexual health services (in males, MSM or youth) include income [32], age [31,32], ethnicity [27,31], education [31], substance use [32,33], HIV status [27,29,33], prior STI [29,33], risk sex [27,31,33-35], perceived risk of HIV [34], and health seeking behavior [36]. Men living in rural areas may be more likely to find sex partners online and be willing to participate in online interventions [36-38]. There is little discussion of potential harms of online sexual health interventions in general or for MSM in particular. The main concerns relate to their inaccessibility by people with no or limited Internet access (the so-called “first-level” digital divide) or by people who have less facility with using the Internet (the “second-level” digital divide) [39]. Recently, Rosser et al emphasized the importance of considering how different age groups of MSM approach technology and use the Internet in the design of Internet-based sexual health services [40]; however, the impact of technology use on acceptability of online services is largely unknown.

The British Columbia (BC) Centre for Disease Control is developing a program, GetCheckedOnline, for Internet-based testing for STIs and HIV and is planning a targeted promotion of the service to MSM in the Vancouver area. The program has been developed through focus groups with MSM and youth who have indicated high acceptability of the service [41,42] and if successful, the intent is to expand the program on a broader geographic scale. The primary objective of this study was to assess the acceptability of Internet-based testing in a national sample of MSM (based on intention to use) and

associated characteristics. In so doing, we aimed to assess acceptability and potential reach of Internet-based testing among MSM with varying sociodemographic characteristics, with greater reported risk of infection for STI and HIV, and facing existing barriers to testing access. A secondary objective of the study was to describe the perceived advantages and disadvantages of Internet-based testing. By identifying factors associated with intention to use and perceptions of Internet-based testing prior to implementation of the program, we will then be better positioned to further refine the service, target its promotion to particular subgroups, and develop strategies to promote acceptance and adoption of the service [43].

Methods

Survey

Sex Now is a national online survey of gay and bisexual men, administered every 12-18 months in Canada [44]. Sex Now content is developed iteratively by a panel of gay men's health researchers, with the aim of responding to evolving needs of the community. Face validity of the questionnaire is ensured through focus groups, interviews, and pilot testing by local gay men. Standard survey domains include relationships, sexual styles, sexual behaviors, sexual health, anti-gay or discriminatory experiences (especially in workplace), substance use, sexual health knowledge, Internet experience, health care access, community participation, mental health issues, and sociodemographics. Questions are available in both French and English.

Participants from the 2011-12 cycle were recruited through dating/sex-seeking websites (6356/8388, 75.8%), gay/bisexual community-based organizations (833/8388, 10.0%), and word-of-mouth (729/8388, 8.7%). The survey is described as a survey of "sex between men" and relies on self-selection. Responses were collected from August 26, 2011, to February 21, 2012. To limit multiple entries, submissions were restricted to one response per Internet Protocol address, and data were rigorously examined to screen out multiple or suspect submissions.

A subset of questions relevant to the BC Internet-based STI/HIV testing model were added to the 2011-12 questionnaire for the purposes of the present study. The following question domains were added: barriers to clinic-based testing; acceptability of Internet-based testing; perceived benefits, risks, and barriers of Internet-testing; and potential factors influencing uptake of Internet-based testing including privacy concerns, use of health services, and use of technology. We used the concept of intention or willingness to use the technology, which is an established metric for measuring acceptance of technology and considered predictive of actual use in theoretical models of technology acceptance [43,45]. As we were also interested in informing strategies for acceptance and adoption, we considered diffusions of innovations theory [46,47] and use of online/mobile technologies to potentially be important predictors of intention to use Internet-based testing among MSM, based on evidence emerging from recent analyses of online sex-seeking behavior [40]. Due to a lack of established metrics to measure these latter concepts, we constructed questions that were validated through the process described above. For all variable definitions, see Multimedia Appendix 1.

Measures

The primary outcome of interest was intention to use Internet-based STI/HIV testing, measured through a 5-point Likert scale response to the following question: "Suppose you could get tested by printing out an order form from a website that you could take to a lab, then get your results online. How likely is it that you would use this service? [Very likely, likely, unlikely, very unlikely, would never use this service; with 'not applicable' option]". Participants were also asked to identify a single greatest perceived benefit and drawback of the aforementioned service from a pre-determined list. Response options were determined based on expert knowledge of local STI clinical services and literature describing other Internet-based STI testing models.

Analysis

The sample was restricted to participants residing in Canada (8388/8497, 98.72%), of which 94.64% (7938/8388) completed the question on intention to use Internet-based testing. All respondents were male. Thirty-eight explanatory variables were selected from the questionnaire a priori based on the literature review described above and were grouped into four broad groups of interest: (A) sociodemographics (14 variables), (B) Internet and technology usage (6 variables), (C) STI/HIV and risk (8 variables), and (D) health care access and testing (10 variables).

To achieve the primary objective of identifying characteristics correlated with intention to use Internet-based STI/HIV testing, we first explored the distribution of responses (5-point Likert scale) across all 38 explanatory variables. Logistic regression was then used to model associations between all explanatory variables and this outcome (dichotomized: very likely/likely versus unlikely/very unlikely/never; those who chose "not applicable" were excluded). A full multivariable model was fit with all 38 explanatory variables, and a final model was selected using Bayesian information criterion (BIC), which is comparable to Akaike information criterion (AIC) but imposes stricter penalties for inclusion of additional explanatory variables (ie, generates a more parsimonious model) [48]. Correlation between explanatory variables was examined, and the covariate considered most relevant to the research question was included in multivariable models for highly collinear sets.

Age groups (less than 30 years, 30 years of age and older) and sexual orientation (gay, bisexual) were identified a priori as subgroups of interest and hence were included in all multivariable models. We hypothesized that other explanatory variables would differ across these subgroups and explored statistical interactions between explanatory variables and age and sexual orientation. Each variable was entered into two bivariable models, including age and sexual orientation respectively. First-level multiplicative interaction terms were added, and interaction terms that were statistically significant at $P<.10$ were carried forward in analysis. Stepwise regression was used to select interaction terms for inclusion in the full

multivariable model, such that all remaining interaction terms were significant at $P<.15$ [49].

To achieve the secondary objective, perceived benefits and drawbacks of the service were summarized using descriptive statistics for the total sample, and among men who delayed or avoided STI or HIV testing in the past 12 months (STI testing only if HIV positive).

All analysis was completed using R version 2.15.2. BIC model selection was performed using the stepAIC function in MASS package version 7.3-22 [50].

Ethics Approval

The survey protocol was approved by the independent Research Ethics Board of the Community-Based Research Centre and also by the Behavioural Research Ethics Board at the University of British Columbia.

Results

Summary of Sample

Characteristics of the sample are shown in Multimedia Appendix 2. The average age was 43 years (range 13-84 years), and 64.50% (5410/8388) self-identified as gay and 32.42% (2719/8388) as bisexual. Further, 57.13% (4792/8388) had completed a college or university degree, and 71.46% (5994/8388) reported annual incomes ≥ Canadian $30K. The sample was predominantly urban (4897/8388, 58.38%), though a significant proportion resided in suburban (2214/8388, 26.39%) or rural/remote (1245/8388, 14.84%) settings. Respondents included residents of all ten Canadian provinces and all three territories, with the distribution generally representative of total regional populations with the exception of British Columbia (greater proportion) and Quebec (smaller proportion) [51]. Respondents represent 71.61% (1173/1638) of the forward sortation areas (first three characters of the postal code) of Canada [52]. Most men reported being “out” about their sexuality generally (5295/8388, 63.13%), but fewer were out at work (3881/8388, 46.27%), and fewer still spent most of their free time with other gay men (1867/8388, 22.26%).

Intention to Use Internet-Based STI/HIV Testing

Of the total sample, 71.53% (5678/7938) indicated that they were likely (2422/7938, 30.51%) or very likely (3256/7938, 41.02%) to use Internet-based STI/HIV testing. Across the full 5-point response scale, intention to use Internet-based testing was right-skewed towards very likely to use, with little variation across subgroups (data not shown). Dichotomized intention to use Internet-based testing was similarly high across nearly all covariates examined, generally ranging 67-77%, with few exceptions: Latino men (88/109, 80.7%), very early purchasers of new technology (270/338, 79.9%), those not at all satisfied with health care services (290/361, 80.3%), and those who delayed or avoided testing in the past 12 months for privacy concerns (944/1128, 83.69%), access issues (1021/1225, 83.35%), or distance from clinic (357/429, 83.2%) (see Multimedia Appendix 2). Intention to use Internet-based testing was lower than 67% in only two subgroups: HIV-positive men (263/476, 55.3%) and regular users of party drugs (105/159, 66.0%).

Crude and adjusted odds ratios (AOR) are shown in Multimedia Appendix 2. In the full model, the following explanatory variables retained statistical ($P<.05$) associations with greater intention to use Internet-based testing: group A (3/14 variables), age <30 years, eastern provinces, less “out” about sexuality at work; group B (4/6 variables), use of Internet to cruise for sex partners, use of Internet to search for sexual health information, greater mobile phone usage, and early uptake of new technology; group C (2/8 variables), unprotected anal intercourse with unknown/discordant HIV status partner, and HIV negative status; group D (4/10 variables), last medical appointment >6 months ago or never, poorer satisfaction with health care services available, delayed/avoided testing because of privacy concern, and delayed/avoided testing because of access issue. Four interaction terms with age and nine with sexual orientation were included in the full multivariable model, as shown in Multimedia Appendix 3.

The final model, as selected by BIC, is shown in Figure 1. Notably, of the nine variables positively associated with greater intention to use Internet-based testing in the final BIC model, three correspond to group B, Internet and technology usage: use Internet to cruise for sex partners (AOR 1.46, 95% CI 1.25-1.70), use Internet to search for sexual health information (AOR 1.36, 95% CI 1.23-1.51), and mobile phone usage (AOR 1.19, 95% CI 1.13-1.24). Three variables corresponded to group D, health care access: not “out” to primary health care provider (AOR 1.24, 95% CI 1.10-1.41), delayed/avoided testing due to privacy concerns (AOR 1.77, 95% CI 1.49-2.11), and delayed/avoided testing due to access issues (AOR 1.65, 95% CI 1.40- 1.95). No interaction terms were selected into the final BIC model.

Perceived Benefits and Drawbacks of Service

The most frequent perceived benefits of the BC Internet-based STI/HIV testing program were greater privacy (2249/8388, 26.81%), convenience in general (1701/8388, 20.28%) and specifically, ability to get tested whenever (1048/8388, 12.49%). The greatest perceived drawbacks were the inability to see a doctor or nurse (1507/8388, 17.97%), wanting to talk to someone about results (1430/8388, 17.05%), not wanting results online (1084/8388, 12.92%), and low trust in the service generally (973/8388, 11.60%). The particular benefits and drawbacks perceived by survey respondents showed very little variation in sensitivity analyses (see Tables 1 and 2).

Figure 1. Correlates of intention to use Internet-based sexually transmitted infection testing selected by Bayesian Information Criterion in a survey sample of Canadian gay and bisexual men (N=7938). [a]Referent group Sexual Orientation - Gay; [b]Mobile phone usage measured on 3-point continuous scale; [c]Referent group HIV status - negative.

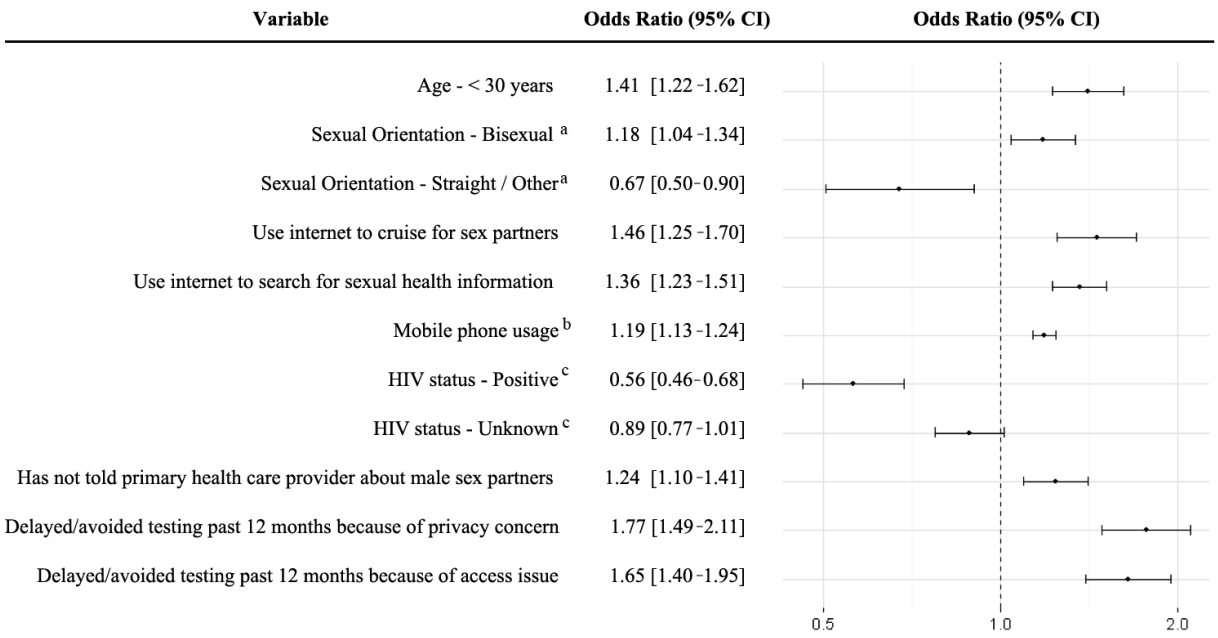


Table 1. Greatest perceived benefit to Internet-based STI and HIV testing among the survey sample of Canadian gay and bisexual men.

Benefit[a]	Total (N=8388)	Among those who intend to use service (n=5678)	Among those delaying/avoiding testing[b] (n=4947)
	n (%)	n (%)	n (%)
Greater privacy	2249 (26.81)	1834 (32.30)	1551 (31.35)
Convenient	1701 (20.28)	1340 (23.60)	920 (18.60)
Get tested whenever	1048 (12.49)	755 (13.30)	596 (12.05)
No nurse/doctor	823 (9.81)	669 (11.78)	553 (11.18)
Save time	618 (7.37)	484 (8.52)	323 (6.53)
No waiting for app't	448 (5.34)	301 (5.30)	249 (5.03)
No worry about running into someone you know	245 (2.92)	166 (2.92)	185 (3.74)
Other[c]	56 (0.67)	18 (0.32)	36 (0.73)
No particular benefit	1200 (14.31)	111 (1.95)	534 (10.79)

[a]Respondents were asked to choose one *greatest* benefit.

[b]Defined as those who reported no STI test in the last 12 months, OR any delay in testing in the last 12 months (for HIV-positive respondents); or no STI test AND no HIV test in the last 12 months, OR any delay in testing in the last 12 months (for HIV-negative/unknown status respondents).

[c]Not specified.

Table 2. Greatest perceived drawback to Internet-based STI and HIV testing among the survey sample of Canadian gay and bisexual men.

Drawback[a]	Total (N=8388)	Among those who do not intend to use service (n=2260)	Among those delaying/avoiding testing[b] (n=4947)
	n (%)	n (%)	n (%)
Wouldn't see doctor/nurse	1507 (17.97)	451 (19.96)	735 (14.86)
Want to talk to someone about results	1430 (17.05)	384 (16.99)	799 (16.15)
Don't want results online	1084 (12.92)	407 (18.01)	652 (13.18)
Low trust in service	973 (11.60)	340 (15.04)	590 (11.93)
No printer	169 (2.01)	43 (1.90)	118 (2.39)
Other[c]	159 (1.90)	67 (2.96)	100 (2.02)
No particular drawback	3066 (36.55)	568 (25.13)	1953 (39.48)

[a]Respondents were asked to choose one greatest drawback.

[b]Defined as those who reported no STI test in the last 12 months, OR any delay in testing in the last 12 months (for HIV-positive respondents); or no STI test AND no HIV test in the last 12 months, OR any delay in testing in the last 12 months (for HIV-negative/unknown status respondents).

[c]Not specified.

Discussion

Principal Findings

Overall, we found that intention to use Internet-based testing for HIV and STI is high (5678/7938, 71.53%) and wide-ranging within this large online sample of gay and bisexual men in Canada, with little variation by participant characteristics. Our study suggests that Internet-based testing has the potential to reach nearly all subgroups of gay and bisexual men, including men at risk of STIs and HIV (as intention was 74.46% [1770/2377] among men reporting unprotected anal intercourse with an unknown or serodiscordant partner, and 73.70% [283/384] among men reporting an STI or hepatitis C diagnosis, in the past year) and facing current barriers to accessing testing (72.90% [2930/4019] among men not tested for HIV in the past year, and 83.2% [357/429] to 83.69% [944/1128] among men reporting delaying or avoiding testing in the past year because of privacy concerns, access issues, or distance to testing services).

On multivariable analysis, men who reported current barriers to accessing appropriate health care and STI/HIV testing were more likely to intend to use Internet-based testing (3/10 variables retained in the final model). In our sample, a large proportion of participants (4217/8388, 50.27%) reported not having disclosed to their primary health care provider that they were sexually active with men. Not being "out" to a primary care provider has been associated with undiagnosed HIV infection and less frequent rates of HIV testing [53,54]; we found greater intention to use Internet-based testing in this group, which may help to bridge this gap. Intent to use Internet-based testing was also more likely among the small but important proportion of participants in our sample who reported delaying or avoiding testing in the past 12 months because of privacy concerns (1140/8388, 13.59%) or because of access issues such as not knowing where to get a test or needing to wait for an appointment (1243/8388, 14.82%).

Men who identified as HIV positive (667/8388, 7.95% of our sample) demonstrated less intention to use Internet-based testing, which we postulate is related to adequate STI testing access through routine care, which confirms previous qualitative findings from our group [41], or a greater appeal of Internet-based testing for HIV testing among HIV negative men. However, men whose HIV status was unknown (1966/8388, 23.44%) also demonstrated less intention to use Internet-based testing after adjustment for other characteristics, which is concerning as this is a population of MSM who are not currently engaged in HIV testing. Unlike other studies, we did not observe an association with behavioral measures such as sexual risk or substance use, after adjusting for covariates [27,29,31,33,34,36]. Encouragingly, we did not find that sociodemographic variables such as ethnicity, income, education, and residence were influential on intention to use Internet-based testing, which differs from previously published studies in this field [27,31,32,37,38,40]. Within this large online sample of gay and bisexual men, Internet-based testing programs such as GetCheckedOnline may not exacerbate existing health inequities along these sociodemographic lines.

One reason for these differences may be that the characteristics of MSM that we considered allowed for better explanation of the variability in intention to use Internet-based testing. In addition to variables related to facing current barriers to access appropriate health care, we found variables related to Internet and technology use were most influential in our final model (3/6 retained): use of the Internet to cruise for sex partners or to search for health information, and mobile phone usage. Internet sex-seeking has long been a primary motivation for developing Internet-based sexual health interventions [55]; that a majority of MSM look for sex partners online underscores the importance of the Internet as a health service delivery venue (7430/8388, 88.58% in our online sample and typically over 50% of MSM in venue-based samples [56]). Given that use of the Internet to search for health information and use of mobile phones for more than phone calls were also associated with intention to use, our findings suggest that ease and facility with online technologies may be an important influence on uptake of Internet-based testing (ie, the "second-level" digital divide) [39]. This appears independent of the influence of age, which

may be related to greater acceptability of online services by persons born in the era of digital technology compared to older persons [40]. While we found a consistent and significant gradient between diffusion of innovations and intention to use in our full model, this was not retained in the final model.

As hypothesized, we found that age and sexual orientation were influential on intention to use Internet-based testing and retained in the final model (greater intention among men <30 years; compared to gay men, bisexual men were more likely and straight/other men were less likely to intend to use Internet-based testing). The association with intention to use Internet-based testing varied across subgroups of men by age and sexual orientation. Notwithstanding the primary conclusion from our results—that intention to use Internet testing is high across varying subgroups of men—the statistical interactions described here suggest that where more nuanced decisions regarding service promotion and delivery are required, program planners must attend to the potentially different (sometimes opposite) intentions and needs of subgroups of gay and bisexual men, particularly those related to sexual orientation and identity.

The perceived benefits and drawbacks identified by Canadian gay and bisexual men in this study did not significantly differ in sensitivity analyses, and overall, men perceived more benefits than drawbacks: the percentages identifying no particular benefit or drawback were 14.31% (1200/8388) and 36.55% (3066/8388) respectively. The most common perceived benefits were greater privacy, convenience in general, and specifically, being able to test at any time; the most common perceived drawbacks were not seeing a doctor or nurse, not being able to talk to someone about the results, not wanting results online, or low trust in the service. These were also the most common perceived benefits and drawbacks from earlier focus groups focused on the GetCheckedOnline program model [41].

Limitations

Our study had a number of limitations. These findings are not generalizable to all gay, bisexual, and other men who have sex with men in Canada, due to the online nature of this convenience sample, recruited primarily from sex-seeking websites. Furthermore, as participants were recruited for an online survey (and likely have some degree of ease with Internet use or predisposition to online services), we may have overestimated intention to use Internet-based testing, including potentially among some subgroups of interest (eg, MSM with poor Internet access, men of minority ethnicities where English may be a second language). However, the relative differences between subgroups within our sample may be accurate for all men who have sex with men in Canada. As an open-access survey, it is possible that multiple submissions from the same individual may have occurred; however, we believe this to be unlikely due to the lack of incentive and time required to complete the survey. As a self-completed survey, it is possible that recall bias may have affected responses to questions spanning the prior 12 months. Finally, while measuring intention to use a service is considered predictive of actual use, it will be important to measure actual diffusion and uptake of GetCheckedOnline among MSM once implemented.

Conclusions

In summary, we observed high intention to use Internet-based testing among MSM in Canada given a brief description of the GetCheckedOnline program model. The high intention to use observed in our study appears most related to the perceived benefits of greater privacy, convenience, and being able to test at any time. Importantly, those who reported facing current barriers to appropriate health care and testing less frequently had higher intended use of Internet testing; intention to use was also more likely among individuals reporting greater use of technology. Our findings differ from that of other studies that have assessed the characteristics associated with acceptability of online sexual health services among youth or MSM (ie, little observed impact of sociodemographic or sexual risk characteristics). Our findings affirm that barriers to health care and technology use are important variables to consider in the design, implementation, and evaluation of online sexual health services. As we go forward, considering the requirements of less technology-savvy MSM and how the service could be more appealing to MSM of unknown HIV status will also be important.

Acknowledgments

The authors would like to acknowledge the contribution of Olivier Ferlatte and members of the Investigaytors Program at the Community-Based Research Centre for survey development, translation, and analytic support. We would also like to acknowledge the contributions of the Clinical Integration Working Group and Community Consultation Working Group for the Online Sexual Health Services program at the BC Centre for Disease Control for their perspectives shared during the development of GetCheckedOnline, which helped inform this study. Finally the authors acknowledge Michael Otterstatter and Robert Balshaw for their advice on the statistical analysis methods used in this study. This work was supported by the Canadian Institutes of Health Research, Grant No. PHE-114129.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Definitions of all explanatory variables.

[PDF File (Adobe PDF File), 126KB - jmir_v15i11e254_app1.pdf]

Multimedia Appendix 2

Characteristics of survey sample of Canadian gay and bisexual men and intention to use Internet-based testing by key variables, across the four domain groups (N=8388).

[PDF File (Adobe PDF File), 138KB - jmir_v15i11e254_app2.pdf]

Multimedia Appendix 3

Interactions between explanatory variables with age and sexual orientation in the full model.

[PDF File (Adobe PDF File), 106KB - jmir_v15i11e254_app3.pdf]

References

1. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [Medline: 18366952]
2. McFarlane M, Bull SS. Use of the Internet in STD/HIV prevention. In: Aral SO, Douglas JM. editors. Behavioral Interventions for Prevention and Control of Sexually Transmitted Diseases. New York: Springer Science+Business Media; 2007:214-231.
3. Noar SM, Willoughby JF. eHealth interventions for HIV prevention. AIDS Care 2012;24(8):945-952. [doi: 10.1080/09540121.2012.668167] [Medline: 22519523]
4. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
5. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008 Oct;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
6. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the Internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
7. van den Broek IV, Hoebe CJ, van Bergen JE, Brouwers EE, de Feijter EM, Fennema JS, et al. Evaluation design of a systematic, selective, Internet-based, Chlamydia screening implementation in the Netherlands, 2008-2010: implications of first results for the analysis. BMC Infect Dis 2010;10:89 [FREE Full text] [doi: 10.1186/1471-2334-10-89] [Medline: 20374635]
8. Gaydos CA, Barnes M, Aumakhan B, Quinn N, Agreda P, Whittle P, et al. Can e-technology through the Internet be used as a new tool to address the Chlamydia trachomatis epidemic by home sampling and vaginal swabs? Sex Transm Dis 2009 Sep;36(9):577-580 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181a7482f] [Medline: 19543145]
9. Novak DP, Karlsson RB. Simplifying chlamydia testing: an innovative Chlamydia trachomatis testing approach using the Internet and a home sampling strategy: population based study. Sex Transm Infect 2006 Apr;82(2):142-7; discussion 152 [FREE Full text] [doi: 10.1136/sti.2005.016832] [Medline: 16581742]
10. Koekenbier RH, Davidovich U, van Leent EJ, Thiesbrummel HF, Fennema HS. Online-mediated syphilis testing: feasibility, efficacy, and usage. Sex Transm Dis 2008 Aug;35(8):764-769. [doi: 10.1097/OLQ.0b013e31816fcb0a] [Medline: 18496472]
11. Woodruff A, Levine DK, Klausner J. Expanded STDtest.org provides San Franciscans with free testing and secure online results for five STDs. Presented at: National STD Prevention Conference; March 10-13, 2008; Chicago, Illinois URL: https://cdc.confex.com/cdc/std2008/techprogram/P15291.HTM
12. Kwan KS, Jachimowicz EA, Bastian L, Marshall L, Mak DB. Online chlamydia testing: an innovative approach that appeals to young people. Med J Aust 2012 Sep 3;197(5):287-290. [Medline: 22938127]
13. van Bergen JE, Fennema JS, van den Broek IV, Brouwers EE, de Feijter EM, Hoebe CJ, et al. Rationale, design, and results of the first screening round of a comprehensive, register-based, Chlamydia screening implementation programme in the Netherlands. BMC Infect Dis 2010;10:293 [FREE Full text] [doi: 10.1186/1471-2334-10-293] [Medline: 20925966]
14. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
15. Farrell, Janine. BC Centre for Disease Control. 2013. Health Equity and GetChecked: How can we make an online tetsing service that works for everyone? URL: http://bclovebytes.files.wordpress.com/2012/08/heia-report-2012-final1.pdf [accessed 2013-08-13] [WebCite Cache ID 6IqeXRvVo]
16. Frohlich KL, Potvin L. Transcending the known in public health practice: the inequality paradox: the population approach and vulnerable populations. Am J Public Health 2008 Feb;98(2):216-221. [doi: 10.2105/AJPH.2007.114777] [Medline: 18172133]
17. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002;26(6):494-503. [Medline: 12437024]
18. Baur C. An analysis of factors underlying e-health disparities. Camb Q Healthc Ethics 2008;17(4):417-428. [doi: 10.1017/S0963180108080547] [Medline: 18724881]

19. Chang BL, Bakken S, Brown SS, Houston TK, Kreps GL, Kukafka R, et al. Bridging the digital divide: reaching vulnerable populations. J Am Med Inform Assoc 2004;11(6):448-457 [FREE Full text] [doi: 10.1197/jamia.M1535] [Medline: 15299002]
20. Gibbons MC, Casale CR. Reducing disparities in health care quality: the role of health IT in underresourced settings. Med Care Res Rev 2010 Oct;67(5 Suppl):155S-162S. [doi: 10.1177/1077558710376202] [Medline: 20829238]
21. Kreps GL. Disseminating relevant health information to underserved audiences: implications of the Digital Divide Pilot Projects. J Med Libr Assoc 2005 Oct;93(4 Suppl):S68-S73 [FREE Full text] [Medline: 16239960]
22. Wolitski RJ, Fenton KA. Sexual health, HIV, and sexually transmitted infections among gay, bisexual, and other men who have sex with men in the United States. AIDS Behav 2011 Apr;15 Suppl 1:S9-17. [doi: 10.1007/s10461-011-9901-6] [Medline: 21331797]
23. Rosenberger JG, Reece M, Novak DS, Mayer KH. The Internet as a valuable tool for promoting a new framework for sexual health among gay men and other men who have sex with men. AIDS Behav 2011 Apr;15 Suppl 1:S88-S90. [doi: 10.1007/s10461-011-9897-y] [Medline: 21331800]
24. Summary: Estimates of HIV Prevalence and Incidence in Canada, 2011. Ottawa, Ontario: Centre for Communicable Diseases and Infection Control, Public Health Agency of Canada; 2012. URL: http://www.phac-aspc.gc.ca/aids-sida/publication/survreport/assets/pdf/estimat2011-eng.pdf [accessed 2013-10-30] [WebCite Cache ID 6KlOfVQ9B]
25. Report on Sexually Transmitted Infections in Canada: 2008.: Centre for Communicable Diseases and Infection Control, Public Health Agency of Canada; 2013. URL: http://www.phac-aspc.gc.ca/std-mts/report/sti-its2008/PDF/10-047-STI_report_eng-r1.pdf [WebCite Cache ID 6IqexMCEx]
26. Ogilvie GS, Taylor DL, Trussler T, Marchand R, Gilbert M, Moniruzzaman A, et al. Seeking sexual partners on the internet: a marker for risky sexual behaviour in men who have sex with men. Can J Public Health 2008;99(3):185-188. [Medline: 18615938]
27. Sharma A, Sullivan PS, Khosropour CM. Willingness to take a free home HIV test and associated factors among Internet-using men who have sex with men. J Int Assoc Physicians AIDS Care (Chic) 2011;10(6):357-364 [FREE Full text] [doi: 10.1177/1545109711404946] [Medline: 21527425]
28. Kubicek K, Carpineto J, McDavitt B, Weiss G, Kipke MD. Use and perceptions of the internet for sexual information and partners: a study of young men who have sex with men. Arch Sex Behav 2011 Aug;40(4):803-816 [FREE Full text] [doi: 10.1007/s10508-010-9666-4] [Medline: 20809373]
29. Mimiaga MJ, Tetu AM, Gortmaker S, Koenen KC, Fair AD, Novak DS, et al. HIV and STD status among MSM and attitudes about Internet partner notification for STD exposure. Sex Transm Dis 2008 Feb;35(2):111-116. [doi: 10.1097/OLQ.0b013e3181573d84] [Medline: 18007274]
30. Mimiaga MJ, Fair AD, Tetu AM, Novak DS, Vanderwarker R, Bertrand T, et al. Acceptability of an Internet-based partner notification system for sexually transmitted infection exposure among men who have sex with men. Am J Public Health 2008 Jun;98(6):1009-1011. [doi: 10.2105/AJPH.2006.098467] [Medline: 17901442]
31. Hooper S, Rosser BR, Horvath KJ, Oakes JM, Danilenko G, Men's INTernet Sex II (MINTS-II) Team. An online needs assessment of a virtual community: what men who use the internet to seek sex with men want in Internet-based HIV prevention. AIDS Behav 2008 Nov;12(6):867-875 [FREE Full text] [doi: 10.1007/s10461-008-9373-5] [Medline: 18401701]
32. Morris SR, Little SJ, Cunningham T, Garfein RS, Richman DD, Smith DM. Evaluation of an HIV nucleic acid testing program with automated Internet and voicemail systems to deliver results. Ann Intern Med 2010 Jun 15;152(12):778-785 [FREE Full text] [doi: 10.7326/0003-4819-152-12-201006150-00005] [Medline: 20547906]
33. Rhodes SD, Hergenrather KC, Yee LJ, Ramsey B. Comparing MSM in the southeastern United States who participated in an HIV prevention chat room-based outreach intervention and those who did not: how different are the baseline HIV-risk profiles? Health Educ Res 2008 Feb;23(1):180-190 [FREE Full text] [doi: 10.1093/her/cym015] [Medline: 17412716]
34. Mackellar DA, Hou SI, Whalen CC, Samuelsen K, Sanchez T, Smith A, WHBS Study Group. Reasons for not HIV testing, testing intentions, and potential use of an over-the-counter rapid HIV test in an internet sample of men who have sex with men who have never tested for HIV. Sex Transm Dis 2011 May;38(5):419-428. [doi: 10.1097/OLQ.0b013e31820369dd] [Medline: 21183863]
35. Samal L, Saha S, Chander G, Korthuis PT, Sharma RK, Sharp V, et al. Internet health information seeking behavior and antiretroviral adherence in persons living with HIV/AIDS. AIDS Patient Care STDS 2011 Jul;25(7):445-449 [FREE Full text] [doi: 10.1089/apc.2011.0027] [Medline: 21682586]
36. Schnarrs PW, Rosenberger JG, Satinsky S, Brinegar E, Stowers J, Dodge B, et al. Sexual compulsivity, the Internet, and sexual behaviors among men in a rural area of the United States. AIDS Patient Care STDS 2010 Sep;24(9):563-569. [doi: 10.1089/apc.2010.0117] [Medline: 20731609]
37. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. AIDS Educ Prev 2011 Apr;23(2):118-127. [doi: 10.1521/aeap.2011.23.2.118] [Medline: 21517661]
38. Bowen AM, Horvath K, Williams ML. A randomized control trial of Internet-delivered HIV prevention targeting rural MSM. Health Educ Res 2007 Feb;22(1):120-127 [FREE Full text] [doi: 10.1093/her/cyl057] [Medline: 16849391]
39. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002 2002 Apr 01:7.

40. Rosser BR, Wilkerson JM, Smolenski DJ, Oakes JM, Konstan J, Horvath KJ, et al. The future of Internet-based HIV prevention: a report on key findings from the Men's INTernet (MINTS-I, II) Sex Studies. AIDS Behav 2011 Apr;15 Suppl 1:S91-100 [FREE Full text] [doi: 10.1007/s10461-011-9910-5] [Medline: 21360127]
41. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
42. Shoveller J, Knight R, Davis W, Gilbert M, Ogilvie G. Online sexual health services: examining youth's perspectives. Can J Public Health 2012;103(1):14-18. [Medline: 22338322]
43. Centers for Disease ControlPrevention (CDC). Acute HIV infection - New York City, 2008. MMWR Morb Mortal Wkly Rep 2009 Nov 27;58(46):1296-1299 [FREE Full text] [Medline: 19940835]
44. Sex Now Survey / Sondage Sexe au present. 2013. URL: http://www.cbrc.net/sexnow [accessed 2013-09-26] [WebCite Cache ID 6JueixDIv]
45. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
46. Berwick DM. Disseminating innovations in health care. JAMA 2003 Apr 16;289(15):1969-1975. [doi: 10.1001/jama.289.15.1969] [Medline: 12697800]
47. Haider M, Kreps GL. Forty years of diffusion of innovations: utility and value in public health. J Health Commun 2004;9 Suppl 1:3-11. [doi: 10.1080/10810730490271430] [Medline: 14960400]
48. Burnham KP. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods & Research 2004 Nov 01;33(2):261-304. [doi: 10.1177/0049124104268644]
49. Hosmer D, Lemeshow S. Applied Logistic Regression. 2 ed. New York: Wiley & Sons, Inc; 2000.
50. Ripley B, Hornik K, Gebhardt A, Firth D. Modern applied statistics with S. In: Modern applied statistics with S. New York: Springer; 2002.
51. Statistics Canada. Population by sex and age group, by province and territory. 2013. URL: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo31a-eng.htm [accessed 2013-08-13] [WebCite Cache ID 6IqfKHs1J]
52. Postal Code Conversion File. Ottawa, ON: Statistics Canada; 2008. URL: http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=92F0153X&lang=eng
53. Millett GA, Ding H, Marks G, Jeffries WL, Bingham T, Lauby J, et al. Mistaken assumptions and missed opportunities: correlates of undiagnosed HIV infection among black and Latino men who have sex with men. J Acquir Immune Defic Syndr 2011 Sep 1;58(1):64-71. [doi: 10.1097/QAI.0b013e31822542ad] [Medline: 21654500]
54. Wall KM, Khosropour CM, Sullivan PS. Offering of HIV screening to men who have sex with men by their health care providers and associated factors. J Int Assoc Physicians AIDS Care (Chic) 2010;9(5):284-288. [doi: 10.1177/1545109710379051] [Medline: 20841438]
55. Klausner JD, Wolf W, Fischer-Ponce L, Zolt I, Katz MH. Tracing a syphilis outbreak through cyberspace. JAMA 2000 Jul 26;284(4):447-449. [Medline: 10904507]
56. Trussler T, Banks P, Marchand R, Robert W, Gustafson R, Hogg R. ManCount Sizes-up the Gaps: a sexual health survey of gay men in Vancouver. Vancouver: Vancouver Coastal Health; 2010. URL: http://www.mancount.ca/files/ManCount_Report2010.pdf [WebCite Cache ID 6KxKz5POZ]

Abbreviations

AIC: Akaike information criterion
AOR: adjusted odds ratio
BIC: Bayesian information criterion
HIV: human immunodeficiency virus
MSM: men who have sex with men
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 20.08.13; peer-reviewed by J Miranda, C Fairley; comments to author 19.09.13; revised version received 26.09.13; accepted 02.10.13; published 14.11.13.

Please cite as:
Gilbert M, Hottes TS, Kerr T, Taylor D, Fairley CK, Lester R, Wong T, Trussler T, Marchand R, Shoveller J, Ogilvie G
Factors Associated With Intention to Use Internet-Based Testing for Sexually Transmitted Infections Among Men Who Have Sex With Men
J Med Internet Res 2013;15(11):e254
URL: http://www.jmir.org/2013/11/e254/
doi:10.2196/jmir.2888
PMID:24240644

Original Paper

Using Multiple Imputations to Accommodate Time-Outs in Online Interventions

Susan M Shortreed[1,2], PhD; Andy Bogart[1], MS; Jennifer B McClure[3], PhD

[1]Group Heatlh Research Institute, Biostatistics Unit, Seattle, WA, United States

[2]University of Washington, Department of Biostatistics, Seattle, WA, United States

[3]Group Health Research Intitute, Seattle, WA, United States

Corresponding Author:
Susan M Shortreed, PhD
Group Heatlh Research Institute
Biostatistics Unit
1730 Minor Avenue, Suite 1600
Seattle, WA, 98101
United States
Phone: 1 206 287 2088
Fax: 1 206 287 2871
Email: shortreed.s@ghc.org

Abstract

Background: Accurately estimating the period of time that individuals are exposed to online intervention content is important for understanding program engagement. This can be calculated from time-stamped data reflecting navigation to and from individual webpages. Prolonged periods of inactivity are commonly handled with a time-out feature and assigned a prespecified exposure duration. Unfortunately, this practice can lead to biased results describing program exposure.

Objective: The aim of the study was to describe how multiple imputations can be used to better account for the time spent viewing webpages that result in a prolonged period of inactivity or a time-out.

Methods: To illustrate this method, we present data on time-outs collected from the Q^2 randomized smoking cessation trial. For this analysis, we evaluate the effects on intervention exposure of receiving content written in a prescriptive versus motivational tone. Using multiple imputations, we created five complete datasets in which the time spent viewing webpages that resulted in a time-out were replaced with values estimated with imputation models. We calculated standard errors using Rubin's formulas to account for the variability due to the imputations. We also illustrate how current methods of accounting for time-outs (excluding timed-out page views or assigning an arbitrary viewing time) can influence conclusions about participant engagement.

Results: A total of 63.00% (1175/1865) of participants accessed the online intervention in the Q^2 trial. Of the 6592 unique page views, 683 (10.36%, 683/6592) resulted in a time-out. The median time spent viewing webpages that did not result in a time-out was 1.07 minutes. Assuming participants did not spend any time viewing a webpage that resulted in a time-out, no difference between the two message tones was observed (ratio of mean time online: 0.87, 95% CI 0.75-1.02). Assigning 30 minutes of viewing time to all page views that resulted in a time-out concludes that participants who received content in a motivational tone spent less time viewing content (ratio of mean time online: 0.86, 95% CI 0.77-0.98) than those participants who received content in a prescriptive tone. Using multiple imputations to account for time-outs concludes that there is no difference in participant engagement between the two message tones (ratio of mean time online: 0.87; 95% CI 0.75-1.01).

Conclusions: The analytic technique chosen can significantly affect conclusions about online intervention engagement. We propose a standardized methodology in which time spent viewing webpages that result in a time-out is treated as missing information and corrected with multiple imputations.

Trial Registration: Clinicaltrials.gov NCT00992264; http://clinicaltrials.gov/ct2/show/NCT00992264 (Archived by WebCite at http://www.webcitation.org/6Kw5m8EkP).

(J Med Internet Res 2013;15(11):e252) doi:10.2196/jmir.2781

KEYWORDS

online interventions; engagement; time spent online; multiple imputations; automatic time-out; smoking cessation; utilization; behavioral research; Internet

Introduction

Tracking Exposure Time to Content

As Internet-based behavioral interventions become more prevalent, it is increasingly important that researchers understand how people interact with these programs, including the time participants spend viewing individual content pages and interacting with the program overall [1-3]. Exposure time is one of several important proxies of engagement and could be an important mediator of the programs' intended effects on participants' knowledge, attitudes, and behavior.

Exposure time can be tracked by monitoring when each webpage is opened or exited or when the browser itself is closed. More sophisticated software can further assess activity on a particular webpage by tracking keystrokes or mouse clicks, but no software is able to distinguish when a user is actively reading or viewing a page versus engaged in other activities in their surroundings. Moreover, there are limitations on tracking activities such as viewing content in separate browsers or windows or even working concurrently in other open programs or applications. In all cases, the result will appear to be long periods of inactivity on the program webpage.

A common strategy for dealing with these extended periods of inactivity has been to time out the program after a prespecified time (eg, 30 minutes) [4-11]. This strategy makes sense as a means for closing out the program, but it would be misleading to rely on the time-stamped data from these timed-out periods as an indicator of how long participants were actually exposed to the program content in the open webpage. Other researchers have allowed long page views with no time-out feature, but then truncate the assumed actual viewing time after the fact for analytic purposes [5-7,12,13]. As these two approaches are equivalent for the purpose of measuring time spent online, we treat them identically and refer to each as a "time-out."

Unfortunately, neither of the approaches above is ideal when trying to estimate the time participants were actively viewing online content. Each will likely either over- or underestimate the true viewing time. The actual length of time an individual spent engaged with the webpage is unknown, resulting in missing information. Consequently, excluding all page views that result in a time-out is the same as a complete case analysis and assigning an arbitrary length of time is the same as a single, uninformed imputation method. It is well-known that complete case analyses can result in bias and a reduction in power, as can single imputation [14-16]. As an alternative analytic approach, we recommend using standard missing data methods, in particular multiple imputations (MI), to accommodate long periods of inactivity or time-outs when analyzing time spent online.

Multiple imputations is a flexible and straightforward approach to accommodating missing data, which uses available observed information to predict values for missing information. Standard software exists and simple formulas can be used to incorporate multiple imputations into an analysis. We outline how to implement multiple imputations methods, reviewing standard formulae, to accommodate page views that resulted in a time-out. As an example, we use data collected from a randomized trial of an online smoking cessation intervention called the "Questions about Quitting" (Q^2) trial [10,17]. Using data from this trial, we demonstrate how the method chosen for dealing with time-out data can significantly affect conclusions drawn about program exposure.

The Questions About Quitting Trial

The Q^2 trial was a collaboration between the Group Health Research Institute in Seattle, Washington, and the University of Michigan Center for Health Communications Research in Ann Arbor, Michigan. Detailed information about the study design and methods have been published elsewhere [17]. In brief, adult smokers were recruited from a large regional health plan population and invited to participate in a randomized clinical smoking cessation trial; however, participants did not have to have an interest in quitting smoking to enroll. The primary aim of this full factorial randomized trial [18] was to assess the effects of contrasting levels of four specific design features or factors, on smokers' abstinence and utilization of adjunct treatment (counseling and pharmacotherapy) available to them through their health insurance. The effects of the contrasting levels of each design factor on program engagement were also explored and have been published [10].

Participants in this trial were randomized to one of 16 different combinations of the levels of the four design factors, with half of the participants assigned to one of two contrasting levels of each factor. Randomization was stratified by a baseline measure of a participant's readiness to quit smoking. The four factors and the two contrasting levels of each were message tone (prescriptive vs motivational); navigation autonomy (dictated vs not dictated); proactive email reminders (yes vs no); and availability of testimonials (yes vs no). Here, we focus on comparing the impact of the two contrasting levels of message tone on program engagement, as measured by total time spent viewing online intervention content assessed during the first two months after study enrollment. Half of the participants were randomized to receive intervention content written in a prescriptive message tone, and half were randomized to an intervention written in a motivational tone. Intervention content written in a prescriptive tone was didactic and directly advised smokers to quit smoking and specified how to achieve this goal. In contrast, motivational messaging was written in a tone consistent with the main principles of motivational interviewing (express empathy, develop discrepancy, roll with resistance, support autonomy, and self-efficacy) [19].

The Q^2 program collected automated tracking data each time participants visited the intervention website. This automated collection process recorded the date and time each participant

visited the website and individual date/time stamps every time a content page was accessed or left by logging out of the intervention website, closing the browser, or moving to a different intervention webpage or an external webpage in the same browser window. The Q^2 online intervention included an automatic time-out feature that logged participants out of the program after 30 minutes of inactivity.

Methods

Multiply Imputing Page View Times

Missing information is often classified according to the assumed missing data generating process, that is, the determinants that affect the probability that a particular data element is missing or observed. There are three general missing data generating processes: missing completely at random (MCAR), missing at random (MAR), or not missing at random (NMAR) [15,20]. MCAR assumes the probability that a data element is missing is independent of both observed and unmeasured information. This is unlikely to occur in practice and is the only situation in which a complete case analysis is unbiased (a reduction in power always occurs). The less restrictive MAR generating process assumes that the probability of a data element being missing depends on observed information, while NMAR means that the probability of missingness is dependent on both observed and unmeasured information.

Multiple imputations is a flexible and straightforward approach for accommodating missing data. Imputation methods estimate predictive models using observed information and replace missing data elements with samples from the estimated predictive models. Multiple imputations methods are preferred over single imputation [15,16,21] and repeatedly utilize estimated predictive models to create several complete datasets. Each complete dataset is then analyzed as if all information was observed and information is combined across each of the completed datasets.

There are two common approaches to estimate predictive models when multivariate imputation models are needed (ie, when more than one variable contains missing data or one longitudinal variable has missing information over time). One approach assumes a joint predictive distribution over all recorded variables [22,23], and the other method estimates separate conditional predictive models for each variable with missing information separately [24-26]. The second method is called multiple imputations by chained equations (MICE) or fully conditional specification and is growing in popularity due to its computational efficiency and flexibility. The MICE procedure can easily accommodate binary, categorical, and continuous variables as well as more complex data challenges such as bounded variable values and imputing information for subsets of individuals. For these reasons, we use MICE to impute missing page view times (ie, times for page views that timed out).

MICE methods cycle through each variable with missing information estimating regression models for each variable. Missing values are then replaced with samples from these regression-based predictive distributions, which include the appropriate random error. There are several built-in and stand-alone software packages that implement the MICE procedure [27-30]. MICE algorithms begin by imputing all missing information with naive values (eg, median of observed values of variable); then, the first variable (variable 1) containing missing information is considered (usually the variable with the least amount of missing data). A regression-based predictive model is estimated using observed values of variable 1 and observed and naively imputed values of all other variables selected as predictors. Usually all other variables are used as predictors, unless the analyst chooses to restrict the set of predictors [31,32]. The naively imputed values from variable 1 are replaced with imputations drawn from this predictive model, and the procedure continues on to the second variable with missing information (variable 2). A predictive model is estimated using the observed values of variable 2, the observed and newly imputed values of variable 1, and the observed and naively imputed values of all other predictors. The naively imputed values of variable 2 are then replaced with imputations drawn from this newly estimated imputation model. The imputation process cycles through all the variables that contain missing information replacing the naively imputed missing values with draws from newly estimated imputation models. When the MICE algorithm has cycled through all of the variables with missing information, this is called one “iteration”. The cycle is then repeated, replacing the imputed values from the first iteration with imputations from newly estimated predictive models in the second iteration. Several iterations of MICE are used to ensure that the imputations have “stabilized”, such that the order in which the variables were cycled through no longer affects the imputation values [24,31].

The iterative nature of the MICE algorithm provides both strengths and weaknesses. While the MICE procedure has proven useful in practice, it does not have the solid theoretical justification of alternative imputation methods. For example, convergence (ie, imputation values that “stabilize”) is not guaranteed [24,25]. It is also possible that conditional imputation models will be estimated such that there exists no joint multivariate distribution that is consistent with all conditional distributions. While these drawbacks may give rise to valid theoretical concerns, it appears that they are generally not a concern in practice [21,24,33,34], and MICE is increasingly being used to accommodate missing data in analyses [31,35-37].

Once M completed datasets have been created, each completed dataset is used to calculate the estimate of interest (see #1 in Multimedia Appendix 1), where the subscript m is used to denote that the estimate corresponds to the m-th completed dataset. The average of the M estimates (see #2 in Multimedia Appendix 1) is used as the estimate for the parameter of interest. Rubin developed a straightforward formula for estimating the standard errors of the multiple imputations estimators that accounts for the traditional sampling variability of the estimator and the added variability due to the imputation process [15,38,39]. Rubin’s formula can be used to calculate the standard error for most standard estimators. It is a function of the M complete data standard errors ($W_{1,..M}$) and the variability between the complete data estimates across the M imputations (B_M). Let W_M be the standard error of the complete data estimator in the m-th

imputed dataset, then Rubin's formula for the standard error of the imputation estimator appears as in #3 in Multimedia Appendix 1 [15,38]. In practice, analysts usually use 5-10 imputations as this has been shown to be sufficient to correctly capture the variability in the imputation estimator [39].

We generated five complete datasets with all missing page view times replaced with samples from estimated conditional imputation models. Imputation models were assumed to be normal distributions after log transforming the page view times with means and appropriate standard deviations estimated from linear regression models. We structured the data in a wide format with each person representing one row in the dataset and multiple webpage views represented by multiple columns. A new imputation model was estimated for each repeated page view. We used observed page view times for estimating imputation models and only imputed times for those page views that were observed but that resulted in an automatic time-out. Linear regression models were used to specify the mean of each of the conditional predictive distributions with the following predictors: baseline participant information (participant demographics, smoking history, beliefs about smoking, and readiness to quit), randomized arm, and the number of minutes spent on the first core content page viewed by the participant. Additionally, we used, as predictors, information about the type of webpage viewed, such as the content addressed in the webpage (getting ready to quit, quitting, and staying quit) and the type of page viewed (eg, introduction page, testimonial).

Effect of Content Tone on Engagement

We calculated the total number of intervention visits, individual page views, and total number of page views that resulted in a time-out. We summarized the distribution of the time in minutes that participants spent viewing intervention content excluding all timed-out page views. After imputing missing page view times, total time spent online was calculated for each participant by adding up the number of minutes spent on an intervention webpage. In order to evaluate the impact of assigning an arbitrary value for time spent viewing pages that resulted in a time-out, we varied the number of minutes assigned to page views that timed out from near zero to 30 minutes.

We then compared the contrasting factor levels of message tone on the total time spent viewing intervention content using a zero-inflated Poisson (ZIP) model [40,41]. We used a ZIP model because the distribution of total time spent online had a larger proportion of zeros than expected from a Poisson distribution; study subjects who were never exposed to the intervention content all spent exactly zero total minutes online, causing a notable point mass in the distribution at zero. We included in the logistic portion of the ZIP, which models the "excess" zeros in the population, an intercept term. In the Poisson part of the ZIP model, we included the randomized factor level and the baseline readiness to quit measure that was used to stratify randomization. We report the estimates from the Poisson part of the ZIP model. Generally, estimates obtained from Poisson models are interpreted as incidence rate ratios, but when all subjects share a common period of exposure, as in the Q^2 trial, estimates can be interpreted as the ratio of mean event counts comparing the two contrasting factor levels. Thus, we report the ratio of the mean number of minutes spent online for individuals who received the content in a motivational tone to those who received the prescriptive message tone. We used Stata Version 12 for all analyses, including imputing missing page view times [30,42].

Results

The Q^2 trial enrolled 1865 current smokers; 1175 (63.00%, 1175/1865) participants accessed the online intervention at least once. The intervention content was viewed on a total of 1691 separate visits, resulting in 6592 unique page views. A total of 683 (10.36%, 683/6592) of these page views automatically timed out after 30 minutes of inactivity, and 550 (46.81%, 550/1175) participants had at least one page view that resulted in a time-out. Figure 1 shows the distribution of the time spent on page views that did not result in a time-out; the median observed time spent on an intervention page was 1.07 minutes (interquartile range 0.47-2.27). This suggests that assigning 30 minutes to all page views that resulted in a time-out would overestimate the time participants spent viewing online intervention content.

Figure 2 presents the estimated ratios of mean time spent online for those who received content in a prescriptive tone compared to those who received content in a motivational tone when the value assigned to the time spent viewing webpages that resulted in a time-out is varied from near zero to 30 minutes. While the ratio of means estimate was stable around 0.87, the width of the 95% confidence intervals (CI) around the estimate vary as the time assigned to time-outs changes. Assigning a value close to zero (0.00001 minutes) for time-outs resulted in an estimate of 0.87 with a 95% CI 0.75-1.02 that includes one (ie, fail to reject the null hypothesis that there are no differences in participant engagement between the two factor levels at a 0.05 significance level). Alternatively, assigning a value of 30 minutes to page views that automatically timed out resulted in an estimate of 0.86 with a 95% CI 0.77-0.97 that excludes one, leading to the conclusion that participants assigned to the prescriptive tone viewed content for significantly fewer minutes than those assigned to the motivational tone.

Averaged across the five completed datasets (ie, time-outs replaced with imputed page view times), the average total time spent viewing intervention content was 12.3 minutes. The total number of minutes spent viewing the intervention ranged from less than 1 minute to greater than 180 minutes, with a median of 7.0 minutes. Comparing the mean cumulative number of minutes spent viewing intervention content among those who viewed content in a prescriptive tone versus a motivational tone resulted in a ratio of means of 0.87 (95% CI 0.75-1.01; P=.06). Thus, participants who had content presented in a prescriptive tone spent 13% less time viewing online intervention content, although this difference was not statistically significant at the .05 level.

Figure 1. Distribution of minutes spent viewing an intervention page, excluding page views that resulted in an automatic time-out.

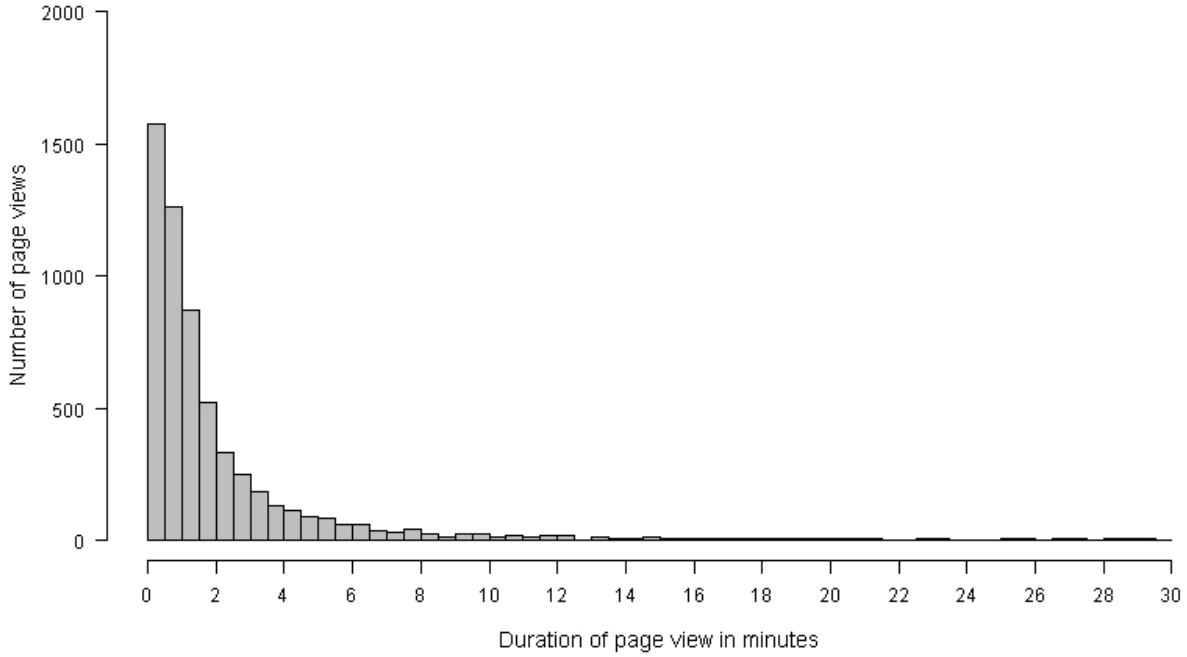
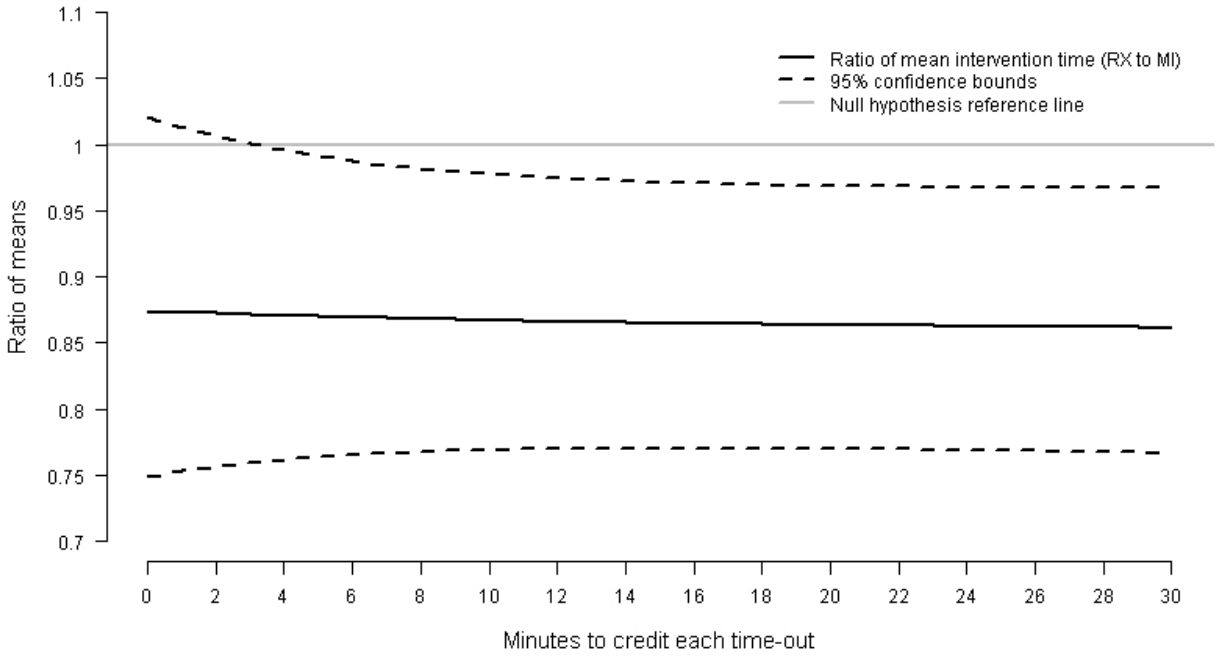


Figure 2. Sensitivity of model results to assigning an arbitrary time spent online to page views that resulted in a time-out (estimate from the zero-inflated Poisson model for the ratio of the mean time spent online comparing individuals who received content in a prescriptive [RX] tone versus a motivational tone [MI]).



Discussion

Principal Findings

The number of available Internet-based behavioral and educational intervention programs has exploded over the past decade. As researchers seek to understand how to optimize the design of these programs to be most effective, it is imperative that researchers examine to what extent participants are exposed to and engage with the programs and to what extent this interaction influences intervention outcomes. Even with the advent of more sophisticated means for tracking program interactivity, there will continue to be periods of time which, either by design or happenstance, involve no direct

human-computer interactions resulting in extended periods of "inactivity". As our case example illustrates, how these data are handled analytically can significantly alter the conclusions drawn about how much time participants actually spent viewing the content. In turn, this could affect analyses designed to explore whether or not program exposure mediated the observed treatment effects.

Conclusions

We propose a standard methodology whereby researchers utilize the MI processes outlined in this paper for managing extended periods of inactivity or time-out data. The decision to use this methodology should be made a priori, as one cannot know ahead of time how much of an impact assigning an arbitrary value to time-outs will have on study conclusions. Researchers are encouraged to employ multiple imputations when examining exposure to online intervention content in the future.

Acknowledgments

This research was funded by the National Cancer Institute (R01 CA138598, J McClure, PI). We are grateful to the contributions of the many study team members at Group Health Research Institute and the University of Michigan. The intervention evaluated in this study was developed by researchers at the Group Health Research Institute and University of Michigan.

Conflicts of Interest

Dr Shortreed has received funding from research grants awarded to Group Health Research Institute by Bristol Meyers Squibb. Mr Bogart and Dr McClure have no conflicts of interest to declare.

Multimedia Appendix 1

Q^2 screenshot.

[PNG File, 139KB - jmir_v15i11e252_app1.png]

References

1. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
2. Baker TB, Gustafson DH, Shaw B, Hawkins R, Pingree S, Roberts L, et al. Relevance of CONSORT reporting criteria for research on eHealth interventions. Patient Educ Couns 2010 Dec;81 Suppl:S77-S86 [FREE Full text] [doi: 10.1016/j.pec.2010.07.040] [Medline: 20843621]
3. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting Internet intervention research. Cognitive Behaviour Therapy 2011 Jun 12;40(2):82-97.
4. Peterson ET. Web Site Measurement Hacks. Sebastopol, CA: O'Reilly Media; 2005.
5. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
6. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39. [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
7. Crenshaw K, Curry W, Salanitro AH, Safford MM, Houston TK, Allison JJ, et al. Is physician engagement with Web-based CME associated with patients' baseline hemoglobin A1c levels? The Rural Diabetes Online Care study. Acad Med 2010 Sep;85(9):1511-1517 [FREE Full text] [doi: 10.1097/ACM.0b013e3181eac036] [Medline: 20736679]
8. Crutzen R, Roosjen JL, Poelman J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. Health Promot Int 2013 Mar;28(1):36-42. [doi: 10.1093/heapro/das008] [Medline: 22377974]
9. Graham AL, Cha S, Papandonatos GD, Cobb NK, Mushro A, Fang Y, et al. Improving adherence to web-based cessation programs: a randomized controlled trial study protocol. Trials 2013;14:48 [FREE Full text] [doi: 10.1186/1745-6215-14-48] [Medline: 23414086]
10. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013;15(3):e69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]
11. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
12. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]

13. Zbikowski SM, Jack LM, McClure JB, Deprey M, Javitz HS, McAfee TA, et al. Utilization of services in a randomized trial testing phone- and web-based interventions for smoking cessation. Nicotine Tob Res 2011 May;13(5):319-327 [FREE Full text] [doi: 10.1093/ntr/ntq257] [Medline: 21330267]
14. Allison PD. Missing Data. Thousand Oaks, CA. USA: Sage; 2002.
15. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Second ed. New York, NY. USA: J Wiley & Sons; 2002.
16. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, editors. Washington, DC: The National Academies Press; 2010.
17. McClure JB, Derry H, Riggs KR, Westbrook EW, St John J, Shortreed SM, et al. Questions about quitting (Q2): design and methods of a Multiphase Optimization Strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. Contemp Clin Trials 2012 Sep;33(5):1094-1102. [doi: 10.1016/j.cct.2012.06.009] [Medline: 22771577]
18. Wu CFJ, Hamada M. Experiments: Planning, Analysis, and Parameter Design Optimization. New York: Wiley and Sons, Inc; 2000.
19. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 2nd ed. New York: The Guilford Press; 2002.
20. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med 2012 Oct 4;367(14):1355-1360. [doi: 10.1056/NEJMsr1203730] [Medline: 23034025]
21. Blankers M, Koeter MW, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]
22. Schafer JL. Analysis of Incomplete Multivariate Data. London, UK: Chapman & Hall; 1997.
23. Schafer JL. Imputation of missing covariates under a multivariate linear mixed model. Technical report, Dept. of Statistics, Penn State University 1997:1-24.
24. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology 2001;27(1):85-95.
25. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation 2006;76(12):1049-1064.
26. van Buuren S. Multiple imputations of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007 Jun;16(3):219-242. [doi: 10.1177/0962280206074463] [Medline: 17621469]
27. Raghunathan TE, Solenberger P, Van Hoewyk J. IVEware: Imputation and Variance Estimation software user guide. Survey Research Center, Institute for Social Research, University of Michigan 2001:1-73.
28. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. Am Stat 2007 Feb;61(1):79-90 [FREE Full text] [doi: 10.1198/000313007X172556] [Medline: 17401454]
29. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45(3):1-67.
30. Royston P, White IR. Multiple Imputations by Chained Equations (mice): Implementation in Stata. Journal of Statistical Software 2011;45(1):1-20.
31. Stuart EA, Azur M, Frangakis C, Leaf P. Multiple imputations with large data sets: a case study of the Children's Mental Health Initiative. Am J Epidemiol 2009 May 1;169(9):1133-1139 [FREE Full text] [doi: 10.1093/aje/kwp026] [Medline: 19318618]
32. Shortreed SM, Laber EB, Pineau J, Murphy SA. Imputations methods for the clinical antipsychotic trials of intervention and effectiveness study. School of Computer Science, McGill University 2010:1-26.
33. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
34. Lee KJ, Carlin JB. Multiple imputations for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol 2010 Mar 1;171(5):624-632 [FREE Full text] [doi: 10.1093/aje/kwp425] [Medline: 20106935]
35. Raghunathan TE, Siscovick DS. A multiple imputation analysis of a case-control study of the risk of primary cardiac arrest among pharmacologically treated hypertensives. Applied Statistics 1996;45:335-352.
36. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. Stat Methods Med Res 1999 Mar;8(1):17-36. [Medline: 10347858]
37. Nevalainen J, Kenward MG, Virtanen SM. Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. Stat Med 2009 Dec 20;28(29):3657-3669. [doi: 10.1002/sim.3731] [Medline: 19757484]
38. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. J Am Stat Assoc 1986;81(394):366-374.
39. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: J Wiley & Sons; 1987.

40. Böhning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated poisson model the decayed, MissingFilled teeth index in dental epidemiology. J R Statist Soc A 1999;162:195-209.
41. Preisser JS, Stamm JW, Long DL, Kincade ME. Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. Caries Res 2012;46(4):413-423 [FREE Full text] [doi: 10.1159/000338992] [Medline: 22710271]
42. StataPress. Stata Statistical Software: Release 12. College Station, TX: Stata Corp; 2011.

Abbreviations

MAR: missing at random
MCAR: missing completely at random
MI: multiple imputations
MICE: multiple imputations by chained equations
NMAR: not missing at random
ZIP: zero-inflated Poisson model

Edited by G Eysenbach; submitted 18.06.13; peer-reviewed by M Blankers, C Morrison; comments to author 19.08.13; revised version received 05.09.13; accepted 13.10.13; published 21.11.13.

Please cite as:
Shortreed SM, Bogart A, McClure JB
Using Multiple Imputations to Accommodate Time-Outs in Online Interventions
J Med Internet Res 2013;15(11):e252
URL: http://www.jmir.org/2013/11/e252/
doi:10.2196/jmir.2781
PMID:24263289

Original Paper

The Effects of Preference for Information on Consumers' Online Health Information Search Behavior

Yan Zhang[1], PhD

University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Yan Zhang, PhD
University of Texas at Austin
1616 Guadalupe Street
Austin, TX, 78701
United States
Phone: 1 512 471 9448
Fax: 1 512 471 3971
Email: yanz@ischool.utexas.edu

Abstract

Background: Preference for information is a personality trait that affects people's tendency to seek information in health-related situations. Prior studies have focused primarily on investigating its impact on patient-provider communication and on the implications for designing information interventions that prepare patients for medical procedures. Few studies have examined its impact on general consumers' interactions with Web-based search engines for health information or the implications for designing more effective health information search systems.

Objective: This study intends to fill this gap by investigating the impact of preference for information on the search behavior of general consumers seeking health information, their perceptions of search tasks (representing information needs), and user experience with search systems.

Methods: Forty general consumers who had previously searched for health information online participated in the study in our usability lab. Preference for information was measured using Miller's Monitor-Blunter Style Scale (MBSS) and the Krantz Health Opinion Survey-Information Scale (KHOS-I). Each participant completed four simulated health information search tasks: two look-up (fact-finding) and two exploratory. Their behaviors while interacting with the search systems were automatically logged and ratings of their perceptions of tasks and user experience with the systems were collected using Likert-scale questionnaires.

Results: The MBSS showed low reliability with the participants (Monitoring subscale: Cronbach alpha=.53; Blunting subscale: Cronbach alpha=.35). Thus, no further analyses were performed based on the scale. KHOS-I had sufficient reliability (Cronbach alpha=.77). Participants were classified into low- and high-preference groups based on their KHOS-I scores. The high-preference group submitted significantly shorter queries when completing the look-up tasks (P=.02). The high-preference group made a significantly higher percentage of parallel movements in query reformulation than did the low-preference group (P=.04), whereas the low-preference group made a significantly higher percentage of new concept movements than the high-preference group when completing the exploratory tasks (P=.01). The high-preference group found the exploratory tasks to be significantly more difficult (P=.05) and the systems to be less useful (P=.04) than did the low-preference group.

Conclusions: Preference for information has an impact on the search behavior of general consumers seeking health information. Those with a high preference were more likely to use more general queries when searching for specific factual information and to develop more complex mental representations of health concerns of an exploratory nature and try different combinations of concepts to explore these concerns. High-preference users were also more demanding on the system. Health information search systems should be tailored to fit individuals' information preferences.

(J Med Internet Res 2013;15(11):e234) doi:10.2196/jmir.2783

KEYWORDS

preference for information; health information; consumer search behavior; search engines

Introduction

Searching for health information online is one of the most popular uses of the Web in the United States across all age groups [1]. When facing health threats, information seeking can enable patients to improve their ability to manage problems and make informed decisions [2] or to make psychosocial and emotional adjustments to illnesses [3]. In everyday life situations, information seeking is a means of gaining knowledge about health behaviors and disease prevention [1,4]. Whether and how users proceed with information seeking and their engagement in the activity, such as the selection of sources, the scope of sources investigated, the types and amount of information sought, and the depth of investigation are affected not only by their demographics (eg, age, gender, and socioeconomic status), knowledge levels (eg, computer and health literacy) and contextual factors (eg, the complexity of a health problem), but also by the individual's personality characteristics, such as locus of control, self-efficacy, and preference for information [3,5-8].

As health care moves from a paternalistic model to a shared decision-making model [8], an unprecedented need is imposed on patients to acquire health-related information. As a result, preference for information, among various personality factors, has drawn much attention from researchers and medical practitioners alike [8,9]. Preference for information is an individual's general tendency toward engaging in health information-seeking behavior. From a stress and coping perspective, it is an enduring personality trait that affects people's use of information seeking as a means of coping with stressful health conditions [10]. Based on this trait, individuals can be classified into "monitors" and "blunters". Monitors are people who are alert and sensitive to the environment and who actively scan it for information to help them cope with stress. Blunters are people who tend to avoid information or distract themselves from it. Miller's Monitor-Blunter Style Scale (MBSS) helps identify monitors and blunters [11]. From the personal control perspective, preference for information (or informational involvement) and behavioral involvement are considered to be the two most common approaches for people to gain a sense of control and a belief that they can alter or affect outcomes. The Krantz Health Opinion Survey-Information Scale (KHOS-I) was developed to measure individuals' informational involvement [12].

Empirical studies have provided evidence in support of the power of preference for information, measured by the MBSS, in predicting people's information-seeking behavior in various contexts of patient-provider interactions, such as people undergoing cancer screening [13], women before gynecological surgery [14], patients in cancer treatments with a palliative intention [15], soldiers with combat-related post-traumatic stress disorder [16], and women with multiple sclerosis [17]. A common finding is that monitors, compared to blunters, have more doubts about medical procedures, desire more information, and ask more questions of providers. Consequently, monitors tend to have significantly more knowledge about medical procedures and their medical situation. The relationship between preference for information and information seeking seems to hold in stressful nonmedical situations as well. For example, Bar-Tal and Spitzer [18] found that, in stressful interpersonal conflict situations, undergraduate students with a monitoring style were more likely to seek information and support from others to solve problems.

Preference for information is also related to people's interactions with written information. Examining the information needs of women with multiple sclerosis, Baker [17] found that more monitors than blunters rated as relevant a pamphlet providing disease-related information (fatigue or treatment for acute attacks), regardless of whether the information was general or specific. Koo et al [19] found that the monitoring versus the blunting style could predict patients' interest in reading about and seeking written information concerning their prescription medicines, with monitors being more than twice as likely to be interested in reading such information.

Similarly, preference for information, measured by the KHOS-I, was found to predict people's health-related information-seeking behavior. In an early study on college students who visited a college medical office with complaints such as headaches, colds, and flu, Krantz et al [12] found that participants who received higher KHOS-I scores asked more questions during the visit. Barsevick and Johnson [20] found that higher KHOS-I scores also predicted the number of questions that women undergoing a colposcopy asked their providers.

Preference for information is not only related to people's information-seeking behavior, but also to their cognitive and emotional states. The monitoring style is often associated with higher concern and anxiety levels and higher demands for assurance. For example, Miller [13] found that, in cancer screening, patients with a monitoring coping style were more concerned and distressed about their cancer risk. Caldwell [21] found that patients who scored highly on the KHOS-I were more anxious in a preoperative setting than those who scored low, due to a tendency to focus on the negative aspects of the threatening situation. Mahler and Kulik [22] identified a similar correlation in male coronary-artery-bypass patients: those with higher KHOS-I scores experienced more social interaction and emotional problems and, as a result, desired more information to help reduce uncertainty and emotional arousal.

Not surprisingly, monitors and blunters are affected differently by information. Although the findings are not conclusive [23], it appears that individuals who prefer to have information and are given information tend to become less anxious, while those who have a low preference for information but are given information tend to become more anxious [13,24-29]. For example, Morgan et al [30] found that, before a colonoscopy, patients given information congruent with their coping style experienced significantly less self-reported anxiety after the information intervention and spent less time in recovery. In contrast, patients given information not congruent with their coping style maintained their pre-intervention anxiety level. Patients are also more likely to take action when information interventions are congruent with their dispositional preference for information. Williams-Piehota et al [31] found that when provided with detailed reassuring messages, monitors were more likely to obtain mammograms and, when provided with

more concise and simple messages, they were less likely to obtain mammograms. In contrast, blunters were more motivated by the less detailed messages.

As reviewed, most studies on preference for information have focused on its relationship with information-seeking behavior in the context of patient-physician interaction or information interventions offered by providers (eg, number of questions asked and the amount of information sought), as well as its relationship with patients' cognition (eg, uncertainty) and emotions (eg, anxiety, level of stress, and satisfaction with treatments). Few studies have examined whether it has an impact on general consumers' health information search behavior—the behavior while interacting with search systems - for example, whether people with a higher preference for information explore more search results. There is also a lack of research on how this personality trait impacts the other two important aspects of an information search experience: people's perceptions of search tasks (as representations of information needs) and their experience with search systems [32]. These are important research topics, as more than 80% of US adult Web users search online for health information and the need for a personalized health information search experience keeps increasing [33]. This study intends to fill these gaps by addressing three research questions:

1. Does the participants' preference for information affect their behavior when they interact with search engines in completing look-up versus exploratory tasks?
2. Does preference for information affect participants' perceptions of task difficulty, the mental effort required to complete the tasks, and satisfaction with their performance?
3. Does preference for information affect participants' experience with search systems?

Knowledge gained through this investigation will not only help improve the current understanding of consumers' health information search behavior, but will also shed light on how search systems can be tailored to individual consumers' information preferences and needs.

Methods

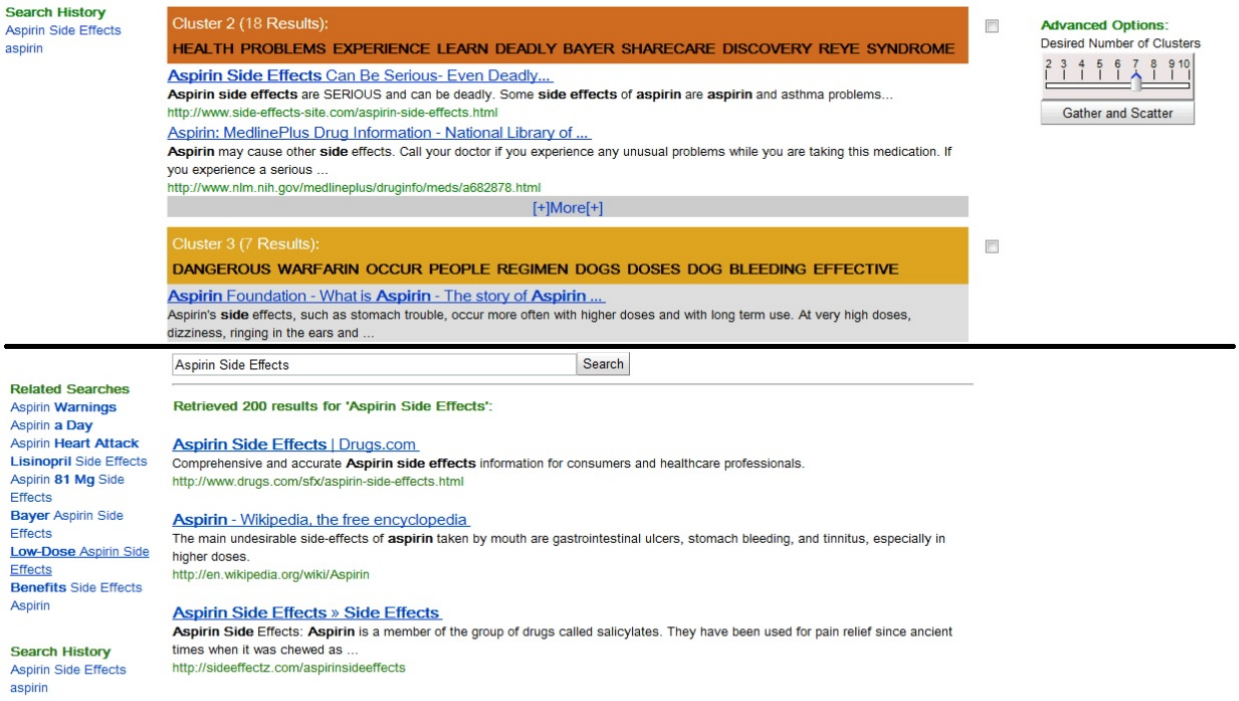
Participants

Participant observation in an experimental setting was used for data collection. This method is appropriate because, in order to examine the effects of preference for information on participants' behavior while they interact with search engines, it is necessary to control the search tasks and systems used. A convenient sample of 40 participants who were general consumers of health information (ie, who had searched previously on the Web for health information) was recruited by a message sent to a mailing list of students, faculty, staff, and alumni at a large research university. The age of the 40 participants ranged from 18-55 (mean 25.0, SD 8.5), with 10% (4/40) being between the ages of 18 and 20, 78% (31/40) between 20 and 30, and 13% (5/40) between 30 and 55. Each participant was compensated with US$15.

Platforms

The participants used one of two interfaces (Figure 1): (1) a classic Web search engine interface or (2) a Scatter/Gather-enabled search interface. Both interfaces are based on the Bing API (application programming interface); when a user types keywords into either interface, they are sent to Bing and the results are retrieved from Bing (Microsoft's Web search engine). The classic interface resembles general Web search engines. It features a simple search box for keywords and the search results are presented as a relevance-ranked list.

Figure 1. Screenshot of search results - the upper half of the image shows the scatter/gather interface and the lower half shows the classic interface.



The Scatter/Gather interface has the same search box, but the results are grouped into clusters based on their topic similarities and the clusters are ranked by size (ie, the number of results contained in each cluster). A set of keywords is also presented along with each cluster [34]. The Scatter/Gather-enabled interface was chosen because its clustering function, as well as the keywords, may help to reduce the difficulty that consumers have with medical terminology.

Tasks

Two types of tasks were defined: simple look-up and exploratory. Look-up tasks are tasks aimed at finding particular health-related facts, whereas exploratory tasks are oriented toward learning, investigating, and making sense of specific health issues. This classification was adopted because it can effectively characterize the goal and complexity of a health information request [35]. Four simulated search tasks were created—two look-up and two exploratory tasks—as shown in Table 1.

To ensure that the tasks reflected those that general health information consumers would search for in real life, the first three tasks were derived from questions posted in the health section of "Yahoo! Answers", a social Q&A site where consumers post questions and/or answer questions posted by peers [36]. The last task was adapted from a task set designed for testing a Medline-based medical document collection [37]. In addition, in two pilot sessions, we asked participants to comment on the tasks. Both agreed that the tasks were something that they were likely to do in real life.

Table 1. Search tasks.

Task	
Look-up tasks	
	A friend of yours is an athlete. Now he wants to increase his muscle mass. He has been training without creatine, but would like to start a regimen. He is seeking your advice on this. You decide to find out what the side effects of taking creatine are.
	A heart attack is a medical emergency and prompt treatment increases the chance for survival. According to the American Heart Association, heart attacks cause 1 out of every 5 deaths. According to the National Institutes of Health (NIH) more than 1.2 million heart attacks occur each year in the United States and about 460,000 of these are fatal. Approximately 300,000 people die annually from heart attacks before they can receive medical treatment. To be prepared for possible emergencies, you decide to find out what to do when a person around you has a heart attack.
Exploratory tasks	
	Imagine that one of your close family members has lived with diabetes for years. Recently, he was also diagnosed with hypertension. You decided to do some research on the clinical associations between the two conditions so that you are able to effectively discuss with him about various implications of this diagnosis.
	Imagine that you recently began suffering from migraines. You heard about two possible treatments for migraine headaches, beta-blockers, and/or calcium channel blockers, and you decided to do some research about them. At the same time, you want to explore whether there are other options for treating migraines without taking medicines.

Data Collection Procedure

The data collection took place in a private lab. Upon arrival, a moderator gave the participants an overview of the study and asked each participant to read and sign an informed consent form. After this, the participants completed a questionnaire reporting demographics, as well as their experience with Web and health information searches. Then they were asked to complete the MBSS and the KHOS-I scales. The MBSS consists of four hypothetical stress situations, each of which is followed by eight declarative statements, with four reflecting monitoring behavior and four blunting behavior. Participants were instructed to check all the statements that applied to them [11]. The KHOS-I consists of seven items and measures people's desire to be informed and their desire to gather information. Participants rated their answers in a forced-choice format as "agree" or "not agree." Higher scores indicated a greater need for seeking information on issues regarding their health [12]. The two measures have been used in a number of studies and demonstrated acceptable reliability and validity [6,8].

After completing the two scales, participants were randomly assigned to one of the two interfaces. As a result, 20 participants used the classic search interface and 20 used the Scatter/Gather interface. Each participant, using one or the other interface, completed all four tasks. The order in which the tasks were presented was randomized to reduce learning effects. Before the search began, participants watched a video tutorial demonstrating the basic functions of the interface to which they were assigned. During the search, when they checked out a website from the results list, they were prompted to rank the relevance and usefulness of the site on a 7-point Likert scale (1=not relevant, 7=relevant; 1=not useful, 7=useful). The ratings, along with the search queries and websites visited, were logged automatically by the search systems.

After completing each task, participants completed a short questionnaire assessing the difficulty of the task, the mental effort it required, and their level of satisfaction with their performance on a 5-point Likert scale. Camtasia software recorded each search session in video format. After completing all four tasks, participants filled out a questionnaire assessing their overall experience with the system that they had used. The questionnaire consisted of statements about users' perceptions of the ease of use and usefulness of the system, their understanding of how the system worked, their levels of enjoyment and engagement, and their intentions toward using the system in the future. The rating scale was a 5-point Likert scale (1=strongly disagree with the statement, 5=strongly agree). The items measuring the ease of use and usefulness were

adapted from prior studies [38,39]. The remaining aspects were each measured by a single item. At the end of the session, participants were asked to comment on search tasks and their behavior of performing the tasks. Each session lasted 1-1.5 hours.

Data Analysis

The independent variable was preference for information, measured by the MBSS and the KHOS-I. For the MBSS, separate monitor and blunter scores were calculated [11]. Each participant's monitoring score was determined by adding the number of monitoring items checked (M) and the blunting score was determined by adding the number of blunting items checked (B). The Kuder-Richardson Formula 20 test [40] was used to examine the internal consistency of the two subscales. The results suggest that neither subscale was reliable with participants in this study (Monitoring subscale: Cronbach alpha=.53; Blunting subscale: Cronbach alpha=.35). As a result, no further analyses were performed based on the MBSS measurements, similar to what has been done in previous research [27,41].

The KHOS-I score was determined by adding the number of items participants checked indicating a desire for information. The Kuder-Richardson 20 coefficient for KHOS-I was 0.77, indicating an adequate internal consistency for this scale with participants in this study. Thus, further analyses were performed based on the KHOS-I scale. Descriptive statistics showed that the mean score of the KHOS-I for the sample in this study was 4.3 (SD 2.1). Participants were categorized into two groups using a median split [42]: those scoring above the median were in the high-preference group and those scoring equal to or below the median were in the low-preference group. The two groups differed significantly on the KHOS-I score (t_{38}=7.7, P=.001).

The dependent variables included the participants' search behavior, their perceptions of the tasks and task performance, and their experience with search systems. Search behavior was operationalized by typical actions involved in a search process [32,43], including (1) session length, (2) query behavior (the number of queries submitted, query length, and query reformulation), and (3) accessing of results (the number of sites viewed, the number of sites rated as relevant, and the number of sites rated as useful; a site is considered relevant or useful when the rating is greater than 4 on a 1-7 Likert scale). These data were recorded in transaction logs. The analysis of query reformulation followed the topology developed by Rieh and Xie [44] and modified by Zhang et al [45]. The resulting topology included five types of query reformulations based on semantic changes: (1) specification (participants specify the meaning of the previous query), (2) generalization (participants generalize the meaning of the previous query), (3) parallel movement (participants replace one concept in the previous query with a new concept and the two queries have a partial overlap in meaning), (4) new concept movement (participants change to new concepts and the new query does not overlap with the previous query), and (5) rephrasing (participants rephrase the previous query by changing the form of the query without changing the meaning, such as correcting misspellings and rephrasing the previous query into a question). Two independent coders, the author and a trained graduate student, coded all the query reformulation instances; the inter-coder reliability was 98.5%. Discrepancies were resolved by discussion.

The participants' perceptions of the search tasks and task performance were measured by their ratings on task difficulty, mental effort, and task performance after completion of each task. Their experience with the search system was measured by ratings on the user experience questionnaire administered at the end of each search session. These data were imported into SPSS for statistical analysis.

A series of t tests suggests that the two interface groups did not differ in demographics (including age, gender, computer experience, and health information search experience), nor in any of the measurements on search behavior, perceptions of tasks, and experience with search systems [45]. In addition, two-way ANOVAs (analyses of variance) suggested that there were no significant interactions between interfaces and information preferences (as measured by KHOS-I). Because the focus of this paper is on the impact of preference for information, to simplify the presentation, the two interface groups were pooled together for further data analysis. Then, t tests were used to investigate the impact of preference for information. The level of statistical significance was set at .05.

Results

Characteristics of the Participants

Table 2 summarizes the demographics of the two information preference groups—high-preference and low-preference—as well as their KHOS-I scores and their experience with health information searches. A chi-square test indicated that the two groups did not differ by gender (χ^2_1=1.44, P=.31). As well, t tests indicated that they also did not differ in age, Web experience, or experience with health information searches (years and frequency of searching for health information). In addition, all of the participants had or were in the process of getting a college degree and the two groups did not differ in their education levels.

Table 2. Demographics and health information search (HIS) experience by preference for information.

	Low preference	High preference	*t* (df)	*P* value
KHOS-I score, mean (SD)	3.0 (1.7)	6.6 (0.5)	−7.72 (38)	<.001
Male	8	7	-	-
Female	18	7	-	-
Age (years), mean (SD)	24.9 (8.0)	25.4 (9.6)	−0.18 (38)	.86
Web experience (years), mean (SD)	12.5 (4.3)	14.6 (4.7)	−1.47 (38)	.15
HIS experience (years), mean (SD)	3.9 (1.2)	3.2 (1.6)	1.57 (38)	.12
HIS frequency (times/month), mean (SD)	2.9 (0.7)	2.6 (1.1)	0.71 (38)	.48

Information Search Behavior

Overview

Participants' behavior while interacting with search engines was measured in light of three aspects: session length (task completion time), query formulation, and accessing of results.

Task Completion Time

Table 3 summarizes the two groups' mean task completion times for each type of task. The *t* tests indicated that preference for information did not have an impact on the task completion time for either the look-up or exploratory tasks.

Table 3. Task completion times (in seconds).

	Low preference Mean (SD)	High preference Mean (SD)	*t* (df)	*P* value
Look-up	452.4 (186.0)	382.4 (125.9)	1.26 (38)	.22
Exploratory	556.6 (211.9)	549.3 (151.7)	0.12 (38)	.91

Query Formulation

Table 4 shows the average number of queries submitted by the two groups in completing each type of task and the average query length. The *t* tests indicated that the two groups differed only in the average length of queries submitted to solve look-up tasks, with the low-preference group submitting significantly longer queries than the high-preference group (t_{38}=2.42, *P*=.02).

Table 4. Number of queries submitted and query length.

		Low preference Mean (SD)	High preference Mean (SD)	*t* (df)	*P* value
Look-up					
	No. of queries	2.1 (1.4)	2.4 (1.0)	−0.67 (38)	.51
	Query length[a]	5.2 (1.6)	4.1 (1.1)	2.42 (38)	.02
Exploratory					
	No. of queries	3.4 (1.5)	3.8 (1.8)	−0.80 (38)	.43
	Query length[a]	3.5 (1.0)	3.8 (0.8)	−0.85 (38)	.40

[a]Number of search words/terms

Table 5. The patterns of query reformulation.

	Low preference, % Mean (SD)	High preference, % Mean (SD)	*t* (df)	*P* value
Look-up				
Specification	18.4 (30.0)	26.7 (28.8)	−0.84 (38)	.40
Generalization	7.7 (21.5)	5.9 (15.4)	0.27 (38)	.79
Parallel movement	19.4 (29.0)	33.2 (29.4)	−1.43 (38)	.16
New concept movement	3.9 (9.4)	4.4 (11.3)	−0.17 (38)	.87
Rephrasing	19.8 (36.8)	8.3 (21.4)	1.07 (38)	.29
Exploratory				
Specification	26.8 (23.4)	24.8 (20.2)	0.27 (38)	.79
Generalization	13.2 (15.9)	19.3 (27.6)	−0.89 (38)	.38
Parallel movement	16.9 (18.9)	32.9 (27.6)	−2.18 (38)	.04
New concept movement	35.1 (29.7)	11.6 (20.0)	2.65 (38)	.01
Rephrasing	8.0 (14.3)	11.3 (18.0)	−0.64 (38)	.53

Table 5 shows the patterns of query reformulation. Because participants differed in the number of queries they submitted, percentages were used to normalize the data for comparisons of their behavioral patterns. The *t* tests indicated that the two groups did not differ in their query reformulation behavior when completing the look-up tasks. In contrast, when completing the exploratory tasks, the high-preference group performed a significantly higher percentage of parallel movements (t_{38}=2.18, *P*=.04) and a significantly lower percentage of new concept movements (t_{38}=2.65, *P*=.01).

Accessing of Results

Three aspects of the access of results were examined: the number of sites viewed by the participants, the percentage of the sites rated as relevant, and the percentage of the sites rated as useful. Table 6 shows the statistics. The *t* tests indicated that the two groups accessed an equal number of results in solving both types of tasks and they also reported equal percentages of sites as relevant and useful.

Table 6. Number of sites visited, the percentage of sites rated as relevant, and the percentage of sites rated as useful.

		Low preference Mean (SD)	High preference Mean (SD)	*t* (df)	*P* value
Look-up					
	No. of results	4.2 (2.4)	3.7 (1.8)	0.69 (38)	.50
	Relevant (%)	80.3 (20.6)	83.4 (21.5)	−0.44 (34)	.66
	Useful (%)	70.2 (27.8)	71.4 (21.2)	−0.13 (35)	.90
Exploratory					
	No. of results	4.8 (2.1)	5.1 (2.0)	−0.33 (38)	.75
	Relevant (%)	84.2 (18.1)	72.3 (25.9)	1.68 (36)	.10
	Useful (%)	73.1 (24.0)	57.9 (27.4)	1.79 (36)	.08

Perceptions of Tasks and Task Performance

Table 7 shows the two groups' perceptions of task difficulty, the mental effort required to complete the tasks, and their satisfaction with their performance. The *t* tests indicated that the two groups did not differ in their perceptions of simple look-up tasks, but one difference was found for exploratory tasks: the high-preference group perceived the exploratory tasks to be more difficult than the low-preference group (t_{38}=2.01, *P*=.05).

Table 7. Perceptions of task difficulty and task performance.

		Low preference Mean (SD)	High preference Mean (SD)	*t* (df)	*P* value
Look-up					
	Task difficulty[a]	2.1 (0.6)	2.0 (0.6)	0.69 (38)	.50
	Mental effort[b]	2.4 (0.7)	2.4 (0.6)	−0.11 (38)	.91
	Satisfaction[c]	4.2 (0.4)	4.1 (0.6)	0.52 (38)	.60
Exploratory					
	Task difficulty[a]	2.3 (0.4)	2.6 (0.6)	−2.01 (38)	.05
	Mental effort[b]	2.6 (0.5)	2.6 (0.6)	0.04 (38)	.97
	Satisfaction[c]	4.0 (0.5)	4.0 (0.7)	0.20 (38)	.85

[a]Rated on a 5-point scale (1-very easy, 5-very difficult)

[b]Rated on a 5-point scale (1-very small amount, 5–very large amount)

[c]Rated on a 5-point scale (1-very disappointed, 5–very satisfied)

User Experience With the Systems

Table 8 shows the two group's experience with the search systems used in the study. The *t* tests indicated that the two groups differed in their perceptions of the systems' usefulness, with the low-preference group rating the search systems as more useful (t_{38}=2.18, *P*=.04).

Table 8. User experience[a].

	Low preference Mean (SD)	High preference Mean (SD)	*t* (df)	*P* value
Ease of use	4.3 (0.5)	4.2 (0.6)	0.47 (38)	.64
Usefulness	4.1 (0.6)	3.5 (0.9)	2.18 (38)	.04
Understand how it works	3.8 (0.7)	3.6 (0.8)	0.59 (38)	.56
Enjoyment	3.8 (0.9)	3.6 (0.9)	0.70 (38)	.49
Engagement	3.3 (0.8)	3.2 (1.2)	0.29 (38)	.77
Future use	3.5 (0.9)	3.6 (1.1)	−0.50 (38)	.62

[a]Rated on a 5-point scale (1-strongly disagree, 5-strongly agree)

Discussion

Summary

This study explores the impact of a personality factor, preference for information, on the health information search behavior of general consumers, on their perceptions of tasks and performance, and on their experience with search systems. In prior studies, preference for information was mainly investigated in the context of patient-provider interactions, in order to predict patients' behavior of seeking information from providers and to inform interventions that can help patients cope with stress [30,46]. This study makes a contribution by going beyond this traditional context to the context of consumers' interaction with Web-based search engines for health information. This extension is important since searching for health information has become one of the most popular online activities [47] and information found online has an increasingly significant impact on consumers' health care decisions [1]. The discussion is organized around two themes: (1) the measurement of preference for information, and (2) the impact of preference for information on consumers' health information search behavior and the implications for system design. The limitations of this study are also discussed.

Measurement of Preference for Information

Participants' preference for information was measured by the MBSS and the KHOS-I. Both the monitoring and blunting subscales had poor internal consistency, indicating low reliability of the MBSS in measuring consumers' tendency to seek or to avoid information in the context of searching for health information in the Web environment. Several factors may contribute to this result. First, the sample size was small. The study involved only 40 participants. Second, the MBSS has inherent limitations. As critiqued by other researchers, scenarios in the MBSS are hypothetical. Particularly, the scenario of "being held hostage by a group of armed terrorists" is too far removed from most people's life experiences [48,49]. Moreover, the validity of the scale has been questioned [50]. For example, Barsevick and Johnson [20] found that the MBSS was not a sensitive indicator of preference for information for patients undergoing colposcopy. A third factor may be the nature

of the context of this study. In most prior studies, the MBSS was applied to patients in life-threatening medical situations (eg, cancer and heart disease) and/or undergoing stressful medical procedures (eg, colposcopy and biopsy) [13,20,27]. In this study, participants were general health information consumers and the scenarios were of an everyday nature and less life-threatening. The applicability of the MBSS in predicting people's preference for information in such a context merits more investigation as more and more consumers go online for health information.

In contrast, the KHOS-I had a sufficient level of reliability (Cronbach alpha=.77). This might have been because the KHOS-I was initially designed to measure individuals' tendency to seek information in routine and general health care contexts, which is well reflected in the statements in the scale, for example: "I usually ask the doctor or nurse lots of questions about the procedures during a medical exam." Similar levels of reliability of this scale were found not only for groups with specific conditions, such as cancer, [9,51], myocardial infarction [52], and dental problems [46], but also for general health information consumers, such as undergraduate students [12]. In this study, the mean KHOS-I was 4.3 (scale 0-7), indicating that participants in this study had a comparatively high preference for information, which may be accounted for by the participants' young age and their overall high level of education.

We also found that the MBSS and the KHOS-I were not correlated, which is consistent with two earlier studies [20,49]. One potential explanation for this is the low reliability of the MBSS for the participants in this study. The other reason could be that the two scales may well measure different constructs [20].

The Impact of Preference for Information on Consumers' Search Behavior, Perceptions of Tasks, and Experience With Search Systems

Information searching involves three major elements: the user, the task, and the system [32]. To understand the impact of preference for information on health information searches, we measured participants' search behavior (session length, query formulation, and accessing of results) and assessed their perceptions of tasks (for difficulty, mental effort, and performance) and their experience with the search systems used in the study (for ease of use, usefulness, understanding of the systems' working mechanisms, enjoyment, engagement, and projected future use of the system).

Participants with different levels of preference for information differed significantly on several of these measurements. First, the length of queries differed. The average length of queries submitted by the high-preference group was significantly shorter than those submitted by the low-preference group when completing the look-up tasks. A possible explanation is that participants with a low preference for information were eager to get the right answer as quickly as possible, so that they attempted to make the search queries as specific as possible, whereas the participants with a high preference were willing to do some exploration and thus submitted more general queries. An examination of the actual queries revealed that many queries submitted by the low-preference group were complete questions (eg, "what to do if someone has a heart attack"), rather than keywords, which further supports this speculation.

A second difference was the pattern of query reformulation. When completing the exploratory tasks, the high-preference group made a significantly higher percentage of parallel movements than did the low-preference group, whereas the low-preference group made a significantly higher percentage of new concept movements than the high-preference group. In parallel movements, participants replace a concept in the previous query, so that the two queries have partial overlap in meaning, and in new concept movements, participants change to a new concept and the new query does not overlap with the previous query. This result suggests that, when exploring a health-related topic, people with a high preference for information may be more likely to take steps to gradually explore relationships between concepts, whereas those with a low preference may be more likely to investigate concepts one by one. This result further indicates that people with a high preference for information might develop a more complex mental representation of the medical problem at hand, whereas people with a low preference might simplify the problem and thus develop a comparatively less complex conceptual representation. Prior studies on patient-provider interactions have revealed that patients with a high preference for information often ask more questions of the provider [12,15,20]. However, few studies investigated the nature of the questions, how the questions were related to one another, and whether more concepts were involved in questions imposed by high-preference patients. In future studies, qualitative studies are needed to shed light on these research questions, which will also help interpret the results of this study.

Along the same lines, the finding that patients with a high preference for information tended to ask more questions of providers naturally leads to an expectation that when searching for health information, particularly with exploratory tasks, high-preference participants may submit more queries to the system and visit more search results than their low-preference counterparts. However, such results were not observed in this study. A possible explanation is the experimental nature of the study. Participants were performing assigned tasks, rather than their own tasks. A naturalistic approach to data collection, such as transaction log analysis, would help elucidate the relationships between preference for information and the number of search queries and search results visited. Another explanation is that high-preference participants, more so than low-preference patients, may possess a greater ability to process retrieved information. Therefore, they did not submit more queries and examine more results, but acquired more information. Future studies may test this speculation by comparing the learning outcomes of low- and high-preference groups after a search session.

A third difference between the groups was the participants' perceptions of task difficulty. The high-preference group perceived the exploratory tasks to be significantly more difficult than did the low-preference group. This relationship did not hold with look-up tasks, which involved seeking factual information. This finding may also be attributed to the possibly

more complex mental representations that the high-preference group developed for the exploratory tasks.

The fourth difference was the participants' perceptions of the usefulness of the search systems. The high-preference group perceived the systems to be less useful than did the low-preference group, which seems consistent with their perceptions of the difficulty of exploratory tasks. It is possible that the high-preference group's perceptions of the task difficulties made them less satisfied with the utility of the systems for addressing their needs. Comparable results have been found in the context of patient-provider interactions. For example, Timmermans et al [15] reported that, in cancer treatment with a palliative intention, high monitors reported having more doubts about the treatment decision and being less satisfied with the information received, while high blunters expressed fewer doubts and more satisfaction.

These results indicate that preference for information has an impact on consumers' interactions with search systems for health information. As reviewed, prior studies have consistently suggested that information interventions in patient-provider encounters are most effective when they are congruent with receivers' preference for information [53]. It is natural to postulate that health information search systems may be most useful and effective when they are tailored to individuals' information preferences. Some system design implications can be drawn from the results of this study. For example, for those with a low preference for information, the system could provide a "natural" user interface that allows them to write queries in natural language (ie, long queries or queries in question format) rather than artificial keywords [54], to accommodate their need for imposing very specific queries. At the same time, the system should improve its ability in processing long queries [55]. When presenting results, the system could present the most specific query results at the top of the results list or recommend a list of more specific queries. For those with a high preference, the system could allow them to explore relationships between concepts and recommend new but related concepts (based on medical thesauri or on the mining of query logs) to accommodate their propensity to develop complex networks of concepts. When recommending queries, both more general and more specific queries can be provided to allow flexibility in exploration. Moreover, systems can offer functions to present search results in visual ways that can enable users to explore relationships among the concepts involved in the results (eg, a tree-map view or a network view of concepts). In this study, the Scatter/Gather interface clustered results based on topic similarity and provided a set of keywords to represent each cluster, but failed to illustrate relationships between concepts. This may be one of the reasons that Scatter/Gather interface did not differ from the basic search interface in supporting searches.

Limitations and Future Studies

There are limitations to this study. First, the sample consisted primarily of people with high education and high computer literacy; thus, the generalizability of the results is limited. Future studies should extend the sample to people with low computer and health literacy, as well as to patients with particular conditions, such as cancer and diabetes. Such studies can inform the tailoring of information systems to the needs of underserved groups. Second, a limited number of search behavior variables was measured, which directly limits our understanding of the scope of the impact of preference for information on consumers' health information search behavior. In future studies, researchers should examine consumer behaviors in relation to, for example, the content examined (eg, evidence-based medical research vs user-generated content) and the types of sites visited (eg, commercial sites vs academic sites), to examine whether preference for information has an impact. Third, the tasks used in this study were classified as look-up and exploratory tasks. Future studies can look into other ways of classifying health-related search tasks, such as by the goals of searches (eg, seeking diagnosis, treatment, or medical facilities) [56]. In addition, only four search tasks, two in each category, were involved in this study. Future studies could attempt to include a larger number of search tasks to further reduce the possible impact of other task features, such as the topic, on people's search behavior.

Conclusions

The personality trait, preference for information, showed an impact on general consumers' search behavior for health information, their perceptions of task difficulties, and their experience with search systems. Compared to people with a low preference for information, those with a high preference exerted greater efforts in information searching. These efforts were manifested not so much at the behavioral level (eg, submitting more search queries or checking out more search results), but more at the conceptual level, with those with a high preference being more likely to use more general queries when searching for specific factual information and to develop more complex mental representations of health concerns of an exploratory nature and try different combinations of concepts to explore these concerns. Consequently, high-preference users were also more demanding on the system. These findings suggest that system developers should take into consideration users' preference for information in designing health information search systems. To further advance our knowledge about consumers' health information search behavior and to inform the design of more effective systems, the influence of other personality factors, such as locus of control, on information searches should also be examined.

Acknowledgments

The author wants to thank all the participants for their contributions to the study. The study was partially supported by an Alumni Fellowship from the School of Information at the University of Texas at Austin.

Conflicts of Interest

None declared.

References

1. Fox S. Pew Internet & American Life Project. 2011. The social life of health information URL: http://www.pewinternet.org/~/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2013-07-13] [WebCite Cache ID 6I5Oz1aTW]
2. van der Molen B. Relating information needs to the cancer experience: 1. Information as a key coping strategy. Eur J Cancer Care (Engl) 1999 Dec;8(4):238-244. [Medline: 10889622]
3. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Libr Assoc 2005 Jul;93(3):353-362 [FREE Full text] [Medline: 16059425]
4. Zhang Y. Consumer health information searching process in real life settings. In: Proc. Am. Soc. Info. Sci. Tech. 2013 Jan 24 Presented at: Proceedings of the ASIST; 2012; Baltimore, MD p. 1-10. [doi: 10.1002/meet.14504901046]
5. Lenz ER. Information seeking: a component of client decisions and health behavior. ANS Adv Nurs Sci 1984 Apr;6(3):59-72. [Medline: 6426379]
6. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
7. Zhang Y. Beyond quality and accessibility: Source selection in consumer health information searching. J Am Soc Inf Sci Technol 2014:1-17 (forthcoming).
8. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. Patient Educ Couns 2011 Mar;82(3):346-354. [doi: 10.1016/j.pec.2010.12.008] [Medline: 21239134]
9. Sutherland HJ, Llewellyn-Thomas HA, Lockwood GA, Tritchler DL, Till JE. Cancer patients: their desire for information and participation in treatment decisions. J R Soc Med 1989 May;82(5):260-263 [FREE Full text] [Medline: 2754680]
10. Glanz K, Schwartz MD. Stress, coping, and health behavior. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education. San Francisco, CA: Jossey-Bass; 2008:211-236.
11. Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Pers Soc Psychol 1987 Feb;52(2):345-353. [Medline: 3559895]
12. Krantz DS, Baum A, Wideman M. Assessment of preferences for self-treatment and information in health care. J Pers Soc Psychol 1980 Nov;39(5):977-990. [Medline: 7441487]
13. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [Medline: 8625088]
14. Steptoe A, O'Sullivan J. Monitoring and blunting coping styles in women prior to surgery. Br J Clin Psychol 1986 May;25 (Pt 2):143-144. [Medline: 3730650]
15. Timmermans LM, van Zuuren FJ, van der Maazen RW, Leer JW, Kraaimaat FW. Monitoring and blunting in palliative and curative radiotherapy consultations. Psychooncology 2007 Dec;16(12):1111-1120. [doi: 10.1002/pon.1177] [Medline: 17345558]
16. Solomon Z, Arad R, Mikulincer M. Monitoring and blunting: Implications for combat-related post-traumatic stress disorder. J. Traum. Stress 1991 Apr;4(2):209-221. [doi: 10.1002/jts.2490040205]
17. Baker LM. A study of the nature of information needed by women with multiple sclerosis. Library & Information Science Research 1996 Dec;18(1):67-81. [doi: 10.1016/S0740-8188(96)90031-6]
18. Bar-Tal Y, Spitzer A. The effect on coping of monitoring, blunting, and the ability to achieve cognitive structure. J Psychol 1999 Jul;133(4):395-412. [doi: 10.1080/00223989909599751] [Medline: 10412220]
19. Koo M, Krass I, Aslani P. Enhancing patient education about medicines: factors influencing reading and seeking of written medicine information. Health Expect 2006 Jun;9(2):174-187. [doi: 10.1111/j.1369-7625.2006.00381.x] [Medline: 16677196]
20. Barsevick AM, Johnson JE. Preference for information and involvement, information seeking and emotional responses of women undergoing colposcopy. Res Nurs Health 1990 Feb;13(1):1-7. [Medline: 2305105]
21. Caldwell LM. The influence of preference for information on preoperative stress and coping in surgical outpatients. Appl Nurs Res 1991 Nov;4(4):177-183. [Medline: 1772249]
22. Mahler HI, Kulik JA. The development and validation of three videos designed to psychologically prepare patients for coronary bypass surgery. Patient Educ Couns 1995 Feb;25(1):59-66. [Medline: 7603934]
23. Hilzenrat N, Yesovitch R, Shrier I, Stavrakis M, Deschênes M. The effect of information level and coping style on pain and anxiety in needle liver biopsy. Can J Gastroenterol 2006 Sep;20(9):597-600 [FREE Full text] [Medline: 17001402]
24. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract 2000 Sep;49(9):796-804. [Medline: 11032203]
25. Miller SM, Mangan CE. Interacting effects of information and coping style in adapting to gynecologic stress: should the doctor tell all? J Pers Soc Psychol 1983 Jul;45(1):223-236. [Medline: 6886967]

26. Davis TM, Maguire TO, Haraphongse M, Schaumberger MR. Undergoing cardiac catheterization: the effects of informational preparation and coping style on patient anxiety during the procedure. Heart Lung 1994 Apr;23(2):140-150. [Medline: 8206772]
27. Rees CE, Bath PA. Information-seeking behaviors of women with breast cancer. Oncol Nurs Forum 2001 Jun;28(5):899-907. [Medline: 11421149]
28. Watkins LO, Weaver L, Odegaard V. Preparation for cardiac catheterization: tailoring the content of instruction to coping style. Heart Lung 1986 Jul;15(4):382-389. [Medline: 3636298]
29. Ludwick-Rosenthal R, Neufeld RW. Preparation for undergoing an invasive medical procedure: interacting effects of information and coping style. J Consult Clin Psychol 1993 Feb;61(1):156-164. [Medline: 8450101]
30. Morgan J, Roufeil L, Kaushik S, Bassett M. Influence of coping style and precolonoscopy information on pain and anxiety of colonoscopy. Gastrointest Endosc 1998 Aug;48(2):119-127. [Medline: 9717776]
31. Williams-Piehota P, Pizarro J, Schneider TR, Mowad L, Salovey P. Matching health messages to monitor-blunter coping styles to motivate screening mammography. Health Psychol 2005 Jan;24(1):58-67. [doi: 10.1037/0278-6133.24.1.58] [Medline: 15631563]
32. Marchionini G. Information seeking in electronic environments. Cambridge: Cambridge University Press; 1997.
33. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]
34. Gong X, Ke W, Khare R. Studying scatter/gather browsing for web search. In: Proc. Am. Soc. Info. Sci. Tech. 2013 Jan 24 Presented at: Proceedings of the American Society for Information Science and Technology; 2012; Baltimore, MD p. 1-4. [doi: 10.1002/meet.14504901328]
35. Zhang Y. Toward a layered model of context for health information searching: An analysis of consumer-generated questions. J Am Soc Inf Sci Tec 2013 Apr 05;64(6):1158-1172. [doi: 10.1002/asi.22821]
36. Zhang Y. Dissertation submitted to the University of North Carolina at Chapel Hill. 2009. The construction of mental models of information-rich web spaces: The development process and the impact of task complexity URL: https://www.ischool.utexas.edu/~yanz/YanZhang_Dissertation.pdf [accessed 2013-10-14] [WebCite Cache ID 6KNERaG20]
37. Mu X, Ryu H, Lu K. Supporting effective health and biomedical information retrieval and navigation: a novel facet view interface evaluation. J Biomed Inform 2011 Aug;44(4):576-586. [doi: 10.1016/j.jbi.2011.01.008] [Medline: 21277997]
38. Capra R, Marchionini G, Oh JS, Stutzman F, Zhang Y. Effects of structure and interaction style on distinct search tasks. 2007 Presented at: Proceedings of the 7th ACM/IEEE-CS JCDL; 2007; Vancouver, Canada p. 422-451. [doi: 10.1145/1255175.1255267]
39. Zhang Y. Searching for specific health-related information in MedlinePlus: Behavioral patterns and user experience. J Am Soc Inf Sci Technol 2014:1-16 (forthcoming). [doi: 10.1002/asi.22957]
40. Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika 1937 Sep;2(3):151-160. [doi: 10.1007/BF02288391]
41. Schwartz MD, Lerman C, Miller SM, Daly M, Masny A. Coping disposition, perceived risk, and psychological distress among women at increased risk for ovarian cancer. Health Psychol 1995 May;14(3):232-235. [Medline: 7641664]
42. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods 2002 Mar;7(1):19-40. [Medline: 11928888]
43. Jansen BJ, Spink A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Information Processing & Management 2006 Jan;42(1):248-263. [doi: 10.1016/j.ipm.2004.10.007]
44. Rieh SY, Xie H. Analysis of multiple query reformulations on the web: The interactive information retrieval context. Information Processing & Management 2006 May;42(3):751-768. [doi: 10.1016/j.ipm.2005.05.005]
45. Zhang Y, Broussard R, Ke W, Gong X. The evaluation of a Scatter/Gather interface for supporting distinct health information search tasks. J Am Soc Inf Sci Technol 2014:1-16 (forthcoming).
46. López-Jornet P, Camacho-Alonso F, Sanchez-Siles M. Patient information preferences and behaviour in relation to oral biopsies. Br J Oral Maxillofac Surg 2012 Dec;50(8):e115-e118. [doi: 10.1016/j.bjoms.2011.12.010] [Medline: 22300552]
47. Zickuhr K. Pew Internet & American Life Project. 2010. Generations 2010 URL: http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf [accessed 2013-07-13] [WebCite Cache ID 6I5P4w0ba]
48. Muris P, Van Zuuren F, De Jong P, De Beurs E, Hanewald G. Monitoring and blunting coping styles: The Miller behavioural style scale and its correlates, and the development of an alternative questionnaire. Personality and Individual Differences 1994 Jul;17(1):9-19. [doi: 10.1016/0191-8869(94)90257-7]
49. Garvin BJ, Kim CJ. Measurement of preference for information in US and Korean cardiac catheterization patients. Res Nurs Health 2000 Aug;23(4):310-318. [Medline: 10940956]
50. Ross CJ, Maguire TO. Informational coping styles: a validity study. J Nurs Meas 1995;3(2):145-158. [Medline: 8935771]
51. Hagerty RG, Butow PN, Ellis PA, Lobb EA, Pendlebury S, Leighl N, et al. Cancer patient preferences for communication of prognosis in the metastatic setting. J Clin Oncol 2004 May 1;22(9):1721-1730. [doi: 10.1200/JCO.2004.04.095] [Medline: 15117995]

52. Garvin BJ, Moser DK, Riegel B, McKinley S, Doering L, An K. Effects of gender and preference for information and control on anxiety early after myocardial infarction. Nurs Res 2003 Dec;52(6):386-392. [Medline: 14639085]
53. Kiesler DJ, Auerbach SM. Optimal matches of patient preferences for information, decision-making and interpersonal behavior: evidence, models and interventions. Patient Educ Couns 2006 Jun;61(3):319-341. [doi: 10.1016/j.pec.2005.08.002] [Medline: 16368220]
54. Hearst MA. 'Natural' search user interfaces. Commun. ACM 2011 Nov 01;54(11):60. [doi: 10.1145/2018396.2018414]
55. Bian J, Liu Y, Agichtein E, Zha H. Finding the right facts in the crowd: factoid question answering over social media. 2008 Presented at: Proceedings of WWW '08; April 21 - 25, 2008; Beijing, China p. 467-476. [doi: 10.1145/1367497.1367561]
56. Xie B, Wang M, Feldman R, Zhou L. Exploring older and younger adults' preferences for health information and participation in decision making using the Health Information Wants Questionnaire (HIWQ). Health Expect 2012 Jul 11:1-14. [doi: 10.1111/j.1369-7625.2012.00804.x] [Medline: 22783800]

Abbreviations

KHOS-I: Krantz Health Opinion Survey-Information scale
MBSS: Monitor-Blunter Style Scale

Edited by G Eysenbach; submitted 19.06.13; peer-reviewed by J Zhang, M Rethlefsen, H Zhai; comments to author 12.07.13; revised version received 16.07.13; accepted 15.09.13; published 26.11.13.

Please cite as:
Zhang Y
The Effects of Preference for Information on Consumers' Online Health Information Search Behavior
J Med Internet Res 2013;15(11):e234
URL: http://www.jmir.org/2013/11/e234/
doi:10.2196/jmir.2783
PMID:24284061

Original Paper

The Virtual Skeleton Database: An Open Access Repository for Biomedical Research and Collaboration

Michael Kistler[1], MSc; Serena Bonaretti[1,2], PhD; Marcel Pfahrer[3]; Roman Niklaus[3]; Philippe Büchler[1], PD, PhD

[1]Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

[2]UCSF School of Medicine, Radiology and Biomedical Imaging, University of California, San Francisco, CA, United States

[3]Engineering and Information Technology, Bern University of Applied Sciences, Biel, Switzerland

Corresponding Author:
Michael Kistler, MSc
Institute for Surgical Technology and Biomechanics
University of Bern
Stauffacherstrasse 78
Bern, 3014
Switzerland
Phone: 41 316315959
Fax: 41 316315960
Email: michael.kistler@istb.unibe.ch

Abstract

Background: Statistical shape models are widely used in biomedical research. They are routinely implemented for automatic image segmentation or object identification in medical images. In these fields, however, the acquisition of the large training datasets, required to develop these models, is usually a time-consuming process. Even after this effort, the collections of datasets are often lost or mishandled resulting in replication of work.

Objective: To solve these problems, the Virtual Skeleton Database (VSD) is proposed as a centralized storage system where the data necessary to build statistical shape models can be stored and shared.

Methods: The VSD provides an online repository system tailored to the needs of the medical research community. The processing of the most common image file types, a statistical shape model framework, and an ontology-based search provide the generic tools to store, exchange, and retrieve digital medical datasets. The hosted data are accessible to the community, and collaborative research catalyzes their productivity.

Results: To illustrate the need for an online repository for medical research, three exemplary projects of the VSD are presented: (1) an international collaboration to achieve improvement in cochlear surgery and implant optimization, (2) a population-based analysis of femoral fracture risk between genders, and (3) an online application developed for the evaluation and comparison of the segmentation of brain tumors.

Conclusions: The VSD is a novel system for scientific collaboration for the medical image community with a data-centric concept and semantically driven search option for anatomical structures. The repository has been proven to be a useful tool for collaborative model building, as a resource for biomechanical population studies, or to enhance segmentation algorithms.

(J Med Internet Res 2013;15(11):e245) doi:10.2196/jmir.2930

KEYWORDS

medical informatics; Internet; image processing; computer-assisted; demographic analysis; statistical models

Introduction

Beyond the obvious variability in appearance, a large anatomical variation among the human population exists. For example, even for similar looking people, the human skeleton has a large variability in bone shape, internal structure, and mechanical strength. Following the systematic use of modern imaging techniques in the medical routine, the interest in studying human variability grew in the medical research community. Roughly two decades ago, Cootes et al [1] introduced the first statistical model to quantify anatomical shapes [2]. This technique became very popular in the medical image analysis community and has been used in a wide range of applications: novel implant design [3,4], surgical planning [5,6], and improving image segmentation

[7,8] and image registration [9]. More recently, these models have been extended to combine the description of the anatomical shape with mechanical information to analyze the risk of fracture in a population [10,11] or to design orthopedic implants [12].

A recurrent problem that is faced when applying statistical shape model (SSM) to specific clinical indications is the large amount of data required to build a valid model. A repository for images and processed data is therefore vital for the fast development of new applications or image processing algorithms based on this technology. However, the existing solutions for storing medical images are not appropriate for building statistical models. In the clinical environment, picture archiving and communication system (PACS) was introduced in the 1980s to manage clinical images of patients [13,14]. The digital imaging and communications in medicine (DICOM) is the standard file format used in PACS, which ensures that the proper metadata concerning the patient and the image technique are available to the radiologist. This system is ideal for archiving clinical data but is ineffective if the user wants to retrieve a large amount of data corresponding to the same anatomical site. In addition, due to legal restrictions and data protection laws, it is not possible for researchers to freely access the large database of medical imagery stored in clinical institutions. Therefore, significant efforts are necessary to obtain data, to process them, and to build the models. However, even after this time-consuming process, collections of data are often lost or mishandled in research institutions, resulting in replication of work, even within the same institution. A centrally organized system can limit or even suppress these barriers, which hinders fast and innovative approaches in medical image research.

Inspired by PACS and institutional repository software [15-17], several projects have been developed for the biomedical research community. For example, the Extensible Neuroimaging Archive Toolkit (XNAT) [18] has developed an open source solution for the neuroimaging research community. Recently, the Midas media archiving platform [19] was developed to extend the functionality of DSpace [15] repository software. This platform aims to simplify the collaboration between researchers and to provide more flexibility on the supported file format and data structure. Several other projects are building infrastructures to host data collections in their respective research domain. For example, the Living Human Digital Library [20] aims to provide a collection of raw and processed data on the anatomo-functional characteristics of the human musculoskeletal system at dimensional scales spanning from the whole body down to the molecules. However, this system is limited to a small number of human bodies generated during the project and does not allow the user to upload their own datasets. The Internet Brain Segmentation Repository [21] provides manually guided expert segmentation along with magnetic resonance brain image data. All these projects provide extremely important data to the scientific community, but they either consider a limited number of subjects, are specific to a certain organ or pathology, or present a mix of patients and pathology. A general database where medical image datasets are stored and further processed toward statistical shape modeling is required to model the variability of the complete human anatomy.

The objective of the Virtual Skeleton Database (VSD) is to provide and maintain an open access repository for scientific researchers to archive, search, preserve, trace, and share data resources for statistical models. In this paper, the authors describe the main concepts implemented in the system and demonstrate its operation on exemplary use cases.

Methods

General Concepts

The VSD can be described as a catalog of data that is accessible and organized with a focus on, but not limited to, building statistical anatomical models. Therefore, the VSD was built around the concept of “data objects”, which constitute the basic element of the system. These data objects can be any type of image file, processed data, or models. This approach provides a large flexibility to the system in terms of data formats, data organization, and data collaboration. The design is intended to have the capacity to provide the same basic functionality as a PACS, but it is more flexible and can integrate additional research data, such as labeled images and statistical models.

Technology

The architecture of the VSD is based on the classical multitier architecture (see Figure 1). The topmost layer provides multiple interfaces: the Web portal for managing and sharing the data, Web-based Distributed Authoring and Versioning protocol (WebDAV) for direct interaction with different file systems, and a Representational State Transfer interface (REST), an application programming interface for interaction with other scientific applications while the administration functionality is provided through a Web application. The middle tier provides all the necessary application logic but also contains the functionality covering security aspects. Data storage and data access are provided on the next tier. Aside from the relational database system for storing the information, the system integrates other tools and frameworks for specific purposes; ClearCanvas is used to handle DICOM files, Insight Toolkit, Visual Toolkit (ITK/VTK) for image processing, Statismo for statistical shape models, and Fuseki for the integration of ontologies. The scalable data layer (database and data) is equipped with a mechanism to provide a full synchronization between two functional redundant systems.

The development of the VSD was based mainly on Microsoft technologies. The system relies on a Microsoft Structured Query Language (SQL) database and the Microsoft Internet Information Service (IIS) Web-server. For the application development, Microsoft Razor, which is built on the Microsoft .NET technology, was chosen to enable consistent development of a responsive Web service. The Razor application framework was complemented by the established frameworks Bootstrap [22] and jQuery [23] for a smooth and dynamic user experience on the website (see Multimedia Appendix 1).

Registration

Standard user and object rights management is required to control access, visibility, and sharing of the objects stored on the VSD. However, since legal and ethical aspects play an important role when medical images are involved, a two-step

registration procedure and user management were implemented. The system's registration procedure was designed to build the VSD as a research-only repository (see Figure 2). In the first step, academic institutions can register as research units. After approval of the research unit by the VSD administrator, each research unit can administer their members and is responsible for controlling access to the research data and ensuring the proper rights and licenses are attached to their uploaded datasets.

Medical Images

The system provides the user with multiple options for the image upload. A standard Web upload form and a Java applet are available. While the standard browser upload is most convenient for small images, the Java applet is able to submit files larger than 2GB to the server. Additionally, if close integration in the local system is required, the VSD can be accessed through the WebDAV, which allows for customized upload workflows. The VSD integrates the most common image formats used by the medical image analysis community, but theoretically any kind of medical image format could be integrated. The VSD can read classical DICOM files through the ClearCanvas framework, and image types recognized by the ITK library (MetaImage, Niftii, and Analyze) are accepted (see Figure 3). During the upload process, metadata are extracted from the images, preview images are generated, and a de-identification of the medical image is performed. The de-identification consists of file type-specific header manipulation and the replacement of the original file name by a generic VSD file name, preserving the information only for the original contributor. To keep the storage system efficient, the uploading functionality has mechanisms to avoid duplicated objects on the disk and supports multiple versions of the objects.

To assign proper and consistent metadata to the images, the VSD features a publishing concept where uploaded data have to be reviewed before they become available to the other users. During this process, the user is asked to review the uploaded dataset and to provide the missing information. First, the dataset has to be characterized by annotating the anatomical structure depicted in the image. For this metadata, the anatomical terms are taken from the comprehensive Foundational Model of Anatomy (FMA) ontology [24]. This ontology includes more than 75,000 anatomical terms and 450,000 direct relations between classes. Additional metadata like age, gender, and image modality can be filled to enrich the meta-information content of the image. If a related dataset already exists on the VSD, the user can reference this data by linking both datasets together. For example, different image modalities of the same patient and anatomical structure can be linked and therefore grant a direct access between both datasets. As a last step, a Creative Commons license can be attached to a dataset to specify the intended offline usage. To control the sharing of the data on the website, predefined permission sets are available to specify the access rights. After a final activation, the data are accessible to other users of the VSD.

Figure 1. Schematics of the multitier architecture of the VSD, which consists of different types of interfaces, an application logic and security layer, and data access and data storage.

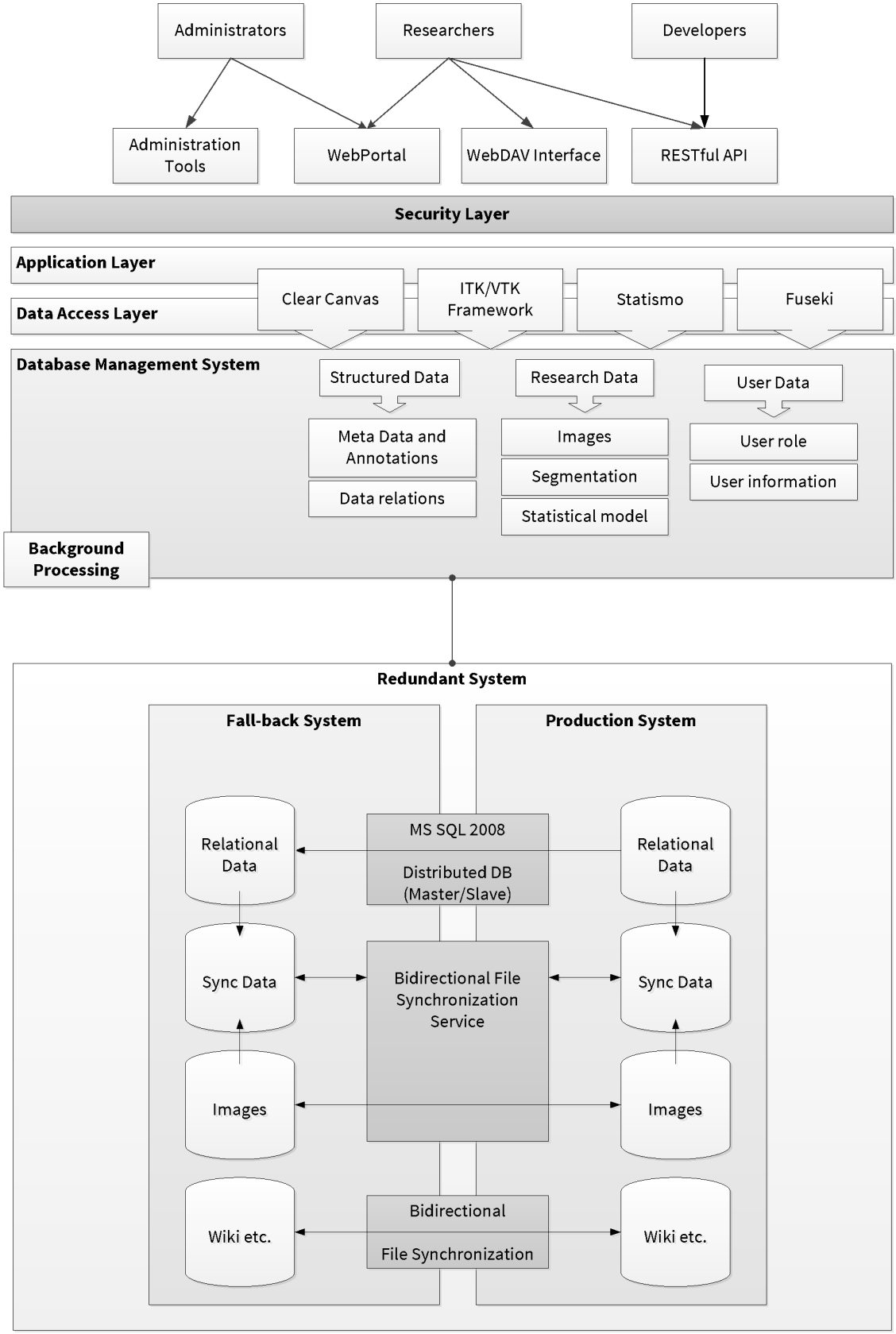


Figure 2. Registration process - Top: registration of a new research unit and appointment of a research unit administrator. Bottom: individual users can register to a research unit and have to be accepted by the respective administrator.

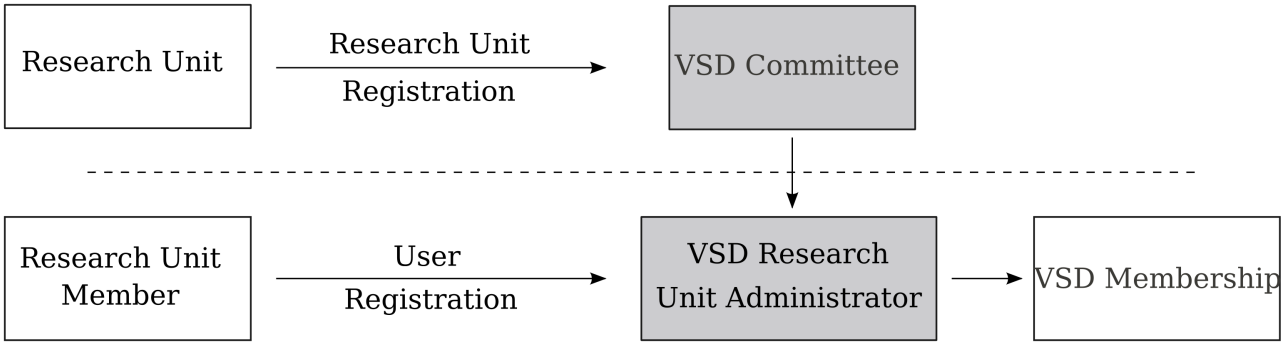
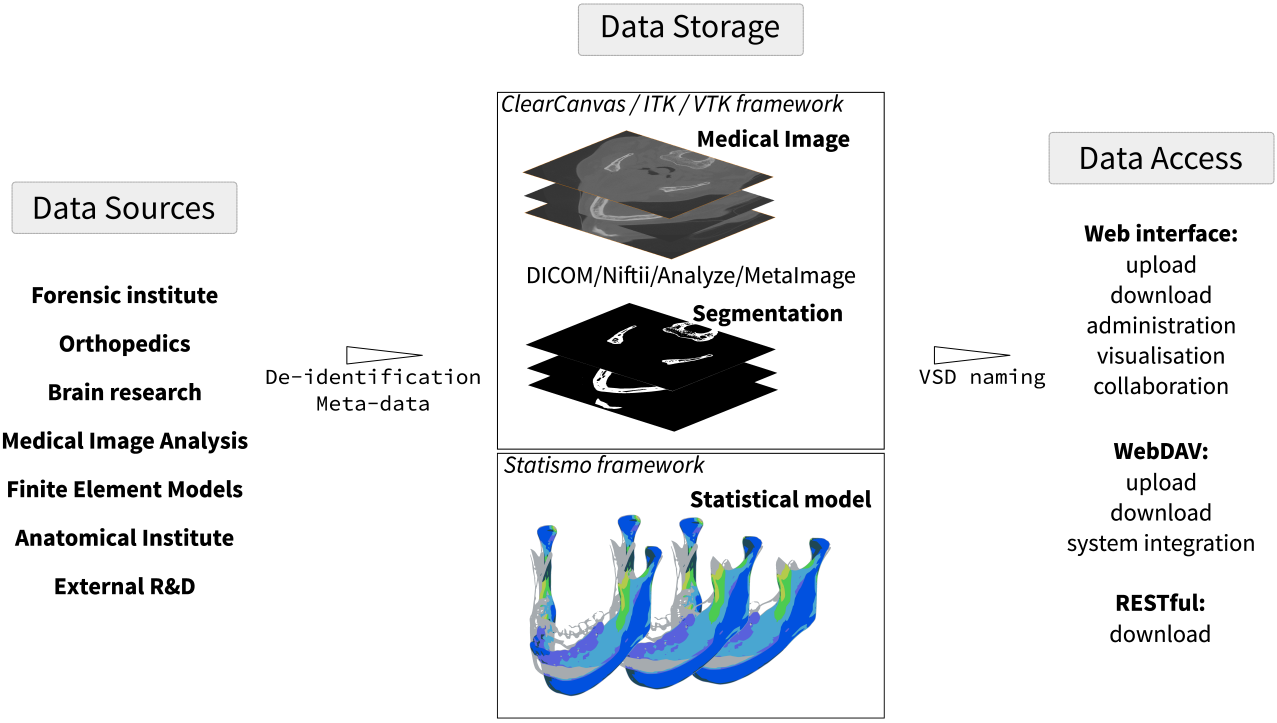


Figure 3. Conceptual overview of the VSD - data from various sources are stored in the VSD and can be accessed, organized, and downloaded.

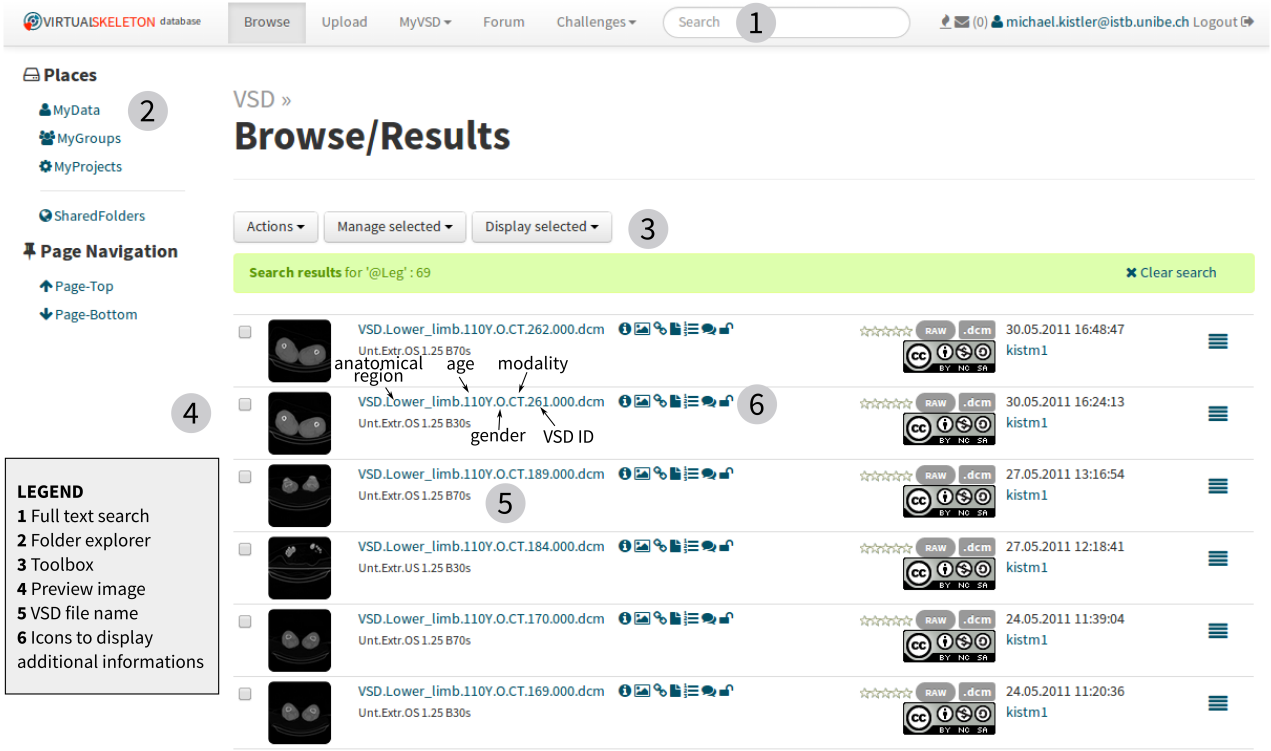


Data Organization

Users can search the available online data catalog to find the datasets relevant for their projects (see Figure 4 and Multimedia Appendices 2 and 3). The system allows the creation of virtual folders to organize the available datasets, which provides the user with the flexibility to freely create their personal workspace. Data objects are not physically moved or duplicated, but the system creates a reference to the data object, retaining the original file permission and ownership. For collaboration or other purposes, the user can share parts of their workspace by changing the permission of the folder accordingly. To simplify the collaboration within a research unit, the system provides a default shared group folder, which is accessible and manageable by all members of the research unit.

Several functions provide additional information to the users and help them identify the relevant datasets. Each dataset can be rated for quality and tagged with personal comments. To comment on the object publicly, a discussion thread can be initiated in the integrated VSD forum (see Multimedia Appendices 4 and 5). Additionally, the search function allows users to find specific metadata including ontology terms. The ontology search is using SPARQL Protocol and RDF Query Language (SPARQL), which queries the semantic information contained in the ontology. This feature enables a powerful search where anatomical substructures can be identified within stored data objects. For example, a user can search for images containing the body part “Leg”. With a conventional index-based search, only data annotated as “Leg” can be retrieved (zero search results in the VSD). With the semantics of the ontology, the VSD will find and return data that are or contain the Leg (69 datasets found in the VSD).

Figure 4. VSD Web interface to browse data with folder explorer: MyData=Location of the user's uploaded data; MyGroups=Default collaboration folder accessible to all group members; MyProjects=Folders to organize data into personal projects; SharedFolders=Folders of other VSD members which are shared to the user.



Retrieve Images

An individual dataset as well as a folder structure can be downloaded from the VSD for processing on a local machine. The data are wrapped for download into a compressed archive file to reduce file size and to preserve the folder structure from the VSD. For local access to folders on the VSD, the user can mount the WebDAV folder as a network drive on their computer.

Segmented Images

Images that contain segmentation labels can be uploaded the same way as medical images. The system identifies how many anatomical structures are labeled in the segmentation image, the de-identification is performed, and the thumbnail images for preview are extracted. During the review process, all the labels of anatomical structures have to be specified and the segmentation technique can be described. The uploaded segmentation data can then be linked to original medical images or to other segmentations of the same structure, which have been performed using a different technique.

Statistical Models

To standardize the building and storage of shape models, the open source library "Statismo" has been developed [25]. An important aspect of this library is that it proposes a simple file format to store shape model, based on the Hierarchical Data Format (HDF5). The VSD recognizes shape models stored in HDF5 and is able to extract important information about the model, such as the datasets used or the registration technique used. After the upload, the review process is similar to that of the images, asking the user to annotate the anatomical region and to specify the license and access rights.

Three exemplary projects using key functionalities of the VSD are presented. First, the creation of a statistical model of the cochlea for surgery and implant optimization is described. As a second example, a statistical appearance model of the femur is used to analyze fracture risk within a population. Finally, an online segmentation challenge hosted by the VSD is presented.

Results

International Collaboration on Statistical Shape Model of the Cochlea

The goal of this international collaboration project, the Hear-EU project [26], is to improve cochlear implants. This surgical procedure tries to overcome hearing loss by implanting a device that accesses the cochlea of the inner ear. Clinical expertise and detailed information about critical structures close to the surgical site are needed for such a surgery. Since this surgical intervention is very complex and a vast variety of patient anatomies exist, the project hypothesizes that a comprehensive knowledge of the involved anatomical structures would enable improvements for implant design and for surgical planning. In this context, the VSD acts as a data hub (see Figure 5; map adapted from [27]) to assist the building of a statistical shape model of both the middle and the inner ear. High resolution computed tomography (CT) images with a data size of over 5GB are uploaded to the VSD and will be processed by different institutions across Europe. The resulting segmentation images will be subsequently re-uploaded to the VSD. Having both the images and processed data available will enable the Technical

University of Denmark to create a statistical shape model of the regions of interest. Finally, the shape models will be available on the VSD and can be used by different teams to optimize surgical procedure as well as the functional outcome of the surgery on a patient-specific basis.

Biomechanical Femur Population Study

The authors searched the VSD for available lower extremities datasets to conduct a study on the fracture risk on this population. A total of 72 female and 72 male subjects were isolated and retrieved together with their available metadata. The population consisted mainly of Caucasian (114) supplemented with Asian (28) and African (2) subjects. The female subpopulation's body mass index (mean 26.4, SD 5.5) was slightly higher compared to the male (mean 25.5, SD 4.9) subjects, while the male represented an older subpopulation than the female group.

For each gender, the statistical appearance model was calculated using the image-based approach described by Bonaretti et al [28] and polyaffine demons [29]. From both appearance models, 1000 instances were created and biomechanical finite element simulations were performed with a loading situation corresponding to normal walking [30,31]. The bone intensity values were converted into material properties [32,33] and mapped onto each node of the FE (finite element) mesh of the second order tetrahedral elements.

The relationships between the fracture risk in the femoral neck and morphometric measurement was performed (see Table 1). Six geometrical measures were identified [34,35] (see Figure 6) as possible risk predictors. As a mechanical parameter, the average Young's modulus in the femoral neck was calculated. All predictor parameters were centered and normalized to perform the regression and evaluate their influence in femur fracture risk.

For both genders, body weight was identified as the most relevant predictor for fracture risk while the mechanical measures showed a lower influence of fracture risk in general. In contrast, among geometrical parameters, the neck-shaft angle could be identified as playing a predominant role for both populations, whereas the order of importance of the parameters shows clear difference between men and women. The complete study results are presented by Bonaretti [28].

Segmentation Challenge for Brain Tumors

The VSD was chosen by the Multimodal Brain Tumor Segmentation (BRATS) challenge organizer to host their data and challenge. The data were publicly available to any team around the world to develop and train their segmentation algorithms. In addition, the VSD was used to evaluate the submission of the competitor during the challenge, which took place during the annual conference of the Medical Image Computing and Computer Assisted Intervention (MICCAI) Conference 2012. Hosting such a challenge required several functionalities, which illustrates most of the features of the system. The VSD hosted multicenter magnetic resonance (MR) imaging sets and their related ground truth, which were defined by inter-reader agreement from clinicians. The clinical segmentations were used to determine the accuracy of the segmentation of tumor subregions. A total of 80 brain MR and segmentations were available to the competitors for training purposes (see Figure 7). An additional 30 challenge datasets which consist only of MR data were available for download. Their respective ground truth segmentation was hosted on the VSD but the download was protected through appropriate file permission. The users uploaded their segmentation through the Web interface, reviewed the uploaded segmentation, and started an automatic evaluation process. The VSD automatically identified the ground truth corresponding to the uploaded segmentations. The evaluation of the different standard parameters used to evaluate the quality of the segmentation (such as Dice coefficients) runs in the background and takes less than 1 minute per segmentation. Individual and overall results of the evaluation were automatically published on the VSD webpage and were downloadable as a comma separated values (CSV) file for further statistical calculations (see Multimedia Appendices 6 and 7). The VSD has evaluated more than 3000 segmentations and had over 80 registered users for the BRATS 2012 group. The service was successfully used for the challenge in 2012 when around 700 evaluations were handled by the system within the last 2 hours of the competition. The results were presented in the proceedings of MICCAI 2012 [36], and the data are now open to anyone who wants to evaluate new segmentation tools against previous BRATS evaluation sets. So far, around 15,000 additional segmentations have been successfully benchmarked by the VSD.

Figure 5. Collaborative network, interactions, and workflows for the Hear-eu project on cochlear implants.

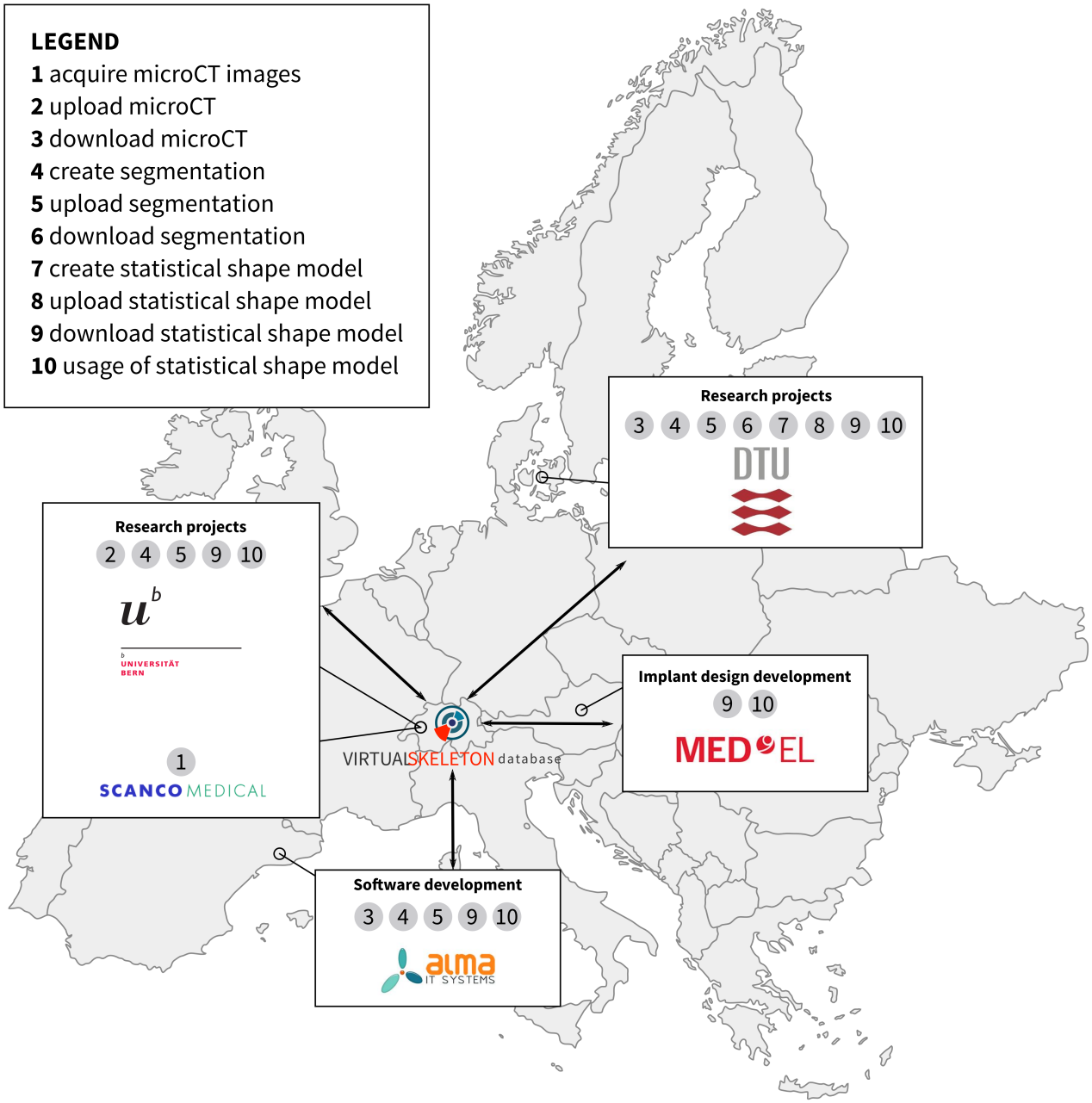


Figure 6. Geometrical measures: (1) femur length, (2) neck length, (3) neck width, (4) femoral head width, (5) neck-shaft angle, (6) anteversion angle.

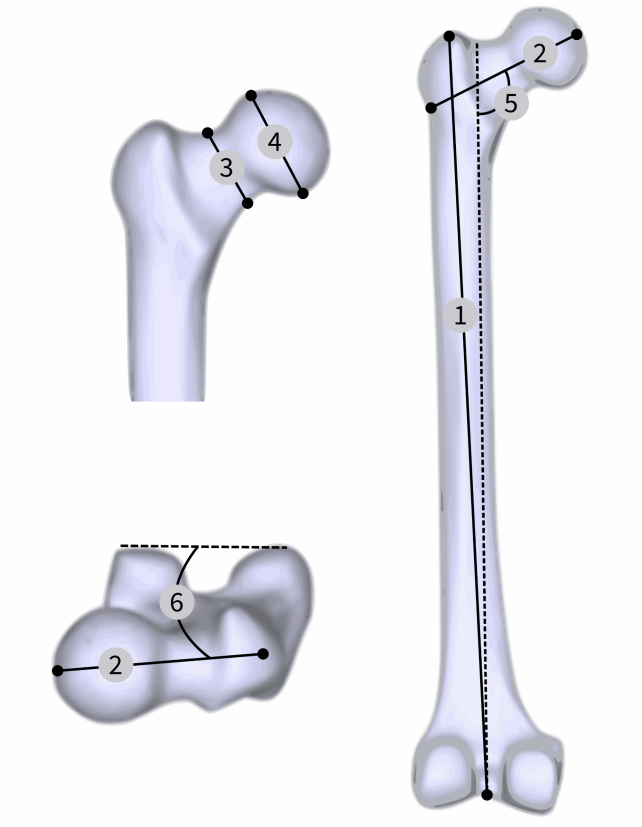


Figure 7. Exemplary multimodal MR images of a high-grade Glioma patient 0001 with truth and label result from an automatic segmentation of a BRATS participant (red=cranio spinal fluid, green=gray matter, blue=white matter, yellow=necrotic, turquoise=active tumor, pink=edema).

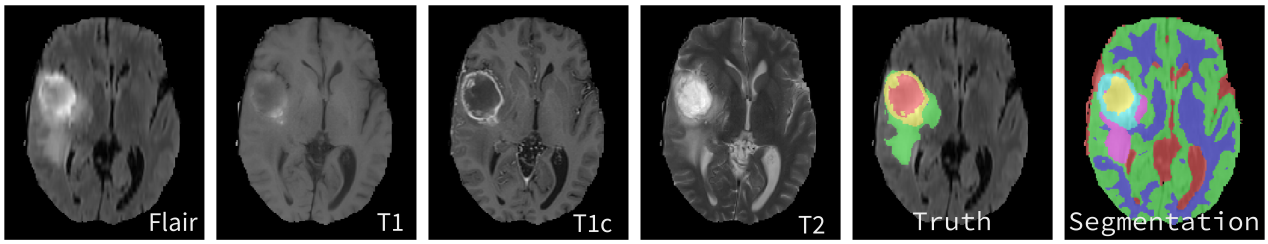


Table 1. Measured bone geometry and bone mechanical properties for men and women.

Parameter	Male mean (SD)	Female mean (SD)
Femur length, cm	42.6 (2.9)	40.8 (2.5)
Neck length, cm	10.5 (0.8)	9.9 (0.5)
Neck width, cm	3.9 (0.3)	3.6 (0.2)
Head width, cm	5.0 (0.5)	4.5 (0.5)
Neck-shaft angle, degree	128 (4)	130 (4)
Anteversion angle, degree	15 (7)	19 (7)
Young's modulus, GPa	2.7 (0.7)	3.4 (0.5)
Body weight, kg	64 (14)	60 (14)

Discussion

Principal Findings

The VSD features various innovative approaches to provide an efficient and easy-to-use open access repository to the medical image and modeling community focusing on the creation of statistical models. The standard image and segmentation file types (DICOM, MetaImage, Analyze, and Niftii) can be read and stored by the system. Additionally, the Statismo framework with a standard file type for statistical models has been included, which allows statistical models to be shared, hosted, and collaborated on via the VSD. The complete data used in the process of building a statistical shape model can be stored and tracked through the related data information in the VSD (see Multimedia Appendices 8 and 9). Built in as generic a way as possible, the VSD hosts a catalog of data objects, which can be individually organized by users according to their needs. The anatomical structures are annotated using a comprehensive ontology allowing the integrated search algorithm to identify anatomical structures within datasets, which is not possible with a standard index-based search.

Access to the VSD is open to research units and their members. Users are invited to contribute by uploading images and processed data, which will increase the data pool available on the VSD. By adding their data to the VSD, the users profit from the permanent accessibility provided through the Web service. This approach supports the current trend towards open access and re-usability of research results, which should be accompanied by the policy of scientific journals (see Multimedia Appendix 10). This facilitated access to data should enhance productivity and research quality through competition and stimulate improvements by the research community. The three example projects presented highlight the benefits of the VSD and its contribution to statistical model building, population-based research, and object identification in medical images.

International Collaboration on Statistical Shape Model of the Cochlea

With the VSD, it is possible to streamline the process of statistical shape model building for accurate population-based models in a multidisciplinary project setting. All parties have access to the same dataset and the metadata. The individual processing tasks can be distributed while the research data is stored on the VSD. An institution that has access to high resolution medical data can provide researchers with medical image analysis background with data for segmentation. Once these segmentations are available on the VSD, the statistical shape model can be engineered and provided through the VSD. The statistical model then can be implemented into surgical planning, segmentation algorithms, or implant design improvements. Since every processing step is stored on the system, this approach ensures a high level of reproducibility of the model building and quality control.

Biomechanical Femur Population Study

The VSD provides a service to overcome the limitation of the small number of available subjects in mechanical studies by promoting statistical shape model, hosting the respective images and processed data. This approach enables population-based studies. These population-based investigations allow not only study of the mechanical behavior of a large amount of bones, but also the comparison among different groups of populations. They allow study of the shape and volumes of anatomical structures and, consequently, clinically relevant parameters. As presented, for geometrical parameters of the femur, the neck-shaft angle plays a predominant role for fracture risk of both genders. Additionally, differences between men and women could be identified due to the use of a population-based model. Similar approaches could be used to perform virtual evaluation of orthopedic implants and optimize their design to market-specific needs.

Segmentation Challenge for Brain Tumors

Accurate segmentation is a critical step in visualization for treatment planning or later to monitor the therapeutic outcome. Manual segmentation from experienced experts is generally considered to contain the most reliable information. However, a lot of effort goes into the development of algorithms to automate this process. However, for almost every anatomical structure, those algorithms have to be adapted or trained. As a consequence, these techniques have to be constantly enhanced. To catalyze the innovation, researchers need image data and the opportunity to compare their results. The VSD provides the data repository and an evaluation platform, which are both needed to enhance those algorithms. The usefulness of such a platform in order to improve the sensitivity and specificity of the identification of glioblastoma is documented by the activity and registration numbers on the VSD. Due to the object-centric architecture of the system, the VSD is able to host any similar setup for other anatomical structures.

Limitations and Further Directions

Although already populated with an initial set of data, the collection still needs to grow significantly. Constant efforts are being made to increase the amount of data, which are also expected to increase with the dissemination activity and number of users. Additionally, the VSD is expanding the supported file type based on requests from users. Finally, an additional interface to exchange data with the system is planned based on the REST interface, which will allow for online preparation of statistical models.

To deal with the delicate aspect of data privacy, the VSD provides automatic tools such as de-identification, metadata stripping, and file name replacement. However, these tools cannot prevent the possible identification of a patient associated with objects visible in the medical images themselves. In some case, a special plate or implant, a rare malformation, or the image of the head is sufficient to recognize the subject. Therefore, the owner of the data is responsible for ensuring that they have the right to share the images and that a proper anonymization has been performed by removing identifiable regions of the image. However, the functionality on the VSD to give the data a two-level control—permission and license—allows the user to ensure that images are accessible only to individuals allowed to work with their data.

Conclusions

The VSD allows users to very flexibly manage and organize their data and online workspace. The VSD provides researchers with access to a growing community-filled data repository. The authors showed a novel system for scientific collaboration in the medical image community with a unique object-centric concept and semantically driven search option for anatomical structures for medical images, processed data, and statistical models. However, the clientele is not limited to these communities since the system can and will deliver functionality for other file types and their applications, for example, finite element models. We have presented three applications that benefit from such a Web application. The VSD is a valuable solution especially in the context of the growing needs of open access, re-usability, and reproducibility of scientific work.

Acknowledgments

The work was partially funded by Swiss National Science Foundation through the NCCR CoMe and by the European Commission within the EU-CHIC project. The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement n° 600841.

The authors want to thank Marcel Lüthi for the help with the Statismo framework and Stefan Bauer for technical support with the BRATS challenge.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Front page of the Virtual Skeleton Database (VSD).

[PNG File, 191KB - jmir_v15i11e245_app1.png]

Multimedia Appendix 2

List view of the full body dataset number 0001.

[PNG File, 163KB - jmir_v15i11e245_app2.png]

Multimedia Appendix 3

Expanded view of the full body dataset showing additional information and previews.

[PNG File, 185KB - jmir_v15i11e245_app3.png]

Multimedia Appendix 4

Forum start page.

[PNG File, 85KB - jmir_v15i11e245_app4.png]

Multimedia Appendix 5

Forum thread of an object showing the related object ratings and comments.

[PNG File, 91KB - jmir_v15i11e245_app5.png]

Multimedia Appendix 6

Result table of the BRATS 2012 challenge.

[PNG File, 116KB - jmir_v15i11e245_app6.png]

Multimedia Appendix 7

Raw results table of the 2012 BRATS challenge.

[CSV File, 106KB - jmir_v15i11e245_app7.csv]

Multimedia Appendix 8

Top part of the details page.

[PNG File, 109KB - jmir_v15i11e245_app8.png]

Multimedia Appendix 9

Bottom half of the details page.

[PNG File, 109KB - jmir_v15i11e245_app9.png]

Multimedia Appendix 10

Presentation at the Open Repository Conference 2013 in Charlottetown (CA).

[PDF File (Adobe PDF File), 3MB - jmir_v15i11e245_app10.pdf]

References

1. Cootes TF, Cooper DH, Taylor CJ, Graham J. A trainable method of parametric shape description. : BMVA Press; 1991 Presented at: British Machine Vision Conference; 1991; London Limited p. 8.1-8.8 URL: http://www.bmva.org/bmvc/1991/bmvc-91-008.html [doi: 10.5244/C.5.8]
2. Cootes T, Hill A, Taylor C, Haslam J. Use of active shape models for locating structures in medical images. Image and Vision Computing 1994 Jul;12(6):355-365. [doi: 10.1016/0262-8856(94)90060-4]
3. Hawkes DJ, Barratt D, Blackall JM, Chan C, Edwards PJ, Rhode K, et al. Tissue deformation and shape models in image-guided interventions: a discussion paper. Med Image Anal 2005 Apr;9(2):163-175. [doi: 10.1016/j.media.2004.11.007] [Medline: 15721231]
4. Kozic N, Weber S, Büchler P, Lutz C, Reimers N, González Ballester MA, et al. Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets. Med Image Anal 2010 Jun;14(3):265-275. [doi: 10.1016/j.media.2010.02.008] [Medline: 20359938]
5. Zheng G, Gollmer S, Schumann S, Dong X, Feilkas T, González Ballester MA. A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images. Med Image Anal 2009 Dec;13(6):883-899. [doi: 10.1016/j.media.2008.12.003] [Medline: 19162529]
6. Schumann S, Liu L, Tannast M, Bergmann M, Nolte LP, Zheng G. An integrated system for 3D hip joint reconstruction from 2D X-rays: a preliminary validation study. Ann Biomed Eng 2013 Oct;41(10):2077-2087. [doi: 10.1007/s10439-013-0822-6] [Medline: 23670657]
7. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. Med Image Anal 2009 Aug;13(4):543-563. [doi: 10.1016/j.media.2009.05.004] [Medline: 19525140]
8. Bauer S, Wiest R, Nolte LP, Reyes M. A survey of MRI-based medical image analysis for brain tumor studies. Phys Med Biol 2013 Jul 7;58(13):R97-129. [doi: 10.1088/0031-9155/58/13/R97] [Medline: 23743802]
9. Albrecht T, Luthi M, Vetter T. A statistical deformation prior for non-rigid imageshape registration. Anchorage: IEEE; 2008 Presented at: IEEE Conference on Computer VisionPattern Recognition; 2008; Anchorage p. 1-8. [doi: 10.1109/CVPR.2008.4587394]
10. Bryan R, Nair PB, Taylor M. Use of a statistical model of the whole femur in a large scale, multi-model study of femoral neck fracture risk. J Biomech 2009 Sep 18;42(13):2171-2176. [doi: 10.1016/j.jbiomech.2009.05.038] [Medline: 19646700]
11. Bryan R, Nair PB, Taylor M. Influence of femur size and morphology on load transfer in the resurfaced femoral head: A large scale, multi-subject finite element study. J Biomech 2012 Jul 26;45(11):1952-1958. [doi: 10.1016/j.jbiomech.2012.05.015] [Medline: 22704609]
12. Bou-Sleiman H, Ritacco LE, Nolte LP, Reyes M. Minimization of intra-operative shaping of orthopaedic fixation plates: a population-based design. Med Image Comput Comput Assist Interv 2011;14(Pt 2):409-416. [Medline: 21995055]
13. Liu BJ, Huang HK, Cao F, Zhou MZ, Zhang J, Mogel G. Informatics in radiology (infoRAD): A complete continuous-availability PACS archive server. Radiographics 2004;24(4):1203-1209 [FREE Full text] [doi: 10.1148/rg.244035051] [Medline: 15256640]
14. Hood MN, Scott H. Introduction to picture archive and communication systems. J Radiol Nurs 2006 Sep;25(3):69-74. [doi: 10.1016/j.jradnu.2006.06.003]
15. DSpace, an open-source solution for accessing, managing and preserving scholarly works. URL: http://www.dspace.org/ [accessed 2013-07-30] [WebCite Cache ID 6IUdfqB0p]
16. Staples T, Wayland R, Payette S. The Fedora Project. D-Lib Magazine 2003 Apr;9(4):-. [doi: 10.1045/april2003-staples]
17. EPrints. The original institutional repository solution. URL: http://www.eprints.org/us/ [accessed 2013-07-19] [WebCite Cache ID 6IE9pVcpm]
18. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics 2007 Jan;5(1):11-34. [Medline: 17426351]

19. Jomier J, Aylward SR, Marion C, Lee J, Styner M. A digital archiving systemdistributed server-side processing of large datasets. In: Proc. SPIE. 2009 Presented at: Medical Imaging 2009: Advanced PACS-based Imaging Informatics and Therapeutic Applications; 2009; Lake Buena Vista URL: http://dx.doi.org/10.1117/12.812227 [doi: 10.1117/12.812227]
20. Living human project. URL: http://www.livinghuman.org/ [accessed 2013-07-19] [WebCite Cache ID 6IE3fXdz7]
21. Internet Brain Segementation Repository. 2009. URL: http://www.cma.mgh.harvard.edu/ibsr/index.html [accessed 2013-10-24] [WebCite Cache ID 6KcBA2Cy5]
22. Bootstrap. 2013. Sleek, intuitive, and powerful front-end framework for faster and easier web development URL: http://getbootstrap.com/2.3.2/ [accessed 2013-07-22] [WebCite Cache ID 6IIlURr0W]
23. JQuery. The Write Less, Do More, JavaScript Library URL: http://jquery.com/ [accessed 2013-07-30] [WebCite Cache ID 6IUeDur35]
24. Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. J Biomed Inform 2003 Dec;36(6):478-500. [doi: 10.1016/j.jbi.2003.11.007] [Medline: 14759820]
25. Lüthi M, Albrecht T, Gass T, Goksel O, Kistler M, Bousleiman H, et al. Statismo - A framework for PCA based statistical models. The Insight Journal 2012;1:1-18 [FREE Full text]
26. Ballester Gonzalez MA. 2013. Hear-EU - High-resolution image-based computational inner ear modelling for surgical planning of cochlear implantation URL: http://www.hear-eu.eu/ [accessed 2013-07-19] [WebCite Cache ID 6IEA0rt1y]
27. Dalet D. D-Maps. URL: http://d-maps.com/carte.php?num_car=30074&lang=en [accessed 2013-07-22] [WebCite Cache ID 6IIldkzvT]
28. Bonaretti S. Statistical models of shape and density for population-based analysis of bone mechanics with applications to fracture risk assessment and implant. In: Thesis. Bern: University of Bern; 2012.
29. Seiler C, Pennec X, Ritacco L, Reyes M. Femur specific polyaffine model to regularize the log-domain demons registration. In: Proceedings of SPIE. 2011 Presented at: Medical Imaging 2011; 2011; Lake Buena Vista. [doi: 10.1117/12.878316]
30. Heller MO, Bergmann G, Kassi JP, Claes L, Haas NP, Duda GN. Determination of muscle loading at the hip joint for use in pre-clinical testing. J Biomech 2005 May;38(5):1155-1163. [doi: 10.1016/j.jbiomech.2004.05.022] [Medline: 15797596]
31. Speirs AD, Heller MO, Duda GN, Taylor WR. Physiologically based boundary conditions in finite element modelling. J Biomech 2007 Jan;40(10):2318-2323. [doi: 10.1016/j.jbiomech.2006.10.038] [Medline: 17166504]
32. Morgan EF, Bayraktar HH, Keaveny TM. Trabecular bone modulus-density relationships depend on anatomic site. J Biomech 2003 Jul;36(7):897-904. [Medline: 12757797]
33. Schileo E, Taddei F, Malandrino A, Cristofolini L, Viceconti M. Subject-specific finite element models can accurately predict strain levels in long bones. J Biomech 2007 Jan;40(13):2982-2989. [doi: 10.1016/j.jbiomech.2007.02.010] [Medline: 17434172]
34. Bergmann G, Graichen F, Rohlmann A, Westerhoff P, Heinlein B, Bender A, et al. Design and calibration of load sensing orthopaedic implants. J Biomech Eng 2008 Apr;130(2):021009. [doi: 10.1115/1.2898831] [Medline: 18412496]
35. Howden R. Gray's Anatomy. London: Chancellor Press; 1985.
36. Menze B, Jakab A, Bauer S, Reyes M, Prastawa M, van Leemput K. BRATS 2012. 2012. MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation URL: http://www2.imm.dtu.dk/projects/BRATS2012/ [accessed 2013-10-25] [WebCite Cache ID 6KcDf9s60]

Abbreviations

BRATS: Multimodal Brain Tumor Segmentation
CSV: comma separated values
CT: computed tomography
DICOM: digital imaging and communications in medicine
FMA: foundational model of anatomy
HDF5: Hierarchical Data Format
IIS: Microsoft Internet Information Service
ITK: Insight Toolkit
MICCAI: medical image computing and computer assisted intervention
MR: magnetic resonance
PACS: picture archiving and communication system
REST: representational state transfer interface
SPARQL: SPARQL protocol and RDF Query Language
SQL: Microsoft Structured Query Language
SSM: statistical shape model
VSD: Virtual Skeleton Database
VTK: Visual Toolkit
WebDAV: Web-Based Distributed Authoring and Versioning protocol
XNAT: Extensible Neuroimaging Archive Toolkit

Edited by G Eysenbach; submitted 30.08.13; peer-reviewed by RZ Lai, G Sakellaropoulos; comments to author 19.09.13; revised version received 02.10.13; accepted 08.10.13; published 12.11.13.

Please cite as:
Kistler M, Bonaretti S, Pfahrer M, Niklaus R, Büchler P
The Virtual Skeleton Database: An Open Access Repository for Biomedical Research and Collaboration
J Med Internet Res 2013;15(11):e245
URL: http://www.jmir.org/2013/11/e245/
doi:10.2196/jmir.2930
PMID:24220210

Original Paper

Evaluation of the Educational Value of YouTube Videos About Physical Examination of the Cardiovascular and Respiratory Systems

Samy A Azer[1*], MD, MEd, MPH, PhD; Hala A AlGrain[1*]; Rana A AlKhelaif[1*]; Sarah M AlEshaiwi[1*]

Curriculum Development and Research Unit, Department of Medical Education, King Saud University, Riyadh 11461, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Samy A Azer, MD, MEd, MPH, PhD
Curriculum Development and Research Unit
Department of Medical Education
King Saud University
PO Box 2925
Riyadh 11461,
Saudi Arabia
Phone: 966 114699178
Fax: 966 114699174
Email: azer2000@optusnet.com.au

Abstract

Background: A number of studies have evaluated the educational contents of videos on YouTube. However, little analysis has been done on videos about physical examination.

Objective: This study aimed to analyze YouTube videos about physical examination of the cardiovascular and respiratory systems. It was hypothesized that the educational standards of videos on YouTube would vary significantly.

Methods: During the period from November 2, 2011 to December 2, 2011, YouTube was searched by three assessors for videos covering the clinical examination of the cardiovascular and respiratory systems. For each video, the following information was collected: title, authors, duration, number of viewers, and total number of days on YouTube. Using criteria comprising content, technical authority, and pedagogy parameters, videos were rated independently by three assessors and grouped into educationally useful and non-useful videos.

Results: A total of 1920 videos were screened. Only relevant videos covering the examination of adults in the English language were identified (n=56). Of these, 20 were found to be relevant to cardiovascular examinations and 36 to respiratory examinations. Further analysis revealed that 9 provided useful information on cardiovascular examinations and 7 on respiratory examinations: scoring mean 14.9 (SD 0.33) and mean 15.0 (SD 0.00), respectively. The other videos, 11 covering cardiovascular and 29 on respiratory examinations, were not useful educationally, scoring mean 11.1 (SD 1.08) and mean 11.2 (SD 1.29), respectively. The differences between these two categories were significant (P<.001 for both body systems). The concordance between the assessors on applying the criteria was 0.89, with a kappa score >.86.

Conclusions: A small number of videos about physical examination of the cardiovascular and respiratory systems were identified as educationally useful; these videos can be used by medical students for independent learning and by clinical teachers as learning resources. The scoring system utilized by this study is simple, easy to apply, and could be used by other researchers on similar topics.

(J Med Internet Res 2013;15(11):e241) doi:10.2196/jmir.2728

KEYWORDS

YouTube; learning resources; medical education; cardiovascular system physical examination; respiratory system physical examination; medical curriculum; clinical skills; self-directed learning; competency

Introduction

Current medical curricula are paying increasing attention to learning how to conduct physical examinations and to early exposure of students to clinical skills. The aim is to highlight physical examination as a core competency for students and to ensure that the content and delivery of this component in the curriculum has been mastered before graduation [1,2]. Learning about physical examination of the cardiovascular and respiratory systems, as is the case with other elements of clinical skills, cannot be achieved from reading textbooks or attending lectures. Such skills are usually developed through observation of clinicians performing examinations and through practice on simulated and real patients [3-5].

For many years, students relied on their clinical teachers as one of the main sources for learning such skills [3,4]. However, with the introduction of problem-based learning (PBL) and self-directed learning into most medical curricula, more emphasis has been placed on changing the learning and teaching pedagogy [5-7]. In these courses, students use a range of learning resources, including review papers, journal articles, textbooks, museum specimens, simulated patients, computer-aided learning programs, and multimedia [8-10]. The Internet has become an easily available resource of up-to-date information worldwide [9,11-15]. Medical students usually rely on Google and YouTube as the first resources in their research [12-14].

YouTube is the largest Internet video-sharing site and is a useful tool in social communication, business, advertising, and news as well as a promising learning resource for students and the general public [16,17]. In 2005, YouTube was created as an arena for personal/social communication and for distribution of commercial content. Although similar video-sharing sites are available for public use, YouTube has become the most popular worldwide. Statistical data offered by YouTube provide evidence of its popularity [17]. For example, YouTube is the largest video site with over 4 billion videos watched around the globe every day and one hour of video is uploaded to YouTube every second [17].

Recently, the quality of YouTube videos has been evaluated in a number of areas related to medical and patient health information and medical skills [18], including first aid information on thermal burns [19], human papilloma virus vaccinations [20], investigation into the mechanisms of elbow dislocation [21], clinical procedures [22], rheumatoid arthritis [23], as a learning resource for nurses [24], surface anatomy [25], cardiopulmonary resuscitation [26], dental education [27], and infantile spasms [28]. However, not all researchers found YouTube videos to be educationally useful videos; on the contrary, a number of researchers warned that there are thousands of videos on YouTube that promote misleading information and could possibly endanger some viewers. For example, there are videos encouraging anorexia as a healthy life-style choice [29] and non-suicidal self-injury videos that may foster normalization of non-suicidal self-injury and potentially reinforce the behavior through regular viewing [30,31].

With these limitations in mind, there is no doubt that educationally designed videos have the advantage of explaining difficult concepts through using simulation, graphic diagrams, dynamic illustrations, analogies, and simulated patients. The learning/teaching benefits of videos will be enhanced if videos are well-designed, explore scientifically correct content, feature a clear presentation, and address students' learning needs. Searching for educationally useful videos on YouTube may be time consuming and requires knowledge from researchers about what makes an educationally useful video [32-35].

No studies assessing the usefulness of YouTube videos on examination of the cardiovascular system (CVS) and the respiratory system (RS) could be found in PubMed. This study is aimed at assessing YouTube videos regarding physical examination of the cardiovascular and respiratory systems.

Methods

Searching YouTube

From November 2, 2011 to December 2, 2011, the YouTube site was searched using the following key words: "cardiovascular system examination", "cardiovascular clinical examination", "cardiovascular physical examination", "respiratory system examination", "respiratory clinical examination", and "respiratory physical examination". When searching YouTube, quotation marks were used with these terms to specify that these terms must be present. Only the first 10 pages of results for each search term were screened for related videos. The reason for not searching videos beyond page 10 of the search term results is that one is less likely to find videos covering the key words and users are less likely to go that far in their search [26]. We observed that videos not related to search words could be seen by the third and the fourth pages. By pages 8 to 10, videos not related to the search were found. Therefore, it was decided to go only as far as the first 10 pages for each search. Videos in the English language only were identified and the related URL was recorded.

The search was conducted by three authors (HAA, RAA, and SMA) independently using the search key words. The search results were evaluated and used to compile a common pool that was used in further analysis. The inclusion criteria were videos covering clinical examination of the cardiovascular or respiratory systems in adults. Videos were excluded if they were (1) not in the English language, (2) in the form of a lecture, (3) an advertisement or news, (4) discussing signs or symptoms of diseases affecting the cardiovascular or respiratory systems, (5) about simulated patients reflecting on their experiences or roles, (6) about patients with cardiovascular or respiratory disorders, (7) about drugs used in treating patients with cardiovascular or respiratory disorders, or (8) in the form of seminars or reviews of cardiovascular or respiratory systems. Duplicated videos were also excluded and repeats were treated as a single file for analysis. The repeat file with the greatest number of hits was used for analysis.

Collected Data

For each video, the following data were collected: title, duration of the video, number of days on YouTube, total number of

viewers, and name of uploader/creator (organization, group of people, one person). The number of "likes" and "dislikes" (a crude scoring system that viewers can use to assign to each video they watched, though it is not necessarily used by all viewers) for each video was recorded. Because the number of days on YouTube varies widely among videos, we decided to calculate viewership/day as a more accurate parameter compared to total number of viewers. The viewership per day is the ratio of number of viewers to the number of days a video is on YouTube. The number of days is calculated from the day of uploading on YouTube up to December 1, 2011. This calculation of viewership/day was conducted for each video. The page number on which a video was placed was recorded. This is because the YouTube search algorithm is designed to show videos according to their relevance to key words used in the search and hence videos on the first three pages are more likely to be watched than videos on lower pages. Thus, the location of the videos may indirectly affect the number of "likes" and "dislikes" given by viewers.

Analysis of Videos

The criteria used in evaluating the videos have been described in detail in an earlier work [25,35] with some modification to suit the study. In summary, the design of the criteria is based on four main domains: video content, technical aspects, authority/creator, and pedagogy used. The items in the criteria are grouped under two categories: major and minor. The major criteria comprise: (1) the video uses simulated patients or patients to demonstrate the examination, (2) contents about clinical examination are scientifically correct, (3) images are clear, (4) the creator and/or organization providing the video are mentioned, and (5) sounds are clear and background is free from noise. The minor criteria comprise: (1) the video covers the topic identified in the title, (2) designed at the level of undergraduate medical science students, (3) time to download is reasonable (about 10-15 minutes at the maximum, not interrupted or challenging to download as reported by the three evaluators), (4) information about the creator is up-to-date, (5) the educational objectives are stated, and (6) the topic is clearly presented. The criteria were used to categorize videos into educationally useful and non-educationally useful videos. We mean by "educationally useful" that a video provides scientifically correct and up-to-date knowledge and clinical instructions/skills accepted by educators in other teaching hospitals about cardiovascular and respiratory examinations. As per the basis of the evaluation criteria, educationally useful videos should fulfill the four domains (scientific content, technical aspects, authority/creator, and pedagogy used). Two scores were allocated for each item in the major criteria and one score was allocated to each item under the minor criteria. If an item was fulfilled, an allocated score was given; if not fulfilled, a zero was given. No half scores were used. As per our previous research work, educationally useful videos should fulfill all major criteria items as the minimum requirements plus at least three items under the minor criteria [35].

Testing the Criteria

To standardize the evaluation of the content of each video and the process of clinical examination, the assessors used the textbook and video by Talley and O'Connor, "Clinical Examination", as a reference to guide their assessment [36]. The content element in the criteria comprised the following: examiner introduces him/herself to the patient; patient is correctly positioned in bed; examination is conducted within time frame; examination covers the sequence of inspection, palpation, percussion, and auscultation; physical signs are correctly elicited; and examination is conducted in a professional manner. Prior to applying the criteria, we piloted its use. A total of 25 videos were randomly selected and used for this purpose. The criteria were applied independently by three assessors (SMA, HAA, and RAA). None of the assessors shared their findings or discussed the outcome of their evaluation. An Excel spreadsheet covering the three evaluations was examined by a fourth researcher (SAA). The agreement among the assessors was in the range of 96-98%. The findings were discussed among the researchers. The criteria were tested again independently by three assessors for another 25 videos. Videos were then rated independently by three assessors (SMA, HAA, and RAA). When videos were difficult to classify or when there was a disagreement among assessors, all researchers reviewed such videos in a meeting and reached a final agreement. This study was approved by the ethics committee at King Saud University College of Medicine.

Data Analysis

The data were entered into Microsoft Excel 2010 and were checked before conducting any analysis. Analysis was conducted with SPSS software (version 18.0 for MS Windows) and was reported via mean, SD, percentage, and minimum and maximum; *t* tests and ANOVA (analysis of variance) were performed to determine significant differences. To assess the degree to which different judges or raters agreed in their assessment decisions, Cohen's kappa for inter-rater reliability was used [37]. Pearson's correlation coefficient (*r*) was calculated to determine if the viewership per day was correlated to the total scores given to each video. This relationship has also been examined to see if there is a correlation between the number of "likes" given by viewers to a video and the total scores given [38,39]. For all calculations, a P value <.05 was considered significant.

Results

Videos on CVS and RS Examinations

A total of 1920 YouTube videos were found on initial search and, on applying the inclusion criteria and visual examination of the videos, only 56 videos were relevant to clinical examination of the CVS (20 videos) and the RS (36 videos) (Table 1). Figure 1 summarizes how the YouTube searches and the number of videos in each category were refined on the basis of the inclusion and exclusion criteria. Examples of screenshots of these videos are shown in Figures 2-4.

Table 1. YouTube videos covering examination of the cardiovascular and respiratory systems (N=56).

Body system	Number of videos	Duration	Total number of days on YouTube	Total viewership	Total scores
	n (%)	minutes (seconds)	n (mean, minimum, maximum)	n (%)	mean (SD)
Cardiovascular system					
Educationally useful	9 (45)	71 (18)	3571 (3968; 20; 1020)	129,350 (29)	14.9 (0.33)
Educationally not useful	11 (55)	73 (30)	5773 (5248; 154; 955)	315,420 (71)	11.1 (1.08)
P value					<.001
Respiratory system					
Educationally useful	7 (19)	56 (48)	5515 (7879; 219; 1673)	413,483 (41)	15.0 (0.00)
Educationally not useful	29 (81)	147 (31)	20592 (7101; 31; 1381)	600,183 (59)	11.2 (1.29)
P value					<.001

Figure 1. Searching YouTube for videos covering examination of the cardiovascular and respiratory systems.

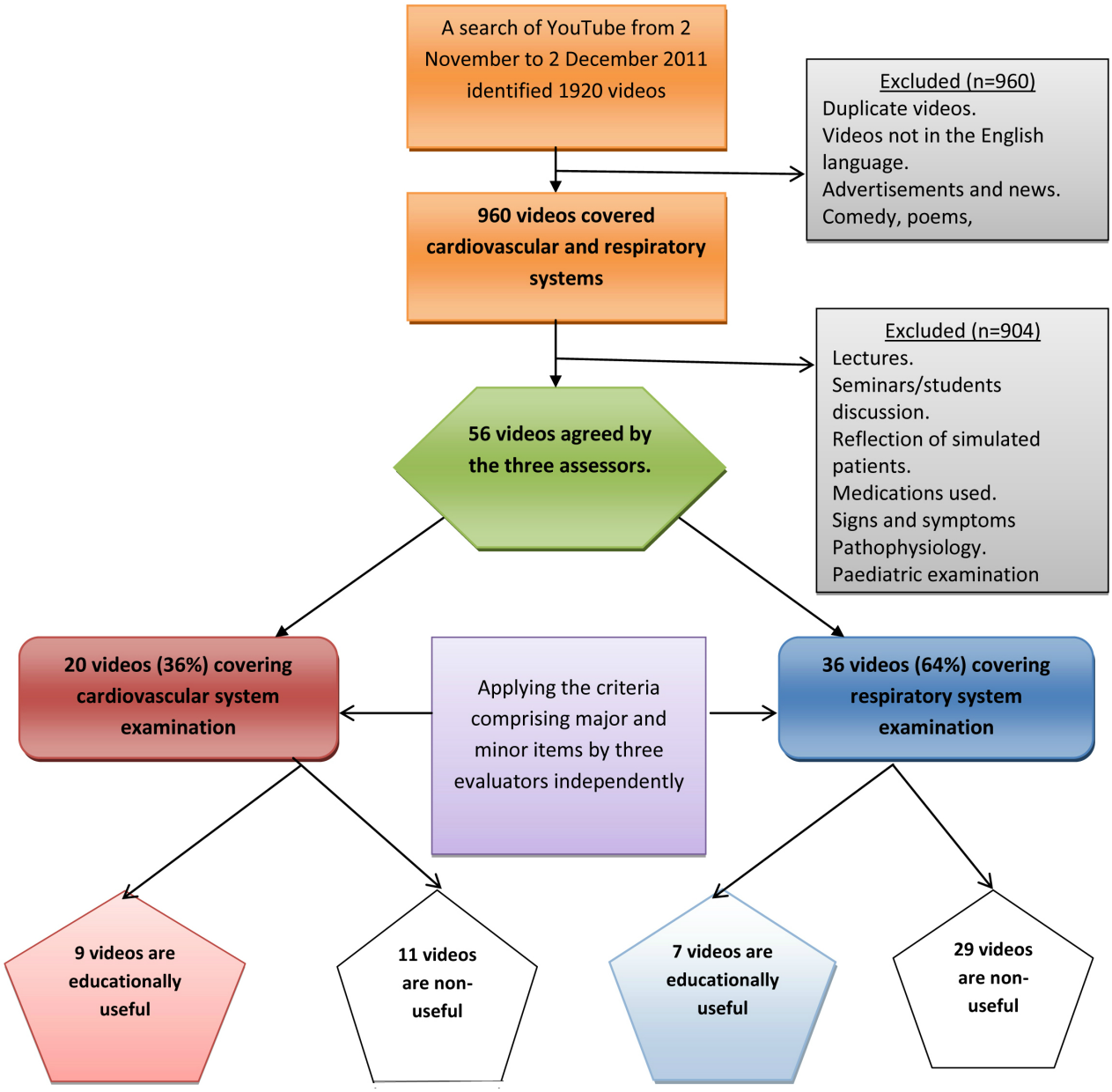


Figure 2. Screenshot of video.

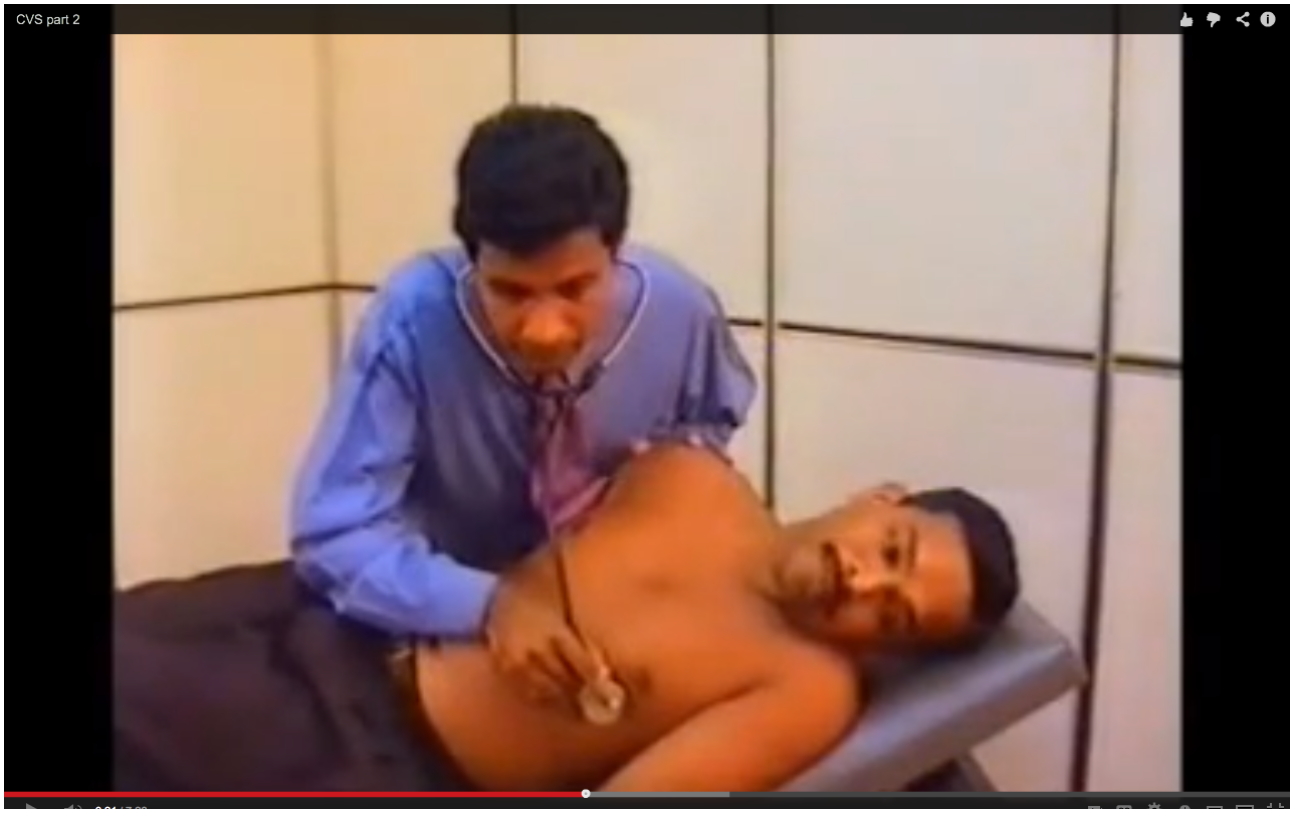


Figure 3. Screenshot of video.

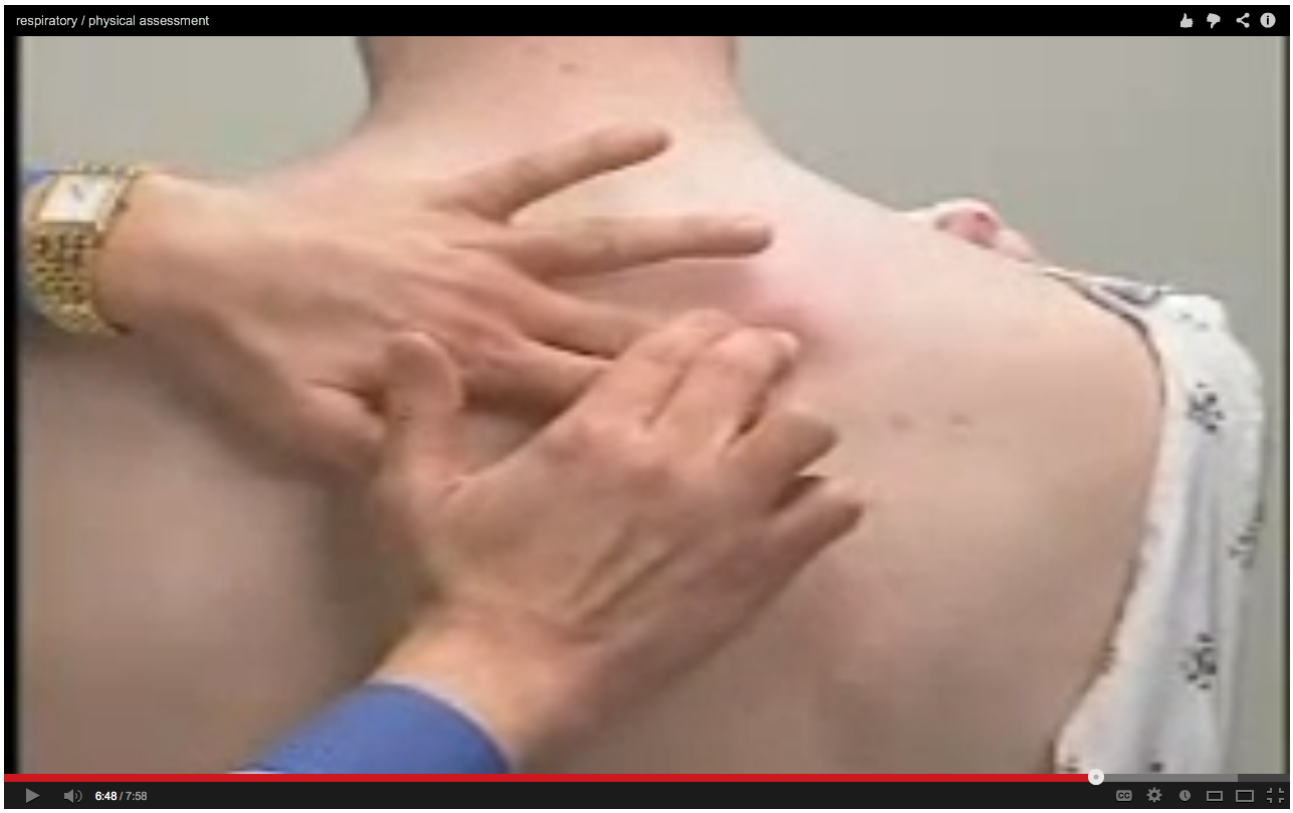
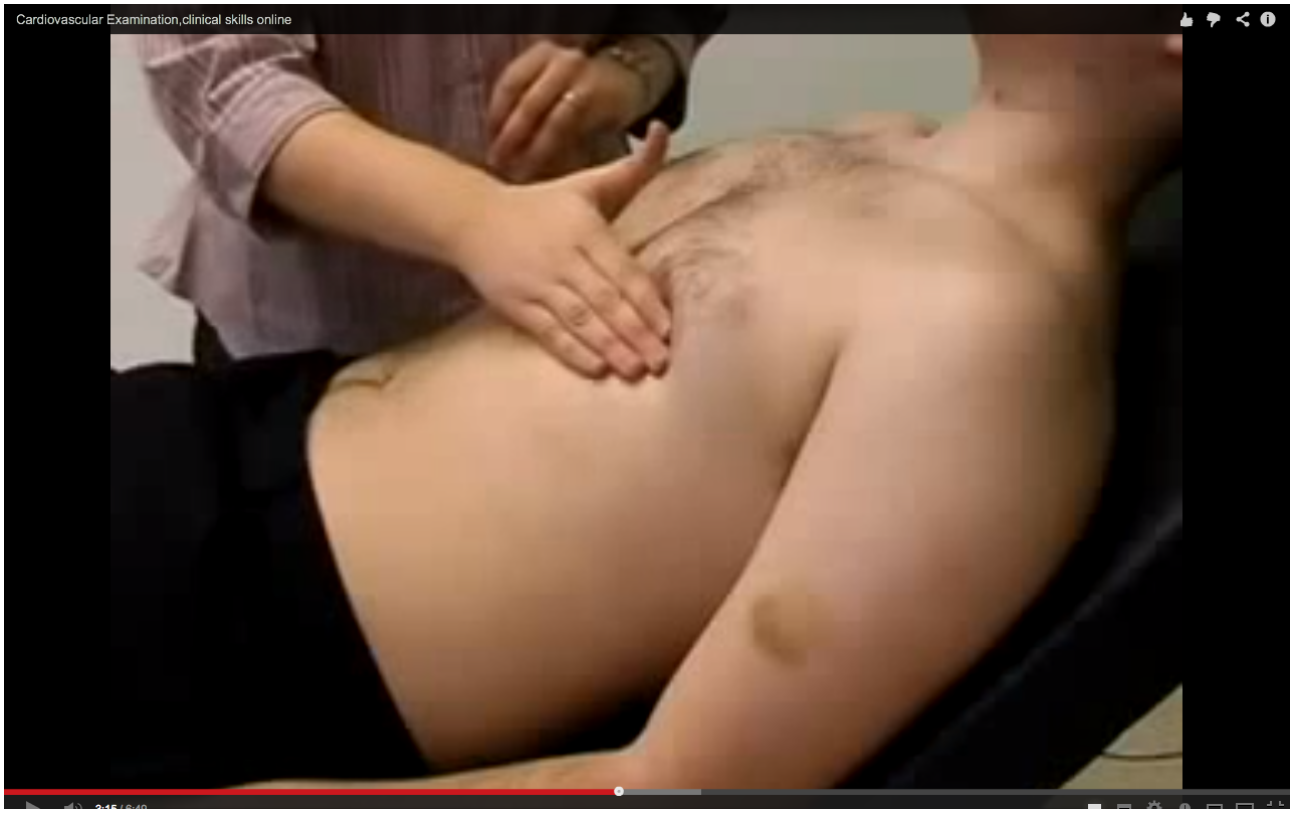


Figure 4. Screenshot of video.



Analysis of Videos

The total duration of these video clips was 349 minutes, 11 seconds. The use of the criteria for categorizing the videos into useful and non-useful videos revealed that there were 9 CVS and 7 RS videos that provided useful information on clinical examination. These videos scored mean 14.9 (SD 0.33) and mean 15.0 (SD 0.00), respectively. The total duration of useful videos was 128 minutes, 6 seconds (71 minutes, 18 seconds for the CVS videos and 56 minutes, 48 seconds for the RS videos). The total viewers of all videos were 1,458,436. For both body systems, useful videos were viewed by 542,833 viewers (37.22% of total viewers, 542,833/1,458,436). Table 2 reveals detailed information about the 56 videos included in the study and the multimedia links show examples of useful videos about examination of the cardiovascular system (Multimedia Appendix 1) and the respiratory system (Multimedia Appendix 2).

Useful videos were created by doctors or professional bodies; a link to an organization such as ImedrxTV, geekymedics123, MedicalStudentApps, mosaicism105, or the name of the creator and his or her credentials, such as Professor S. Femando from the Open University of Sri Lanka, were shown. Some videos were linked to universities and known teaching institutes such as Manchester Medical School, St. George's University of London, Central Manchester University Hospitals, the National Health Service, and the University of Wisconsin School of Medicine and Public Health.

Non-educationally useful videos scored mean 11.1 (SD 1.08) for CVS examination and mean 11.2 (SD 1.29) for the RS examination. Compared to educationally useful videos, the differences were significant (P<.001 for both CVS and RS videos) (Table 1). For both body systems, non-useful videos were viewed by 915,603 (62.78% of total viewers, 915,603/1,458,436). Non-educationally useful videos failed due to a number of reasons. The majority of the non-educationally useful videos failed to fulfill one of the major criterion items. Among those, more than 80% (33/40) were due to the image lacking clarity or no mention of the creator of the video. The concordance between the assessors on applying the criteria was 0.89, with a kappa score >.86.

The correlations between the total video scores and the number of viewers/day were not significant for the CVS videos (r=.06, P=.059) and the RS videos (r=.027, P=.827). Similarly, no correlation was found between likes or dislikes and the total number of scores given to CVS videos (r=.06, P=.059) or RS videos (r=.02, P=.832). No significant correlation was found between the video location on the first pages and the scores for like or dislike.

Table 2. Details about the 56 videos identified in the study.

No.	Title	URL	Days on YouTube	Viewership/day	Total score
1[a]	Cardiovascular system	http://www.youtube.com/watch?v=v12EAuyxs6w	721	27.56	15
2	Macleod's Clinical Examination, Cardiovascular System	http://www.youtube.com/watch?v=8h8h7ee9D84	219	21.84	15
3	Cardiovascular Examination, clinical skills online	http://www.youtube.com/watch?v=E6EOCY-OszE	894	37.68	15
4	Cardiovascular Examination - OSCE guide	http://www.youtube.com/watch?v=SJ3UwKkLyy0	266	69.80	15
5	Cardiovascular Examination	http://www.youtube.com/watch?v=XiM2fn-VDg9A	367	111.06	15
6	Cardiovascular Exam	http://www.youtube.com/watch?v=69LuXu42Ahc	1020	9.44	14
7	Cardiovascular Examination	http://www.youtube.com/watch?v=iiPvQcyEfEY	44	43.75	15
8	CVS Examination Part 1	http://www.youtube.com/watch?v=CwM8viJ4M6Q&feature=mfu_in_order&list=UL	20	4.05	15
9	CVS Examination Part 2	http://www.youtube.com/watch?v=MQLJv-1qP_s	20	2.85	15
10[b]	Respiratory System	http://www.youtube.com/watch?v=vD7b-Ery014	721	9.48	15
11	Respiratory System Examination	http://www.youtube.com/watch?v=YK82HlDrvWw	665	18.59	15
12	Respiratory Examination	http://www.youtube.com/watch?v=hWGzi5h2UR8	1673	172.05	15
13	Respiratory / physical assessment	http://www.youtube.com/watch?v=IepL5u5lAtE	1183	55.09	15
14	Macleod's Clinical Examination, Examination of the Respiratory System	http://www.youtube.com/watch?v=akr40RXu_H8	219	36.27	15
15	04a.Physical Exam –Thorax - part 1/2	http://www.youtube.com/watch?v=axjRJtXc-Xc	527	31.54	15
16	04a.Physical Exam –Thorax - part 2/2	http://www.youtube.com/watch?v=j0lKzPyM7_k	527	31.73	15
17[c]	Physical Assessment of Thorax Lungs and Cardiovascular	http://www.youtube.com/watch?v=cKMRO-jEaowc	405	12.01	10
18	CVS examination.wmv	http://www.youtube.com/watch?v=C7c-K6Jltaw	294	3.31	12
19	Cardiovascular Examination	http://www.youtube.com/watch?v=hXU24g95wJU	448	183.23	11
20	Cardiovascular Physical.mpg	http://www.youtube.com/watch?v=McOA-JxQWb5A	717	1.91	10
21	The One Minute Cardiac Examination	http://www.youtube.com/watch?v=EwL43Oww-PYY	569	18.631	13
22	Cardiovascular Examination	http://www.youtube.com/watch?v=BYZJN3Zlllg	666	58.89	10
23	Cardiovascular Examination	http://www.youtube.com/watch?v=-8Hi1PjZam4	903	79.76	12
24	Jugular Venous Pressure (JVP)	http://www.youtube.com/watch?v=4YBX-aWWG3Ns	431	55.14	12
25	Inspection Of Chest	http://www.youtube.com/watch?v=SkWrl7t2CE8	154	2.01	12
26	Examination: General Cardiovascular	http://www.youtube.com/watch?v=XR-Ja7Vina10	231	0.94	11
27	Cardiovascular examination	http://www.youtube.com/watch?v=U-fyQfjL9aQ	955	83.75	10
28[d]	Examination of Respiratory System Part 1 General Exam	http://www.youtube.com/watch?v=4xPxX94XANo	81	6.15	11
29	04b.Anterior and posterior Thorax and Axilla	http://www.youtube.com/watch?v=_xt3nXnEo-qY	528	214.60	12

No.	Title	URL	Days on YouTube	Viewership/day	Total score
30	Respiratory Physical.mpg	http://www.youtube.com/watch?v=hkxH0fy-cRVE	717	6.79	13
31	Clinical Skills Session – Resp. Exam	http://www.youtube.com/watch?v=uRg7zg1wCec	1381	154.79	11
32	9-Thorax and Lungs-Examining the Anterior Thorax and Lungs	http://www.youtube.com/watch?v=bYpJ0DYtxnE	162	2.03	11
33	Respiratory Examination	http://www.youtube.com/watch?v=a-4BUharW-MA	666	51.55	12
34	HOW Examination of Respiratory System Part 1 of 2	http://www.youtube.com/watch?v=OX1Y21K41rs	67	1.73	10
35	HOW Examination of Respiratory System Part 2 of 2	http://www.youtube.com/watch?v=ol44OFQMHz0	67	1.97	10
36	Medical Gallery - Loyola Full Thorax Exam Part 1	http://www.youtube.com/watch?v=hTagtppnCIk	468	1.82	12
37	Medical Gallery - Loyola Full Thorax Exam Part 2	http://www.youtube.com/watch?v=TvLX0EP7fc0	468	0.30	11
38	Deb Video 6 - Respiratory Exam.mpg	http://www.youtube.com/watch?v=ud6rV3kVd-jI&feature=mfu_in_order&list=UL	722	10.77	10
39	Deb Video 7 - Respiratory Exam.mpg	http://www.youtube.com/watch?v=WLt5kY15BBE	722	16.47	8
40	Respiratory Mohamed	http://www.youtube.com/watch?v=VBTrAz_Lw-JE	397	0.78	11
41	[Respiratory Mohamed] Part 2	http://www.youtube.com/watch?v=Qc-TAsQew8Cc	397	0.75	11
42	[Respiratory Mohamed] Part 3	http://www.youtube.com/watch?v=6WtQqf-Ss9D0	31	25.42	11
43	[Respiratory Mohamed] Part 4	http://www.youtube.com/watch?v=M2bmOrL_ouI	31	9.19	11
44	[Respiratory Mohamed] Part 5	http://www.youtube.com/watch?v=PY8pA0cB2ko	31	9.29	11
45	Examination of the thoracic and respiratory system 1	http://www.youtube.com/watch?v=3emIM6BsWdw	1295	18.25	13
46	Examination of the thoracic and respiratory system 2	http://www.youtube.com/watch?v=vez8Vzm70kc	1295	12.84	13
47	Examination of the thoracic and respiratory system 3	http://www.youtube.com/watch?v=ygP03M0yI-wE	1295	27.10	13
48	Examination of the thoracic and respiratory system 4	http://www.youtube.com/watch?v=ABuAMGHrB-VM	1295	12.95	11
49	Examination of the thoracic and respiratory system 5	http://www.youtube.com/watch?v=ifnqBzmid-DY	1295	7.85	11
50	Examination of the thoracic and respiratory system 6	http://www.youtube.com/watch?v=Us5M1gCxBC8	1295	4.96	11
51	Examination of the thoracic and respiratory system 7	http://www.youtube.com/watch?v=S9ew7uZip-bA	1295	7.78	11
52	Examination of the thoracic and respiratory system 8	http://www.youtube.com/watch?v=saWuYMkXQGM	1295	43.14	11
53	Examination of the thoracic and respiratory system 9	http://www.youtube.com/watch?v=GNcmP-cLZHi4	1295	7.48	11
54	Examination of the thoracic and respiratory system 10	http://www.youtube.com/watch?v=TkA0CqLuYfM	1295	7.70	11
55	Respiratory Examination (part 1) for OSCEs	http://www.youtube.com/watch?v=Odkd-JZ_ppOI	353	24.81	10

No.	Title	URL	Days on YouTube	Viewership/day	Total score
56	Respiratory Examination (part 2) for OSCEs	http://www.youtube.com/watch?v=OY-DeI2HEyZw&feature=related	353	20.20	10

[a]From 1 to 9 – educationally useful videos on cardiovascular examination (n=9)

[b]From 10 to 16 – educationally useful videos on respiratory examination (n=7)

[c]From 17 to 27 – non-educationally useful videos on cardiovascular examination (n=11)

[d]From 28 to 56 – non-educationally useful videos on respiratory examination (n=29)

Discussion

Principal Findings

The aim of this study was to conduct an analysis of YouTube videos about CVS and RS physical examinations and to categorize these videos for educational purposes. Several methods have been described in evaluating videos [40-43]. However, none of these systems were useful for classifying videos about clinical skills. The system used in this study is simple, easy to apply, and covers four key elements, namely: scientific content, technicality, authority, and pedagogy parameters. It has also been tested in two previous studies and has shown a high level of inter-rater correlation while covering the key elements required for an educationally useful video [25,35].

The findings in this study indicated that there are 9 educationally useful videos on CVS examination and 7 on RS examination. These videos provided approximately 71 minutes, 18 seconds of CVS examination and 56 minutes, 48 seconds of RS examination that can be used in clinical learning and teaching purposes. Useful videos were linked mainly to universities or educational institutes. This indicates the involvement of universities and teaching institutes in promoting the use of educational videos as a resource to learners. This is particularly important with the shift of most universities toward self-directed learning and student-centered programs.

YouTube's search engine, despite all precautions taken to target the search results to videos about examination of these two body systems, delivered more than 1800 videos that were not related to the search terms. Furthermore, the lack of correlation between the total score given to a video and the location of the video when searched (the page number) highlights the fact that the YouTube search algorithm is not well calibrated and unrelated videos usually appear despite the search filter provided by YouTube. Also, the lack of correlation between the total score given to a video to the "like" and "dislike" numbers as well as to viewership/day suggests the possibility that many viewers were watching substandard videos in their learning. Assuming that the majority of these viewers were medical/health students and trainees, these results highlight the need for YouTube to improve its search algorithm system to generate more accurate lists of videos that match with the search terms.

Much emphasis has been placed on physical examination in current medical curricula. To learn such skills, students usually rely on observing clinicians conducting examinations and then practicing on simulated patients and real patients. Recently, Duvivier et al (2012) in a study from Maastricht University reported that on average students devote 20% of their self-study time to skills training on physical examination [3]. They found that students use textbooks, examination guidelines, scientific articles, the Internet, videos/DVD, and scoring forms from previous OSCEs as their learning resources. Although the Internet and videos were among these resources, the study did not explore the sources of these videos and whether YouTube videos were used. Also, the study did not specify how much time they spend on average on resources such as videos and the Internet [3].

There is no substitute for witnessing a physical examination or a clinical procedure being performed live. Neither static images nor a description in a book outlining these techniques can offer the same impact as personal experience. Videos, however, are a medium that can transfer the experience and help in mastering such skills through repeated watching of techniques used and information provided. Online videos have become a routine and important tool in a student's preparation for clinical skills. Attending clinical skills sessions, reading textbooks along with watching online videos, and practicing skills learned on patients have become important learning strategies in most clinical schools [44,45]. Based on cognitive psychological research, the use of videos will help expose students to the techniques of clinical examination, approaches for examining patients, and how to manage the sequence of technical steps in such examinations [45-50].

Considering the increasing number of learners using the Internet as their primary source of information, medical educators and clinical teachers should recognize the importance of YouTube in education and invest in using Web 2.0 in learning and teaching activities such as clinical teaching [8,13-14]. Although there are other links on the Web that provide free videos, we decided to examine the videos on YouTube for a number of reasons. First, YouTube is popular and usually preferred by users compared to other websites. Second, it is relatively easy to share videos on YouTube. Third, there is continuous improvement in the design of the YouTube site and each video is accompanied by useful data that reflect the evolution of online social networking and can be of use to researchers and viewers. Finally, YouTube has succeeded in providing social networking and enabling discussion among viewers, which could be useful to viewers as they share useful videos with each other.

This study shows that clinical teachers who are competent in clinical teaching and clinical educators, such as those who created the videos we identified in this study, can offer great service to medical and health students worldwide by placing their work on sharing websites such as YouTube. This can be part of knowledge transfer and scholarly work as it is created

by academics, shared, and peer-reviewed [50]. Recently, YouTube launched YouTube EDU, an area of YouTube where video creators must possess high-quality credentials and provide evidence of significant mass of resources before they qualify to have their content included in YouTube EDU. This may have a major impact on improving successful search strategies particularly for YouTube videos designed for educational purposes.

Limitations

This study represents a snapshot of available resources during November/December 2011, and, since then, there may have been more videos uploaded and made available. Given the continuous upload of videos on YouTube and the volume of new videos added to the system on a daily basis, further studies are needed to assess whether new videos of high quality and coverage have been added.

The small number of videos used in this study is a limitation to this study. The study was limited to videos in the English language and only those covering examination of adults, which may have contributed to the smaller number of videos found. However, the majority of videos uploaded on YouTube are in the English language. Also, there is the possibility that some videos were not labeled as such and thus were unidentifiable under our search terms, despite all our precautions including conducting the search of YouTube over 30 days independently by three researchers and using six search keywords to compile a common pool. The search in this study was limited to YouTube and there is the possibility that videos on other websites, such as those of medical and other health professional societies and medical journals, were not included.

Conclusions

This is possibly the first study assessing the educational value of YouTube videos about physical examination of the CVS and RS systems. Despite the small number of videos identified and found educationally useful, these videos can be used by medical students for independent learning and by clinical teachers as learning resources. The scoring system utilized by this study is simple, easy to apply, and could be used by other researchers on similar topics. The authors encourage other researchers to assess the tool and contribute to its improvement.

Acknowledgments

This work was funded by the College of Medicine Research Center, Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia.

The authors would like to thank Diana Azer for reviewing and editing proofs of the paper. Samy A Azer is a professor of the Department of Medical Education and Chair of the Curriculum Development and Research Unit at King Saud University, Riyadh, Saudi Arabia.

Part of this study was presented as a poster at the Association for Medical Education in Europe (AMEE) conference, incorporating the 4th SIFEM conference, which was held from August 25 to 29 2012, Lyon, France.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Useful cardiovascular video.

[MPG File, 24MB - jmir_v15i11e241_app1.mpg]

Multimedia Appendix 2

Useful respiratory video.

[MPG File, 26MB - jmir_v15i11e241_app2.mpg]

References

1. Anderson RC, Fagan MJ, Sebastian J. Teaching students the art and science of physical diagnosis. Am J Med 2001 Apr 1;110(5):419-423. [Medline: 11286962]
2. GME (2003). London; General Medical Council. Tomorrow's doctors URL: http://www.gmc-uk.org/education/undergraduate/tomorrows_doctors.asp [accessed 2013-10-17] [WebCite Cache ID 6KRGp5Qdq]
3. Duvivier RJ, van Geel K, van Dalen J, Scherpbier AJ, van der Vleuten CP. Learning physical examination skills outside timetabled training sessions: what happens and why? Adv Health Sci Educ Theory Pract 2012 Aug;17(3):339-355 [FREE Full text] [doi: 10.1007/s10459-011-9312-5] [Medline: 21710301]
4. Bhoopatkar H, Weam A. Medical students describe their patterns of practising clinical examination skills outside time tabled sessions. Med Teach 2008;30(3):334. [doi: 10.1080/01421590701753567] [Medline: 18484465]

5. Holzinger A, Kickmeier-Rust MD, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education 2009 Feb;52(2):292-301. [doi: 10.1016/j.compedu.2008.08.008]
6. Azer SA. Navigating problem-based learning. Marrickville, N.S.W: Elsevier Australia; 2008.
7. Hessinger M, Holzinger A, Leitner D, Wassertheurer S. Hemodynamic models for education in physiology. Mathematics and Computers in Simulation 2008 Dec;79(4):1039-1047. [doi: 10.1016/j.matcom.2008.02.015]
8. Lemley T, Burnham JF. Web 2.0 tools in medical and nursing school curricula. J Med Libr Assoc 2009 Jan;97(1):50-52 [FREE Full text] [doi: 10.3163/1536-5050.97.1.010] [Medline: 19159003]
9. Azer SA, Guerrero AP, Walsh A. Enhancing learning approaches: practical tips for students and teachers. Med Teach 2013 Jun;35(6):433-443. [doi: 10.3109/0142159X.2013.775413] [Medline: 23496121]
10. Weiner SA, Stephens G, Nour AY. Information-seeking behaviors of first-semester veterinary students: a preliminary report. J Vet Med Educ 2011;38(1):21-32. [Medline: 21805932]
11. Mirando MA, Bewley JM, Blue J, Amaral-Phillips DM, Corriher VA, Whittet KM, et al. Extension Education Symposium: reinventing extension as a resource--what does the future hold? J Anim Sci 2012 Oct;90(10):3677-3692 [FREE Full text] [doi: 10.2527/jas.2011-5074] [Medline: 21805932]
12. Keene AB, Shiloh AL, Dudaie R, Eisen LA, Savel RH. Online testing from Google Docs™ to enhance teaching of core topics in critical care: a pilot study. Med Teach 2012;34(12):1075-1077. [doi: 10.3109/0142159X.2012.716553] [Medline: 22931149]
13. Kingsley K, Galbraith GM, Herring M, Stowers E, Stewart T, Kingsley KV. Why not just Google it? An assessment of information literacy skills in a biomedical science curriculum. BMC Med Educ 2011;11:17 [FREE Full text] [doi: 10.1186/1472-6920-11-17] [Medline: 21518448]
14. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
15. Baudains C, Metters E, Easton G, Booton P. What educational resources are medical students using for personal study during primary care attachments? Educ Prim Care 2013 Sep;24(5):340-345. [Medline: 24041098]
16. Alexa Traffic Rank for YouTube. 2011 URL: http://www.alexa.com/search?q=youtube.com%2C&p=&r= [accessed 2013-10-17] [WebCite Cache ID 6KRKxdovo]
17. Goodwin D. YouTube now serving 4 billion videos daily. Search engine watch URL: http://searchenginewatch.com/article/2141050/YouTube-Now-Serving-4-Billion-Videos-Daily [accessed 2013-10-17] [WebCite Cache ID 6KRK3SbrR]
18. Gabarron E, Fernandez-Luque L, Armayones M, Lau AY. Identifying measures used for assessing quality of YouTube videos with patient health information: a review of current literature. Interact J Med Res 2013;2(1):e6. [doi: 10.2196/ijmr.2465] [Medline: 23612432]
19. Butler DP, Perry F, Shah Z, Leon-Villapalos J. The quality of video information on burn first aid available on YouTube. Burns 2012 Dec 26:1. [doi: 10.1016/j.burns.2012.10.017] [Medline: 23273651]
20. Ache KA, Wallace LS. Human papillomavirus vaccination coverage on YouTube. Am J Prev Med 2008 Oct;35(4):389-392. [doi: 10.1016/j.amepre.2008.06.029] [Medline: 18675530]
21. Schreiber JJ, Warren RF, Hotchkiss RN, Daluiski A. An online video investigation into the mechanism of elbow dislocation. J Hand Surg Am 2013 Mar;38(3):488-494. [doi: 10.1016/j.jhsa.2012.12.017] [Medline: 23391358]
22. Topps D, Helmer J, Ellaway R. YouTube as a platform for publishing clinical skills training videos. Acad Med 2013 Feb;88(2):192-197. [doi: 10.1097/ACM.0b013e31827c5352] [Medline: 23269305]
23. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
24. Clifton A, Mann C. Can YouTube enhance student nurse learning? Nurse Educ Today 2011 May;31(4):311-313. [doi: 10.1016/j.nedt.2010.10.004] [Medline: 21036430]
25. Azer SA. Can "YouTube" help students in learning surface anatomy? Surg Radiol Anat 2012 Jul;34(5):465-468. [doi: 10.1007/s00276-012-0935-x] [Medline: 22278703]
26. Murugiah K, Vallakati A, Rajput K, Sood A, Challa NR. YouTube as a source of information on cardiopulmonary resuscitation. Resuscitation 2011 Mar;82(3):332-334. [doi: 10.1016/j.resuscitation.2010.11.015] [Medline: 21185643]
27. Knösel M, Jung K, Bleckmann A. YouTube, dentistry, and dental education. J Dent Educ 2011 Dec;75(12):1558-1568 [FREE Full text] [Medline: 22184594]
28. Fat MJ, Doja A, Barrowman N, Sell E. YouTube videos as a teaching tool and patient resource for infantile spasms. J Child Neurol 2011 Jul;26(7):804-809. [doi: 10.1177/0883073811402345] [Medline: 21551373]
29. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
30. Lewis SP, Heath NL, St Denis JM, Noble R. The scope of nonsuicidal self-injury on YouTube. Pediatrics 2011 Mar;127(3):e552-e557 [FREE Full text] [doi: 10.1542/peds.2010-2317] [Medline: 21339269]

31. Lewis SP, Heath NL, Sornberger MJ, Arbuthnott AE. Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. J Adolesc Health 2012 Oct;51(4):380-385. [doi: 10.1016/j.jadohealth.2012.01.013] [Medline: 22999839]
32. Srivastava G, Roddy M, Langsam D, Agrawal D. An educational video improves technique in performance of pediatric lumbar punctures. Pediatr Emerg Care 2012 Jan;28(1):12-16. [doi: 10.1097/PEC.0b013e31823ed5ed] [Medline: 22193693]
33. Aronson ID, Plass JL, Bania TC. Optimizing educational video through comparative trials in clinical environments. Educ Technol Res Dev 2012 Jun 12;60(3):469-482. [Medline: 22904607]
34. Cardoso AF, Moreli L, Braga FT, Vasques CI, Santos CB, Carvalho EC. Effect of a video on developing skills in undergraduate nursing students for the management of totally implantable central venous access ports. Nurse Educ Today 2012 Aug;32(6):709-713. [doi: 10.1016/j.nedt.2011.09.012] [Medline: 22036273]
35. Azer SA, Aleshaiwi SM, Algrain HA, Alkhelaif RA. Nervous system examination on YouTube. BMC Med Educ 2012;12:126 [FREE Full text] [doi: 10.1186/1472-6920-12-126] [Medline: 23259768]
36. Talley NJ, O'Connor S. Clinical examination: a systematic guide to physical diagnosis. Marrickville, N.S.W: Elsevier Australia; 2006.
37. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
38. Riffenburgh RH. Statistics in Medicine, Third Edition. Oxford: Academic Press; 2012.
39. Veney JE, Kros JF, Rosenthal DA. Statistics for Health Care Professionals: Working With Excel (Public Health/Epidemiology and Biostatistics). San Francisco: Jossey-Bass; 2009.
40. Huang C, Fu T, Chen H. Text-based video content classification for online video-sharing sites. Journal of the American Society for Information Science and Technology 2010;61:891-906. [doi: 10.1002/asi.21291]
41. Camm CF, Sunderland N, Camm AJ. A quality assessment of cardiac auscultation material on YouTube. Clin Cardiol 2013 Feb;36(2):77-81. [doi: 10.1002/clc.22080] [Medline: 23172251]
42. Koya KD, Bhatia KR, Hsu JT, Bhatia AC. YouTube and the expanding role of videos in dermatologic surgery education. Semin Cutan Med Surg 2012 Sep;31(3):163-167. [doi: 10.1016/j.sder.2012.06.006] [Medline: 22929353]
43. Cardoso AF, Moreli L, Braga FT, Vasques CI, Santos CB, Carvalho EC. Effect of a video on developing skills in undergraduate nursing students for the management of totally implantable central venous access ports. Nurse Educ Today 2012 Aug;32(6):709-713. [doi: 10.1016/j.nedt.2011.09.012] [Medline: 22036273]
44. Krawczyk A, Lau E, Perez S, Delisle V, Amsel R, Rosberger Z. How to inform: comparing written and video education interventions to increase human papillomavirus knowledge and vaccination intentions in young adults. J Am Coll Health 2012;60(4):316-322. [doi: 10.1080/07448481.2011.615355] [Medline: 22559091]
45. Smith W, Rafeek R, Marchan S, Paryag A. The use of video-clips as a teaching aide. Eur J Dent Educ 2012 May;16(2):91-96. [doi: 10.1111/j.1600-0579.2011.00724.x] [Medline: 22494307]
46. Casado MI, Castaño G, Arráez-Aybar LA. Audiovisual material as educational innovation strategy to reduce anxiety response in students of human anatomy. Adv Health Sci Educ Theory Pract 2012 Aug;17(3):431-440. [doi: 10.1007/s10459-011-9307-2] [Medline: 21678089]
47. Srivastava G, Roddy M, Langsam D, Agrawal D. An educational video improves technique in performance of pediatric lumbar punctures. Pediatr Emerg Care 2012 Jan;28(1):12-16. [doi: 10.1097/PEC.0b013e31823ed5ed] [Medline: 22193693]
48. Aronson ID, Plass JL, Bania TC. Optimizing educational video through comparative trials in clinical environments. Educ Technol Res Dev 2012 Jun 12;60(3):469-482. [Medline: 22904607]
49. Martens MJ, Duvivier RJ, van Dalen J, Verwijnen GM, Scherpbier AJ, van der Vleuten CP. Student views on the effective teaching of physical examination skills: a qualitative study. Med Educ 2009 Feb;43(2):184-191. [doi: 10.1111/j.1365-2923.2008.03283.x] [Medline: 19161490]
50. P Mennin Martha Cole McGrew S. Scholarship in teaching and best evidence medical education: synergy for teaching and learning. Med Teach 2000;22(5):468-471. [doi: 10.1080/01421590050110722] [Medline: 21271958]

Abbreviations

CVS: cardiovascular system
PBL: problem-based learning
RS: respiratory system

Edited by G Eysenbach; submitted 23.05.13; peer-reviewed by N Allen, F Aminpour, A Holzinger, C Mather, L Woodham, R Robinson; comments to author 16.06.13; revised version received 13.07.13; accepted 06.09.13; published 13.11.13.

Please cite as:
Azer SA, AlGrain HA, AlKhelaif RA, AlEshaiwi SM
Evaluation of the Educational Value of YouTube Videos About Physical Examination of the Cardiovascular and Respiratory Systems
J Med Internet Res 2013;15(11):e241
URL: http://www.jmir.org/2013/11/e241/
doi:10.2196/jmir.2728
PMID:24225171

Original Paper

Smartphones as Multimodal Communication Devices to Facilitate Clinical Knowledge Processes: Randomized Controlled Trial

Christoph Pimmer[1], Mag; Magdalena Mateescu[2], MSc; Carmen Zahn[2], PhD; Urs Genewein[3], MD, MME

[1]Institute for Information Systems, School of Business, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Basel, Switzerland

[2]Institute for Research and Development of Collaborative Processes, School of Applied Psychology (APS), University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

[3]Hightech Research Center, Cranio-Maxillofacial Surgery, University Hospital Basel, Basel, Switzerland

Corresponding Author:
Christoph Pimmer, Mag
Institute for Information Systems
School of Business
University of Applied Sciences and Arts Northwestern Switzerland FHNW
Peter Merian-Strasse 86
Basel, 4002
Switzerland
Phone: 41 61 279 18 49
Fax: 41 61 279 17 98
Email: Christoph.Pimmer@fhnw.ch

Abstract

Background: Despite the widespread use and advancements of mobile technology that facilitate rich communication modes, there is little evidence demonstrating the value of smartphones for effective interclinician communication and knowledge processes.

Objective: The objective of this study was to determine the effects of different synchronous smartphone-based modes of communication, such as (1) speech only, (2) speech and images, and (3) speech, images, and image annotation (guided noticing) on the recall and transfer of visually and verbally represented medical knowledge.

Methods: The experiment was conducted from November 2011 to May 2012 at the University Hospital Basel (Switzerland) with 42 medical students in a master's program. All participants analyzed a standardized case (a patient with a subcapital fracture of the fifth metacarpal bone) based on a radiological image, photographs of the hand, and textual descriptions, and were asked to consult a remote surgical specialist via a smartphone. Participants were randomly assigned to 3 experimental conditions/groups. In group 1, the specialist provided verbal explanations (speech only). In group 2, the specialist provided verbal explanations and displayed the radiological image and the photographs to the participants (speech and images). In group 3, the specialist provided verbal explanations, displayed the radiological image and the photographs, and annotated the radiological image by drawing structures/angle elements (speech, images, and image annotation). To assess knowledge recall, participants were asked to write brief summaries of the case (verbally represented knowledge) after the consultation and to re-analyze the diagnostic images (visually represented knowledge). To assess knowledge transfer, participants analyzed a similar case without specialist support.

Results: Data analysis by ANOVA found that participants in groups 2 and 3 (images used) evaluated the support provided by the specialist as significantly more positive than group 1, the speech-only group (group 1: mean 4.08, SD 0.90; group 2: mean 4.73, SD 0.59; group 3: mean 4.93, SD 0.25; $F_{2,39}$=6.76, P=.003; partial η^2=0.26, 1–β=.90). However, significant positive effects on the recall and transfer of visually represented medical knowledge were only observed when the smartphone-based communication involved the combination of speech, images, and image annotation (group 3). There were no significant positive effects on the recall and transfer of visually represented knowledge between group 1 (speech only) and group 2 (speech and images). No significant differences were observed between the groups regarding verbally represented medical knowledge.

Conclusions: The results show (1) the value of annotation functions for digital and mobile technology for interclinician communication and medical informatics, and (2) the use of guided noticing (the integration of speech, images, and image annotation) leads to significantly improved knowledge gains for visually represented knowledge. This is particularly valuable in situations involving complex visual subject matters, typical in clinical practice.

(J Med Internet Res 2013;15(11):e263) doi:10.2196/jmir.2758

KEYWORDS

mobile health; mobile phone; telemedicine; educational technology; learning; problem solving; multimedia; audiovisual aids

Introduction

Interclinician Communication and Mobile Phones

Interclinician communication is a key component of health care systems. The significance becomes clear in light of its impact on patient care: poor communication between clinicians results in enormous costs and, more importantly, a high number of adverse clinical outcomes and deaths [1-4]. Typical forms of communication between medical professionals are shaped by the particularities of clinical work and can be characterized as instant and synchronous [2,3,5], interdisciplinary and interprofessional (between actors holding different domains and levels of knowledge [6,7]), and mobile (ie, between physically/locally mobile hospital workers [3,8] who increasingly communicate using mobile technologies). In particular, cellphones and smartphones are becoming increasingly popular in clinical settings, with adoption rates of up to 98% [9-12]. According to recent reviews, in light of these characteristics and recent developments, there is little known about how mobile phone-based communication can contribute to effective interclinician communication [13-15]. In many of the existing studies that have explored mobile clinical communication technologies, the methodological design was reported to be of lower quality and based on users' perceptions [13]. Additionally, only a few randomized controlled experiments were identified [13,16].

Rich Communication Modes: Speech, Images, and Annotation

In recent publications, mobile phones have been considered as potentially efficient tools that enable instant location-independent communication [10,14,17-20]. In addition to supporting immediate communication in the form of speech, a number of studies have demonstrated how mobile phones and smartphones allow the ability to capture, exchange, and interpret images [14,21]. Comparisons to standard methods have demonstrated the suitability of mobile phones for use in assessment and diagnosis, such as when using computed tomography (CT), computed tomography angiography (CTA), and noncontrast computed tomography (NCCT) soft tissue or ophthalmic images [22-28]. A number of clinical information systems allow for the annotation of clinical images [29-31]; annotation refers to marking, drawing, highlighting, labeling, or otherwise describing (and enriching) aspects of visual material that should be the focus of attention. The creation of annotations is an essential part of clinical communication environments [32], and is considered a "fundamental task for clinicians, medical educators, or basic scientists" [30].

Mobile Communication and Knowledge Exchange

The aim of synchronous interclinician communication involves the building of a shared understanding and a "just-in-time grounding" [2] between clinical actors with varying levels and domains of knowledge and expertise for the well-being of patients. During this process, knowledge is exchanged and new knowledge is created. Less knowledgeable actors learn from the communication and may transfer what they have learned to future patient cases. In other words, interclinician communication represents a valuable opportunity for medical actors to learn from one another for present and for future patient cases, similar to problem-based learning (eg, [33]). Regarding knowledge exchange and learning, some research emphasizes the value of using images in mobile phone-based communication. For example, it has been suggested that mobile image messaging can serve as an instructional tool that allows for instant feedback and further insights for less knowledgeable actors [34]. In a study involving doctor-to-doctor consultations that were based on images taken and sent by means of mobile phones (in the form of Multimedia Messaging Service, MMS), it was concluded that multimedia consultation had a positive effect on patient management and led to an improvement of the service. In addition, 86% of the residents reported "the multimedia information contributed to their ability to independently handle similar cases in [the] future" [35].

We conclude, despite the widespread use and advancements of mobile technology that enables rich communication modes, that there is a surprising paucity of evidence that demonstrates the actual value of mobile technology for effective interclinician communication, knowledge exchange, and learning.

Objectives and Hypothesis

To address this research gap, we delineated 2 sets of hypotheses based on cognitive and sociocognitive science approaches. The first set of hypotheses relates to theories of dual coding and multimedia learning [36-39] and the second set relates to the notion of guided noticing in communication [40,41].

Concerning cognitive science, it has been shown that images support human information processing and understanding. This is not only because images convey different information than words (concrete vs abstract), but because text and images are processed in 2 different channels or modes [36-39]. The dual channel processing assumption asserts that humans have separate channels for processing words and images and that they use them both to construct coherent mental knowledge representations [36,38,39]. Empirical evidence has demonstrated picture superiority effects (ie, that people can recall information better from both words and images than from words alone) [37,42-44]. Moreover, a vast body of research on learning with multimedia [37,43] suggests that when words (written words or speech) and images are presented together and in a well-designed and integrated manner (eg, redundant and spatially contingent), they support the construction of coherent mental representations and, thus, learning (eg, learning about the functioning of physical systems) [45]. This idea is especially true for less knowledgeable learners (eg, [46]) and in situations in which test items were presented in image-based form [47]. However, research on dual coding and multimedia learning was primarily conducted in experimental studies by using simple content. There is little evidence that demonstrates learning effects in applied and more complex situations, such as that of

interclinician communication and knowledge exchange. According to multimedia theory, we formulated our first assumption for the experimental design of mobile phone-based medical communication.

Mobile phone-based medical communication, in which speech and images are combined to enhance the understanding of complex subject matter, leads to a better understanding by a less knowledgeable actor when compared to the speech-only condition. Therefore, knowledge recall and transfer of verbally and visually represented knowledge should increase when compared to the speech-only condition. Therefore, in our specific case of mobile communication and less knowledgeable actors, we hypothesize that mobile communication involving integrated speech and images leads to better recall and better transfer of visually represented and verbally represented medical knowledge than the speech-only condition.

For our second set of hypotheses, we derived our assumptions from the concept of guided noticing. Guided noticing describes how specific digital technology affordances can be used in distributed collaborative settings to direct joint attention to notice specific elements in visual media (eg, videos, images). Guided noticing can involve highlighting (eg, annotation), naming, and commenting; thus, it allows the assignment of typological representations (culturally meaningful categories) to topological representations (eg, images). This is especially important when experts interpret visualizations in detail and explain their interpretations to others. Here, guided noticing can be a way to establish common ground [48] in communication between professionals (eg, in educational settings), for example, by using annotations. The creation of annotations is considered as a grounding or communication act by users (rather than designers) who intend to close gaps in common ground required to complete a task [32]. Guided noticing can also serve to develop domain expertise and a professional vision [49] (eg, the training of diagnostic skills with x-rays [41]). Professional vision was empirically investigated and characterized by Goodwin [49] as "...socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group." Case studies of professional practices (eg, developing coding schemes or highlighting) with domain experts (eg, archeologists in field research, lawyers in courtroom) were conducted analyzing these practices in great detail. Yet, empirical research on guided noticing [41] is still sparse. In our own previous research on history learning with advanced video tools, we identified episodes of guided noticing and analyzed them in relation to the successful acquisition of visual knowledge [50]. Our present study extends the existing qualitative research by adding original quantitative data on the effects of highlighting (by annotation) for guided noticing in the specific area of medical knowledge.

Accordingly, we formulated the assumption that communication that is based on guided noticing, in which speech (typological representations) is linked with images (topological representations) in the form of annotation, leads to better knowledge recall and transfer by the less knowledgeable actor compared with speech-only communication or integrated speech and images without annotation. For our specific research of mobile phone-based communication in the context of less knowledgeable actors, we propose a second set of hypotheses that mobile communication that is based on guided noticing leads to better recall and better transfer of visually represented and verbally represented medical knowledge than speech-only communication or that of integrated speech and images without annotation.

Methods

Experimental Design

To test our hypotheses, an intervention study was designed. The experiment was conducted from November 2011 to May 2012 at the University Hospital Basel (Switzerland) with 42 medical students who were obtaining their master's degrees. All medical students with a master's curriculum were eligible to participate.

Before starting the experiment, informed written consent was obtained from the participants. In the first step, an experimenter asked the participants to assume the role of an emergency assistant. They briefly analyzed a patient case about a subcapital fracture of the fifth metacarpal bone [51,52] to initiate a consultation with a hand specialist. The patient case included a short text with the initial information about anamnesis and status, as well as a radiological image and photographs showing the limited functional capabilities (impaired fist closure and extension deficit). We chose this case because it is a frequently encountered, yet complex, case for novices. The associated clinical reasoning and treatment involves a great deal of medical and clinical knowledge, including the consideration of radiological, functional, and sociodemographic indicators.

In the second step, the participants were put in communication with a hand surgeon by means of an iPhone 4. To initiate the phone consultation, they were required to briefly characterize the patient case. The specialist provided pertinent standardized advice according to the 3 following experimental conditions/groups: group 1 received verbal explanations (speech only) from the specialist, group 2 received verbal explanations and the radiological image and photographs from the specialist, and group 3 received verbal explanations, the radiological image and the photographs, and annotations (by drawing structures/angle elements) on the radiological image from the specialist (speech, images, and image annotation as guided noticing). Although visual information was varied as described in groups 2 and 3, the verbal information (speech) remained constant in all 3 groups. The case information was prepared by a hand surgeon (UG) and, thereafter, the specialist's role was assumed by MM.

The technical setting was based on 2 widely available tools: Skype for the communication and Google Drive for the sharing and annotation of the images. Google Drive allows real-time collaboration, inter alia, in the form of sharing and annotating images on desktop and mobile technologies. To avoid variances in the annotations throughout the experiments, the images were preannotated and then displayed in a predefined sequence by the specialist as integral part of the communication setting.

After the conversation with the expert, the participants were asked to complete 2 tasks. Task 1 (knowledge recall) required

the participants to write a brief description of the case in which they summarized and justified the relevant points of the conversation (including the diagnosis) in a given time period of 90 seconds (ie, recall of verbally represented knowledge). Second, they were required to draw relevant angles/structures in the radiological image and to estimate the angle size (ie, recall of visually represented knowledge). Task 2 was to measure knowledge transfer. The participants were then given a second case that they had to solve in a similar manner. They were asked to analyze the case information and to develop and justify a diagnosis without specialist support. They wrote a short description of the case including the diagnosis (to determine the transfer of verbally represented knowledge) and drew the relevant angles/structures, and estimated the angle size (to determine the transfer of visually represented knowledge).

In the last step, the participants completed an additional questionnaire about their previous knowledge and experience with fractures. In addition, they answered questions related to the task and experimental conditions. These questions were used as control variables to ensure that there were no differences between the 3 groups.

Outcome Measures

After the intervention, the participants were requested to estimate the (self-perceived) usefulness of the support provided by the specialist. We also tested the actual recall and transfer of visually and verbally represented medical knowledge. Sample solutions were elaborated by the research team, including medical (hand surgery), and education and psychology experts. For categories and descriptions, see Table 1. The participants' answers were jointly rated by 2 raters. A negotiated coding approach was applied for the coding procedure because the raters had limited expertise in medical and health sciences and, although they were experienced coders, they were not familiar with applying medical coding schemes. Negotiated approaches have been proposed previously (eg, for transcript analyses when familiarity with a new coding scheme is low) [53]. Negotiated coding was enhanced by resolving disagreements upon discussion with a hand surgeon.

To test the hypotheses with a 2-sided significance level of 5% and a power of 80%, a sample size of 45 participants was calculated, a priori, using G*Power-software [54]. The participants were randomly assigned to 1 of 3 groups using a 3-arm parallel design. A computer-generated list of random numbers was used to distribute the participants according to the principles of simple randomization with a 1:1 ratio.

Statistical Methods

The data were analyzed using a 3×2 repeated measures ANOVA with a between-subjects factor (group 1, group 2, and group 3) and a within-subjects factor (task 1 and task 2). The effect size (partial η^2) and the observed power (1–β) for variables that addressed the recall and transfer of verbally and visually represented knowledge were reported. A 1-way ANOVA was calculated for the support offered by the specialist variable. Differences between conditions were assessed using a post hoc comparison. Statistical significance was determined by values $P<.05$ and all analyses were conducted using SPSS version 19 (IBM Corp, Armonk, NY, USA).

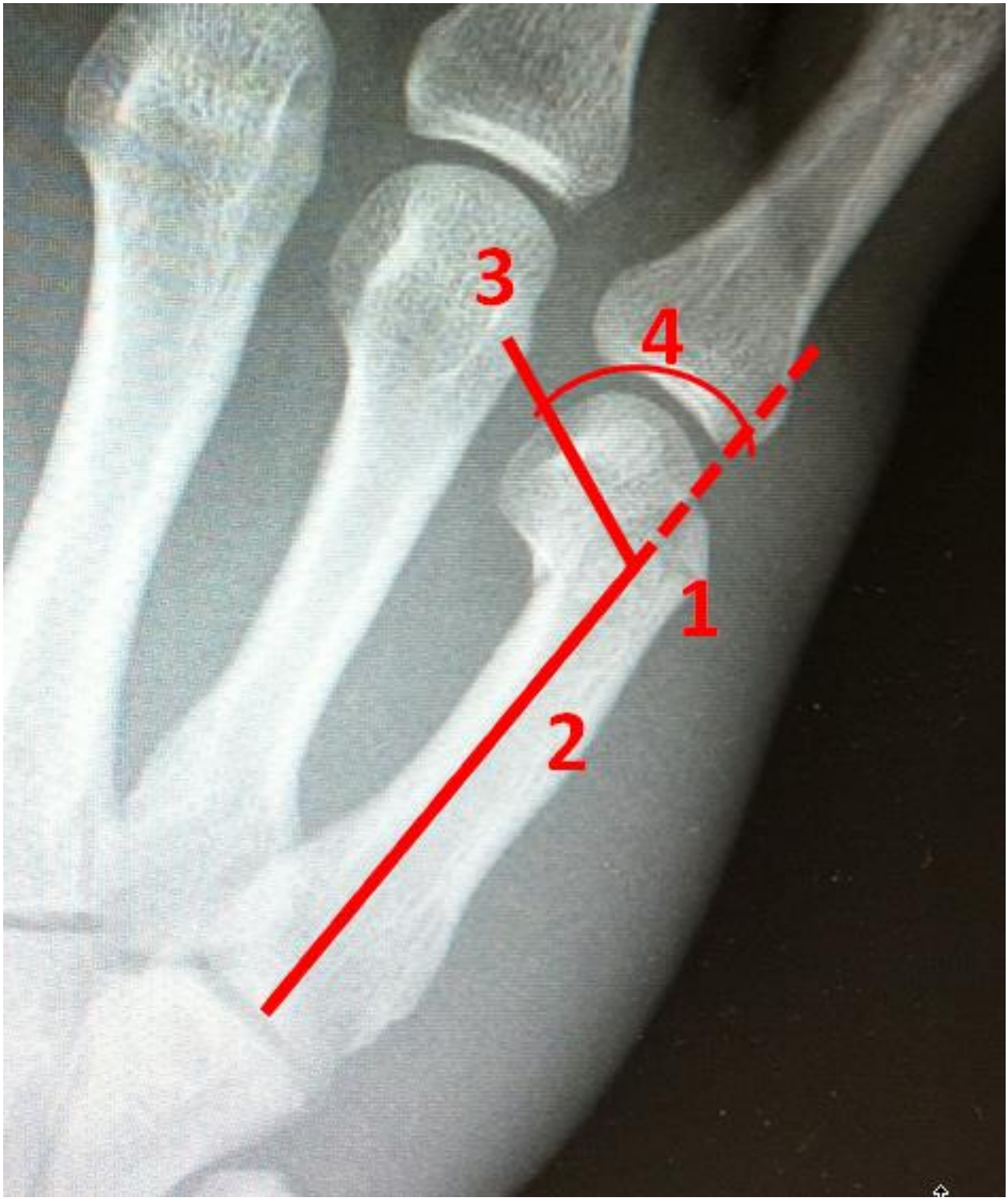
Ethical Considerations

Ethical approval was sought from the regional ethical review board. Because no patients were involved and the participants were considered health care professionals, the board ruled the experiment exempt from further ethical approval and trial registration requirements. Nevertheless, we consulted an expert outside the research team, a professor of ethics at a Swiss university who was part of a separate Swiss ethical board for ethical advice. The confidentiality of the participants was ensured and written informed consent was obtained from every participant before the experiment.

Table 1. Measurements of verbally and visually represented knowledge.

Category	Description	Examples
Recall and transfer of visually represented knowledge		
Correctness/completeness of drawn angle elements	Positioning of the following elements: (1) point of intersection, (2) basis line, (3) fracture line, and (4) position of angle	See Figure 1
Correctness of drawn angle size	Measurement of the angle size drawn by the participant	eg, 75 degrees
Correctness of estimated angle size	Angle size estimated by the participant	eg, 75 degrees
Recall and transfer of verbally represented knowledge		
Correctness/completeness of functional/radiological characteristics	Measurement of written information, including identification of fracture, type of fracture and position, type of functional limitations (extension deficit and rotation error), and degree of functional restrictions and angulation of the fracture	Fracture of metacarpal V, extension deficit of 10 degrees, and rotation error of 10 degrees, representing 2 functional deficits
Correctness/completeness of individual sociodemographic patient characteristics	Measurement of written information including age, profession, and dominant hand	Young, employed patient, dominant hand
Correctness/completeness of overall written answers	Summarizes both measurements	

Figure 1. X-ray with fracture as used in the study.

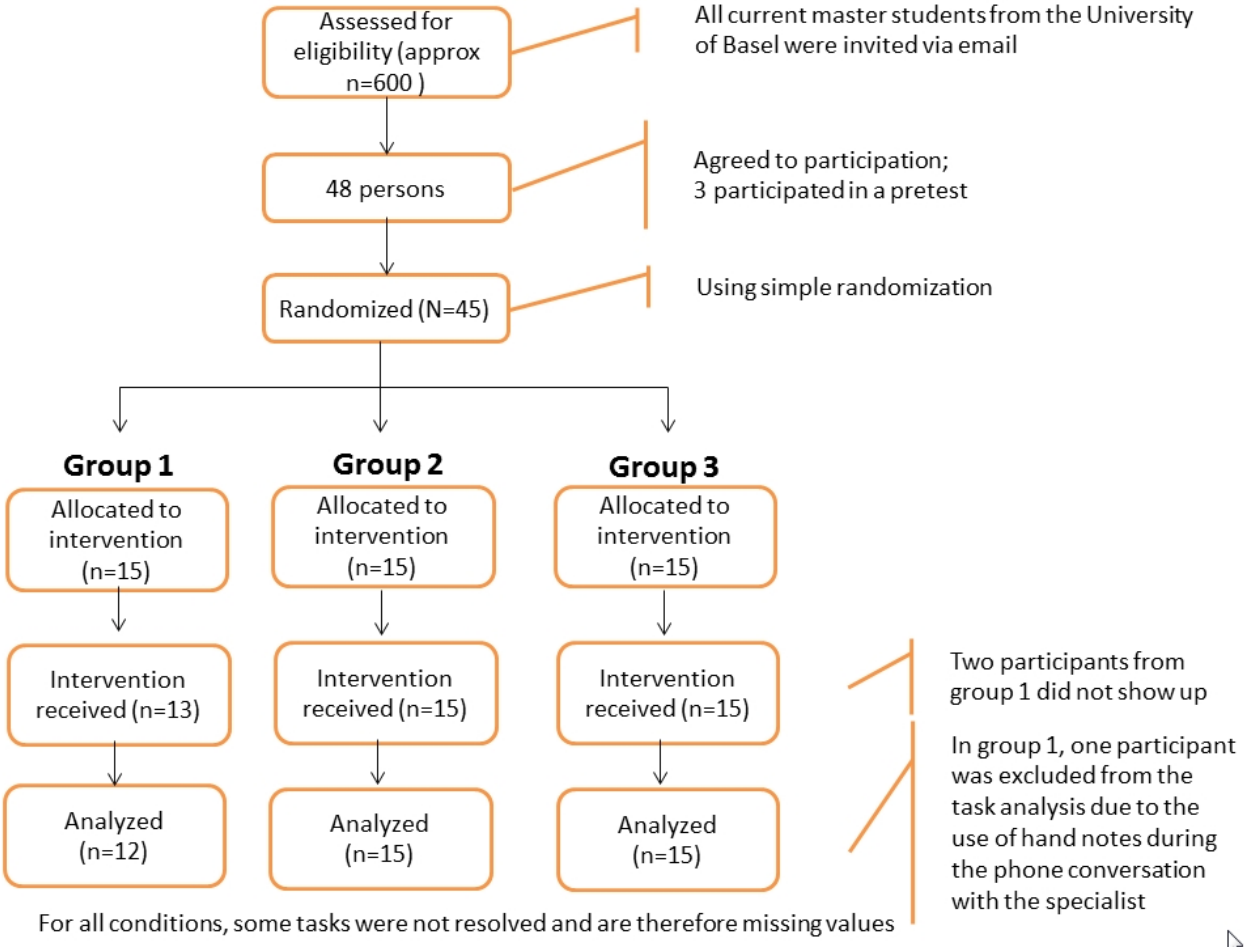


Results

Participant Flow

In all, 48 participants agreed to participate in the experiment. Pretests were conducted with 3 participants to validate the technical feasibility and time limits. The remaining 45 persons were randomized and assigned to the 3 groups. For a flow diagram of participants, see Figure 2.

Figure 2. Participant flow diagram.



Recruitment

Eligible participants were invited by means of email and direct invitation by the medical faculty of the University of Basel in November 2011, using a flyer with basic information. The participants did not receive monetary compensation, but a draw for an iPad was used as an incentive.

Baseline Demographic Characteristics and Prior Knowledge for Each Group

All 42 participants (23 male, 19 female) were between ages 19 and 38 years (mean 24.57, SD 2.95) and were medical students obtaining their master's degrees (see Table 2). To determine the bias caused by potentially different levels of prior knowledge in the groups, we asked the participants about their interest in surgical topics ($F_{2,39}$=0.10, P=.91), their knowledge on the evaluation of fractures ($F_{2,39}$=2.17, P=.13), prior participation in the treatment of fractures of metacarpal bones ($F_{2,39}$=0.96, P=.39), and prior experience with diagnosis and treatment of fractures of metacarpal bones ($F_{2,39}$=0.58, P=.57). No significant differences between the groups were found (for average scores and standard deviation see Table 2).

Outcomes

Support Offered by the Specialist

ANOVA revealed significant differences between group 1 and groups 2 and 3. No significant differences between groups 2 and 3 were observed; a test of between-subjects contrasts yielded the following: $F_{2,39}$=6.76, P=.003; partial η^2=0.26 and 1–β=.90. These results suggest that the students in groups 2 and 3 placed significantly more value on the support offered by the specialist. The students' evaluations of the support offered by the specialist are displayed in Table 3.

Table 2. Demographic characteristics and prior knowledge of participants (N=42).

Demographic characteristics	Group 1 (n=12)	Group 2 (n=15)	Group 3 (n=15)
Age, mean (SD)	24.17 (2.12)	25.33 (4.39)	24.13 (1.30)
Sex, n (%)			
Male	7 (17)	8 (19)	8 (19)
Female	5 (11)	7 (17)	7 (17)
Years of study, mean (SD)	5 (1)	5 (1)	4 (1)
Prior knowledge, mean (SD)			
Experience with touchscreen[a]	4.42 (1.16)	3.53 (1.77)	4.13 (1.50)
Interest in surgical topics[b]	3.17 (1.19)	3.07 (1.39)	3.27 (1.16)
Estimated knowledge on the evaluation of fractures (compared to peer students)[b]	2.58 (0.67)	2.20 (1.08)	2.93 (1.03)
Prior participation in the treatment of fractures of metacarpal bones[b]	1.25 (0.62)	1.13 (0.52)	1.53 (1.13)
Prior experience in diagnosis and treatment of fractures of metacarpal bones[b]	1.25 (0.87)	1.20 (0.77)	1 (0.00)

[a]Scale: 1=no use; 5=daily use.

[b]Likert scale from 1-5.

Table 3. Measurement of specialist support.

Measure	Group, mean (SD)			Mean difference (95% CI)		
	G1 (n=12)	G2 (n=15)	G3 (n=15)	G1–G2	G1–G3	G2–G3
Support offered by the specialist	4.08 (0.90)	4.73 (0.59)	4.93 (0.26)	–0.65[a] (–1.25, –0.05)	–0.85[a] (–1.45, –0.25)	–0.20 (–0.76, 0.36)

[a]Mean difference was determined as significant at the $P<.05$ level.

Visually Represented Knowledge

We used the following 3 measures to assess the recall and transfer of visually represented medical knowledge: correctness/completeness of drawn angle elements, correctness of drawn angle size, and correctness of estimated angle size (see Table 4). For all these measures, we performed a 2×3 ANOVA with a within-subjects factor (task 1 vs task 2) and a between-subjects factor (group 1 vs group 2 vs group 3). As predicted, for the correctness/completeness of drawn angle elements in both recall and transfer tasks, group 3 (supported with guided noticing) scored significantly higher than the other 2 groups as revealed by the test of between-subjects contrasts ($F_{2,37}$=11.32, $P<.001$; partial η^2=0.38, 1–β=.99). The means, standard deviations, and post hoc contrasts are shown in Table 4.

To test the extent to which the experimental conditions affected the correctness of drawn angle size and the correctness of estimated angle size, we calculated *z* scores and then performed an ANOVA on the standardized data. As hypothesized, group 3 displayed better performance on both measurements: test of between-subjects contrasts relating to the correctness of drawn angle size ($F_{2,37}$=8.81, $P<.001$; partial η^2=0.32, 1–β=.96) and the test of between-subjects contrasts relating to the correctness of estimated angle size ($F_{2,37}$=7.67, $P<.001$; partial η^2=0.30, 1–β=.93). More precisely, for knowledge recall, post hoc tests showed that the ability to draw correct angle sizes and the ability to estimate the size of the drawn angle was significantly higher for group 3 when compared with the other 2 groups. However, a similar result was not supported in the transfer task, in which significant differences were observed only between groups 1 and 3.

Table 4. Recall and transfer of visually represented medical knowledge.

Submeasures of visually represented medical knowledge	Group, mean (SD)			Adjusted difference, mean (95% CI)		
	G1 (n=11)	G2 (n=15)	G3 (n=14)	G1–G2	G1–G3	G2–G3
Correctness/completeness of drawn angle elements						
Task 1: Recall	1.23 (1.54)	1.83 (1.21)	3.29 (1.14)	–0.61 (–1.88, 0.67)	–2.06[a] (–3.35, –0.76)	–1.45[a] (–2.65, –0.26)
Task 2: Transfer	1.82 (1.52)	2.03 (1.32)	3.57 (0.47)	–0.22 (–1.38, 0.95)	–1.75[a] (–2.93, –0.58)	–1.53[a](–2.62, –0.45)
Correctness of drawn angle size [b]						
Task 1: Recall	0.54 (.91)	0.36 (.95)	–0.78 (0.63)	0.18 (–0.66, 1.01)	1.32[a] (0.47, 2.17)	1.14[a] (0.36, 1.92)
Task 2: Transfer	0.48 (0.98)	0.15 (1.01)	–0.56 (0.76)	0.33 (–0.59, 1.25)	1.05[a] (0.11, 1.98)	0.71 (–0.15, 1.57)
Correctness of estimated angle size [b]						
Task 1: Recall	0.51 (0.92)	0.37 (0.96)	–0.74 (0.69)	0.14 (–0.73, 1.02)	1.25[a] (0.38, 2.12)	1.11[a] (0.29, 1.93)
Task 2: Transfer	0.45 (0.98)	0.15 (1.05)	–0.55 (0.75)	0.30 (–0.65, 1.25)	1.00[a] (0.06, 1.95)	0.70 (–0.18, 1.59)

[a]The mean difference was determined to be significant (P<.05).

[b]Small numbers indicate better performance.

Verbally Represented Knowledge

We used the following 2 submeasures to assess the recall and transfer of verbally represented medical knowledge: the correctness/completeness of functional/radiological characteristics and the correctness/completeness of individual sociodemographic patient characteristics (see Table 2). Statistical analysis was performed on the general measure of correctness and completeness of overall written answers (Table 5). Scores that assessed the recall and transfer of verbally represented knowledge were analyzed by means of a 2×3 repeated ANOVA with a within-subjects factor (task 1 vs task 2) and a between-subjects factor (group 1 vs group 2 vs group 3). The results showed no significant differences between the 3 groups, with respect to the correctness and completeness of verbally represented knowledge for both the recall and transfer tasks. The calculated tests of between-subjects effects (factor 1) did not reach statistical significance with a level of P<.05 ($F_{2,39}$=0.58, P=.58; partial η^2=0.03, 1–β=.13).

Table 5. Correctness/completeness of overall written answers for recall and transfer.

Measure	Group, mean (SD)			Mean difference (95% CI)		
	G1 (n=12)	G2 (n=15)	G3 (n=15)	G1–G2	G1–G3	G2–G3
Task 1: Recall	5.92 (2.07)	6.33 (2.02)	6.67 (2.58)	–0.42 (–2.60, 1.76)	–0.75 (–2.93, 1.43)	–0.33 (–2.39, 1.72)
Task 2: Transfer	3.00 (2.34)	3.00 (2.54)	3.67 (2.55)	0.00 (–2.41, 2.41)	–0.67(–3.08, 1.74)	–0.67 (–2.94, 1.61)

Performance Between Task 1 and 2

With respect to verbally represented knowledge (correctness/completeness of overall written answers), we observed significant differences in participants' performance between task 1 and 2 using a test of within-subjects effects ($F_{1,39}$=43.72, P<.001; partial η^2= 0.53, 1–β=1), whereby the performance decreased in task 2, independent of the experimental condition (task 1: mean 6.30, SD 0.39; task 2: mean 3.22, SD 0.38; adjusted difference: mean 3.08, 95% CI 2.14-4.03). We observed no significant differences with respect to visually represented knowledge relating to the correctness/completeness of drawn angle elements ($F_{1,37}$=3.37, P=.06); the correctness of drawn angle ($F_{2,37}$= 0.02, P=.89); and the correctness of estimated angle ($F_{1,36}$=0.04, P=.84). We observed a tendency toward improved performance when assessing the variable correctness/completeness of drawn angle elements in task 2 (see Table 6).

Table 6. Visually represented knowledge: performance between tasks 1 (knowledge recall) and 2 (knowledge transfer).

Measure	Task, mean (SD)		Mean difference (G1–G2) (95% CI)
	Task 1: Recall	Task 2: Transfer	
Correctness/completeness of drawn angle elements	2.17 (1.51)	2.51 (1.38)	–0.36 (–0.73, 0.07)
Correctness of drawn angle size[a]	0.01 (1.01)	–0.01 (1.00)	0.20 (–0.27, 0.31)
Correctness of estimated angle size[a]	0.01 (1.01)	–0.01 (1.00)	0.30 (–0.26, 0.32)

[a]Small numbers indicate better performance.

Discussion

Principal Results

In light of the rapid adoption of mobile technology, such as smartphones, for clinical communication, this study systematically compares different synchronous mobile phone-based communication modes, including (1) speech only, (2) speech and images, and (3) speech, images, and image annotation (guided noticing). Using an experimental approach, we investigated to what extent these modes affect knowledge processes in clinical communication as indicated by (1) recall and (2) transfer of visually and verbally represented knowledge by a less knowledgeable medical actor. Our variations and hypotheses were informed by psychological theories from related research in the cognitive and sociocognitive sciences.

In the conditions where speech was integrated with images (groups 2 and 3), participants evaluated the support provided by the specialist as significantly more positive compared to the speech-only condition (group 1). However, in measuring actual knowledge gains, significant positive effects on the recall and transfer of visually represented medical knowledge were only measured in group 3, which integrated speech, images, and image annotation (guided noticing). With respect to verbally represented medical knowledge, no significant differences were measured. In other words, the presentation of visual information did not contribute to the recall and retention of verbally represented knowledge.

These findings contribute original data to at least 2 broad discourses: (1) the practice of clinical communication and medical informatics, and (2) psychological, cognitive, and sociocognitive learning and communication theories.

Implications for Medical/Clinical Communication and Medical Informatics

Unlike most of the other studies that have investigated the contributions of mobile technologies to clinical communication and knowledge exchange, we used a randomized controlled experimental design. The findings suggest that, despite increased self-perceived effectiveness of the use of visual communication modes (compared to a speech-only mode), the understanding and knowledge gains of less knowledgeable medical actors are only enhanced if annotations are used to integrate speech with images. The finding that the actual effectiveness of mobile multimodal communication is different than the self-perceived effectiveness is critical and may negatively affect patient treatment, such as during a specialist consultation; if a requesting doctor were to deem phone-based advice from the specialist (without annotation) as sufficient, this could result in poor decision making with regard to the requesting doctor's further treatment of the patient. In addition, the present results challenge studies in which positive outcomes of knowledge gains of rich smartphone-based consultations were determined using only self-perceived evaluation (eg, [35]). The present study also highlights the value of annotation for clinical communication and medical informatics. Hospital management may consider this mode of communication when developing or buying new communication solutions, including mobile solutions. Senior clinicians may learn from our findings that less knowledgeable actors, students, and younger colleagues are much more likely to benefit from explanations that integrate speech and visual information by means of labeling and annotating, which is a practice that should not be limited to mobile communication. Recently, it was shown how doctors who create rich multimodal representations by using gestures to connect speech with different computational and bodily representations can improve the understanding of less knowledgeable colleagues [55]. These aspects of communicating, which are highly relevant for clinical communication and learning, are widely neglected in the current clinical practice and should be more thoroughly considered in the future.

Theoretical Implications and Discussion

To our knowledge, our experimental study is the first to examine the dual channel and multimedia theories in applied, complex, and authentic situations, such as clinical communication. Our first set of hypotheses (that mobile communication involving integrated speech and images leads to better recall and better transfer of visually represented and verbally represented medical knowledge than the speech-only condition) was based on the dual channel and multimedia theories: we assumed that synchronous mobile communication which integrated images and speech (group 2; ie, contingent multimedia information) would enhance a less knowledgeable actor's understanding of the subject matter compared with the speech-only condition (group 1). However, we found that using images in group 2 did not lead to significantly improved learning compared to the speech-only condition (group 1). This result challenges other studies from the related field of cognitive research (eg, [56]). A possible explanation lies in our specific experimental clinical setting. During the preparation phase, the images were shown to all participants (including those in the first condition) to initiate the specialist consultation. It is possible that the participants from the speech-only group could later recall these images during the communication phase, thus compensating to some extent for not having the real images available. In other

words, the participants in the speech-only condition might have invested more mental effort by imagining the images.

Our second set of hypotheses (that mobile communication based on guided noticing leads to better recall and better transfer of visually represented and verbally represented medical knowledge than speech-only communication or that of integrated speech and images without annotation) was derived from the concept of guided noticing [40]. We assumed that speech, images, and annotations (guided noticing) in group 3 would lead to knowledge gains with respect to the recall and transfer of both verbally and visually represented medical knowledge (compared with the other conditions). Our results confirm an increased ability to recall visually represented medical knowledge in the guided noticing condition for all 3 submeasures when compared with the other 2 conditions. With regard to the knowledge transfer, the correctness/completeness submeasure of drawn angle elements was significantly affected compared with the other conditions. However, the correctness of the drawn and estimated angle was significantly superior in group 3 (guided noticing) only compared to group 1 (the speech-only condition) and not when compared to group 2 (speech and images).

Taken together, these results might be readily explained by the fact that the correctness/completeness of drawn angle elements submeasure (see Figure 1) requires a complex and comprehensive understanding of the image and knowledge that cannot be compensated for over time or by increased mental effort with imagination. With this difference in mind, the results demonstrate the importance of guided noticing as a concept and we recommend this mode of communication for the clinical practice.

In summary, the simple integration of images and speech in mobile phone-based communication did not lead to significantly improved knowledge gains (compared with speech-only communication). Our results confirmed the value of guided noticing in communication. Moreover, they support the view that guided noticing is particularly important in situations involving complex visual subject matter, typical in clinical communication.

Strengths, Limitations, Generalizability, and Future Research

In addition to the strengths indicated (ie, using an experimental randomized design in authentic and complex settings), our study is based on firm theoretical underpinnings. Our design draws on and is discussed in terms of a broad theoretical spectrum, combining both cognitive (dual channel/multimedia theory) and sociocognitive (guided noticing) perspectives. This is an approach we deem suitable to address the complexities of clinical communication. However, the findings need to be interpreted and generalized with respect to several limitations. First, we recruited medical students as participants. This is a limitation because students represent a specific group of clinical actors. Although they have accumulated some clinical experiences in the course of their education, they represent a group with limited medical expertise and experience. Second, we tested the hypotheses exclusively against a specific clinical case, a specialist consultation. From a practical viewpoint, our study was conducted in a laboratory setting in a hospital (ie, in an environment that was relatively stable compared to the hectic, noisy, interrupted, and chaotic contexts of real clinical communication), where learning and teaching additionally depend on the nonlinear interplay of a variety of different factors [57,58]. In real clinical settings, it remains unclear, for example, to what extent clinicians are willing to use annotations and extended phone-based explanations. Similarly, to what extent an additional device (eg, a mobile phone) may contribute to cognitive overload in an environment where multiple sources already claim the full attention of a clinician needs to be explored [59]. A further limitation of our experimental design is that prior knowledge was not evaluated in the form of a test, but only as a self-reported measurement. It also needs to be acknowledged that the negotiated coding may represent a limitation because the widely recognized Guidelines for Reporting Reliability and Agreement Studies (GRRAS) recommend using the mean of 2 or more raters to increase reliability [60]. Although we fully agree with GRRAS, in our case we deem the negotiated coding approach a legitimate technique, as recently proposed in the research literature [53]. Moreover, negotiated coding was strengthened by resolving disagreement and uncertainty with a hand surgeon.

In conclusion, the present study yields interesting insights into the knowledge-based effects of multimodal, phone-based communication, but it cannot provide definite accounts of all the observed phenomena. Future research addressing these questions may (1) use different user groups (eg, residents), (2) use different case representations (eg, MRI), (3) test the system not only in authentic but in real clinical settings (eg, by evaluating prior and post hoc knowledge), and (4) capture knowledge-based effects over a longer period of time.

Acknowledgments

The authors thank all research participants, the departments involved, and Andreas Brenner for their support of this study. The authors also thank the project sponsors and partners, the CTI - the Swiss Confederation’s Innovation Promotion Agency, AMTS, Agfa Healthcare, University Hospital Basel, Hightech Research Center of Cranio-Maxillofacial Surgery University of Basel, and the University of Applied Sciences Northwestern Switzerland for their support of this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [61].

[PDF File (Adobe PDF File), 1MB - jmir_v15i11e263_app1.pdf]

References

1. Parker J, Coiera E. Improving clinical communication: A view from psychology. J Am Med Inform Assoc 2000;7(5):453-461 [FREE Full text] [Medline: 10984464]
2. Coiera E. When conversation is better than computation. J Am Med Inform Assoc 2000;7(3):277-286 [FREE Full text] [Medline: 10833164]
3. Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. BMJ 1998 Feb 28;316(7132):673-676 [FREE Full text] [Medline: 9522794]
4. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. Acad Med 2004 Feb;79(2):186-194. [Medline: 14744724]
5. Fitzpatrick LA, Melnikas AJ, Weathers M, Kachnowski SW. Understanding communication capacity. Communication patterns and ICT usage in clinical settings. J Healthc Inf Manag 2008;22(3):34-41. [Medline: 19267030]
6. Varpio L, Hall P, Lingard L, Schryer CF. Interprofessional communication and medical error: a reframing of research questions and approaches. Acad Med 2008 Oct;83(10 Suppl):S76-S81. [doi: 10.1097/ACM.0b013e318183e67b] [Medline: 18820507]
7. Bezemer J, Cope A, Kress G, Kneebone R. “Do you have another Johan?” Negotiating meaning in the operating theatre. Applied Linguistics Review 2011;2:313-334. [doi: 10.1515/9783110239331.313]
8. Bardram JE, Bossen C. Mobility work: The spatial dimension of collaboration at a hospital. Comput Supported Coop Work 2005 Apr;14(2):131-160. [doi: 10.1007/s10606-005-0989-y]
9. Norman N. Mobile Learning for the NHS: Research Report. Brighton, UK: Epic; 2011. URL: http://www.workforce.southcentral.nhs.uk/libraries__elearning/elearning/mlearning_research_report_2012.aspx [accessed 2013-06-10] [WebCite Cache ID 6HGuBOcY3]
10. Ramesh J, Carter AO, Campbell MH, Gibbons N, Powlett C, Moseley H, et al. Use of mobile phones by medical staff at Queen Elizabeth Hospital, Barbados: evidence for both benefit and harm. J Hosp Infect 2008 Oct;70(2):160-165. [doi: 10.1016/j.jhin.2008.06.007] [Medline: 18701190]
11. Baumgart DC. Smartphones in clinical practice, medical education, and research. Arch Intern Med 2011 Jul 25;171(14):1294-1296. [doi: 10.1001/archinternmed.2011.320] [Medline: 21788549]
12. Franko OI, Tirrell TF. Smartphone app use among medical providers in ACGME training programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
13. Wu RC, Tran K, Lo V, O'Leary KJ, Morra D, Quan SD, et al. Effects of clinical communication interventions in hospitals: a systematic review of information and communication technology adoptions for improved communication between clinicians. Int J Med Inform 2012 Nov;81(11):723-732. [doi: 10.1016/j.ijmedinf.2012.05.014] [Medline: 22727613]
14. Prgomet M, Georgiou A, Westbrook JI. The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review. J Am Med Inform Assoc 2009;16(6):792-801 [FREE Full text] [doi: 10.1197/jamia.M3215] [Medline: 19717793]
15. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
16. Ortega GR, Taksali S, Smart R, Baumgaertner MR. Direct cellular vs. indirect pager communication during orthopaedic surgical procedures: a prospective study. Technol Health Care 2009;17(2):149-157. [doi: 10.3233/THC-2009-0540] [Medline: 19564680]
17. Spiegelman J, Detsky AS. Instant mobile communication, efficiency, and quality of life. JAMA 2008 Mar 12;299(10):1179-1181. [doi: 10.1001/jama.299.10.1179] [Medline: 18334693]
18. Ammenwerth E, Buchauer A, Bludau B, Haux R. Mobile information and communication tools in the hospital. Int J Med Inform 2000 Jan;57(1):21-40. [Medline: 10708253]
19. Horwitz LI, Detsky AS. Physician communication in the 21st century: to talk or to text? JAMA 2011 Mar 16;305(11):1128-1129. [doi: 10.1001/jama.2011.324] [Medline: 21406650]
20. Hanada E, Fujiki T, Nakakuni H, Sullivan CV. The effectiveness of the installation of a mobile voice communication system in a university hospital. J Med Syst 2006 Apr;30(2):101-106. [Medline: 16705994]
21. Knobloch K, Gohritz A, Vogt PM. Cell phone-based multimedia messaging service in reconstructive microsurgery: a novel telemedicine application. Plast Reconstr Surg 2009 Jun;123(6):220e-222e. [doi: 10.1097/PRS.0b013e3181a3f53b] [Medline: 19483557]
22. Ng WH, Wang E, Ng I. Teleradiology multimedia messaging service in the provision of emergency neurosurgical service. In: Kumar S, Krupinski E, editors. Teleradiology. Heidelberg: Springer; 2008:77-87.
23. Tsai HH, Pong YP, Liang CC, Lin PY, Hsieh CH. Teleconsultation by using the mobile camera phone for remote management of the extremity wound: a pilot study. Ann Plast Surg 2004 Dec;53(6):584-587. [Medline: 15602257]

24. Kumar S, Wang EH, Pokabla MJ, Noecker RJ. Teleophthalmology assessment of diabetic retinopathy fundus images: smartphone versus standard office computer workstation. Telemed J E Health 2012 Mar;18(2):158-162. [doi: 10.1089/tmj.2011.0089] [Medline: 22304438]
25. Boyce Z, Gilmore S, Xu C, Soyer HP. The remote assessment of melanocytic skin lesions: a viable alternative to face-to-face consultation. Dermatology 2011;223(3):244-250. [doi: 10.1159/000333363] [Medline: 22095005]
26. Filip M, Linzer P, Šámal F, Tesař J, Herzig R, Školoudík D. Medical consultations and the sharing of medical images involving spinal injury over mobile phone networks. Am J Emerg Med 2012 Jul;30(6):961-965. [doi: 10.1016/j.ajem.2011.05.007] [Medline: 21795006]
27. Mitchell JR, Sharma P, Modi J, Simpson M, Thomas M, Hill MD, et al. A smartphone client-server teleradiology system for primary diagnosis of acute stroke. J Med Internet Res 2011;13(2):e31 [FREE Full text] [doi: 10.2196/jmir.1732] [Medline: 21550961]
28. Ng WH, Wang E, Ng I. Multimedia Messaging Service teleradiology in the provision of emergency neurosurgery services. Surg Neurol 2007 Apr;67(4):338-341. [doi: 10.1016/j.surneu.2006.10.056] [Medline: 17350396]
29. Caruso RD, Postel GC, McDonald CS, Aronson BA, Christensen JD. Software-annotated, digitally photographed, and printed MR images: suitability for publication. Acad Radiol 2002 Mar;9(3):346-351. [Medline: 11887950]
30. Goede PA, Lauman JR, Cochella C, Katzman GL, Morton DA, Albertine KH. A methodology and implementation for annotating digital images for context-appropriate use in an academic health care environment. J Am Med Inform Assoc 2004;11(1):29-41 [FREE Full text] [doi: 10.1197/jamia.M1247] [Medline: 14527971]
31. Jang SH, Kim WY. Defining a new annotation object for DICOM image: a practical approach. Comput Med Imaging Graph 2004 Oct;28(7):371-375. [doi: 10.1016/j.compmedimag.2004.06.004] [Medline: 15464876]
32. Coiera E. Communication spaces. J Am Med Inform Assoc 2013 Sep 4. [doi: 10.1136/amiajnl-2012-001520] [Medline: 24005797]
33. Kilroy DA. Problem based learning. Emerg Med J 2004 Jul;21(4):411-413 [FREE Full text] [doi: 10.1136/emj.2003.012435] [Medline: 15208220]
34. Lam TK, Preketes A, Gates R. Mobile phone photo messaging assisted communication in the assessment of hand trauma. ANZ J Surg 2004 Jul;74(7):598-602. [doi: 10.1111/j.1445-2197.2004.03069.x] [Medline: 15230801]
35. Farber N, Haik J, Liran A, Weissman O, Winkler E. Third generation cellular multimedia teleconsultations in plastic surgery. J Telemed Telecare 2011;17(4):199-202. [doi: 10.1258/jtt.2010.100604] [Medline: 21508079]
36. Clark JM, Paivio A. Dual coding theory and education. Educ Psychol Rev 1991 Sep;3(3):149-210. [doi: 10.1007/BF01320076]
37. Mayer RE. Applying the science of learning to medical education. Med Educ 2010 Jun;44(6):543-549. [doi: 10.1111/j.1365-2923.2010.03624.x] [Medline: 20604850]
38. Paivio A. Mental imagery in associative learning and memory. Psychological Review 1969;76(3):241-263. [doi: 10.1037/h0027272]
39. Baddeley AD, Hitch GJ. Working memory. In: The Psychology of Learning and Motivation: Advances in Research and Theory. New York: Academic Press; 1974:8-89.
40. Pea R, Mills M, Rosen J, Dauber K, Effelsberg W, Hoffert E. The diver project: interactive digital video repurposing. IEEE Multimedia 2004 Jan;11(1):54-61. [doi: 10.1109/MMUL.2004.1261108]
41. Pea RD. Video-as-data and digital video manipulation techniques for transforming learning sciences research, education, and other cultural practices. In: Weiss J, Nolan J, Hunsinger J, Trifonas P, editors. The International Handbook of Virtual Learning Environments. Dordrecht: Springer; 2006:1321-1393.
42. Fletcher J, Tobias S. The multimedia principle. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005:117-134.
43. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005:31-48.
44. Baddeley A. Working memory: theories, models, and controversies. Annual Review of Psychology 2012;63(1):1-29. [doi: 10.1146/annurev-psych-120710-100422]
45. Hegarty M. Multimedia learning about physical systems. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005:447-465.
46. Mayer RE, Gallini JK. When is an illustration worth ten thousand words? Journal of Educational Psychology 1990;82(4):715-726. [doi: 10.1037/0022-0663.82.4.715]
47. Brünken R, Steinbacher S, Schnotz W, Leutner D. Mentale Modelle und Effekte der Präsentations- und Abrufkodalität beim Lernen mit Multimedia. Zeitschrift für Pädagogische Psychologie 2001 Mar 1;15(1):16-27. [doi: 10.1024//1010-0652.15.1.16]
48. Clark HH. Using Language. Cambridge: Cambridge University Press; 1996.
49. Goodwin C. Professional vision. American Anthropologist 1994 Sep;96(3):606-633. [doi: 10.1525/aa.1994.96.3.02a00100]
50. Zahn C, Pea R, Hesse FW, Rosen J. Comparing simple and advanced video tools as supports for complex collaborative design processes. Journal of the Learning Sciences 2010 Jul 19;19(3):403-440. [doi: 10.1080/10508401003708399]

51. Kuokkanen HO, Mulari-Keränen SK, Niskanen RO, Haapala JK, Korkala OL. Treatment of subcapital fractures of the fifth metacarpal bone: a prospective randomised comparison between functional treatment and reposition and splinting. Scand J Plast Reconstr Surg Hand Surg 1999 Sep;33(3):315-317. [Medline: 10505445]
52. Abdon P, Mühlow A, Stigsson L, Thorngren KG, Werner C, Westman L. Subcapital fractures of the fifth metacarpal bone. Arch Orthop Trauma Surg 1984;103(4):231-234. [Medline: 6508512]
53. Garrison DR, Cleveland-Innes M, Koole M, Kappelman J. Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. The Internet and Higher Education 2006 Jan;9(1):1-8. [doi: 10.1016/j.iheduc.2005.11.001]
54. Baguley T. Understanding statistical power in the context of applied research. Appl Ergon 2004 Mar;35(2):73-80. [doi: 10.1016/j.apergo.2004.01.002] [Medline: 15105068]
55. Pimmer C, Pachler N, Genewein U. Reframing clinical workplace learning using the theory of distributed cognition. Acad Med 2013 Sep;88(9):1239-1245. [doi: 10.1097/ACM.0b013e31829eec0a] [Medline: 23887014]
56. Mayer RE, Anderson RB. Animations need narrations: An experimental test of a dual-coding hypothesis. Journal of Educational Psychology 1991;83(4):484-490. [doi: 10.1037/0022-0663.83.4.484]
57. Pimmer C, Pachler N, Genewein U. Contextual dynamics in clinical workplaces: learning from doctor-doctor consultations. Med Educ 2013 May;47(5):463-475. [doi: 10.1111/medu.12130] [Medline: 23574059]
58. Dornan T, Boshuizen H, King N, Scherpbier A. Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. Med Educ 2007 Jan;41(1):84-91. [doi: 10.1111/j.1365-2929.2006.02652.x] [Medline: 17209896]
59. Laxmisan A, Hakimzada F, Sayan OR, Green RA, Zhang J, Patel VL. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. Int J Med Inform 2007;76(11-12):801-811. [doi: 10.1016/j.ijmedinf.2006.09.019] [Medline: 17059892]
60. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud 2011 Jun;48(6):661-671. [doi: 10.1016/j.ijnurstu.2011.01.016] [Medline: 21514934]
61. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

CT: computed tomography
CTA: computed tomography angiography
GRRAS: Guidelines for Reporting Reliability and Agreement Studies
NCCT: noncontrast computed tomography

Edited by G Eysenbach; submitted 17.06.13; peer-reviewed by I Mavridis, M Kalz; comments to author 12.07.13; revised version received 17.09.13; accepted 07.10.13; published 27.11.13.

Please cite as:
Pimmer C, Mateescu M, Zahn C, Genewein U
Smartphones as Multimodal Communication Devices to Facilitate Clinical Knowledge Processes: Randomized Controlled Trial
J Med Internet Res 2013;15(11):e263
URL: http://www.jmir.org/2013/11/e263/
doi:10.2196/jmir.2758
PMID:24284080

Original Paper

Net Improvement of Correct Answers to Therapy Questions After PubMed Searches: Pre/Post Comparison

Kathleen Ann McKibbon[1*], MLS, PhD; Cynthia Lokker[1*], MSc, PhD; Arun Keepanasseril[1], MDS, MSc; Nancy L Wilczynski[1*], MSc, PhD; R Brian Haynes[1*], MSc, PhD, OC

McMaster University, Department of Clinical Epidemiology and Biostatistics, Health Information Research Unit, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Kathleen Ann McKibbon, MLS, PhD
McMaster University
Department of Clinical Epidemiology and Biostatistics
Health Information Research Unit
CRL Building
1280 Main Street West
Hamilton, ON, L8S 4K1
Canada
Phone: 1 9055259140 ext 22803
Fax: 1 9055268447
Email: mckib@mcmaster.ca

Abstract

Background: Clinicians search PubMed for answers to clinical questions although it is time consuming and not always successful.

Objective: To determine if PubMed used with its Clinical Queries feature to filter results based on study quality would improve search success (more correct answers to clinical questions related to therapy).

Methods: We invited 528 primary care physicians to participate, 143 (27.1%) consented, and 111 (21.0% of the total and 77.6% of those who consented) completed the study. Participants answered 14 yes/no therapy questions and were given 4 of these (2 originally answered correctly and 2 originally answered incorrectly) to search using either the PubMed main screen or PubMed Clinical Queries narrow therapy filter via a purpose-built system with identical search screens. Participants also picked 3 of the first 20 retrieved citations that best addressed each question. They were then asked to re-answer the original 14 questions.

Results: We found no statistically significant differences in the rates of correct or incorrect answers using the PubMed main screen or PubMed Clinical Queries. The rate of correct answers increased from 50.0% to 61.4% (95% CI 55.0%-67.8%) for the PubMed main screen searches and from 50.0% to 59.1% (95% CI 52.6%-65.6%) for Clinical Queries searches. These net absolute increases of 11.4% and 9.1%, respectively, included previously correct answers changing to incorrect at a rate of 9.5% (95% CI 5.6%-13.4%) for PubMed main screen searches and 9.1% (95% CI 5.3%-12.9%) for Clinical Queries searches, combined with increases in the rate of being correct of 20.5% (95% CI 15.2%-25.8%) for PubMed main screen searches and 17.7% (95% CI 12.7%-22.7%) for Clinical Queries searches.

Conclusions: PubMed can assist clinicians answering clinical questions with an approximately 10% absolute rate of improvement in correct answers. This small increase includes more correct answers partially offset by a decrease in previously correct answers.

(J Med Internet Res 2013;15(11):e243) doi:10.2196/jmir.2572

KEYWORDS

information services; information storage and retrieval; Internet; Medline; physicians; primary health care

Introduction

Medline indexed 760,903 new articles in 2012, bringing their total to just under 20 million articles. The number of journals indexed by Medline has grown by 50% in the past 20 years [1]. During 2012, 2.2 billion Medline searches were done. Although quantification of this information overload in the health care literature is limited [2], it is widely perceived as an obstacle for physicians practicing evidence-based medicine and searching for answers to their clinical questions [3].

The 6S pyramid of evidence from health care research describes a range of tools and resources to assist physicians in accessing or retrieving relevant research evidence. The pyramid is structured so that original studies form the base and are topped by, in ascending order of clinical usefulness, synopses of studies, syntheses (systematic reviews), synopses of syntheses, summaries (evidence-driven online texts), and systems (eg, clinical decision support systems) [4]. In addition to published evidence, colleagues and textbooks are often first-line information resources used by physicians [5-7] because these give answers most efficiently [7]. Although higher levels of evidence (eg, meta-analyses or clinical summaries) are more clinically useful, this kind of information is not available for many clinical questions and physicians often need to search the primary literature [8]. Physicians report substantial use of PubMed or Medline through other vendors. Davies [9] reported that 81% of US physicians in 2007, 77% of UK physicians, and 76% of Canadian physicians used PubMed or Medline occasionally or often to support their practices.

Research has shown that published original studies and reviews can provide clinicians with answers to their clinical questions [10-13] and lead to changes in patient care [13-15]. Medline searches helped medical and nurse practitioner students answer simulated clinical questions [12]. A virtual library containing Medline, textbooks, and clinical guidelines helped physicians find relevant information on clinical questions [10]. A study of 33 emergency department residents, however, found that Google search results gave participants a false sense of security, resulting in a dramatic increase in confidence in their answers. Google searches translated into more correct responses to simulated questions, but also slightly more wrong answers after searching [16].

Other studies have reported negative effects of information searching on physician responses to clinical questions. McKibbon and Fridsma [17] found that 11% of answers to clinical questions went from correct before searching to incorrect after searching when clinicians used their preferred online resources. Hersh and colleagues [12,18] found rates of correct-to-incorrect answers of 4.5% and 10.5% using Medline in 2 studies.

Search filters have been developed to help clinicians search the primary literature. These filters are rigorously developed and validated to increase the yield of clinically relevant articles based on research methods or clinical content. The Health Information Research Unit at McMaster University has developed filters for detecting primary studies for therapy [19,20], diagnosis [21,22], economics [23], prognosis [24,25], etiology [26,27], systematic reviews [28], and studies in mental health [29]. A number of filters have been made available on PubMed in the Clinical Queries interface [30] and the Special Queries feature [31]. A recent study comparing search retrieval from the main PubMed screen and from Clinical Queries found that Clinical Queries returned fewer studies, more of which were methodologically sound [32].

The objective of this pragmatic study was to determine if differences exist in the rate of correct answers to clinical questions when primary care physicians use the PubMed main screen or the Clinical Queries feature of PubMed for searches. Specifically, do searches done by primary care physicians through the PubMed main screen or through Clinical Queries give different rates of correctness of answers to clinical questions related to therapy?

Methods

Standardized Questions

To assess if PubMed provided correct answers to clinical questions related to therapies, standard clinical questions with answers based on recent systematic reviews were developed. The reviews were selected from a database of clinical research from 125 journals preappraised for methodological rigor [33] and rated by a worldwide panel of practicing clinicians for relevance to clinical practice and newsworthiness. Reviews relevant to general practice from the first 6 months of 2011, with clinical relevance and newsworthiness ratings >5 of 7, were assessed to determine whether they reported a definitive answer to the clinical question at hand.

In all, 24 standard questions were devised and iteratively tested on 3 physicians. The physicians were 2 experienced general practitioners and 1 experienced general internist. A fourth general internist also reviewed the questions. The physicians provided input on clinical applicability, perceived difficulty of the question, and relevance to practice for each question. Revised questions were then piloted on 2 general practitioners who provided further feedback. Questions were dropped if they were perceived by the clinicians as being too difficult or easy to answer, not relevant to general practice, or if the answer was perceived to be controversial. The remaining 14 questions are presented in Table 1.

Table 1. Standardized questions provided to general practitioners based on systematic reviews published in early 2011.

Question	Evidence-based answer
1. In adults wishing to quit smoking, is varenicline (Champix) better than bupropion in terms of successful smoking cessation? [34]	Yes
2. Should antidepressants be prescribed for patients >18 years who are diagnosed with minor/subthreshold depression according to standardized criteria? [35]	No
3. In a middle-aged patient who is at high risk for cardiovascular events, does clopidogrel plus aspirin provide safer and more effective protection from cardiovascular events than aspirin alone? [36]	No
4. Over the long term, can daily low-dose aspirin reduce mortality caused by a range of cancers? [37]	Yes
5. Does estrogen therapy increase the risk of kidney stones in otherwise healthy postmenopausal women (>60 years)? [38]	Yes
6. Can maternal depression during pregnancy lead to preterm birth and low birth weight? [39]	Yes
7. Is it safe and effective to progressively increase statin therapy intensity to lower LDL[a] levels and reduce the risk of occlusive vascular events in patients with high LDL levels? [40]	Yes
8. Does dietary supplementation with folic acid to lower homocysteine levels prevent cardiovascular events in high-risk adults? [41]	No
9. For a patient at high risk of cardiovascular events and who is concerned about erectile dysfunction, can you prescribe ACE[b]-inhibitors, angiotensin receptor blockers, or calcium channel blockers without worrying about his sexual functioning? [42]	Yes
10. Compared to other antihypertensive drugs, is hydrochlorothiazide 12.5 to 25 mg/day suitable as first-line drug therapy for the treatment of adult hypertension? [43]	No
11. For an adult patient with type 2 diabetes who needs thiazolidinedione treatment, is pioglitazone a safer treatment than rosiglitazone? [44]	Yes
12. Does treatment of periodontal disease (simple dental scaling and root planing) in pregnant women reduce their risk of preterm delivery? [45]	No
13. In patients with chronic back pain caused by disk degeneration, does spinal fusion surgery result in better long-term benefits than nonsurgical approaches? [46]	No
14. Should I advise patients with asthma to double their regular dose of inhaled corticosteroids as a first step in dealing with an exacerbation? [47]	No

[a]LDL: Low Density Lipoprotein

[b]ACE: angiotensin-converting enzyme

Recruitment

Practicing physicians self-identified as general practitioners, family practitioners, or primary care general internal medicine practitioners who were registered with the McMaster Online Rating of Evidence (MORE) [48] system were emailed invitations to participate in the online research study. Invitations were sent to 528 physicians in November 2011 with up to 2 reminders sent by the end of January 2012. Participants were provided with certification of 1 hour of continuing medical education credit for completing the study.

Survey

Participants were sent an Internet link to the survey that required them to sign into our information production system of high-quality clinical articles using their system passwords, which started the task (Figure 1). After providing consent, physicians were asked to answer the 14 clinical questions with a yes or no answer (Table 1). They were then asked to search for information on 4 of the questions (Figure 2). The 4 questions included 2 that they had initially answered correctly and 2 that they had answered incorrectly; we did not indicate to the participant if his or her answers were correct. Three separate computer-generated randomizations were involved: (1) questions for searching were selected randomly, (2) the questions were sent to PubMed main screen or Clinical Queries randomly (1 correct and 1 incorrect in each), and (3) the order in which the clinicians searched was randomized. The 2 interfaces were conduits to the PubMed search system and all the search algorithms functioned in their usual manner; the entered terms were passed into PubMed with or without Clinical Queries filters.

Because our questions were treatment questions, we used the therapy category of the Research Methodology filter of Clinical Queries. We were interested in clinicians searching for answers to clinical questions; therefore, we used the narrow Clinical Queries. The narrow search filters are designed for clinical care because they retrieve a good proportion of potentially relevant citations while keeping the number of nonrelevant citations to a minimum (sensitivity of 93% and specificity of 97% [19]). The broad clinical filters are designed for researchers and meta-analysts who want to retrieve the highest proportion of relevant citations with less regard to retrieving nonrelevant citations.

For each question, participants were asked to enter search terms into a textbox. If participants were unhappy with the retrieval, they could alter their search terms and submit a new search.

After each of the 4 searches, the first 20 results were presented. The participant was blinded as to which PubMed interface the retrieval came from. They could view the abstract of the article in the same window by selecting the title of the article. Participants were asked to select the top 3 articles most important to forming/supporting their answer. After the 4 searches were performed and articles were selected, the participants were given the 14 questions again and asked to answer them with yes or no. The study was approved by the McMaster University Hamilton Health Sciences/Faculty of Health Sciences Research Ethics Board.

Figure 1. Entry screens asking for answers to 14 clinical questions. Each participant completed this task twice (before and after the search process).

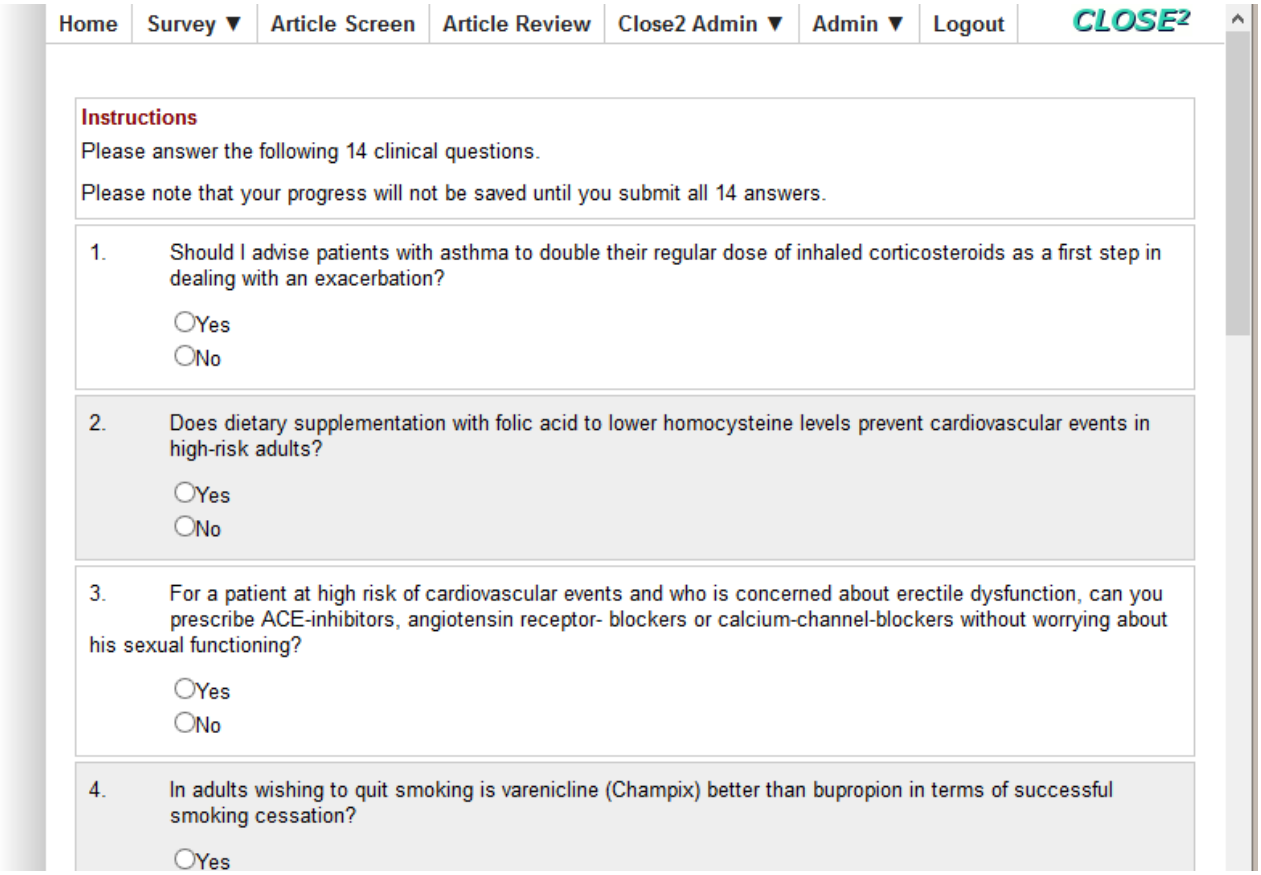
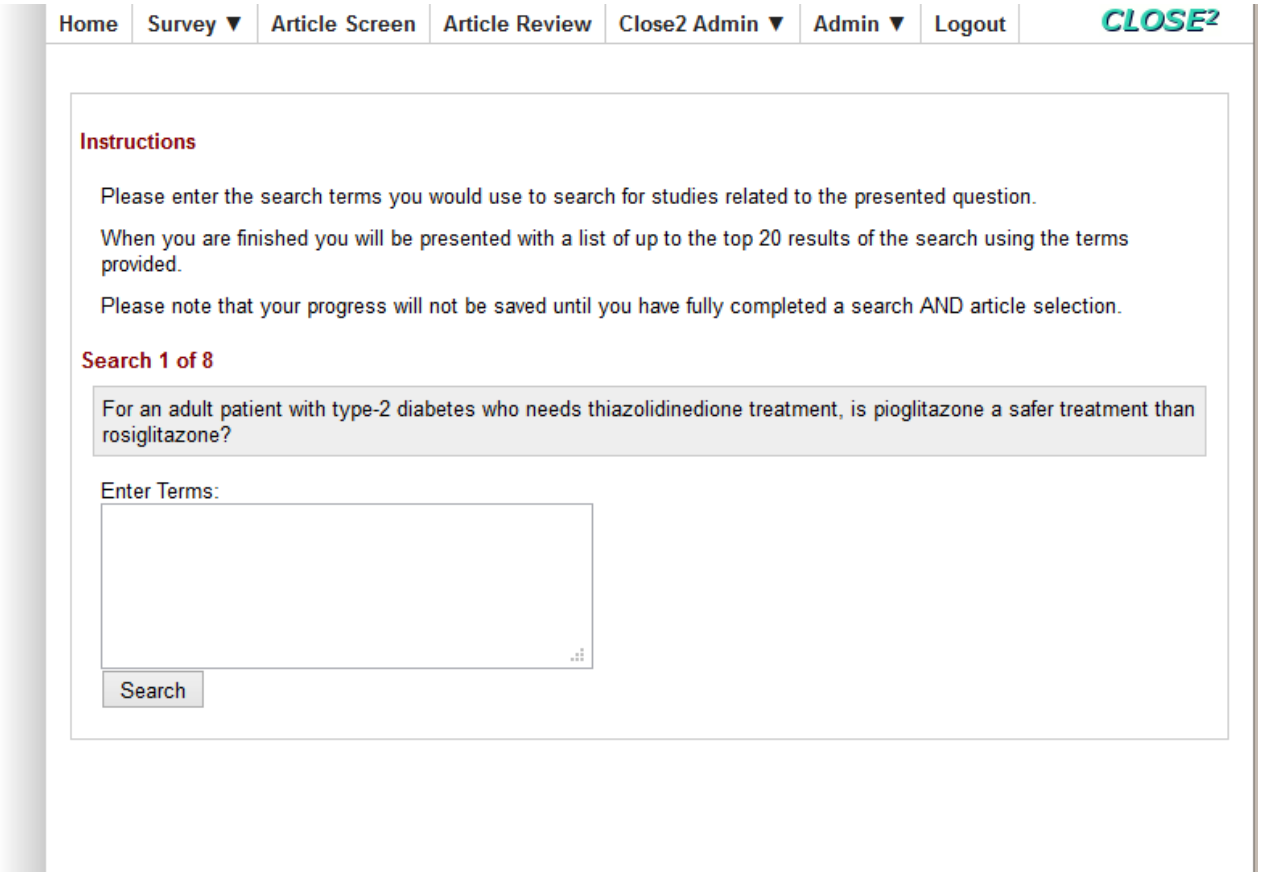


Figure 2. Term entry screen for both searching tasks.



Statistical Analysis

The primary outcome of the study was the difference in the proportion of correct answers before and after searching. Secondary outcomes were the proportion of questions searched that went from incorrect to correct and correct to incorrect, the proportion of questions without searches that went from correct to incorrect and incorrect to correct, and the time taken to complete the project tasks.

Based on previous studies, starting proportions of answers to clinical questions were 27% correct (n=557) [10] and 40% correct (n=46) [17]. These studies found a rate of answers going from correct to incorrect of 7% [18] and 11% [17], respectively.

We anticipated an approximately 10% change in correct to incorrect answers; therefore, we set a 5% absolute difference between search modes as clinically interesting. This gave us a sample size of 522 searches for the correct group and 459 for the incorrect group to detect a 5% difference in search modes with 80% power.

The Mantel-Haenszel test for matched pairs, stratified by participant and by question, was used to determine the odds ratio of changing a response by using Clinical Queries searches versus PubMed main screen searches. A posteriori we recognized that question 6 was a prognosis question rather than a treatment question. Given that the Clinical Queries searches used a therapy filter, we performed our analysis including this question and the sensitivity analysis without this question. In the entire dataset, only 1 of 29 participants (3%) presented with question 6 changed their answer (correct to incorrect with Clinical Queries).

Study Quality

Articles selected by the participants as being relevant to answering their question were independently assessed in duplicate for methodological criteria outlined below. Disagreements were resolved through consensus.

A therapy study is methodologically sound if it meets these 3 criteria:

1. Random allocation of participants to comparison groups;
2. Outcome assessment of at least 80% of those entering the investigation accounted for in one major analysis at any given follow-up assessment; and
3. Analysis consistent with study design.

A systematic review of therapy studies is methodologically sound if it meets these 6 criteria:

1. Explicit statement of clinical topic;
2. Question refers to treatment;
3. Methods are described in report body (not just the abstract);
4. More than one major database searched or Cochrane CENTRAL searched;
5. Explicit inclusion/exclusion criteria; and
6. One or more articles meet criteria set out for therapy studies (listed previously).

Results

Summary

During recruitment, 528 physicians were invited to take part in the study; 143 (27.1%) provided consent, 110 of whom (21.0% of those invited and 77.6% of those who consented) completed the study tasks (24 abandoned the task after the first search and 9 did not perform any searches). Two participants (1.8%) answered all 14 questions correctly; consequently, they were directed to search for only 2 questions. At baseline, participants answered 62.3% (95% CI 59.8%-64.7%) of the questions correctly.

Time to complete the tasks was calculated based on the time the participant signed in to the website until the time they submitted the survey. If the participant logged off without clicking the submit button, the timer continued to count. As such, 16 observations were more than 100 minutes, ranging from 119 to 103,786 minutes (72 days). We selected a cutoff of 100 minutes as a likely point at which the tasks were not completed in 1 sitting. The remaining 95 participants completed the tasks within 6 to 76 minutes (mean 24.5 minutes, 95% CI 21.4-27.5).

Searches

During the study, 440 searches were executed, 222 (50.5%) answered correctly and 218 (49.5%) answered incorrectly initially. For questions selected for searching, baseline responses were 50.0% correct in both groups by design. After searching, responses were correct for 61.4% (95% CI 55.0%-67.8%) of questions for the PubMed main screen group, and 59.1% (95% CI 52.6%-65.6%) for the Clinical Queries group. We found no differences in the rate of answers going from incorrect to correct for the PubMed main screen searches (45/220, 20.5%) compared with the Clinical Queries searches (39/220, 17.7%) (Table 2). Both sets of searches also had an approximate 9% rate of going from correct to incorrect: 21 of 220 (9.5%) for PubMed main screen and 20 of 220 (9.1%) for Clinical Queries (Table 2). Searches resulted in a net gain of 11.4% (95% CI 2.1%-20.4%) in correct answers for PubMed main screen searches and 9.1% (95% CI –0.2% to 18.2%) for Clinical Queries searches.

The odds of changing an answer with a Clinical Queries search versus a PubMed main screen search was not different for questions that were initially correct or initially wrong stratified by user or by question (P>.05) (Table 3). Sensitivity analysis removing question 6 (a prognosis question) did not alter the results.

Table 2. Proportion of the PubMed main screen search group and PubMed Clinical Queries search group that changed answers (correct to incorrect or incorrect to correct) or kept them the same (correct or incorrect).

Search platform	Answers stayed the same				Answers changed			
	Correct		Incorrect		Correct to incorrect		Incorrect to correct	
	Searches, n	% (95% CI)	Searches, n	% (95% CI)	Searches, n	% (95% CI)	Searches, n	% (95% CI)
PubMed main screen (n=220)	90	40.9 (34.4-47.4)	64	29.1 (23.1-35.1)	21	9.5 (5.6-13.4)	45	20.5 (15.2-25.8)
PubMed Clinical Queries (n=220)	91	41.4 (34.9-47.9)	70	31.8 (25.7-38.0)	20	9.1 (5.3-12.9)	39	17.7 (12.7-22.7)

Table 3. Mantel-Haenszel odds ratios for changed answers based on searches through Clinical Queries vs the PubMed main screen.

Starting answer	n	OR (95% CI)
Correct		
Participant	33	0.94 (0.48-1.86)
Question	13	1.14 (0.55-2.35)
Incorrect		
Participant	52	0.79 (0.46-1.37)
Question	13	0.80 (0.47-1.36)

Nonsearched Questions

For questions answered before and after searching but without intervening searches, an average of 65.4% (95% CI 62.8%-68.0%) were correct at baseline, 64.6% (95% CI 62.0%-67.2%) were correct at the end of the study across the 14 questions, 4.0% (95% CI 2.3%-5.6%) went from correct to incorrect, and 3.1% (95% CI 2.2%-4.1%) went from incorrect to correct. There was variability in baseline performance across questions (Table 4). Without searches, the odds of changing an answer from correct to incorrect was lower (OR 0.06, 95% CI 0.05-0.08) than changing from incorrect to correct (OR 0.11, 95% CI 0.08-0.13; P=.002).

Table 4. Responses for questions without searches.

Question	Correct to correct, % (n)	Incorrect to incorrect, % (n)	Correct to incorrect, % (n)	Incorrect to correct, % (n)	Responses, n
1	64 (55)	28 (24)	5 (4)	3 (3)	86
2	83 (70)	8 (7)	1 (1)	7 (6)	84
3	74 (59)	23 (18)	3 (2)	1 (1)	80
4	68 (50)	27 (20)	3 (2)	3 (2)	74
5	25 (16)	66 (43)	5 (3)	5 (3)	65
6	78 (64)	16 (13)	5 (4)	1 (1)	82
7	77 (64)	17 (14)	4 (3)	2 (2)	83
8	82 (75)	10 (9)	4 (4)	3 (3)	91
9	45 (38)	45 (38)	5 (4)	5 (4)	84
10	18 (12)	79 (53)	1 (1)	1 (1)	67
11	64 (53)	31 (26)	2 (2)	2 (2)	83
12	45 (35)	38 (29)	13 (10)	4 (3)	77
13	93 (78)	4 (3)	2 (2)	1 (1)	84
14	45 (33)	49 (36)	3 (2)	4 (3)	74

Study Quality

Clinical Queries were developed to retrieve clinically useful studies based on study design. Therapy filters retrieve citations based on the article being a randomized controlled trial. Therefore, we were interested in determining if the participants identified studies with strong methods (ie, randomized controlled trials or reviews that analyzed randomized controlled trials) when presented with the first 20 retrievals. The participants were asked to identify the 3 most important articles that provided evidence to answer the clinical question they were addressing. Overall, the PubMed main page group tagged 334 articles as important and the Clinical Queries group tagged 321 articles. Table 5 shows the number of treatment articles and systematic review articles tagged as important to the questions asked. Articles selected from the PubMed main screen searches and the Clinical Queries searches did not differ in the number of review treatment articles selected as important or the number of original or review articles that had strong methods.

Table 5. Number of articles with strong methods (randomized controlled trials or systematic reviews of randomized controlled trials) identified as important by the clinician searchers.

Methodologic rigor for articles on treatment identified as influencing decisions	PubMed main screen	Clinical Queries
Proportion of original articles meeting criteria (strong methods for therapy)	45/100 (45.0%)	58/118 (49.1%)
Proportion of review articles meeting criteria (strong methods for therapy)	42/124 (33.8%)	42/124 (33.8%)

Discussion

Although we sought to show that searches with PubMed Clinical Queries were associated with more correct answers to clinical questions than were PubMed main screen searches, we did not find any differences. This may be because we did not meet our sample size of approximately 1000 searches as originally calculated. Another explanation for these results may be the training and experience of the study participants. All were practicing clinicians and were registered with the MORE system, wherein they evaluate and rate clinician-ready health research studies. Also, this study was done on the Internet. The participants likely had strong computer and Internet skills and were probably skilled users of PubMed and the clinical research literature. Therefore, the study participants may be the least likely group of clinicians who could benefit from using Clinical Queries. Naïve users or new clinicians, such as interns and residents, or those clinicians less skilled at the assessment and application of research findings may derive more benefit from the Clinical Queries searches. Time is also a major factor in seeking answers to clinical questions. If the time to seek answers had been more tightly constrained in the study (we did not have time limits on the tasks) we may have found a larger difference between the correctness of the answers found with standalone PubMed searches and searches using Clinical Queries.

However, this study does show that PubMed, either on its own or using Clinical Queries, helps clinicians answer clinical questions with increased accuracy. For questions that clinicians answered twice without searching, the rate of correct answers stayed the same (65.4% correct at first answer and 64.6% correct on second answer). With searching, clinicians improved their rate of correct replies. Their answers went from 50.0% correct

(set by the study) to 60.2% correct (59.1% for PubMed Clinical Queries searches and 61.4% for PubMed main screen, *P*=.60).

Our findings are consistent with other studies that found use of information resources is associated with an increase in accuracy of clinical answers [10,12,17,18]. This increase is often in the range of an absolute 10% improvement. However, the increase in the number of correct answers with searching is almost always a combination of approximately 20% of answers going from an initial incorrect answer to correct at the same time as a 10% rate going from an initial correct answer to an answer that is incorrect.

We also found some change in answers for questions that were not the basis of searches in this study. The steady state of the study participants being correct approximately 65% of the time was almost balanced with 4% of the questions going from correct to incorrect and 3% going from incorrect to correct. This phenomenon of changing answers should be taken into account for studies that are based on outcomes of correct answers to clinical questions.

We have shown that complex searching studies with multiple tasks can be done through the Internet and we were able to recruit clinicians for searching trials. Our participants spent an average of 25 minutes online. During this time, they answered 14 yes/no questions twice, completed 4 PubMed searches, and selected articles of importance to clinical questions. Our methods were strengthened in that we blinded participants to the purpose of the study, kept the clinicians blinded to their initial answers and whether they were using Clinical Queries or not, and performed blinded and duplicate readings in the assessment of the methodological strength of the original and review articles on treatment. We also randomized 3 procedures (choice of question to be searched, order of using PubMed main screen or Clinical Queries searches, and questions that were sent to the 2 searching methods).

The questions we used in the study were based on strong evidence from current systematic reviews, and they were pretested with various physician groups. However, despite these strengths, our questions were not questions that arose in the participants' daily practices.

Future research needs to be done to improve the quality of search tools and their ability to maximize the correct answers to questions while minimizing the answers that go from an initial correct answer to an incorrect answer. Focusing on specific groups of clinicians (eg, those in early years of practice or those with less experience assessing and applying research findings) or in certain situations (eg, constrained time or posing difficult questions outside the domain of the clinician) may also address the potential for automated assistance of PubMed searching. Other research has shown that an interface in PubMed leads to better question answering if the search entry screen required clinicians to enter concepts related to patients or populations, intervention, comparison, and outcome (PICO) aspects of the questions. [49] Comparisons across systems are also warranted, taking into account quality (eg, Google), access (eg, clinicians working inside and outside academic institutions), and cost (eg, UptoDate).

We have shown that complex studies of searching can be done through the Internet. We also have reinforced that clinician searching in PubMed produces an absolute improvement of approximately 10% in clinician ability to correctly answer clinical questions. This 10% improvement is consistent with other similar studies [10,12,17,18] and includes an absolute improvement (incorrect answers to correct answers) of approximately 20% and a decrement of approximately 10% (correct answers to incorrect answers).

Acknowledgments

We thank Nicholas Hobson for computer programming our search interface and data capture. The study was funded by Canadian Institutes of Health Research, MOP 86465. The sponsors did not play a role in the study.

Conflicts of Interest

PubMed Clinical Queries search filters for therapy and reviews were produced by KAM, NLW, and RBH and the Health Information Research Unit, McMaster University.

References

1. US National Library of Medicine, National Institutes of Health. 2012. MEDLINE/PubMed Resources: Key MEDLINE Indicators URL: http://www.nlm.nih.gov/bsd/bsd_key.html [accessed 2013-02-01] [WebCite Cache ID 6E70bQYc5]
2. Hall A, Walton G. Information overload within the health care system: a literature review. Health Info Libr J 2004 Jun;21(2):102-108. [doi: 10.1111/j.1471-1842.2004.00506.x] [Medline: 15191601]
3. Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. Int J Med Inform 2003 Aug;71(1):9-15. [Medline: 12909153]
4. DiCenso A, Bayley L, Haynes RB. ACP Journal Club. Editorial: Accessing preappraised evidence: fine-tuning the 5S model into a 6S model. Ann Intern Med 2009 Sep 15;151(6):JC3-2, JC3. [Medline: 19755349]
5. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc 2006 Jan;94(1):55-60 [FREE Full text] [Medline: 16404470]
6. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10. [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]

7. Dwairy M, Dowell AC, Stahl JC. The application of foraging theory to the information searching behaviour of general practitioners. BMC Fam Pract 2011;12:90 [FREE Full text] [doi: 10.1186/1471-2296-12-90] [Medline: 21861880]
8. Koonce TY, Giuse NB, Todd P. Evidence-based databases versus primary medical literature: an in-house investigation on their optimal use. J Med Libr Assoc 2004 Oct;92(4):407-411 [FREE Full text] [Medline: 15494755]
9. Davies KS. Physicians and their use of information: a survey comparison between the United States, Canada, and the United Kingdom. J Med Libr Assoc 2011 Jan;99(1):88-91 [FREE Full text] [doi: 10.3163/1536-5050.99.1.015] [Medline: 21243061]
10. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321 [FREE Full text] [doi: 10.1197/jamia.M1717] [Medline: 15684126]
11. Westbrook JI, Gosling AS, Westbrook MT. Use of point-of-care online clinical evidence by junior and senior doctors in New South Wales public hospitals. Intern Med J 2005 Jul;35(7):399-404. [doi: 10.1111/j.1445-5994.2005.00836.x] [Medline: 15958109]
12. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Rose L, Friedman CP. Factors associated with successful answering of clinical questions using an information retrieval system. Bull Med Libr Assoc 2000 Oct;88(4):323-331 [FREE Full text] [Medline: 11055299]
13. Gorman PN, Ash J, Wykoff L. Can primary care physicians' questions be answered using the medical journal literature? Bull Med Libr Assoc 1994 Apr;82(2):140-146 [FREE Full text] [Medline: 7772099]
14. Magrabi F, Coiera EW, Westbrook JI, Gosling AS, Vickland V. General practitioners' use of online evidence during consultations. Int J Med Inform 2005 Jan;74(1):1-12. [doi: 10.1016/j.ijmedinf.2004.10.003] [Medline: 15626631]
15. Schilling LM, Steiner JF, Lundahl K, Anderson RJ. Residents' patient-specific clinical questions: opportunities for evidence-based learning. Acad Med 2005 Jan;80(1):51-56. [Medline: 15618093]
16. Krause R, Moscati R, Halpern S, Schwartz DG, Abbas J. Can emergency medicine residents reliably use the internet to answer clinical questions? West J Emerg Med 2011 Nov;12(4):442-447 [FREE Full text] [doi: 10.5811/westjem.2010.9.1895] [Medline: 22224135]
17. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. J Am Med Inform Assoc 2006 Dec;13(6):653-659 [FREE Full text] [doi: 10.1197/jamia.M2087] [Medline: 16929042]
18. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [Medline: 11971889]
19. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. BMJ 2005 May 21;330(7501):1179 [FREE Full text] [doi: 10.1136/bmj.38446.498542.8F] [Medline: 15894554]
20. Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc 2006 Jan;94(1):41-47 [FREE Full text] [Medline: 16404468]
21. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. BMJ 2004 May 1;328(7447):1040 [FREE Full text] [doi: 10.1136/bmj.38068.557998.EE] [Medline: 15073027]
22. Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. BMC Med 2005;3:7 [FREE Full text] [doi: 10.1186/1741-7015-3-7] [Medline: 15796772]
23. McKinlay RJ, Wilczynski NL, Haynes RB, Hedges Team. Optimal search strategies for detecting cost and economic studies in EMBASE. BMC Health Serv Res 2006;6:67 [FREE Full text] [doi: 10.1186/1472-6963-6-67] [Medline: 16756662]
24. Wilczynski NL, Haynes RB. Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. J Am Med Inform Assoc 2005;12(4):481-485 [FREE Full text] [doi: 10.1197/jamia.M1752] [Medline: 15802476]
25. Wilczynski NL, Haynes RB, Hedges Team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. BMC Med 2004 Jun 9;2:23 [FREE Full text] [doi: 10.1186/1741-7015-2-23] [Medline: 15189561]
26. Haynes RB, Kastner M, Wilczynski NL, Hedges Team. Developing optimal search strategies for detecting clinically sound and relevant causation studies in EMBASE. BMC Med Inform Decis Mak 2005;5:8 [FREE Full text] [doi: 10.1186/1472-6947-5-8] [Medline: 15784134]
27. Wilczynski NL, Haynes RB, Hedges Team. Developing optimal search strategies for detecting clinically sound causation studies in MEDLINE. AMIA Annu Symp Proc 2003:719-723 [FREE Full text] [Medline: 14728267]
28. Wilczynski NL, Haynes RB, Team Hedges. Optimal search strategies for identifying mental health content in MEDLINE: an analytic survey. Ann Gen Psychiatry 2006;5:4 [FREE Full text] [doi: 10.1186/1744-859X-5-4] [Medline: 16556313]
29. Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. BMJ 2005 Jan 8;330(7482):68 [FREE Full text] [doi: 10.1136/bmj.38336.804167.47] [Medline: 15619601]

30. NCBI. 2013. PubMed Clinical Queries URL: http://www.ncbi.nlm.nih.gov/pubmed/clinical [accessed 2013-01-24] [WebCite Cache ID 6DuxZ1ZWe]
31. US National Library of Medicine, National Institutes of Health. PubMed Special Queries URL: http://www.nlm.nih.gov/bsd/special_queries.html [accessed 2013-01-25] [WebCite Cache ID 6DwE2WUJo]
32. Lokker C, Haynes RB, Wilczynski NL, McKibbon KA, Walter SD. Retrieval of diagnostic and treatment studies for clinical use through PubMed and PubMed's Clinical Queries filters. J Am Med Inform Assoc 2011;18(5):652-659 [FREE Full text] [doi: 10.1136/amiajnl-2011-000233] [Medline: 21680559]
33. Health Information Research Unit, McMaster University. 2013. Evidence Updates from the BMJ Evidence Centre URL: http://plus.mcmaster.ca/EvidenceUpdates/ [accessed 2013-01-25] [WebCite Cache ID 6DwEKLHag]
34. Cahill K, Stead LF, Lancaster T. Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst Rev 2011(2):CD006103. [doi: 10.1002/14651858.CD006103.pub5] [Medline: 21328282]
35. Barbui C, Cipriani A, Patel V, Ayuso-Mateos JL, van Ommeren M. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. Br J Psychiatry 2011 Jan;198(1):11-6, sup 1 [FREE Full text] [doi: 10.1192/bjp.bp.109.076448] [Medline: 21200071]
36. Squizzato A, Keller T, Romualdi E, Middeldorp S. Clopidogrel plus aspirin versus aspirin alone for preventing cardiovascular disease. Cochrane Database Syst Rev 2011(1):CD005158. [doi: 10.1002/14651858.CD005158.pub3] [Medline: 21249668]
37. Rothwell PM, Fowkes FG, Belch JF, Ogawa H, Warlow CP, Meade TW. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. Lancet 2011 Jan 1;377(9759):31-41. [doi: 10.1016/S0140-6736(10)62110-1] [Medline: 21144578]
38. Maalouf NM, Sato AH, Welch BJ, Howard BV, Cochrane BB, Sakhaee K, et al. Postmenopausal hormone use and the risk of nephrolithiasis: results from the Women's Health Initiative hormone therapy trials. Arch Intern Med 2010 Oct 11;170(18):1678-1685 [FREE Full text] [doi: 10.1001/archinternmed.2010.342] [Medline: 20937929]
39. Grote NK, Bridge JA, Gavin AR, Melville JL, Iyengar S, Katon WJ. A meta-analysis of depression during pregnancy and the risk of preterm birth, low birth weight, and intrauterine growth restriction. Arch Gen Psychiatry 2010 Oct;67(10):1012-1024 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.111] [Medline: 20921117]
40. Cholesterol Treatment Trialists' (CTT) Collaboration, Baigent C, Blackwell L, Emberson J, Holland LE, Reith C, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet 2010 Nov 13;376(9753):1670-1681 [FREE Full text] [doi: 10.1016/S0140-6736(10)61350-5] [Medline: 21067804]
41. Clarke R, Halsey J, Lewington S, Lonn E, Armitage J, Manson JE, B-Vitamin Treatment Trialists' Collaboration. Effects of lowering homocysteine levels with B vitamins on cardiovascular disease, cancer, and cause-specific mortality: Meta-analysis of 8 randomized trials involving 37 485 individuals. Arch Intern Med 2010 Oct 11;170(18):1622-1631. [doi: 10.1001/archinternmed.2010.348] [Medline: 20937919]
42. Baumhäkel M, Schlimmer N, Kratz M, Hackett G, Hacket G, Jackson G, et al. Cardiovascular risk, drugs and erectile function--a systematic analysis. Int J Clin Pract 2011 Mar;65(3):289-298. [doi: 10.1111/j.1742-1241.2010.02563.x] [Medline: 21314866]
43. Messerli FH, Makani H, Benjo A, Romero J, Alviar C, Bangalore S. Antihypertensive efficacy of hydrochlorothiazide as evaluated by ambulatory blood pressure monitoring: a meta-analysis of randomized trials. J Am Coll Cardiol 2011 Feb 1;57(5):590-600. [doi: 10.1016/j.jacc.2010.07.053] [Medline: 21272751]
44. Loke YK, Kwok CS, Singh S. Comparative cardiovascular effects of thiazolidinediones: systematic review and meta-analysis of observational studies. BMJ 2011;342:d1309 [FREE Full text] [Medline: 21415101]
45. Polyzos NP, Polyzos IP, Zavos A, Valachis A, Mauri D, Papanikolaou EG, et al. Obstetric outcomes after treatment of periodontal disease during pregnancy: systematic review and meta-analysis. BMJ 2010;341:c7017 [FREE Full text] [Medline: 21190966]
46. Brox JI, Nygaard Ø, Holm I, Keller A, Ingebrigtsen T, Reikerås O. Four-year follow-up of surgical versus non-surgical therapy for chronic low back pain. Ann Rheum Dis 2010 Sep;69(9):1643-1648 [FREE Full text] [doi: 10.1136/ard.2009.108902] [Medline: 19635718]
47. Quon BS, Fitzgerald JM, Lemière C, Shahidi N, Ducharme FM. Increased versus stable doses of inhaled corticosteroids for exacerbations of chronic asthma in adults and children. Cochrane Database Syst Rev 2010(12):CD007524. [doi: 10.1002/14651858.CD007524.pub3] [Medline: 21154378]
48. Health Information Research Unit, McMaster University. 2013. McMaster Online Rating of Evidence (MORE) Clinical Relevance Online Rating System URL: http://hiru.mcmaster.ca/more_new/ [accessed 2013-01-25] [WebCite Cache ID 6DwEPhPxd]
49. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]

Abbreviations

MORE: McMaster Online Rating of Evidence
PICO: populations, intervention, comparison, and outcome

Edited by G Eysenbach; submitted 10.02.13; peer-reviewed by M Rethlefsen, L Lafrado, X Zhang; comments to author 24.04.13; revised version received 04.09.13; accepted 11.09.13; published 08.11.13.

Please cite as:
McKibbon KA, Lokker C, Keepanasseril A, Wilczynski NL, Haynes RB
Net Improvement of Correct Answers to Therapy Questions After PubMed Searches: Pre/Post Comparison
J Med Internet Res 2013;15(11):e243
URL: http://www.jmir.org/2013/11/e243/
doi:10.2196/jmir.2572
PMID:24217329

Original Paper

Patients' Reported Reasons for Non-Use of an Internet-Based Patient-Provider Communication Service: Qualitative Interview Study

Cecilie Varsi[1], RN, MS; Deede Gammon[1,2], PhD; Torunn Wibe[3], RN, MS; Cornelia M Ruland[1,4], RN, PhD

[1]Center for Shared Decision Making and Collaborative Care Research, Oslo University Hospital, Oslo, Norway

[2]Norwegian Centre for Integrated Care and Telemedicine, University Hospital in North Norway, Tromsø, Norway

[3]Abildsø Nursing Home, Center for Development of Institutional Care Services, Oslo, Norway

[4]Faculty of Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
Cecilie Varsi, RN, MS
Center for Shared Decision Making and Collaborative Care Research
Oslo University Hospital
Forskningsveien 2b
PO Box 4950 Nydalen
Oslo, 0424
Norway
Phone: 47 23075466
Fax: 47 23075450
Email: cecilie.varsi@rr-research.no

Abstract

Background: The adoption of Internet-based patient–provider communication services (IPPC) in health care has been slow. Patients want electronic communication, and the quality of health care can be improved by offering such IPPCs. However, the rate of enrollment in such services remains low, and the reasons for this are unclear. Knowledge about the barriers to use is valuable during implementation of IPPCs in the health care services, and it can help timing, targeting, and tailoring IPPCs to different groups of patients.

Objective: The goal of our study was to investigate patients' views of an IPPC that they could use from home to pose questions to nurses and physicians at their treatment facility, and their reported reasons for non-use of the service.

Methods: This qualitative study was based on individual interviews with 22 patients who signed up for, but did not use, the IPPC.

Results: Patients appreciated the availability and the possibility of using the IPPC as needed, even if they did not use it. Their reported reasons for not using the IPPC fell into three main categories: (1) they felt that they did not need the IPPC and had sufficient access to information elsewhere, (2) they preferred other types of communication such as telephone or face-to-face contact, or (3) they were hindered by IPPC attributes such as login problems.

Conclusions: Patients were satisfied with having the opportunity to send messages to health care providers through an IPPC, even if they did not use the service. IPPCs should be offered to the patients at an appropriate time in the illness trajectory, both when they need the service and when they are receptive to information about the service. A live demonstration of the IPPC at the point of enrollment might have increased its use.

Trial Registration: ClinicalTrials.gov NCT00971139; http://clinicaltrial.gov/ct2/show/NCT00971139 (Archived by WebCite at http://www.webcitation.org/6KlOiYJrW).

(J Med Internet Res 2013;15(11):e246) doi:10.2196/jmir.2683

KEYWORDS

communication; email; Internet; interviews as topic; patient dropouts; patient non-use; patient preference; professional-patient relations; qualitative research

Introduction

A growing number of Internet-based patient-provider communication services (IPPC) are being offered to patients. These services provide patients the opportunity to have secure email contact with their health care providers over the Internet and can be a valuable supplement to traditional health services. An increasing number of studies indicate that IPPCs can help patients manage their illness better and improve health outcomes [1,2], improve patient centeredness [3], address unmet communication needs in health care [4,5], increase patients' satisfaction [1,6-9], and improve quality of care [9-11]. Thus, better utilization of IPPCs is increasingly becoming part of health care policies [12].

Patients expect access to IPPCs in order to communicate with health care providers [9,13,14]. However, most studies have shown that only a small number of patients who are offered an IPPC actually make use of the service to communicate with their health care providers. A study from four ambulatory practices reported that 3.2% of the patients used an eVisit service [15]. A secure messaging system for diabetics was used by 19% of the patients [2]. An encrypted messaging system was used by 4.3% of parents with chronically ill children [16]. A secure messaging system in internal medicine was used by 31% of the patients [17], and two different secure messaging systems in primary care showed 6% use [18] and 52% use [19]. There is also evidence that patients are concerned about privacy and confidentiality [20,21], security [22], and trust [23], and many patients eventually stop using the service [24,25]. Use of email in the health care sector is not routine [26], even though the interest in using it is steadily increasing [27,28].

The IPPC in the current study is a further development of a secure multicomponent Web-based system for illness management support called WebChoice, which was designed to support cancer patients living at home between treatments and during rehabilitation. WebChoice allows cancer patients to monitor their symptoms and problems, provides individually tailored information and self-management support, and offers IPPC with cancer nurses as well as an e-forum for group discussion with other patients [29]. WebChoice has been tested in a randomized controlled trial (RCT) involving patients with breast or prostate cancer [30], where the patients were randomized into the WebChoice intervention group or a control group that received standard care. Results showed significant group differences in global symptom distress. In addition, patients in the WebChoice group had significant within-group reductions in depression. Self-efficacy and Health Related Quality of Life (HRQoL) significantly deteriorated in the control group, but not in the WebChoice group [30]. Interviews with 10 of the WebChoice users showed that some patients experienced the tool as highly supportive, while others had more ambivalent or conflicting feelings about it [31]; 38% of those who had access to WebChoice used the IPPC component where a study nurse responded to patients' anonymous questions and concerns [32,33]. The IPPC was one of the components of the WebChoice package that patients spent most time using and was also the one they valued most highly [33].

A new study was initiated to determine whether the results from the RCT could be repeated when only the IPPC module was offered to the patients, without the other features of WebChoice. To utilize the full potential of the IPPC as a supplement to traditional health services, we also integrated the IPPC as a part of regular patient care into five hospital specialties and examined the implementation process simultaneously. To adapt the IPPC to patients' needs and care providers' requirements, we used several participatory methods in the design phase of the project, including research-practice networks, focus groups, workshops, heuristic evaluations, and usability testing [34]. Despite our efforts to make the service adaptable to the patients' needs and desires, the IPPC was used by only 22% of the patients to whom it was offered—a participation rate significantly lower than in the previous WebChoice RCT, where patients could communicate anonymously with a study nurse.

This result led the research team to investigate the reasons for non-use in greater detail. A better understanding of user acceptance of IPPCs is essential in order to achieve effective implementation of such services into regular health care. The non-users of the service can provide valuable insights and explanations and thus contribute to the understanding of barriers to use and in the implementation process. We found that very little research has investigated non-use of IPPCs. One of the few studies that have examined patients' reported reasons for non-use of an IPPC is a study among pediatric patients and their families [16]. Little is known about adult patients and their reasons for non-use of IPPC services. One research group investigated patients' barriers to enrollment in a patient portal where one of the components was a secure electronic message system. They found that a lack of awareness of the patient portal or a lack of motivation were the primary barriers to enrollment [22]. Another research group conducted a survey among patients with no experience of e-consultations and found that the most prominent reasons for non-use were that the patients were not aware of the existence of the service, that they preferred to see a doctor, or that their doctor did not offer e-consultations [35].

A better understanding of why patients choose not to use IPPCs is needed for these services to be targeted and adapted to the patients' needs and preferences and subsequently implemented successfully into health care. As argued by Eysenbach, studies of non-use will contribute to the understanding of impact and uptake of eHealth interventions, and therefore should be of great interest to researchers [24]. The aim of the current study was to identify the patients' views of an IPPC and their reported reasons for non-use of the service.

In planning and designing the current study, we examined different models and frameworks suitable for implementation, adoption, and acceptance of new technology in health care [36-40]. These models focus on predicting use of the technology and on how people and organizations adopt and start using the services, while non-users, who are the focus of the current study, are given less attention. In addition, the models focus mainly on health care providers and their organizations and are sparser in the constructs that reflect the patients' point of view. As no existing theory or framework was suitable, we did an open approach to obtain as much richness as possible during the study.

Methods

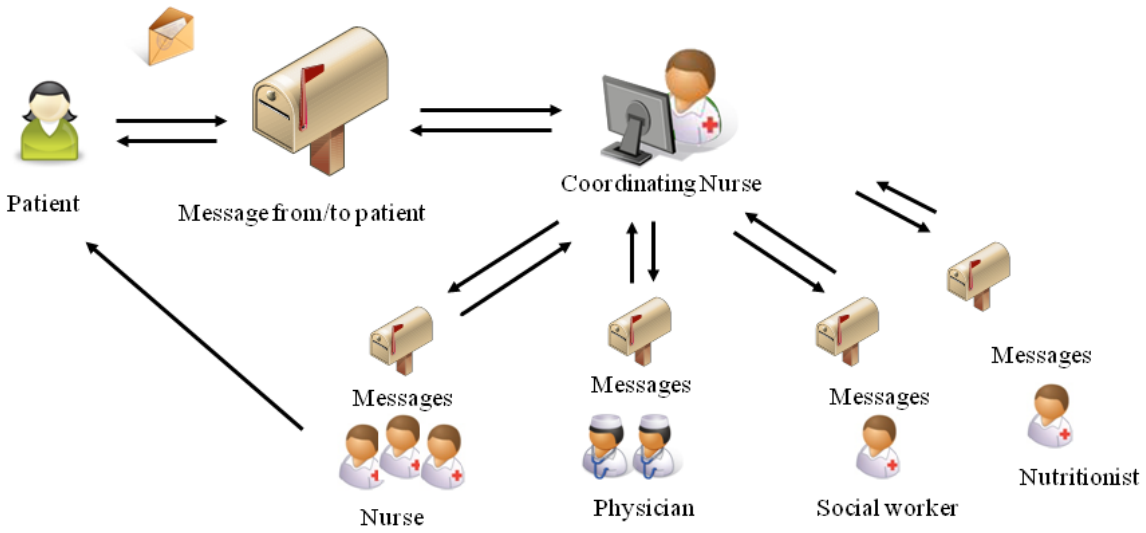
Study Design and Participants

This study is part of a larger study to examine the implementation of the WebChoice IPPC module as part of routine care in five hospital specialties. Participants were recruited at discharge from a hospital stay or an outpatient visit. To be eligible for the main study, participants had to be at least 18 years of age, be able to speak and read Norwegian, have secure Internet access at home using a public key solution for secure electronic identification, and have one of five diagnoses or treatments: (1) liver transplantation, (2) testicular cancer, (3) autologous stem cell transplantation, (4) advanced cancer and participating in a clinical drug trial, or (5) type 1 diabetes. For the current study, the additional inclusion criterion was non-use of the IPPC after they had enrolled in the study and had access to the service for at least 6 months. To get an in-depth understanding of the reasons why the patients did not use the service, we conducted a qualitative study [41] based on individual interviews with non-users of the service. The study was approved by the Regional Committee for Medical and Health Research Ethics and the Data Security Inspectorate in Norway. Written informed consent was obtained from all participants.

Internet-Based Patient–Provider Communication Service

The IPPC in this study is an Internet-based system where patients can send messages to and receive answers from hospital nurses, physicians, nutritionists, and social workers. The IPPC system has a high security level, requiring both patients and health care providers to log into the system by means of strong authentication keys. The IPPC is designed so that patients can get access to advice at the right level of expertise within the same system, without needing to know who the right person to ask is. The message from the patient is received in the mailbox of the coordinating nurse. In this study, the coordinating nurse had expertise on the respective diagnoses and treatments and had access to the patients' medical record at the hospital. The nurse could address the question directly or forward the message to the mailbox of another provider who was in a better position to answer the question (see Figure 1).

Figure 1. IPPC message flow between patients and health care providers.



Recruitment to the Study

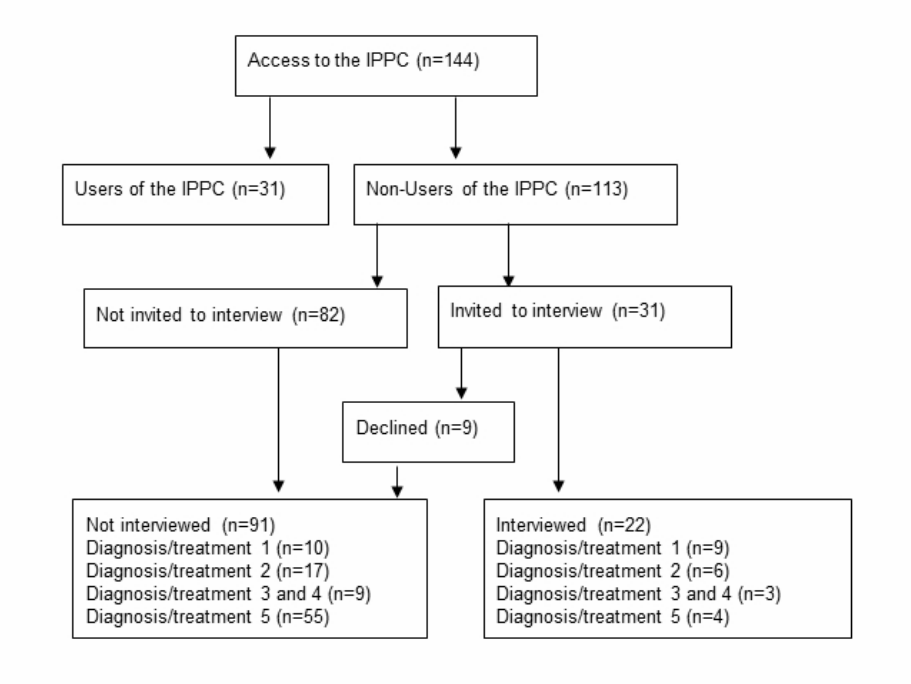
Patients who met the inclusion criteria were invited to participate in the IPPC study by a nurse from the relevant hospital unit. The patients who were interested in participating in the study were referred to a member of the project team who explained the purpose of the study, asked for consent, and filled out the necessary registration forms and baseline questionnaires. The patients also received a brief introduction with information about how to log into and use the IPPC. They were informed that they could send messages with questions and concerns related to their illness and would receive advice and support from hospital health care providers in between and after their hospital admissions. The patients were informed that they could use the IPPC as much as they wanted over the study period, which lasted for 6 months for the patients with diabetes and 8 months for the other patients.

Recruitment to the Interviews

We applied convenience sampling and asked some of the non-users who completed their study period either between January and November 2011, or in December 2012, to participate in an individual interview. Interviewees were selected from two different periods of time, to include patients from all five groups, as the study was conducted at different times at the different study sites. Criteria for participation in the interviews were that the patients had completed their study period. In addition, for practical reasons, they had to live within 180 km (110 miles) of the study hospital or have an appointment there. Due to the number of patients living far away, we subsequently

also approved interviews over the telephone to achieve a sufficient number of interviews. There were 9 patients who declined participation while 22 patients agreed: 3 women and 19 men. The interviews were conducted by the first author either in a meeting room at the hospital, at the first author's workplace, at the responder's workplace, or by telephone. The interviews were recorded with a digital voice recorder and transcribed verbatim (see Figure 2).

Figure 2. Patient recruitment for the interviews. Diagnosis/treatment 1: Liver transplantation. Diagnosis/treatment 2: Testicular cancer. Diagnosis/treatment 3: Autologous stem cell transplantation. Diagnosis/treatment 4: Advanced cancer and participation in clinical drug trials. Diagnosis/treatment 5: Type 1 diabetes.



Measures

A semistructured interview guide based on literature on implementation and non-use of eHealth interventions [16,22,24,35] was developed to gather the non-users' reasons for not using the IPPC. The interview guide contained five themes: (1) the diagnosis and treatment of the patients; (2) how the patients were introduced to the IPPC, their expectations of the service, and what they expected they could use it for; (3) reasons for non-use; (4) factors that could have influenced their use of IPPC; and (5) use of telephone, computers, and the Internet in their everyday life. Information about age was collected from the demographic form that they had completed at the point of inclusion to the main study.

Analysis

The transcripts were analyzed using techniques of qualitative content analysis, inspired by a deductive directed approach, which is applicable when the analysis is investigating assumptions retrieved from prior research [42]. In the first step of the analysis, we identified variables as initial coding categories on the basis of the interview guide. Second, the first author coded the transcripts into the predefined categories using the framework approach of the software program NVivo version 9. Data that could not be coded into the predefined categories were given new categories. Next, the data were discussed within the research team, and the data were coded into the final list of categories. Not all the predefined categories were used, and new categories emerged. Additionally, some categories were combined.

The patients' reasons for not using the IPPC, which fell into three main themes, are presented in the next section. In addition, the patients' information needs, use of telephone, and the Internet in their daily life, as well as their views about the IPPC, are briefly presented.

Results

User Information

In total, 22 patients (19 men and 3 women) participated in the interviews. The patients were between 29 and 71 years old (mean 50, median 51.5). Nine had undergone a liver transplantation, 6 had been diagnosed with testicular cancer, 4 had type 1 diabetes, and 3 had either undergone an autologous stem cell transplantation or participated in a clinical drug trial.

Use of the computer was a daily activity for all the patients in the study, either related to their work or in their leisure time. They were neither novice nor expert users, but reported spending a moderate amount of time on the Internet. More than half of them had searched for health information on the Internet, most at the start of their illness, and the rest had not searched for health information on the Internet at all. Some of them said that they chose to use the telephone, text messages, or email on the basis of what they felt was most appropriate in the specific situation in their daily life. One of the patients said:

> *If I'm in a hurry I send a text message. If I want something confirmed, I'll send an email. And if I have plenty of time, I'll use the phone.*

All of the patients with diabetes had received their diagnoses several years prior to the study. They reported being reconciled with the illness and said that it did not give them cause for concern in their daily life. The patients in the other four groups had in common that they had been quite ill during the disease trajectory, spent time in hospital, and had been through a convalescent period that had lasted for some time after discharge. They also had periods of sick leave before they could return to their ordinary daily lives. Only a few of them recovered quickly. Psychological reactions had also been a part of the illness. They worried about not getting well, about relapses of cancer, and about complications after treatment; sometimes they just had bad days or felt down.

Patients described anxiety before they knew what was wrong with them and felt reassured when they received information about the treatment and follow-up plan. Many of them had to wait for some time before receiving information, but when they got it, they felt that it was sufficient. They also reported being satisfied with the follow-up from health care providers during the hospitalization and after discharge. A few of them experienced that their primary care physician either was too busy or had too little knowledge about their particular situation to offer them sufficient follow-up.

Many of the patients viewed IPPC as a good tool for communication with health care providers at their hospital treatment unit. They felt it was a good service for them, and they were positive about the service. Some of them said that they liked having the opportunity to contact the hospital if they should feel a need for it and that they thought it could be helpful to many people. However, a few of the patients assumed right from the start, when introduced to the IPPC, that they would not use it.

Reasons for Non-Use of the IPPC

Overview

The patients' reported reasons for non-use of the IPPC were consolidated into three different explanations: (1) had sufficient access to information elsewhere, (2) preferred other types of communication, and (3) were prevented by conditions with the IPPC.

Had Sufficient Access to Information Elsewhere

Many of the patients explained their non-use of the IPPC with the fact that they did not have any questions, because they had long-term experience with their illness, they had already received in-depth information about their situation and how to take care of themselves, or they did not have any particular problems after discharge. One of them said:

> *I was given excellent information before I started the chemotherapy. I knew pretty much everything about how I would react and all that. So later on, when my reactions weren't all that bad, I really haven't had many problems.*

The diabetes patients in the study had lived with their disease for many years. They reported that they had limited need of the additional support and follow-up offered by the IPPC since they were past the period in which they had many questions.

Those who had questions handled this in different ways. Patients said that they were used to finding answers to their questions on their own. Some talked to other patients in the same situation as themselves. The majority said they received information and could ask questions either in consultations with the primary care physician or at the hospital. Some said that their questions were not urgent, so they wrote lists of questions to bring to the next encounter. Some patients also reported having a close relationship with their physician, so that they could drop in if they had questions. Some of them also felt that the answers were given to them before they even thought of having any questions. One of the patients said:

> *Why I haven't used it? It simply never occurred to me. I haven't even called my doctor to ask about anything. I haven't really wondered about anything in particular. So it has just never been an issue, it's as simple as that. I feel like I've been given all the answers and all the information I need during my appointments. They took blood tests and X-rays and said everything looks fine.*

Forgetting that the service was available after they returned home from the hospital was another explanation of non-use from many of the patients. A few of the patients reported having too much to think about in the first period after discharge, and some others did not have any questions during the first period at home. When the questions arose at a later point in time, they had forgotten about the service. Lack of time or of motivation was also reported as reasons for non-use. Some patients said that they did not prioritize logging into the IPPC when they had some time available.

Preferred Other Types of Communication

An often expressed reason for non-use of the IPPC was that the patients preferred to talk to the health care providers, either by telephone or in person, and their experience of getting answers quickly enhanced this practice. They also appreciated the opportunity to ask follow-up questions and wanted the health care provider to have that opportunity as well. Even if their question was not an urgent matter, they preferred getting an answer immediately. One patient explained:

> *When a question comes to my mind, I want an answer right away. It's quite possible that I would have gotten an answer just as quickly that way [through the IPPC]. But there's something special about talking to someone. I think that's the most important thing.*

One patient also said that the information seemed more trustworthy when it was explained verbally, instead of written. Some patients also stated that they had felt too ill to use the IPPC when they came home from the hospital:

> *When you're a bit tired, how much can you do? Can you find enough energy to start your computer then? [...] It's been like a part-time job, going to the doctor twice a week. [...] So I don't have the energy to sit down with a research project, to put it plainly. I'm sure it would have been useful for me.*

Prevented by Factors Associated With the IPPC

Not all the patients could remember the moment when they were introduced to the IPPC by the member of the project team. A few of the patients explained the lack of recall with the overwhelming amount of information they had received at the hospital, which had pushed the information about the IPPC into the background. Our study also revealed that not all patients had understood what the IPPC was, how the service was organized, and what they could use it for.

Some of the patients said it would have been more appropriate to introduce the IPPC at an earlier point of time in the disease trajectory. One patient said:

> *I clearly believe that such a service would certainly be perceived as positive for family, friends and especially the patient, in cases where the discovery of the disease is much closer in time than what is the case with me.*

A few of the patients who wanted to use the IPPC, experienced login problems that hindered use. One of the patients said:

> *It was just fine right up until I tried to log in. That just didn't work. [...] Everything crashed. Bam! and both my Internet connection and the program were gone. It was the same every single time. I typed in my password and the program disappeared.*

Some of them also thought that the login procedure was too cumbersome and said that they had forgotten how to log into the IPPC. One of the patients tried unsuccessfully to get in touch with the study's support service. This patient said that when one is ill, one lacks the energy to persevere, and it is therefore easy to give up.

Some patients also said that they did not rely on their questions reaching the correct provider, and therefore preferred to use the telephone. According to one patient:

> *It's better if I call and get a clarification than if I send an email and then walk around wondering 'has he received my email?'*

Some of the study participants also felt an obligation to answer the questionnaires for the main study before it was fair for them to log into the IPPC. Since they lacked the energy to embark on the questionnaires, they also skipped logging in.

None of the patients expressed concerns about the system's security level or concerns that unauthorized persons could get access to the messages in the IPPC. None reported not having the user's manual. No one expressed concerns about bothering health care providers or reported that they felt their questions were too insignificant.

The patients in this study did not have many suggestions for what could have been done to encourage IPPC use. Most of them said they did not know, but some said they would have used it if they had not received sufficient follow-up from their health care providers, or if they had had more questions. One said that a reminder could have helped, and one said that better support with the login problems might have helped.

Discussion

Principal Findings

In this study, we investigated patients' views of an Internet-based patient provider communication service that they could use from home for communication with their hospital health care providers and their reported reasons for non-use of the service. Our results show that the patients' reasons for not using the IPPC can be divided into three main categories: they felt that they had sufficient access to information elsewhere, they preferred other types of communication, or they were prevented by conditions with the IPPC. But even if they did not use the service, they appreciated having the IPPC available. In the discussion of these finding, attentions is directed towards themes perceived as important by the patients and essential in relation to the design and operation of services such as IPPC in the future.

Timing and Targeting

In the current study, the patients were told that they could use the IPPC as they wanted over the study period, with questions related to the disease that they were being treated for at the university hospital. Four of the five patient groups included in the study had recently been treated for serious cancer or liver diseases. They were under close follow-up with frequent encounters and thus had good opportunities to ask questions directly to the health care providers. At the point of inclusion they also were about to be transferred from the university hospital to the primary health care system and thus could have considered it more appropriate to direct their questions to the primary care providers and not to the university hospital where the IPPC was available. The fifth group in the study consisted of experienced diabetes patients who, at the time of inclusion, were past the period in which they had many questions. The current study also revealed that some of the patients felt too sick or exhausted to use the IPPC at the point in time when it was offered to them. This is consistent with other studies that found an association between poor health status and infrequent computer use [25,43]. Another of our earlier studies also found that different user characteristics are associated with different use patterns and that this is important to take into account in the targeting of Web-based support systems to patients with different characteristics [32].

The fact that some of the patients in the current study had not realized what the IPPC actually was, how they could use it, and for what purpose, may be due to information overload. As well as receiving information about the IPPC and the research study at the point of inclusion, the patients had to fill out many documents, and they had received substantial volumes of medical information from their health care providers. In addition, some of the patients said they did not feel in need of the service when they were introduced to it, but when problems or challenges arose later, they had forgotten about the service. Some patients also reported that they were too busy to start using the IPPC. Other studies have also found that this type of system can be forgotten [16,22] or that patients lack the time to use it [16].

Selection of the "wrong" users, that is, those who already are doing so well that they do not feel a need for the technology, is reported to prevent its use [44]. Such attrition can be seen as an indicator of a well-functioning health care service, where the patients get what they need through regular follow-up [44,45]. However, one might ask whether the patients have the sufficient insight into their own situation to evaluate how well they are actually doing [44]. In one of our other studies, we reported how a patient did not understand that his blood values indicated serious kidney failure, but when he used the IPPC this was discovered and action was taken to prevent permanent kidney damage [4]. This indicates that introduction of the IPPC must be targeted to the right patient groups [46] with an appropriate timing in the disease trajectory, both when the patients have questions, when they are capable of using the service, but also when they are receptive to information about the features available and the login procedures. This is consistent with the results in two of our other studies, where patients reported that they wanted the IPPC earlier in the disease trajectory [4,33].

Privacy

Both the current study and earlier studies have shown that patients want to know who is reading their messages and that they want reassurance that the messages reach the right person [20,47]. In the current study, the patients did not know if the coordinating nurse would forward the message to another health care provider or if the message could be read by the whole team connected to the service. Others have also emphasized that patients may not be comfortable sending sensitive information to an office clerk [48]. This may have affected the patients' use of the system and is a possible disadvantage with a triage-based system like this, at least compared to an anonymous service such as WebChoice, where the patients could communicate anonymously with a study nurse [30]. One can presume that patients will direct different types of questions to providers with whom they have already established a trusting relationship, as pointed out by Andreassen et al [23], from the questions they direct to providers who are completely anonymous. The health care providers' knowledge of the patients and access to their medical record is an advantage in some respect, but it can also preclude the anonymity that was offered in the previous WebChoice study. This suggests that the solution one chooses can affect both use and types of questions posted, and that it seems important to take into account the patients' requirement for privacy, anonymity, confidentiality, security, and trust in the design of IPPC services.

Awareness

The results in this study are in line with other studies [22], showing that many of the patients did not use the IPPC because they did not feel a need for it at the time of enrollment. It has been suggested that reminders as "push" factors might encourage patients to use Internet-based interventions [44,45]. In the current study, we tested automated reminders on a small scale at one of the five study sites. The patients who had not made use of the IPPC after 1 or 3 months received a message encouraging them to use the IPPC. However, the non-use rate at this study site remained the same as at the other study sites.

Previous studies have found that patients prefer communication channels related to the type of question they have. Email is often preferred for simple interactions like refilling a prescription and general medical information whereas face-to-face encounters are preferred for more complex interactions like treatment instructions or communication about serious health issues [20,43,49]. However, some studies have also found IPPCs to be suitable for addressing serious concerns, questions, and unmet information needs [4,50]. Therefore, availability of a range of different communication channels is important to meet the different users' needs.

In the current study, patients initiated all communication. Communication initiated by providers might have increased use. Personalizing messages from providers would be in line with proactive follow-up, and targeting and tailoring of messages can be used as a specific strategy for influencing health behaviors [45,46]. Reminders about the availability of the service, combined with disease-specific FAQs (Frequently Asked Questions) may be one way of facilitating meaningful use. However, it is essential to find the right balance between the independent user and the proactive team. In the current study, the access to the IPPC was offered in addition to regular follow-up. If the health care system wants to replace some face-to-face or telephone encounters with Internet-based communication, there is a need for robust organization of the service to ensure quality and safety, and a reimbursement policy must also be in place. There are many ways to utilize IPPCs, and development of comprehensive multifunction patient portals has been proposed because they might increase use [16]. Suggested areas for use are long-term illness management [51], medication management [52], personalizing treatment, tracking patients' progress over time, communicating information about recovery after treatment, enhancing patient education, and giving patients the opportunity to express concerns and receive responses from health care providers [53]. One might think that the self-management activation in the latter services would increase use and satisfaction, but one study revealed that intensity rather than selecting the content of the service did matter [54]. However, many of the patients in the current study stated that they had no questions, so there is a need to ensure that artificial needs are not created. In our previous WebChoice study [33], patients recruited themselves by contacting the research group after receiving information about the study through newspaper advertisements, magazines, and websites, among others. They might thus have been more eager to use the system than the patients in the current study, where all patients fulfilling the inclusion criteria were asked to participate. Some of the patients in the current study said that they knew from the start that they did not want to use the IPPC; they just agreed to participate in order to be helpful and to support the research.

In terms of planning for studies like this, earlier studies have suggested that attrition needs to be taken into account and that Internet-based trials should plan for the worst case scenario of losing half of the participants during the first month of the intervention [45]. Non-use should be taken into account in the same manner.

User-Friendliness

Some of the patients in the current study intended to use the IPPC, but when they tried to log in, they encountered obstacles and gave up. These patients reported having limited amounts of energy, so when they could not get into the system on the first try, many of them did not try again. Earlier studies have also shown how factors related to the systems structure and login procedures affect use [16,55]. In addition, it has been emphasized that a clear and shared understanding of the features available in the communication service is essential and that providing information at the time of enrollment can increase use of the system [22,56]. Other studies have found that lack of training is a barrier to use of email by patients with cancer [57] and that many people will not be able to make use of eHealth technologies without at the same time being offered support in how to use the services [58]. There is thus a need for IPPC systems to be reliable and for sufficient helpdesk services to be offered. Although computer skills are increasing in the population, electronic communication in health care is still new and unfamiliar to many patients. The patients who were offered the IPPC in the current study had experience in using computers, but they had no experience specifically with electronic communication with health care providers. They thus had limited background experience to guide the current use, and one may wonder whether they felt uncertain about deciding what types of questions to ask in the IPPC and what to bring to the face-to-face encounters. We suggest that demonstrating the IPPC live at a computer could be helpful and could increase understanding of the IPPC, empowering patients to make better-informed choices about use of the IPPC.

Reassurance

Contacts with the health care system can be fragile. When patients have established good contact and communication with health care providers during consultations, the threshold for trying a new form of communication can be high, as patients do not know how it will turn out. Some of the patients in the current study said that they would have used the IPPC if the ordinary follow-up program had not worked so well. Findings from earlier studies indicated that patients did not feel a need for new forms of electronic communication when their clinic was responsive to their phone calls [16] and that some people found email more impersonal [16,20]. The patients in this study had different preferences, and some of them said that they would never make use of electronic communication with health care providers, but rather continue using the telephone and face-to-face consultations as before. To make use of the system, patients have to be convinced that the current system is better or can provide an add-on to the regular follow-up services.

Patients in this study did not make use of the IPPC, but they liked having the service available. This corresponds with other studies, in which patients reported being interested in using email to communicate with their health care providers, but their actual IPPC use was low [16]. Some of the patients in the current study stated that they viewed the IPPC as a back-up solution. They said they would have used it if their initial follow-up program had failed, for example, if they had not received satisfactory answers from their health care providers or if new questions or problems had arisen during the disease trajectory. Having the IPPC as an option can provide a sense of reassurance and be of value to patients, even if they choose not to use it. Non-use does not necessarily mean a lack of perceived benefit. Use is not a goal in itself, and the current study has revealed that the reasons why patients do not use the system are not always associated with the system features, but as much with the surrounding factors.

Strengths and Limitations

There are a number of limitations to our study. Our data are qualitative, and we can obtain descriptions of the experiences of those included in the study but cannot statistically compare differences with other reported results. The study was conducted at a single university hospital, and the results may not be representative for other practice settings. Five different groups of patients were included to strengthen the transferability of the study, but four of the five groups consisted of severely ill patients who had recently undergone highly specialized life-saving treatment. These patients were thus quite different from, for example, chronically ill patients who represent many of the patients seeking health care. Three women and 19 men participated in the interviews. The results of the study might have been different if more women had participated. For example, women might bring a different perspective regarding fundamental aspects such as privacy, confidentiality, and security. The gender distribution in this sample reflected the gender distribution in the main study. Retrospectively, we see that women could have been oversampled in order to achieve a more even gender distribution for the interviews. A strength of the study is that the number of interviews was large enough to provide a variety of experiences and to allow sufficient depth in the analyses.

Conclusion

This study offers insights into the reasons why patients who had access to an IPPC did not make use of it. Such knowledge is crucial for implementation of IPPCs to the health care service and can help timing, targeting, and tailoring of the IPPCs to different patient groups. Our findings indicate that patients like having the opportunity to send messages to health care providers through an IPPC, even if they do not make use of the service, and that they think it can be useful to many patients. There is need for more knowledge about how to reach those in need of an IPPC and to determine appropriate timing in the disease trajectory for introducing the service, when patients are receptive to information about how to use the service and for what purposes. Finally, we believe that demonstrating the service on a computer at the time of introduction could have increased the understanding of the service.

Acknowledgments

This study was funded by the Research Council of Norway, grant 191008.

Conflicts of Interest

The last author is the developer of the system but has no ownership rights to the application.

References

1. Goldzweig C, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, et al. Systematic review: secure messaging between providers and patients, and patients' access to their own medical record: evidence on health outcomes, satisfaction, efficiency and attitudes. VA Evidence-based Synthesis Program Reports 2012 Jul:- [FREE Full text] [Medline: 22973584]
2. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: 10.2337/dc08-1771] [Medline: 19366959]
3. Johansen MA, Henriksen E, Horsch A, Schuster T, Berntsen GK. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. part 1: state of the art. J Med Internet Res 2012;14(5):e118 [FREE Full text] [doi: 10.2196/jmir.2214] [Medline: 23032300]
4. Wibe T, Hellesø R, Varsi C, Ruland C, Ekstedt M. How does an online patient-nurse communication service meet the information needs of men with recently diagnosed testicular cancer? ISRN Nurs 2012;2012:260975 [FREE Full text] [doi: 10.5402/2012/260975] [Medline: 23251816]
5. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273. [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]
6. Cornwall A, Moore S, Plant H. Embracing technology: patients', family members' and nurse specialists' experience of communicating using e-mail. Eur J Oncol Nurs 2008 Jul;12(3):198-208. [doi: 10.1016/j.ejon.2007.09.008] [Medline: 18406667]
7. Leong SL, Gingrich D, Lewis PR, Mauger DT, George JH. Enhancing doctor-patient communication using email: a pilot study. J Am Board Fam Pract 2005 May;18(3):180-188 [FREE Full text] [Medline: 15879565]
8. Liederman EM, Lee JC, Baquero VH, Seites PG. Patient-physician Web messaging. The impact on message volume and satisfaction. J Gen Intern Med 2005 Jan;20(1):52-57 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40009.x] [Medline: 15693928]
9. McGeady D, Kujala J, Ilvonen K. The impact of patient-physician Web messaging on healthcare service provision. Int J Med Inform 2008 Jan;77(1):17-23. [doi: 10.1016/j.ijmedinf.2006.11.004] [Medline: 17188564]
10. Rosen P, Kwoh CK. Patient-physician e-mail: an opportunity to transform pediatric health care delivery. Pediatrics 2007 Oct;120(4):701-706 [FREE Full text] [doi: 10.1542/peds.2007-1094] [Medline: 17908755]
11. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375 [FREE Full text] [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
12. Baur S. Helath information technology policy issues. Relevance and implications for eHealth applications. In: Noar SM, editor. eHealth applications: promising strategies for behavior change. New York: Routledge; 2012:246-262.
13. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009 Jun;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]
14. Virji A, Yarnall KS, Krause KM, Pollak KI, Scannell MA, Gradison M, et al. Use of email in a family practice setting: opportunities and challenges in patient- and physician-initiated communication. BMC Med 2006;4:18 [FREE Full text] [doi: 10.1186/1741-7015-4-18] [Medline: 16911780]
15. Jung C, Padman R, Shevchik G, Paone S. Who are portal users vs. early e-Visit adopters? A preliminary analysis. AMIA Annu Symp Proc 2011;2011:1070-1079 [FREE Full text] [Medline: 22195168]
16. Hsiao AL, Bazzy-Asaad A, Tolomeo C, Edmonds D, Belton B, Benin AL. Secure web messaging in a pediatric chronic care clinic: a slow takeoff of "kids' airmail". Pediatrics 2011 Feb;127(2):e406-e413 [FREE Full text] [doi: 10.1542/peds.2010-1086] [Medline: 21220392]
17. Lin CT, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
18. Byrne JM, Elliott S, Firek A. Initial experience with patient-clinician secure messaging at a VA medical center. J Am Med Inform Assoc 2009 Mar;16(2):267-270 [FREE Full text] [doi: 10.1197/jamia.M2835] [Medline: 19074303]
19. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710 [FREE Full text] [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
20. Katz SJ, Moyer CA, Cox DT, Stern DT. Effect of a triage-based E-mail system on clinic resource use and patient and physician satisfaction in primary care: a randomized controlled trial. J Gen Intern Med 2003 Sep;18(9):736-744 [FREE Full text] [Medline: 12950483]

21. Siva C, Smarr KL, Hanson KD, Parikh M, Lawlor K, Ge B. Internet use and e-mail communications between patients and providers: a survey of rheumatology outpatients. J Clin Rheumatol 2008 Dec;14(6):318-323. [doi: 10.1097/RHU.0b013e318190b636] [Medline: 19033871]
22. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
23. Andreassen HK, Trondsen M, Kummervold PE, Gammon D, Hjortdahl P. Patients who use e-mediated communication with their doctor: new constructions of trust in the patient-doctor relationship. Qual Health Res 2006 Feb;16(2):238-248. [doi: 10.1177/1049732305284667] [Medline: 16394212]
24. 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]
25. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]
26. Dixon RF. Enhancing primary care through online communication. Health Aff (Millwood) 2010 Jul;29(7):1364-1369 [FREE Full text] [doi: 10.1377/hlthaff.2010.0110] [Medline: 20606189]
27. Menachemi N, Prickett CT, Brooks RG. The use of physician-patient email: a follow-up examination of adoption and best-practice adherence 2005-2008. J Med Internet Res 2011 May;13(1):e23 [FREE Full text] [doi: 10.2196/jmir.1578] [Medline: 21447468]
28. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki C, Kummervold PE, Rasmussen J, et al. Online communication between doctors and patients in Europe: status and perspectives. J Med Internet Res 2010 Jun;12(2):e20 [FREE Full text] [doi: 10.2196/jmir.1281] [Medline: 20551011]
29. Ruland CM, Jeneson A, Andersen T, Andersen R, Slaughter L, Bente-Schjødt-Osmo, et al. Designing tailored Internet support to assist cancer patients in illness management. AMIA Annu Symp Proc 2007:635-639 [FREE Full text] [Medline: 18693913]
30. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
31. Grimsbø GH, Engelsrud GH, Ruland CM, Finset A. Cancer patients' experiences of using an Interactive Health Communication Application (IHCA). Int J Qual Stud Health Well-being 2012;7:- [FREE Full text] [doi: 10.3402/qhw.v7i0.15511] [Medline: 22582085]
32. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in Web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
33. Ruland CM, Maffei RM, Børøsund E, Krahn A, Andersen T, Grimsbø GH. Evaluation of different features of an eHealth application for personalized illness management support: cancer patients' use and appraisal of usefulness. Int J Med Inform 2013 Jul;82(7):593-603. [doi: 10.1016/j.ijmedinf.2013.02.007] [Medline: 23507561]
34. Ruland CM, Borosund E, Varsi C. User requirements for a practice-integrated nurse-administered online communication service for cancer patients. Stud Health Technol Inform 2009;146:221-225. [Medline: 19592838]
35. Nijland N, van Gemert-Pijnen JE, Boer H, Steehouder MF, Seydel ER. Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. Int J Med Inform 2009 Oct;78(10):688-703. [doi: 10.1016/j.ijmedinf.2009.06.002] [Medline: 19625210]
36. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
37. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-339.
38. Glasgow RE, Klesges LM, Dzewaltowski DA, Estabrooks PA, Vogt TM. Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. Health Educ Res 2006 Oct;21(5):688-694 [FREE Full text] [doi: 10.1093/her/cyl081] [Medline: 16945984]
39. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
40. Rogers EM. Diffusion of innovations. New York: Free Press; 2003.
41. Kvale S, Brinkmann S. InterViews: learning the craft of qualitative research interviewing. Los Angeles: Sage Publications; 2009.
42. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

43. LaVela SL, Schectman G, Gering J, Locatelli SM, Gawron A, Weaver FM. Understanding health care communication preferences of veteran primary care users. Patient Educ Couns 2012 Sep;88(3):420-426. [doi: 10.1016/j.pec.2012.06.004] [Medline: 22766452]
44. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Jul;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
45. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008 Feb;2:57-65 [FREE Full text] [Medline: 19920945]
46. Schmid KL, Rivers SE, Latimer AE, Salovey P. Targeting or tailoring? Mark Health Serv 2008;28(1):32-37 [FREE Full text] [Medline: 18389854]
47. Moyer CA, Stern DT, Dobias KS, Cox DT, Katz SJ. Bridging the electronic divide: patient and provider perspectives on e-mail communication in primary care. Am J Manag Care 2002 May;8(5):427-433 [FREE Full text] [Medline: 12019595]
48. Grenon J. Nurse-Patient Email Communication: Comprehensive Guidelines. Canadian Journal of Nursing Informatics 2010;5(4):1009 [FREE Full text]
49. van Baar JD, Joosten H, Car J, Freeman GK, Partridge MR, van Weel C, et al. Understanding reasons for asthma outpatient (non)-attendance and exploring the role of telephone and e-consulting in facilitating access to care: exploratory qualitative study. Qual Saf Health Care 2006 Jun;15(3):191-195 [FREE Full text] [doi: 10.1136/qshc.2004.013342] [Medline: 16751469]
50. Grimsbø GH, Finset A, Ruland CM. Left hanging in the air: experiences of living with cancer as expressed through E-mail communications with oncology nurses. Cancer Nurs 2011 Mar;34(2):107-116. [doi: 10.1097/NCC.0b013e3181eff008] [Medline: 20921887]
51. Sun S, Zhou X, Denny JC, Rosenbloom TS, Xu H. Messaging to your doctors: understanding patient-provider communications via a portal system. In: CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI 2013: Changing Perspectives; April 27-May 2, 2013; Paris, France p. 1739-1748 URL: http://chi2013.acm.org/
52. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
53. Katzen C, Solan MJ, Dicker AP. E-mail and oncology: a survey of radiation oncology patients and their attitudes to a new generation of health communication. Prostate Cancer Prostatic Dis 2005;8(2):189-193. [doi: 10.1038/sj.pcan.4500797] [Medline: 15852052]
54. Han JY, Hawkins RP, Shaw BR, Pingree S, McTavish F, Gustafson DH. Unraveling uses and effects of an interactive health communication system. J Broadcast Electron Media 2009;53(1):112-133 [FREE Full text] [doi: 10.1080/08838150802643787] [Medline: 23172985]
55. Tjora A, Tran T, Faxvaag A. Privacy vs usability: a qualitative exploration of patients' experiences with secure Internet communication with their general practitioner. J Med Internet Res 2005 May;7(2):e15 [FREE Full text] [doi: 10.2196/jmir.7.2.e15] [Medline: 15998606]
56. Kleiboer A, Gowing K, Holm Hansen C, Hibberd C, Hodges L, Walker J, et al. Monitoring symptoms at home: what methods would cancer patients be comfortable using? Qual Life Res 2010 Sep;19(7):965-968. [doi: 10.1007/s11136-010-9662-0] [Medline: 20446045]
57. Dilts D, Ridner SH, Franco A, Murphy B. Patients with cancer and e-mail: implications for clinical communication. Support Care Cancer 2009 Aug;17(8):1049-1056. [doi: 10.1007/s00520-008-0538-6] [Medline: 19050941]
58. Sheaves B, Jones RB, Williamson GR, Chauhan R. Phase 1 pilot study of e-mail support for people with long term conditions using the Internet. BMC Med Inform Decis Mak 2011;11:20 [FREE Full text] [doi: 10.1186/1472-6947-11-20] [Medline: 21466699]

Abbreviations

FAQs: frequently asked questions
HRQoL: health-related quality of life
IPPC: Internet-based patient-provider communication service
IT: information technology
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 22.04.13; peer-reviewed by S Santana, C Bond; comments to author 15.06.13; revised version received 11.09.13; accepted 07.10.13; published 11.11.13.

Please cite as:
Varsi C, Gammon D, Wibe T, Ruland CM
Patients' Reported Reasons for Non-Use of an Internet-Based Patient-Provider Communication Service: Qualitative Interview Study
J Med Internet Res 2013;15(11):e246
URL: http://www.jmir.org/2013/11/e246/
doi:10.2196/jmir.2683
PMID:24220233

Original Paper

Correlation and Interaction Visualization of Altmetric Indicators Extracted From Scholarly Social Network Activities: Dimensions and Structure

Chun Li Liu[1], Master of Library Science; Yue Quan Xu[2], Master of Library Science; Hui Wu[3], Doctor of Public Health and Preventive Medicine; Si Si Chen[1], Bachelor of Medicine; Ji Jun Guo[4], Master of Medicine

[1]Information Office, Library, China Medical University, Shenyang, Liaoning, China

[2]Information Management Department, School of Computer Science and Information Technology, Northeast Normal University, Changchun, Jilin, China

[3]Department of Social Medicine, School of Public Health, China Medical University, Shenyang, Liaoning, China

[4]Sector of Library Director, Library, China Medical University, Shenyang, Liaoning, China

Corresponding Author:
Chun Li Liu, Master of Library Science
Information Office
Library
China Medical University
No.92 of Beier Road, Heping District, Shenyang, Liaoning
Shenyang, Liaoning, 110001
China
Phone: 86 (024)23256666 ext 5434
Fax: 86 (024)23266229
Email: liuchunliliangxu@163.com

Abstract

Background: Citation counts for peer-reviewed articles and the impact factor of journals have long been indicators of article importance or quality. In the Web 2.0 era, growing numbers of scholars are using scholarly social network tools to communicate scientific ideas with colleagues, thereby making traditional indicators less sufficient, immediate, and comprehensive. In these new situations, the altmetric indicators offer alternative measures that reflect the multidimensional nature of scholarly impact in an immediate, open, and individualized way. In this direction of research, some studies have demonstrated the correlation between altmetrics and traditional metrics with different samples. However, up to now, there has been relatively little research done on the dimension and interaction structure of altmetrics.

Objective: Our goal was to reveal the number of dimensions that altmetric indicators should be divided into and the structure in which altmetric indicators interact with each other.

Methods: Because an article-level metrics dataset is collected from scholarly social media and open access platforms, it is one of the most robust samples available to study altmetric indicators. Therefore, we downloaded a large dataset containing activity data in 20 types of metrics present in 33,128 academic articles from the application programming interface website. First, we analyzed the correlation among altmetric indicators using Spearman rank correlation. Second, we visualized the multiple correlation coefficient matrixes with graduated colors. Third, inputting the correlation matrix, we drew an MDS diagram to demonstrate the dimension for altmetric indicators. For correlation structure, we used a social network map to represent the social relationships and the strength of relations.

Results: We found that the distribution of altmetric indicators is significantly non-normal and positively skewed. The distribution of downloads and page views follows the Pareto law. Moreover, we found that the Spearman coefficients from 91.58% of the pairs of variables indicate statistical significance at the .01 level. The non-metric MDS map divided the 20 altmetric indicators into three clusters: traditional metrics, active altmetrics, and inactive altmetrics. The social network diagram showed two subgroups that are tied to each other but not to other groups, thus indicating an intersection between altmetrics and traditional metric indicators.

Conclusions: Altmetrics complement, and most correlate significantly with, traditional measures. Therefore, in future evaluations of the social impact of articles, we should consider not only traditional metrics but also active altmetrics. There may also be a

transfer phenomenon for the social impact of academic articles. The impact transfer path has transfer, or intermediate, stations that transport and accelerate article social impact from active altmetrics to traditional metrics and vice versa. This discovery will be helpful to explain the impact transfer mechanism of articles in the Web 2.0 era. Hence, altmetrics are in fact superior to traditional filters for assessing scholarly impact in multiple dimensions and in terms of social structure.

(J Med Internet Res 2013;15(11):e259) doi:10.2196/jmir.2707

KEYWORDS

altmetrics; article-level metrics; scholarly social network tools; indicator; dimension; structure

Introduction

The evaluation of an academic paper's influence is important for scientists and academic management mechanisms [1]. Scientific papers are regarded by the scientific community as the formal carriers of recent findings and innovative ideas from scientific experiments [2]. Official periodicals are considered the major medium used in scientific communication [3]. Citation rates per paper and the impact factor per scientific journal have been used as evaluation indicators for measuring academic impact [4,5].

Impact factor is based on journals, not journal articles. It is unlikely that one type of metric (for example, citation counts) can adequately inform evaluations across multiple disciplines, departments, career stages, and job types. In addition, a newly published article requires time to accumulate citations—a citation delay may range from 3 months to 1-2 years, sometimes longer in formal publications. By contrast, only a few days are required to tabulate statistics from viewing, downloading, tags, digs, tweets, and blogs in scientific social networks.

A reasonable evaluation should include not only quantitative assessments but also the peer-review process. The traditional peer-review process has been criticized for its scalability, that is, the inability to cope with an increasingly large number of scientific paper submissions, given the limited number of available reviewers and publication time constraints.

With the development of the open access platform [6,7] and the practical application of academic social networks [8,9], scientific achievements have now been able to spread more rapidly [10-13]. Given these new developments, the open access platforms, social network tools, and other online usage and comment-based statistics have been paving the way to new forms of scientific evaluation, which could complement traditional metrics such as the citation rate and the impact factor.

Hence, researchers and publishers are exploring article-level metrics, which include not only citation rates but also potential extracted indicators such as page view, download, click, note, recommend, tag, post, trackback, and comments [14-17]. By using such multidimensional indicators, we aim to broaden researchers' vision in the field of scientometrics and to provide richly measurable metadata for post peer review. For example, Priem and Costello [18] found Twitter citations are generated considerably more quickly than traditional citations, with 40% occurring within 1 week of the cited resource's publication. In this paper, we call these new indicators "altmetric indicators". Compared to traditional indicators, they are superior in terms of coverage, efficiency, and scalability.

In light of the advantage of altmetrics, many authors have called for its further evaluation. Neylon and Wu [14] noted the unsatisfactory results of traditional methods for measuring impact, and they assert that good filters of quality, importance, and relevance to apply to scientific literature are required. Taraborelli [19] suggested that collaboratively aggregated metadata may help to close the gap between traditional citation-based metrics and usage-based metrics for scientific evaluation. He also proposed that social software could be used to extract large-scale indicators of scientific quality. Priem and Hemminger [3] have likewise stated that citation-based methods poorly evaluate and filter articles and considered an examination of the usage of articles in Web 2.0 services novel and promising. They developed the most comprehensive list of Web 2.0 tools and assessed the potential value and the availability of data. Groth and Gurney [20] used keyword and citation similarity maps to analyze differences between blog posts in chemistry and in academic literature. Weller and Puschmann [21] categorized scientific tweets on Twitter and devised a method for identifying and measuring citations.

Do altmetrics correlate with traditional measures? Some researchers have studied this question and provided evidence that altmetric and traditional indicators correlate significantly. For example, Yan and Gerstein [16] examined the correlation between 18 different metrics, including article usage (HTML views, PDF downloads, XML downloads), citation statistics, blog coverage, social bookmarking, and online ratings in the PLOS Article-Level Metrics. They observed that the number of citations correlates most strongly with access statistics (r=.44), with the highest correlation being with number of PDF downloads (r=.48). Additionally, Priem et al [17] studied the correlations between 19 types of altmetric indicators and concluded that the scholarly bookmarking services Mendeley (r=.26) and CiteULike (r=.16) correlated with citations, while services such as Delicious did not. Li et al [22] investigated 1613 journal papers and studied the correlation between two online reference managers (Mendeley and CiteULike) and two types of citations (WoS and Google Scholar). Their results indicate that the Mendeley user counts significantly correlate with WoS citations, and Mendeley attracted more users than did CiteULike. Eysenbach [23] selected a cumulative number of tweetations (ie, a citation in a tweet) 7 days after article publication as tweetation counts and then calculated the correlation between citations and tweetations. The Pearson correlation coefficients for the citation versus tweetation counts were statistically significant at a 5% level and ranged from .57 to .89. Additionally, Google Scholar citations were more strongly correlated with tweetations than were Scopus citations.

In summary, all previous researchers have focused on demonstrating the performance of altmetric indicators and correlations between traditional and altmetric indicators. However, important questions such as the dimensionality and structure of altmetrics have not been explored. In other words, the overall configuration is unclear and requires further verification. For example, how many dimensions should altmetric indicators be divided into? How does the interactive structure look? Motivated by these questions, we attempt to look into the similarities and the differences between traditional and altmetric indicators. We will represent the interactive structure visually in a social network context.

For our study, it is vital to make sure altmetric indicators have the attributes of openness and maneuverability for samples before conducting an altmetrics study. The publisher platforms where articles are being written, read, and published, such as JMIR, PLOS, and social networks, such as Twitter, CiteULike, blogs, or Mendeley, where articles are being shared, recommended, discussed, and rated, make their data available through standardized application programming interfaces (APIs), which allow authors, editors, and academic administration to select the most meaningful data for a particular use at a particular time. These individuals could thus showcase a wider range of article impact in an immediate, open, and individualized way.

Article-Level Metrics represent a comprehensive set of impact indicators that capture usage, citations, social bookmarking and dissemination activity, media and blog coverage, discussion activity, and ratings. API for Article-Level Metrics is freely and publicly available. More than 150 developers have downloaded the API for data reuse to determine the total impact of articles. Hence, we consider these data to provide a good sample for our study.

On selection of the tests, we considered the options proposed by researchers, such as graduated colors for correlation coefficient matrixes [16,17] and methods that Priem used for data transformations of datasets [17]. However, we did not adopt factor analysis to disclose the clusters of altmetric indicators because the Kaiser-Meyer-Olkin (KMO) is low (KMO=0.45). Instead, we explored a nonmetric multidimensional scaling (MDS) method to reveal the dimensions of alternative metrics after nonparametric testing; presumably, there are social networks relationships between multidimensional altmetrics, so we used a social network analysis to map the altmetrics interactions.

Methods

We downloaded an "Article-Level Metrics" dataset (specifically, a sample of 33,128 academic articles) from the PLOS API website on December 14, 2011. The dataset includes data for a number of metrics, for example, counts of article usage, citation rates, and other types of metrics (eg, social bookmarks, comments, notes, blog posts, and ratings). We noticed that the values of the altmetric variables differ too markedly in dimensions, thus resulting in smaller absolute values weighing less when calculating the distances between values. Therefore, variables were handled as dimensionless with an algorithm "mean of 1" to keep the coefficients of the original variables constant [24].

First, we drew a histogram to discern approximately whether the data followed a normal distribution. In a normal distribution, the 2 "halves" of the histogram appear as mirror images of each other [25]. In a skewed distribution, one tail of the distribution may commonly be considerably longer or drawn out relative to the other tail. For example, in a "skewed right" distribution, the tail is on the right [26,27]. Many statistical tests are based on the assumption that the data are sampled from a normal distribution. However, when the variables are skewed (non-normal), a nonparametric test is appropriate [28]. In this paper, we also performed a one-sample Kolmogorov-Smirnov (K-S) test (a type of nonparametric test).

Second, a correlation, indicated by a correlation coefficient, measures the strength and the direction of a linear relationship between two variables [29]. For an abnormal distribution, it is more advisable to use the Spearman rank correlation than the Pearson correlation. Examining a table of coefficient numbers is impractical because a matrix of 20x20 is large, so graphical visualization tools are suitable. Various methods have been proposed, from heat maps to correlation ellipses [30]. We visualized the correlation matrix using a color graph generated by the Corrplot package in the R programming language.

Third, MDS could generate a visual representation of the subjective dimensions that are not directly indicated in the data [31]. Many applications of this method are available in bioinformatics [32,33] and ecological science [34-36]. A nonmetric MDS analysis enabled us to find a nonparametric monotonic relationship between similarities in the item-item matrix, the Euclidean distances between items, and the location of each item in the low-dimensional space [37]. We explored a nonmetric MDS analysis method with the software package UCINET to determine the types of variables that have a higher degree of similarities.

Fourth, an MDS diagram can reveal the similarities among variables, though not the strength and the structure of the relationships among variables. Visualizing the correlation matrix in a network context is useful. Researchers observe social relationships based on the theory that a social network comprises nodes and ties. Nodes represent individual actors within the network, and ties represent relationships between variables and individuals [38]. We used NetDraw (a social network analysis software package) to visualize the interaction of the variables and its strength. We also aimed to ascertain the relative importance of variables in interconnecting the network. The social network diagram helped us distinguish the number of clusters and the corresponding degrees of clustering.

Results

Right-Skewed Distribution and the Pareto Principle

We used a one-sample K-S test to determine whether the altmetric variables are normally distributed. In general, if $P<.05$, then the data are considered to follow an abnormal distribution [39]. Our results showed that the $P<.001$ for all variables; therefore, we rejected the normality assumption. One way to

determine whether a variable is "significantly skewed" is comparing the numerical value for "skewness" with twice the standard error of skewness, including the range from minus twice to plus twice the standard error of skewness [40]. Because the skewness value falls outside this range, we concluded that the distribution is significantly non-normal and, in this case, positively skewed. Table 1 shows the integration of the results, including the K-S test, the skewness, and the kurtosis of variables. Table 2 lists the legends for B1 to B20.

We also drew histograms and obtained a group of skewed histograms. Because variable B_i is highly skewed throughout testing (the average of the skewness is 1.267, and the average of the kurtosis is 2.033), we log-transformed it into variable D_i (after excluding zeros) to show its distribution more clearly. Figure 1 summarizes the frequency distribution of the cumulative variable D_i. The right tail is longer, and the distribution's mass is concentrated on the left of the figure, thus confirming that the 20 histograms are right-skewed distributions, according to the direction of the tail. We inferred that skewness may be related to the meaning of the variables: the percentage of relative activities mentioned in the articles cannot be less than zero.

As shown in the histogram of downloads and page views (from D to G), the data have two relative peaks that follow a bimodal distribution, similar in appearance to the back of a two-humped camel. This distribution is reminiscent of the Pareto Principle (or the 80-20 rule), that is, approximately 80% of the effects arise from 20% of the causes [41,42]. With reference to the theory of knowledge scatter [43], this pattern suggested that 80% of download counts were generated by 20% of the articles.

Table 1. Integration of the results including K-S, skewness, and kurtosis of variables (N=33,128).

	K-S		Skewness		Kurtosis	
	Z	Asymp sig[a] (2-tailed)	S	SE[b]	K	SE
B1	59.205	<.001	7.271	0.013	88.479	0.027
B2	73.141	<.001	15.342	0.013	299.692	0.027
B3	64.031	<.001	10.778	0.013	218.256	0.027
B4	79.442	<.001	27.596	0.013	1405.772	0.027
B5	83.492	<.001	7.902	0.013	128.074	0.027
B6	70.240	<.001	15.694	0.013	405.590	0.027
B7	81.168	<.001	21.363	0.013	908.977	0.027
B8	94.749	<.001	18.727	0.013	421.814	0.027
B9	92.407	<.001	60.796	0.013	4051.419	0.027
B10	94.019	<.001	17.715	0.013	436.739	0.027
B11	88.133	<.001	29.028	0.013	1075.095	0.027
B12	72.099	<.001	20.820	0.013	916.229	0.027
B13	93.919	<.001	16.017	0.013	359.758	0.027
B14	91.698	<.001	25.092	0.013	1010.583	0.027
B15	93.719	<.001	19.392	0.013	770.375	0.027
B16	91.612	<.001	17.786	0.013	566.946	0.027
B17	93.434	<.001	40.112	0.013	2102.985	0.027
B18	87.564	<.001	17.025	0.013	556.158	0.027
B19	92.924	<.001	32.325	0.013	1447.864	0.027
B20	94.296	<.001	27.080	0.013	1205.024	0.027

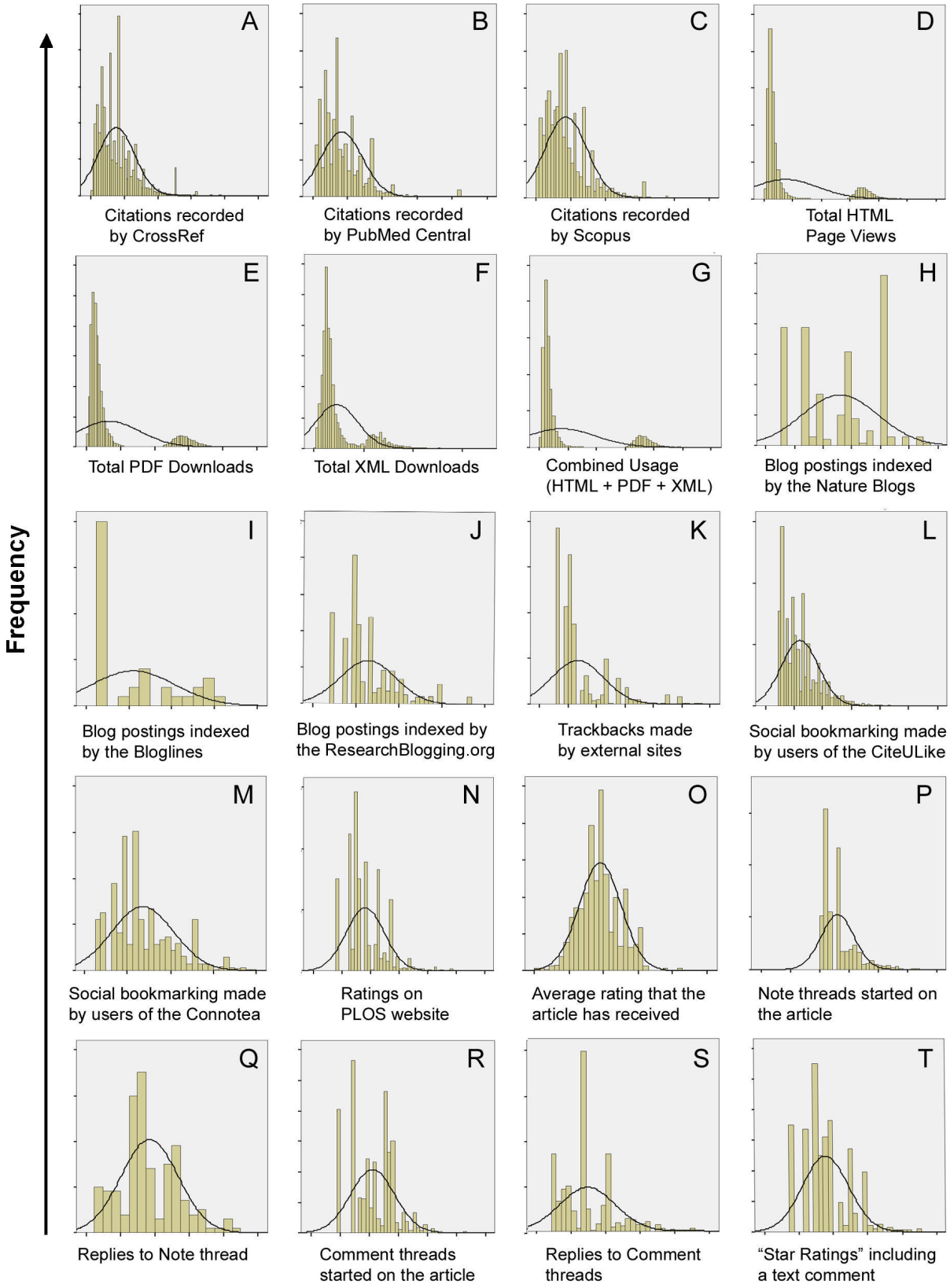
[a]asymptotic significance

[b]standard error

Table 2. Legends for B1 to B20.

Altmetric indicators	Legends
B1	Citations recorded by CrossRef
B2	Citations recorded by PubMed Central
B3	Citations recorded by Scopus
B4	Total HTML page views
B5	Total PDF downloads
B6	Total XML downloads
B7	Combined usage (HTML + PDF + XML)
B8	Blog postings indexed by Nature Blogs
B9	Blog postings indexed by Bloglines
B10	Blog postings indexed by ResearchBlogging.org
B11	Trackbacks made by external sites
B12	Social bookmarking made by users of CiteULike
B13	Social bookmarking made by users of Connotea
B14	Ratings on PLOS website
B15	Average rating that the article has received
B16	Note threads started on the article
B17	Replies to Note thread
B18	Comment threads started on the article
B19	Replies to Comment threads
B20	“Star Ratings” including a text comment

Figure 1. Histograms of frequency distribution for altmetric indicators.



Spearman Correlation Coefficients and Their Visualization in R

We performed the normality test to conclude that neither altmetric variable is normally distributed. We used a Spearman rank order correlation to examine the correlation pattern among altmetric indicators with SPSS 18.0. Tables 3 and 4 present the results of the Spearman rank correlation coefficient for altmetric indicators.

The correlation coefficient can range from -1 to 1, with -1 or 1 indicating a perfect relationship [44]. The Spearman rho between B14 and B15 is 1. B14 represents ratings on the PLOS website, and B15 represents an article's average rating. Therefore, it is unsurprising that the relationship is approximately perfect. The similarities in correlation strength was observed for another pair of variables (B4 and B7) with "rho=1", likely because HTML page views (expressed by B4) accounted for the largest proportion of the combined (HTML + PDF + XML) usage of articles (expressed by B7).

The second strongest correlation, rho=.899, is between total HTML page views (expressed by B4) and total PDF downloads (expressed by B5), possibly because they are two aspects of article usage counts, and people choose view or download with approximately equivalent frequencies. The Spearman rho between B6 and B13 is –.25, so we can predict that as B6 (total XML downloads) increases, B13 (social bookmarking made by Connotea users) will decrease.

The Spearman coefficients from 91.58% of the variable pairs are significant at the .01 level, with one pair of variables (B9 and B12) correlating at the .05 significance level. Approximately no correlation exists between approximately 7.89% variable pairs. B9 (blog postings indexed by the Bloglines) also hardly correlate with eight variables (ie, B13 to B20), possibly implying that Blogline is unpopular and not widely used by researchers and citizen scientists.

The correlation matrix also yields the probability of being incorrect if we assume that the relationship observed in our sample accurately reflects the relationship among variables of altmetric indicators in the actual population from which the sample was drawn, labeled as Sig (2-tailed). We found that 91.58% of the probability value is <.001 (the value is rounded to three digits), well below the conventional threshold of $P<.05$, thus supporting our hypothesis. There is a relationship (ie, the coefficient is not 0) in the predicted direction (positive), and we can generalize the results to the population ($P<.05$).

To show the correlation among altmetrics clearly, we visualized the correlation coefficient matrices with graduated colors and a blue-white-red scale. An R programming package, corrplot, helped map the correlation coefficients to the specified color square. We chose two color series to identify positive and negative correlation coefficients. Blue corresponds to a correlation of approximately 1; red to approximately –1; and white to approximately 0. To economize space, we multiplied the correlation coefficients by 100 and added them to the squares in the color correlation matrix. See Figure 2.

We can readily identify clusters with strong similarities and locate possible redundant indicators. Matching this map with the physical meaning revealed the following: (1) the citation indicators (B1, B2, and B3) and download indicators (B4, B5, B6, and B7) are clustered into two categories, which we call the "citation metrics class" and "download metrics class", respectively; (2) the citation and download indicators are combined into a clustering, which we call the "traditional metrics class"; (3) a group of indicators (B14 to B20) are conjoined into another clustering type, called the "rating, note, and comment metrics class"; and (4) finally, as a general rule, we suggested that all four blog-aggregating services would record different sets of data, so the datasets require comparison and "de-duplication" to obtain a complete picture of blog activity (as recorded by these services), as would all three citation services.

Table 3. Spearman rank correlation coefficient for B1-B11 (N=33,128).

		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
B1	Corr. coefficient	1.000	.599[a]	.738[a]	.322[a]	.378[a]	.153[a]	.338[a]	.050[a]	.019[a]	.088[a]
	Sig (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B2	Corr. coefficient	.599[a]	1.000	.669[a]	.226[a]	.268[a]	.079[a]	.237[a]	.051[a]	.023[a]	.063[a]
	Sig (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B3	Corr. coefficient	.738[a]	.669[a]	1.000	.402[a]	.444[a]	.244[a]	.415[a]	.040[a]	.020[a]	.084[a]
	Sig (2-tailed)	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
B4	Corr. coefficient	.322[a]	.226[a]	.402[a]	1.000	.899[a]	.662[a]	.996[a]	.070[a]	-.002	.156[a]
	Sig (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001	<.001	.784	<.001
B5	Corr. coefficient	.378[a]	.268[a]	.444[a]	.899[a]	1.000	.616[a]	.928[a]	.065[a]	.002	.125[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001		<.001	<.001	<.001	.695	<.001
B6	Corr. coefficient	.153[a]	.079[a]	.244[a]	.662[a]	.616[a]	1.000	.672[a]	.039[a]	-.005	.089[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001	<.001	.399	<.001
B7	Corr. coefficient	.338[a]	.237[a]	.415[a]	.996[a]	.928[a]	.672[a]	1.000	.070[a]	-.001	.153[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001		<.001	.892	<.001
B8	Corr. coefficient	.050[a]	.051[a]	.040[a]	.070[a]	.065[a]	.039[a]	.070[a]	1.000	.019[a]	.158[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
B9	Corr. coefficient	.019[a]	.023[a]	.020[a]	-.002	.002	-.005	-.001	.019[a]	1.000	-.001
	Sig (2-tailed)	<.001	<.001	<.001	.784	.695	.399	.892	<.001		.918
B10	Corr. coefficient	.088[a]	.063[a]	.084[a]	.156[a]	.125[a]	.089[a]	.153[a]	.158[a]	-.001	1.000
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.918	
B11	Corr. coefficient	.072[a]	.061[a]	.073[a]	.063[a]	-.004	-.015[a]	.053[a]	.071[a]	.035[a]	.214[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	.426	.006	<.001	<.001	<.001	<.001
B12	Corr. coefficient	.240[a]	.248[a]	.222[a]	.288[a]	.299[a]	.156[a]	.293[a]	.086[a]	.014[b]	.128[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.014	<.001
B13	Corr. coefficient	.120[a]	.159[a]	.102[a]	.065[a]	.071[a]	-.025[a]	.067[a]	.042[a]	.010	.031[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.071	<.001
B14	Corr. coefficient	.074[a]	.075[a]	.072[a]	.087[a]	.057[a]	.050[a]	.085[a]	.055[a]	.005	.101[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.328	<.001
B15	Corr. coefficient	.074[a]	.076[a]	.072[a]	.087[a]	.057[a]	.050[a]	.085[a]	.055[a]	.005	.101[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.350	<.001
B16	Corr. coefficient	.069[a]	.061[a]	.067[a]	.075[a]	.053[a]	.042[a]	.073[a]	.028[a]	-.001	.070[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.794	<.001
B17	Corr. coefficient	.027[a]	.021[a]	.027[a]	.045[a]	.024[a]	.026[a]	.043[a]	.036[a]	-.002	.058[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.662	<.001
B18	Corr. coefficient	.090[a]	.101[a]	.096[a]	.099[a]	.056[a]	.043[a]	.097[a]	.063[a]	.010	.133[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.063	<.001
B19	Corr. coefficient	.058[a]	.055[a]	.057[a]	.082[a]	.053[a]	.041[a]	.079[a]	.052[a]	.008	.103[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.137	<.001

		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
B20	Corr. coefficient	.050[a]	.049[a]	.049[a]	.062[a]	.043[a]	.035[a]	.060[a]	.037[a]	.009	.073[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.090	<.001

[a]Correlation is significant at the .01 level (2-tailed).

[b]Correlation is significant at the .05 level (2-tailed).

Table 4. Spearman rank correlation coefficient for B12-B20 (N=33,128).

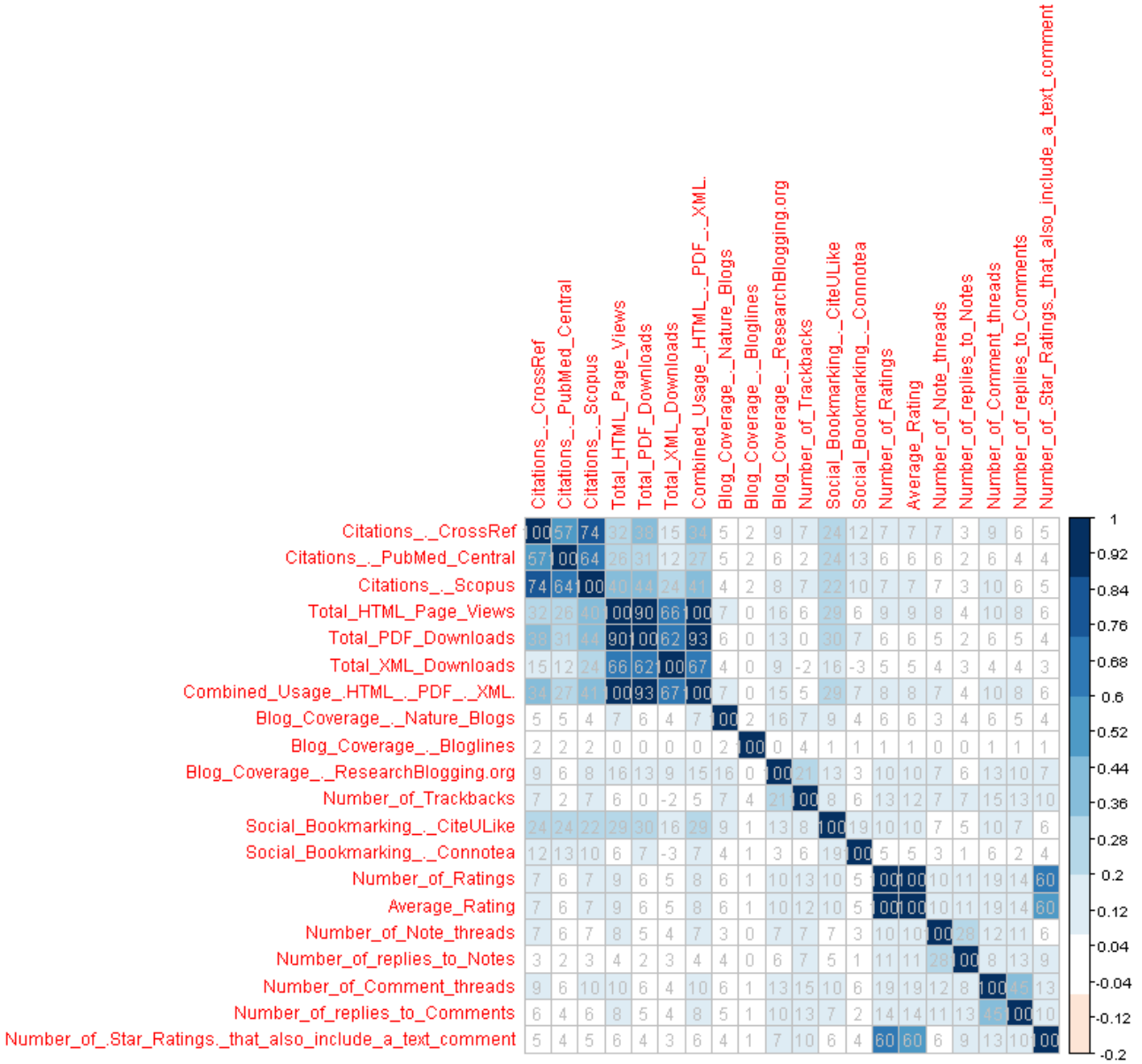
		B11	B12	B13	B14	B15	B16	B17	B18	B19	B20
B1	Corr. coefficient	.072[a]	.240[a]	.120[a]	.074[a]	.074[a]	.069[a]	.027[a]	.090[a]	.058[a]	.050[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B2	Corr. coefficient	.061[a]	.248[a]	.159[a]	.075[a]	.076[a]	.061[a]	.021[a]	.101[a]	.055[a]	.049[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B3	Corr. coefficient	.073[a]	.222[a]	.102[a]	.072[a]	.072[a]	.067[a]	.027[a]	.096[a]	.057[a]	.049[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B4	Corr. coefficient	.063[a]	.288[a]	.065[a]	.087[a]	.087[a]	.075[a]	.045[a]	.099[a]	.082[a]	.062[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B5	Corr. coefficient	-.004	.299[a]	.071[a]	.057[a]	.057[a]	.053[a]	.024[a]	.056[a]	.053[a]	.043[a]
	Sig (2-tailed)	.426	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B6	Corr. coefficient	-.015[a]	.156[a]	-.025[a]	.050[a]	.050[a]	.042[a]	.026[a]	.043[a]	.041[a]	.035[a]
	Sig (2-tailed)	.006	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B7	Corr. coefficient	.053[a]	.293[a]	.067[a]	.085[a]	.085[a]	.073[a]	.043[a]	.097[a]	.079[a]	.060[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B8	Corr. coefficient	.071[a]	.086[a]	.042[a]	.055[a]	.055[a]	.028[a]	.036[a]	.063[a]	.052[a]	.037[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B9	Corr. coefficient	.035[a]	.014[b]	.010	.005	.005	-.001	-.002	.010	.008	.009
	Sig (2-tailed)	<.001	.014	.071	.328	.350	.794	.662	.063	.137	.090
B10	Corr. coefficient	.214[a]	.128[a]	.031[a]	.101[a]	.101[a]	.070[a]	.058[a]	.133[a]	.103[a]	.073[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B11	Corr. coefficient	1.000	.078[a]	.059[a]	.125[a]	.124[a]	.073[a]	.068[a]	.148[a]	.128[a]	.096[a]
	Sig (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B12	Corr. coefficient	.078[a]	1.000	.194[a]	.098[a]	.097[a]	.067[a]	.045[a]	.097[a]	.073[a]	.064[a]
	Sig (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
B13	Corr. coefficient	.059[a]	.194[a]	1.000	.050[a]	.049[a]	.031[a]	.007	.060[a]	.023[a]	.039[a]
	Sig (2-tailed)	<.001	<.001		<.001	<.001	<.001	.215	<.001	<.001	<.001
B14	Corr. coefficient	.125[a]	.098[a]	.050[a]	1.000	1.000[a]	.097[a]	.109[a]	.190[a]	.143[a]	.602[a]
	Sig (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001	<.001
B15	Corr. coefficient	.124[a]	.097[a]	.049[a]	1.000[a]	1.000	.096[a]	.107[a]	.189[a]	.141[a]	.599[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001	<.001
B16	Corr. coefficient	.073[a]	.067[a]	.031[a]	.097[a]	.096[a]	1.000	.283[a]	.116[a]	.112[a]	.065[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001	<.001
B17	Corr. coefficient	.068[a]	.045[a]	.007	.109[a]	.107[a]	.283[a]	1.000	.085[a]	.132[a]	.094[a]
	Sig (2-tailed)	<.001	<.001	.215	<.001	<.001	<.001		<.001	<.001	<.001
B18	Corr. coefficient	.148[a]	.097[a]	.060[a]	.190[a]	.189[a]	.116[a]	.085[a]	1.000	.448[a]	.127[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001	<.001
B19	Corr. coefficient	.128[a]	.073[a]	.023[a]	.143[a]	.141[a]	.112[a]	.132[a]	.448[a]	1.000	.105[a]
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001

		B11	B12	B13	B14	B15	B16	B17	B18	B19	B20
B20	Corr. coefficient	.096[a]	.064[a]	.039[a]	.602[a]	.599[a]	.065[a]	.094[a]	.127[a]	.105[a]	1.000
	Sig (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

[a]Correlation is significant at the .01 level (2-tailed).

[b]Correlation is significant at the .05 level (2-tailed).

Figure 2. Visualization of the correlation matrix in R.



Nonmetric MDS With UCINET and Network Visualization With NetDraw

Nonmetric MDS is often preferred because it tends to provide a better "goodness-of-fit" (stress) statistic, which is correspondingly better with lower stress (0=perfect fit) [45]. Generally, stress levels below 0.1 are considered excellent, while levels above 0.2 are considered unacceptable. Accordingly, a higher RSQ (r-squared) value (1=perfect fit) is better, and RSQ values exceeding 0.6 are usually considered excellent [46]. We conducted nonmetric MDS with UCINET 6. The output map is shown in Figure 3.

The reliability value stress was 0.00424, considerably less than 0.1, and the validity value RSQ was 0.99998, greater than 0.60, which equals an excellent goodness of fit. The map plots each variable, thus permitting us to examine the similarity according to the variables' proximity to each other. We labeled three

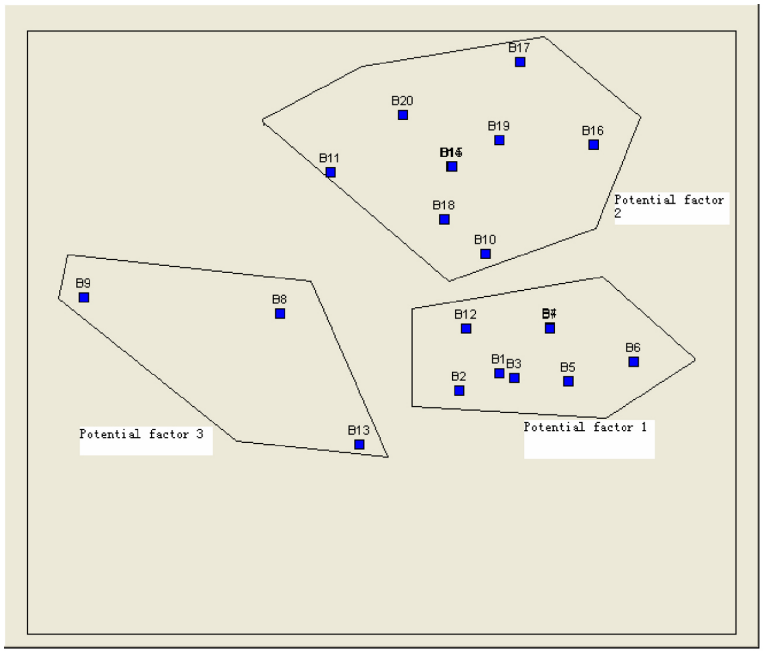
dimensions, or categories, with each dimension implicating a potential factor.

The three clusters and their interpretations are as follows. (1) The first cluster contains B1 to B7 and B12. This cluster has 8 spots, and they are more interconnected. B4 and B7 occupy approximately the same coordinate. This cluster implicates a potential factor of 1, which we call a traditional metrics group because 7 out of 8 indicators in this cluster are citation and download indicators. (2) The second cluster contains B10, B11, and B14 to B20. This cluster has 9 spots, and they are more interconnected. B14 and B15 occupy approximately the same coordinate. This cluster implicates a potential factor of 2, and we call it the trackback, rating, note, and comment metrics group. (3) The third cluster contains B8, B9, and B13. This cluster has 3 spots, yet they are less interconnected, with more diverse networks. This cluster implicates a potential factor of 3, and we call it the blog and social bookmark metrics group.

We know that an MDS graph can represent the relations among nodes, while a network diagram can describe the social structure. Hence, we visualized the results of the nonmetric MDS from a network context with NetDraw (version 2.084, which is distributed with UCINET 6). The network diagram is shown in Figure 4.

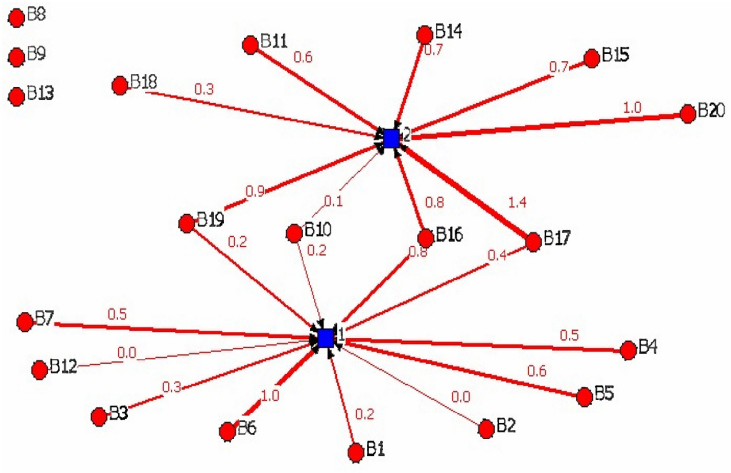
A good drawing of a graph can immediately suggest some of the most important features of the overall network structure. The diagram indicates the following findings: (1) not all nodes are connected, as three nodes (B8, B9, and B13) that are disconnected from the others; (2) two subgroups or local "clusters" of actors are tied to each other, not to other groups, and (3) some actors have many ties, and some, few ties. Four nodes (B10, B16, B17, and B19) have two ties, while the other nodes have one tie or zero ties. These nodes are embedded in the neighborhood by the two clusters; that is, they are important for connecting the two clusters, which we call cluster 1 and cluster 2. Thus, examining the node and the "node network" (ie, "neighborhood") indicates a sense of the structural constraints and opportunities that an actor faces and may help us to understand an actor's role in a social structure. Finally, it indicates that (4) some difference in the strength of the relationship between a multivariable and its center remain. For example, B12 and B2 have a weak relationship with their center, while B6 and B20 have a relatively stronger relationship (that is, "1.0") with their centers, and B17 has the strongest relationship (that is, "1.4") with its center.

Figure 3. MDS diagram of altmetric indicators.



B1 Citations recorded by CrossRef
B2 Citations recorded by PubMed Central
B3 Citations recorded by Scopus
B4 Total HTML Page Views
B5 Total PDF Downloads
B6 Total XML Downloads
B7 Combined Usage (HTML + PDF + XML)
B8 Blog postings indexed by the Nature Blogs
B9 Blog postings indexed by the Bloglines
B10 Blog postings indexed by the ResearchBlogging.org
B11 Trackbacks made by external sites
B12 Social bookmarking made by users of the CiteULike
B13 Social bookmarking made by users of the Connotea
B14 Ratings on PLOS website
B15 Average rating that the article has received
B16 Note threads started on the article
B17 Replies to Note thread
B18 Comment threads started on the article
B19 Replies to Comment threads
B20 "Star Ratings" including a text comment

Figure 4. Social network structure diagram of altmetric indicators.



B1 Citations recorded by CrossRef
B2 Citations recorded by PubMed Central
B3 Citations recorded by Scopus
B4 Total HTML Page Views
B5 Total PDF Downloads
B6 Total XML Downloads
B7 Combined Usage (HTML + PDF + XML)
B8 Blog postings indexed by the Nature Blogs
B9 Blog postings indexed by the Bloglines
B10 Blog postings indexed by the ResearchBlogging.org
B11 Trackbacks made by external sites
B12 Social bookmarking made by users of the CiteULike
B13 Social bookmarking made by users of the Connotea
B14 Ratings on PLOS website
B15 Average rating that the article has received
B16 Note threads started on the article
B17 Replies to Note thread
B18 Comment threads started on the article
B19 Replies to Comment threads
B20 "Star Ratings" including a text comment

Discussion

Principal Findings

Our study is the first to use the MDS and network map to analyze the dimensions and interactions among altmetrics variables. Although MDS diagrams have been used for co-citation [47] and co-word analysis [48], it is still innovative to draw nonmetric MDS diagrams for altmetrics variables. We found three dimensions or metrics groups, that is, traditional metrics (citation and download metrics), active altmetrics (trackback, rating, note, and comment metrics), and inactive altmetrics (blog and social bookmark metrics).

More importantly, we transformed the MDS diagram into a social network graph, whose advantage is that it displays the overall network structure. In research related to altmetrics, authors have developed co-word social network maps for articles published in blogs [20]. Our map represents the MDS diagram in a social network context. We found that the ResearchBlogging.org posts, note threads, and replies to comment threads are the three intermediary metrics between traditional metrics and active altmetrics; in other words, they possess attributes of traditional metrics and active altmetrics.

What do these findings imply? There may be a transfer phenomenon for social impact of academic articles. Then, Figure 4 could be considered an article impact transitive map. Along the impact transfer path, B10, B16, B17, and B19 are the transfer, or intermediate, stations that transport article social impact between active altmetrics and traditional metrics. Aman [49] quantified the extent to which preprints in arXiv accelerate scholarly communication using many subject samples. He found that, in all fields except biology, a significant citation advantage exists in terms of speed and citation rates for articles with a previous preprint version on arXiv. Shuai et al [50] studied whether Wikipedia shapes academic impact and showed that articles mentioned on Wikipedia have higher citations than do unmentioned articles. Our finding of altmetrics interactions posits that an intermediate station and a potential pathway may exist by which impact activator arXiv, Wikipedia, or other open access platforms and social network tools likely help articles attract more online usage, in turn accelerating online social activities such as comment, note, post, rate, or bookmark and thus expanding an article's social influence, reflected in larger citation rates and higher dissemination speed. This results in the observation that altmetrics is the superior way to look at publications.

Another finding is that altmetrics correlate with traditional measures significantly; that the citation and download metrics cluster closely together by the Spearman correlation method is consistent with previous results [17] to some extent. This is exemplified by the correlation between citation counts and access statistics (r=.30); the highest correlation being with number of PDF downloads (r=.44); and the correlations between citations and scholarly bookmarking services CiteULike (r=.24) and Connotea (r=.13). Before studying the correlation of altmetric indicators, we looked more closely at the choice of method for skewed data. However, the Spearman and color square visualization methods we used differ from the methods used in previous research. For example, the Pearson, not the Spearman correlation method, was used by Eysenbach [23], while our study added a color square visualization method to better reveal correlations. Furthermore, Yan [16] found that article access metrics, citation metrics, and social bookmarking metrics broadly cluster, a formation signified by relatively high correlation coefficients among the metrics; we found additional clusters such as the "rating, note, and comment metrics class". Moreover, we came up with a "traditional metrics class", which integrates the "citation metrics class" and the "download metrics class".

Our third contribution is the adoption of the theory of nonparametric testing throughout analysis. Based on the one-sample K-S test and the shapes of the histograms, we concluded that the distribution is significantly non-normal and positively skewed. Priem summarized a group of histograms similarly but did not perform a nonparametric test to prove the abnormal distribution or to compute the skewness [17]. We calculated the Spearman coefficients (obtained by nonparametric measures), not Pearson coefficients, to calculate the correlation strengths. Although the Spearman measure has been used by Yan and Gerstein [16], their sample size (13,000 articles) was smaller than ours (33,128 articles). More importantly, the Spearman coefficient is statistically fit for abnormal datasets. Additionally, the nonmetric MDS employed to detect the similarity of the variables is also a nonparametric test.

Our results also support that altmetric indicators may obey certain rules, for example, the Pareto law. Eysenbach [23] was the first to report the Pareto law for tweetation (one of the altmetric indicators). He mined all tweets containing links to articles in the *Journal of Medical Internet Research* between July 2008 and November 2011. He explored the dynamics, the content, and the timing of tweets based on a subset of 1573 tweets on approximately 55 articles and found an uneven distribution in which the top 20% of the tweet authors, as ranked by number of tweetations, accounted for 63.4% of all tweetations. This tweetation regularity follows a Pareto distribution (80/20 rule). Similarly, our frequency distribution histograms from D to G (four types of altmetric indicators) indicate that the top 20% of articles triggered 80% of download and page views and thus verified that the distribution follows the Pareto law. Therefore, we offer a complementary explanation of the Pareto regularities using altmetric indicators.

Based on our experimental results, we conclude that altmetrics complements traditional statistics and contains approximately three dimensions: traditional, active, and inactive metrics. In summary, our study demonstrates a novel interaction among the altmetrics variables and analyzes articles' social impact transfer mechanism.

Our conclusion that the distribution is significantly non-normal and positively skewed rests primarily on the results obtained with the Article-Level Metrics dataset downloaded from PLOS API. Both Priem [17] and Yan [16] studied altmetrics based on a similar dataset. Our views regarding whether the distribution of variables is normal are consistent with theirs. However, we demonstrated the necessity of nonparameter testing in analyzing the altmetrics dataset.

Limitations

However, as alternative metrics indicators are preset in the dataset, the implication of our study's findings is limited. This study was a preliminary attempt, and we are preparing to test and verify these findings for other types of datasets. The findings of correlations have been confirmed by another dataset concerning altmetric indicators in [22] and [23]. Further research is required on the dimension, structure, and potential impact transfer mechanism.

Conclusions

In conclusion, we studied the dimension and the structure of altmetrics with visual graphics. Our findings provide an important direction regarding the current practices of authors, editors, and academic administrations. Authors should pay more attention to the scholarly social impact that originates from active altmetrics and then participate more in related activities such as rating websites, noting, and commenting on articles. The publishers should attempt to launch an open peer review and consider scientific citizens' perspectives before deciding whether to publish. They should also explore the value and the applications of post-publication interactivity in terms of ratings, notes, or comments. Academic administrations should track the dissemination of published articles (in terms of multiple types of citation, ratings, comments, and notes) and access up-to-date altmetrics data to determine article quality or the impact context for tenure and promotion decisions.

Acknowledgments

The authors would like to acknowledge the contributions from Qi Ruiqun in figure editing, and help from Cara Bertozzi and his colleagues in manuscript editing and proofreading. Thanks also to the Public Library of Science for creating and maintaining Article-Level Metrics as a free and open data source, as well as all other providers of free altmetrics data.

Conflicts of Interest

None declared.

References

1. Abbott A, Cyranoski D, Jones N, Maher B, Schiermeier Q, Van Noorden R. Metrics: Do metrics matter? Nature 2010 Jun 17;465(7300):860-862 [FREE Full text] [doi: 10.1038/465860a] [Medline: 20559361]
2. Alex G, Holmberg K, Christina K, Heather P, Jason P, Nicholas W. Embracing new methods for publishing, finding, discussing, and measuring our research output. In: Proceedings of the American Society for Information Science and Technology. 2011 Presented at: Bridging the Gulf: Communication and Information in Society, Technology, and Work; October 9-13, 2011; New Orleans, Louisiana p. 1-3 URL: http://onlinelibrary.wiley.com/doi/10.1002/meet.2011.14504801205/abstract [doi: 10.1002/meet.2011.14504801205]
3. Jason P, Bradely H. First Monday. 2010 Jul 05. Scientometrics 2.0: New metrics of scholarly impact on the social Web URL: http://firstmonday.org/ojs/index.php/fm/article/view/2874/2570 [accessed 2013-11-03] [WebCite Cache ID 6KqxOKSst]
4. Paul S, Nicholas C. How to improve your impact factor: Questioning the quantification of academic quality. Journal of Philosophy of Education 2011 Mar 10;45(1):1-17 [FREE Full text] [doi: 10.1111/j.1467-9752.2011.00787.x]
5. Reinstein A, Hasselback JR, Riley ME, Sinason DH. Pitfalls of using citation indices for making academic accounting promotion, tenure, teaching load, and merit pay decisions. Issues in Accounting Education 2011 Feb;26(1):99-131 [FREE Full text] [doi: 10.2308/iace.2011.26.1.99]
6. Ramalho Correia AM, Teixeria JC. Reforming scholarly publishing and knowledge communication: From the advent of the scholarly journal to the challenges of open access. Online Information Review 2005;29(4):349-364 [FREE Full text] [doi: 10.1108/14684520510617802]

7. Willinsky J. The nine flavours of open access scholarly publishing. J Postgrad Med 2003;49(3):263-267 [FREE Full text] [Medline: 14597793]
8. Greenhow C. Knowledge Quest.: American Libraries Association; 2009 Mar 30. Social scholarship: applying social networking technologies to research practices URL: https://aasl.metapress.com/content/r282223126950757/resource-secured/?target=fulltext.pdf [accessed 2013-05-09] [WebCite Cache ID 6GTql2lus]
9. Veletsianos G, Kimmons R. Networked Participatory Scholarship: Emergent techno-cultural pressures toward open and digital scholarship in online networks. Computers & Education 2012 Feb;58(2):766-774 [FREE Full text] [doi: 10.1016/j.compedu.2011.10.001]
10. Ebner M, Reinhardt W. Social networking in scientific conferences –Twitter as tool for strengthening a scientific community. In: Proceedings of the 1st International Workshop on Science. 2009 Presented at: Fourth European Conference on Technology Enhanced Learning; Sept. 29-Oct. 2, 2009; Nice, France p. 1-8 URL: http://zh.scribd.com/doc/20363438/Social-networking-in-scientific-conferences-%E2%80%93-Twitter-as-tool-for-strengthen-a-scientific-community
11. Kjellberg S. First Monday. 2010 Aug 02. I am a blogging researcher: Motivations for blogging in a scholarly context URL: http://firstmonday.org/article/view/2962/2580 [accessed 2013-05-09] [WebCite Cache ID 6GTtkbAFg]
12. Kirkup G. Academic blogging: academic practice and academic identity. London Review of Education 2010 Mar;8(1):75-84 [FREE Full text] [doi: 10.1080/14748460903557803]
13. Wang ML. The impact of open access journals on library and information scientists' research in Taiwan. : Faculty of Information Management, UiTM; 2011 Presented at: Asia-Pacific Conference on Library & Information Education & Practice (A-LIEP); June 22-24, 2011; Putrajaya, Malaysia p. 22-24 URL: http://eprints.uitm.edu.my/3624/1/SP_TIO11_61.pdf
14. Neylon C, Wu S. Article-level metrics and the evolution of scientific impact. PLoS Biol 2009 Nov;7(11):e1000242 [FREE Full text] [doi: 10.1371/journal.pbio.1000242] [Medline: 19918558]
15. Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, Rodriguez MA, et al. Clickstream data yields high-resolution maps of science. PLoS One 2009;4(3):e4803 [FREE Full text] [doi: 10.1371/journal.pone.0004803] [Medline: 19277205]
16. Yan KK, Gerstein M. The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics. PLoS One 2011;6(5):e19917-e19917 [FREE Full text] [doi: 10.1371/journal.pone.0019917] [Medline: 21603617]
17. Jason P, Heather A, Bradley H. Altmetrics in the mild: An exploratory study of impact metrics based on social media. In: Metrics 2011.: Association for Information Science and Technology; 2011 Oct 12 Presented at: Symposium on Informetric and Scientometric Research (SIG MET); October 12, 2011; Orleans, USA URL: https://docs.google.com/presentation/d/1JOBBjzKEDmtE5Z6W4ZYrB0kjuqU2O6Hi-hpwmJtVhJY/edit?pli=1#slide=id.i0
18. Priem J, Costello KL. How and why scholars cite on Twitter. Proc Am Soc Info Sci Tech 2011 Feb 03;47(1):1-4 [FREE Full text] [doi: 10.1002/meet.14504701201]
19. Taraborelli D. Soft peer review: social software and distributed scientific evaluation. 2008 Presented at: 8th International Conference on the Design of Cooperative Systems; May 20-23, 2008; Carry-le-Rouet, Provence, France p. 99-110 URL: http://eprints.ucl.ac.uk/8279/
20. Groth P, Gurney T. Studying scientific discourse on the web using bibliometrics: a chemistry blogging case study. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line. 2010 Apr Presented at: Web Science Conference 2010:Extending the Frontiers of Society On-Line; 26-27 April 2010; Raleigh, NC p. 26-27 URL: http://journal.webscience.org/308/2/websci10_submission_48.pdf
21. Weller K, Puschmann C. Twitter for scientific communication: How can citations/references be identified and measured. 2011 Jun Presented at: ACM Web Science Conference 2011; June 14-17 2011; Koblenz, Germany URL: http://journal.webscience.org/500/1/153_paper.pdf
22. Li X, Thelwall M, Giustini D. Validating online reference managers for scholarly impact measurement. Scientometrics 2011 Dec 21;91(2):461-471 [FREE Full text] [doi: 10.1007/s11192-011-0580-x]
23. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
24. Vidakovic B. Chapter 6: Normal distribution. In: Statistics for Bioengineering Sciences: With MATLAB and WinBUGS Support. New York, NY: Springer Publishing Company; Aug 31, 2011:191-230.
25. Carroll C, Paul T, James J, Barry H, William G, Ledi T, et al. NIST/SEMATECH e-Handbook of Statistical Methods. 2003 Jun 01. Histogram Interpretation: Symmetric with Outlier URL: http://www.itl.nist.gov/div898/handbook/ [accessed 2013-11-03] [WebCite Cache ID 6Kr5VtdFC]
26. Howitt D, Cramer D. Shapes of distributions of scores. In: Introduction to Statistics in Psychology. Upper Saddle River, NJ: Prentice Hall, Inc; Nov 25, 2010.
27. Babak S. Exploring numerical variables. In: Robert G, Kurt H, Giovanni P, editors. Biostatistics with R: An introduction to statistics through biological data. California: Springer; Dec 17, 2012:25-38.
28. Motulsky H. The Gaussian distribution. In: Intuitive biostatistics: a nonmathematical guide to statistical thinking. New York, NY: Oxford University Press; 2010.

29. Yeh ST, GlaxoSmithKline, King of Prussia, PA. Exploratory visualization of correlation matrices. 2007 Presented at: NorthEast SAS Users Group (NESUG) conference 2007; November 11-14, 2007; Baltimore, US URL: http://www.nesug.org/proceedings/nesug07/np/np18.pdf
30. David S. Revolutions: a blog dedicated to news and information of interest to members of the R community.: Revolution Analytics; 2009 Mar 23. Visualizing correlation matrices URL: http://blog.revolutionanalytics.com/2009/03/visualizing-correlation-matrices.html [accessed 2013-11-04] [WebCite Cache ID 6Kr97gpkL]
31. Bronstein AM, Bronstein MM, Kimmel R. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. Proceedings of the National Academy of Sciences of the United States of America (PNAS) 2006 Jan 31;103(5):1168-1172 [FREE Full text] [doi: 10.1073/pnas.0508601103]
32. Antoine N, Wlodzislaw D. Interactive data exploration using Mds mapping (2000). 2000 May Presented at: 5th international conference on neural networks and soft computing; May 2000; Zakopane, Poland p. 1-6 URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6425975A7E49D901E17A3A1E94F418FF?doi=10.1.1.35.2761&rep=rep1&type=pdf
33. Taguchi YH, Oono Y. Relational patterns of gene expression via non-metric multidimensional scaling analysis. Bioinformatics 2005 Mar;21(6):730-740 [FREE Full text] [doi: 10.1093/bioinformatics/bti067] [Medline: 15509613]
34. Kenkel N, Orloci L. Applying metric and nonmetric multidimensional scaling to ecological studies: Some new results. Ecology 1986 Aug;67(4):919-928 [FREE Full text] [doi: 10.2307/1939814]
35. János P. Multivariate exploratory analysis of ordinal data in ecology: Pitfalls, problems and solutions. Journal of Vegetation Science 2005;16(5):497-510 [FREE Full text] [doi: 10.1111/j.1654-1103.2005.tb02390.x]
36. Holger K, Walter J. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 2010;37(11):2029-2053 [FREE Full text] [doi: 10.1111/j.1365-2699.2010.02375.x]
37. Wikipedia. Multidimensional scaling URL: http://en.wikipedia.org/wiki/Multidimensional_scaling [accessed 2013-05-12] [WebCite Cache ID 6GYPEABQp]
38. Wasserman S, Katherine F. Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences. Cambridge, UK: Cambridge University Press; Nov 25, 1994.
39. Paul A, Venita D. An Overview of Non-parametric Tests in SAS: When, Why, and How. 2004. URL: http://analytics.ncsu.edu/sesug/2004/TU04-Pappas.pdf [accessed 2013-05-12] [WebCite Cache ID 6GYPluJnz]
40. Determining if skewness and kurtosis are significantly non-normal. New South Wales: School of Psychology University of New England; 2000. URL: http://www.une.edu.au/WebStat/unit_materials/c4_descriptive_statistics/determine_skew_kurt.html [accessed 2013-05-12] [WebCite Cache ID 6GYPwprrt]
41. Beth E, Cristina M, Randolph G, Donald G. Behavioral citation analysis: Toward collection enhancement for users. College & Research Libraries 2006 Jul;67(4):304-324 [FREE Full text]
42. Wilson CS, Tenopir C. Local citation analysis, publishing and reading patterns: Using multiple methods to evaluate faculty use of an academic library's research collection. J Am Soc Inf Sci 2008 Jul;59(9):1393-1408 [FREE Full text] [doi: 10.1002/asi.20812]
43. Tonta Y, Al U. Scatter and obsolescence of journals cited in dissertations of librarianship. Library & Information Science Research 2006;28(2):281-296 [FREE Full text]
44. Edward N, Nan C, Hecht L, Nelson E, Ross J, Fiddler L. Correlation and regression. In: SPSS for Windows Version 11.0: A Basic Tutorial. New York: McGraw-Hill, Inc; 2000.
45. Sean F. A guide for the visually perplexed: visually representing social networks. Stanford: Stanford University; 2004 Jan. URL: http://www.stanford.edu/group/esrg/siliconvalley/docs/networkmemo.pdf [accessed 2013-05-12] [WebCite Cache ID 6GYRnKOwf]
46. Borgatti SP, Everett MG, Freeman LC. Ucinet for Windows: Software for Social Network Analysis. Collegeville: Analytic Technologies, Inc; 2002. URL: http://pages.uoregon.edu/vburris/hc431/Ucinet_Guide.pdf [accessed 2013-05-12] [WebCite Cache ID 6GYSBbor4]
47. Zhao DZ. Towards all-author co-citation analysis. Information Processing and Management: an International Journal - Special issue: Informetrics 2006 Dec;42(6):1578-1591 [FREE Full text] [doi: 10.1016/j.ipm.2006.03.022]
48. Vaughan L, Yang RB, Chen C, Liang WB, Li BY. Extending web co-link analysis to web co-word analysis for competitive intelligence. In: Proceedings of the Annual Conference of the Canadian Association for Information Science. 2010 Presented at: Information Science: Synergy through Diversity; June 2-4, 2010; Montreal, Quebec p. 1-4 URL: http://www.cais-acsi.ca/proceedings/2010/CAIS007_Vaughan_Final.pdf
49. Valeria A. arXiv. 2013 Jun. The potential of preprints to accelerate scholarly communication: A bibliometric analysis based on selected journals URL: http://arxiv.org/ftp/arxiv/papers/1306/1306.4856.pdf [accessed 2013-11-04] [WebCite Cache ID 6Ksc2lHkx]
50. Shuai X, Jiang ZR, Liu XZ, Johan B. A comparative study of academic impact and Wikipedia ranking. In: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries. 2013 Presented at: 13th ACM/IEEE-CS Joint Conference on Digital Libraries; July 22-26, 2013; Indiana p. 25-28 URL: http://www.cs.indiana.edu/~xshuai/papers/acm_wiki.pdf [doi: 10.1145/2467696.2467746]

Abbreviations

API: application programming interface
KMO: Kaiser-Meyer-Olkin
K-S tests: Kolmogorov-Smirnov Test
MDS: multidimensional scale
RSQ: r-squared

Edited by G Eysenbach; submitted 12.05.13; peer-reviewed by D Lee, S Wilson, T Lee; comments to author 15.06.13; revised version received 05.08.13; accepted 19.09.13; published 25.11.13.

Please cite as:
Liu CL, Xu YQ, Wu H, Chen SS, Guo JJ
Correlation and Interaction Visualization of Altmetric Indicators Extracted From Scholarly Social Network Activities: Dimensions and Structure
J Med Internet Res 2013;15(11):e259
URL: http://www.jmir.org/2013/11/e259/
doi:10.2196/jmir.2707
PMID:24275693

Review

Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review

Tara Donker[1], PhD; Katherine Petrie[1], BScPsych (Hons); Judy Proudfoot[1], PhD; Janine Clarke[1], PhD; Mary-Rose Birch[1], MPH; Helen Christensen[1], PhD

Black Dog Institute, University of New South Wales, Sydney, Australia

Corresponding Author:
Tara Donker, PhD
Black Dog Institute
University of New South Wales
Hospital Road
Prince of Wales Hospital
Sydney, 2031
Australia
Phone: 61 20 5988 265
Fax: 61 20 5988 756
Email: t.donker@vu.nl

Abstract

Background: The rapid growth in the use of mobile phone applications (apps) provides the opportunity to increase access to evidence-based mental health care.

Objective: Our goal was to systematically review the research evidence supporting the efficacy of mental health apps for mobile devices (such as smartphones and tablets) for all ages.

Methods: A comprehensive literature search (2008-2013) in MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, PsycINFO, PsycTESTS, Compendex, and Inspec was conducted. We included trials that examined the effects of mental health apps (for depression, anxiety, substance use, sleep disturbances, suicidal behavior, self-harm, psychotic disorders, eating disorders, stress, and gambling) delivered on mobile devices with a pre- to posttest design or compared with a control group. The control group could consist of wait list, treatment-as-usual, or another recognized treatment.

Results: In total, 5464 abstracts were identified. Of those, 8 papers describing 5 apps targeting depression, anxiety, and substance abuse met the inclusion criteria. Four apps provided support from a mental health professional. Results showed significant reductions in depression, stress, and substance use. Within-group and between-group intention-to-treat effect sizes ranged from 0.29-2.28 and 0.01-0.48 at posttest and follow-up, respectively.

Conclusions: Mental health apps have the potential to be effective and may significantly improve treatment accessibility. However, the majority of apps that are currently available lack scientific evidence about their efficacy. The public needs to be educated on how to identify the few evidence-based mental health apps available in the public domain to date. Further rigorous research is required to develop and test evidence-based programs. Given the small number of studies and participants included in this review, the high risk of bias, and unknown efficacy of long-term follow-up, current findings should be interpreted with caution, pending replication. Two of the 5 evidence-based mental health apps are currently commercially available in app stores.

(J Med Internet Res 2013;15(11):e247) doi:10.2196/jmir.2791

KEYWORDS

mobile applications; mobile mental health; mobile phones; self-help; depression; anxiety; stress; substance use

Introduction

Global mobile phone penetration reached 91% at the end of 2012, with 4.3 billion unique mobile subscribers [1] identified. Mobile health (mHealth)—specifically, mental health supported by mobile devices—thus has the potential to be delivered to large numbers of people worldwide. The first mobile software applications or “apps” became available to download on a mobile device in 2008. Since then, penetration has increased rapidly and is anticipated to continue rising. As of September

2012, an estimated 1,520,000 apps had been developed for mobile devices [2], and around 13,600 health apps intended for use by consumers were available for download in Apple's App Store [3]. About 6% of these apps targeted mental health outcomes, while 18% focused on related health issues, such as sleep, stress, relaxation, and smoking behaviors. A survey among the Australian general public indicated that 76% would be interested in using mobile phones for mental health monitoring and self-management [4]. This suggests that mHealth is acceptable and may be a useful vehicle for enhancing access to evidence-based monitoring and self-help for individuals with mild-to-moderate common mental health conditions [4]. Clinical practice guidelines recommend cognitive behavior therapy (CBT) and self-help resources (such as mHealth) as options for psychological treatment for individuals experiencing mild-to-moderate symptoms of anxiety or depression [5]. mHealth apps can be used as stand-alone self-help programs or as a conjunctive treatment modality in guided programs, for example, part of a website or through direct contact with a mental health professional. The app can include treatment components such as cognitive therapy (CT), behavioral activation (BA), psychoeducation, or monitoring of symptoms.

Advantages of mHealth include the improvement of treatment accessibility and participant retention, real-time symptom and activity monitoring and tracking of treatment progress through ecological momentary assessment (EMA), provision of personalized feedback and motivational support, portability and flexibility of use, and the potential to improve adherence to treatment [6-10]. However, there are also disadvantages with using mobile devices for mental health. Technical problems and factors related to telecommunication can arise (eg, battery failures, reliability and sustainability of connections [11]), and issues of data security, patient privacy, and the identification and timely management of crises and risk of harm must be carefully considered when integrating smartphone technology into behavioral health care [12].

Previous research suggests that mental health interventions delivered through mobile apps can be effective in treating a range of mental health disorders, such as depression, stress, anxiety, and smoking cessation [6,7]. However, the thriving development of mental health apps warrants a systematic review of the available evidence base in this growing area. Previous reviews examining evidence-based mental health apps did not incorporate quantitative analyses [12] or included mHealth interventions that were not directly downloadable as an app (such as programs using SMS [short message service] text messaging or Internet-enabled interventions on mobile phones [6,13]). Therefore, the aim of this paper is to systematically review the available evidence-based apps directly downloadable on mobile devices (such as smartphones and tablets) for mental health symptoms or disorders (depression, anxiety, substance use, sleep disorders, suicidal behavior, psychotic disorders, eating disorders, stress, gambling) in children, adolescents, adults, and older individuals.

Methods

Search Strategy and Selection of Studies

A comprehensive literature search in bibliographic databases (MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, PsycINFO, PsycTESTS, and Compendex and Inspec) for relevant articles published from January 1, 2008 (launch date of the first app), to May 30, 2013, was conducted. Terms indicative of mobile apps and mental health disorders were used to search these databases, with the search being limited to "humans", English, and peer-reviewed journals (see Multimedia Appendices 1-3 for the full search string). The identified titles and abstracts were screened for eligibility by 2 independent researchers. Full text copies of all potentially relevant papers, or papers where there was insufficient information in the abstract to determine eligibility, were obtained. Full text articles were further screened and discarded from further analyses if they met exclusion criteria. In addition, references of earlier reviews and reference lists of the included primary articles were examined. Furthermore, key technology journals (Cybertechnology, Behavior and Social Networking; Journal of Medical Internet Research; and Studies in Health Technology and Informatics) were hand-searched. We also reviewed Beacon, a website for evidence-based online programs for mental health, developed and delivered by the Centre for Mental Health Research at the Australian National University. Finally, a search was conducted of prominent individual authors' and researchers' names in the field of mHealth or Internet interventions (see Multimedia Appendix 4) in MEDLINE. Data extraction of relevant articles was completed by 2 independent researchers, with disagreements resolved through discussion or with a third researcher.

We applied strict inclusion criteria in order to investigate any evidence-based mental health apps that could be downloaded from app stores (eg, Google Play for Google Android [14] or the Apple iTunes store [15]). Studies examining the effects of mental health apps on mental health symptoms or disorders (depression, anxiety, substance use, sleep disorders, suicidal behavior, self-harm, psychotic disorders, eating disorders, stress, and gambling) that were directly downloadable on a mobile device (eg, smartphone or tablet) compared with a control group were included. The control group could consist of a wait list, treatment-as-usual, or another treatment. Studies without a control group (pre-post design) were also included. There was no restriction on participant age. Studies were excluded if they did not include an intervention or if mental health symptoms/disorders were not an outcome, and if the intervention was an Internet-based intervention, virtual reality exposure treatment, interactive voice response technology intervention, or a text messaging-only intervention without a mobile application component. Studies were also excluded if the intervention was downloaded on a computer and transferred (eg, through Bluetooth or infrared) to a mobile device, if the intervention targeted a medical disorder (eg, irritable bowel syndrome, diabetes), if the paper provided a description of the mobile application but no outcome data, and if the intervention was developed before 2008. Conference abstracts, protocol

papers, case studies, non-peer reviewed papers, and non-English papers were also excluded.

Quality Assessment

Study quality was assessed according to 6 basic criteria of the Cochrane Risk of Bias Assessment Tool [16]: sequence generation, allocation concealment, blinding of outcome assessors, incomplete outcome data, selective outcome reporting, and other sources of bias. For the third criterion (blinding of outcome), we omitted blinding of participants since blinding participants for treatment allocation is rarely achievable in intervention trials for mental health disorders.

Outcome Measures

Primary outcome measures included reduction of depression symptoms, anxiety symptoms, substance use, sleep disturbance, suicidal behavior (suicide ideation, suicide plans, and attempts), self-harm, psychotic symptoms, symptoms of eating disorders, and gambling, as assessed with validated mental health scales.

Statistical Analyses

When data were available and extractable, intention-to-treat (ITT) within-group and between-group effect sizes (Cohen's *d*) for the intervention group were calculated by taking the difference between the mean pre- and posttest scores (within-group effect size) or the difference of the posttest scores (between-group effect size) and dividing by the pooled standard deviation. Effect sizes of 0.8 can be assumed to be large, while effect sizes of 0.5 are moderate, and effect sizes of 0.2 are small [17]. Where authors provided only *t* test statistics, we computed effect sizes using the formula: $d=t / \text{sort}(df)$ [18]. Hedges' *g* effect sizes were converted to Cohen's *d*. Authors were contacted to provide additional data if needed. Two studies [19,20] did not provide sufficient data to calculate ITT within-group effect sizes.

Results

Selection and Inclusion of Studies

A total of 5464 abstracts in MEDLINE (n=1859), Embase (n=1030), the Cochrane Central Register of Controlled Trials (n=277), PsycINFO (n=1095), PsycTESTS (n=1), and Compendex and Inspec (n=1203) were examined (N=4997 abstracts in total, after removal of duplicates). The majority of records that were excluded addressed nonpsychological technical issues, provided descriptions of mobile apps without outcome data, or were protocol papers or conference abstracts. Of these, 133 full text papers potentially eligible for inclusion were retrieved for further consideration, of which 126 were excluded. Seven trials met inclusion criteria. A further screening for potentially relevant references in recent systematic reviews or meta-analyses and the included studies, individual author names in MEDLINE, and hand-searching of technology journals (Cybertechnology, Behavior and Social Networking; Journal of Medical Internet Research; and Studies in Health Technology and Informatics [January 1, 2008, to May 30, 2013]) and the Beacon website resulted in 95 potentially relevant abstracts and retrieval of 64 additional full text papers for further assessment. Of these, only 1 study met inclusion criteria and was included in the final analysis. In total, 8 trials were identified. These described 5 apps (Mobilyze! [11], mobiletype [21,22], DBT Coach [23], Mobile Stress Management [19,20,24], and Get Happy Program [25]) (see Figure 1 for a flowchart of the screening process). There was a high degree of consensus among raters who screened the titles and abstracts (an interrater reliability of 95.2%).

Characteristics of Included Studies

A total of 227 participants were recruited across all studies. One study [19] did not provide sufficient information about sample size per treatment arm. Of the 8 included studies, 4 trials describing 3 apps assessed depression (Mobilyze!, mobiletype, Get Happy Program), and 3 studies describing 1 app (Mobile Stress Management) assessed stress as a primary outcome measure. Substance use was used as an outcome measure in 1 study (DBT Coach).

Table 1 provides an overview of the included studies (see Multimedia Appendix 5 for the complete version of the table). One study used BA and another used CBT as the therapeutic mode of the intervention. Two studies described a trial delivering emotional self-awareness (ESA), 1 study was based on dialectical behavioral therapy (DBT) and opposite action (ie, emotional regulation skills), and 3 studies described an app delivering stress inoculation training (SIT) as the content of the intervention. Four studies describing 3 trials used an attention-placebo as a control group, 1 study used an active comparison, and 1 study did not specify the nature of the control group. Two studies used a pre-post design without a control group, and all studies except one were feasibility and/or pilot studies. Two studies recruited adults from the community, 1 study recruited from an outpatient clinic, and 2 studies recruited from the workplace. Two studies describing 1 trial recruited adolescents from general practice, and 1 study targeted female university students. Four studies delivered the intervention through a stand-alone mobile app, while 3 studies describing 2 trials used a mobile app alongside a website and EMA to deliver the intervention. One study used a mobile application in conjunction with traditional face-to-face therapy. All included studies delivered the program on a mobile phone, with 1 study also including iPads. Delivery length varied between 6 days and 8 weeks. Five studies assessed posttest outcomes only, whereas 3 studies describing 2 trials undertook follow-up assessments as well (6 weeks and 3 months). Five studies describing 4 apps were guided by mental health professionals through phone or email contact, whereas in 3 studies describing 1 app, participants independently navigated their way through the trial.

Figure 1. Flow diagram of participants.

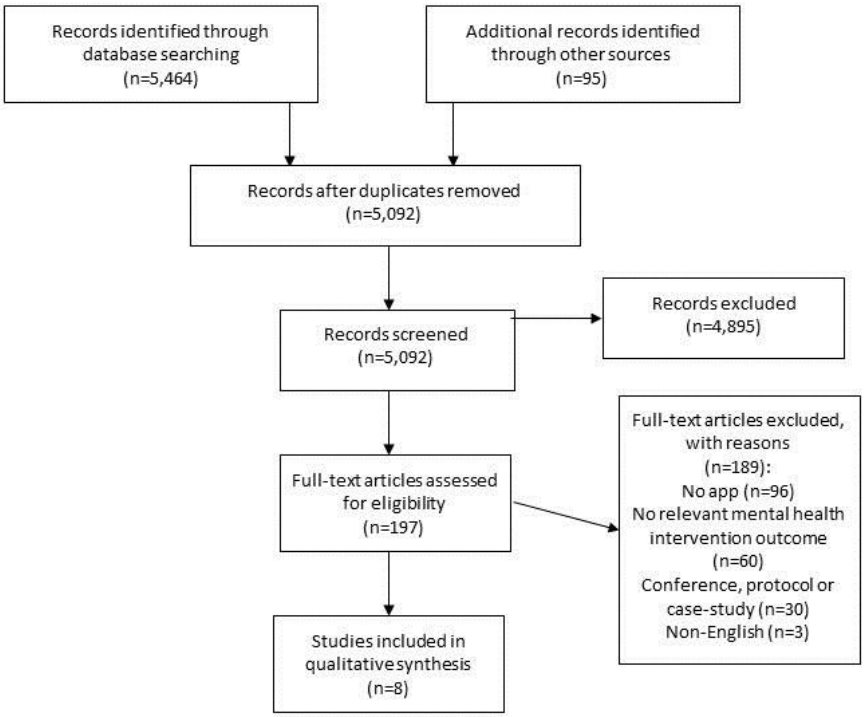


Table 1. Psychosocial studies of applications on mobile devices (Intention-To-Treat). App: Application; BA: Behavioral Activation; BCQ: Behavior Confidence Questionnaire; BDI: Beck Depression Inventory; BPD: Borderline Personality Disorder; BSI: Brief Symptom Inventory; COPE: COPE Inventory; DASS: Depression, Anxiety, and Stress Scale; DBT: Dialectical Behavior Therapy; ESA: Emotional Self Awareness; F2F: Face-to-Face; GAD-7: Generalized Anxiety Disorder-7 item scale; GP: General Practitioner; It: Italian; K10: Kessler Psychological Distress Scale-10 item scale; MDD: Major Depressive Disorder; MHP: Mental Health Professional (psychologist or psychotherapist, GP); MINI: Mini-International Neuropsychiatric Interview; NA: Not Applicable; OA: Opposite Action; PHQ: Patient Health Questionnaire; QIDS-C: Quick Inventory of Depression Symptoms-Clinician Rated; RCT: Randomized Controlled Trial; SIT: Stress Inoculation Training; STAI: State-Trait Anxiety Inventory.

Author (year); name of app	Trial	Primary outcome measure	Study sample	Intervention group	Control group	Delivery type	Delivery length and support	Within[f] and between[g] effect size (Cohen's *d*)
Burns et al (2011); *Mobilyze!*	Pre-post pilot	MDD	Adults from the com-mu-nity	n=8; BA	NA	Mobile app + website + EMA on mobile phone	8 weeks; MHP	PHQ-9: d=1.95[a,e,f] d=2.28[a,e,f] GAD-7: d=1.37[a,e,f]
Kauer et al (2012); Reid et al (2011); *Mobile-type*	RCT	MDD	Adoles-cents from gen-eral prac-tice	n=68; ESA + Individual-ized data summa-ry reports + meet-ing with GP	n=49; Attention control + part of the indi-vidual-ized data summary reports + meeting with GP	Stand-alone mobile app + EMA on mo-bile phone	8 modules over 2-4 weeks; MHP	DASS Stress: d=0.37[a,f] d=0.59[b,e,f] d=0.14[a,g] d=0.22[b,g] DASS Anxiety: d=0.31[b,f] d=0.45[b,d,f] d=0.25[a,g] d=0.07[b,g] DASS Depression: d=0.34 d=0.64[b,e] d=0.11[a,g] d=0.09[b,g] ESA: d=0.31[a] d=0.66[b,d] d=0.09[a,e,g] d=0.58[b,d,g]
Rizvi et al (2011); *DBT Coach*	Pre-post pilot	BPD and substance use	Adults from out-patient clinic	N=21; DBT + OA	NA	Mobile app on mobile phone + F2F DBT	10-14 days; MHP	BDI: d=0.55[a,d,f] BSI: d=0.43[a,d,f] BCQ: d=0.59[a,d,f] Emotional intensity to use substance: d=0.52[c,d,f] Urge to use substance: d=0.29[c,d,f]
Villani et al (2012); *Mobile Stress Management*	RCT	Stress	Female oncology nurses	n=8; SIT	n=8; Attention control	Stand-alone mobile app on mobile phone	8 videos over 4 weeks; no sup-port	NA
Villani et al (2011); *Mobile Stress Management*	RCT	Stress	Female oncology nurses	n=15; SIT	n=15; At-tention control	Stand-alone mobile app on mobile phone	8 videos over 4 weeks; no sup-port	STAI (anxiety trait): d=0.41[a,d,f] COPE (Active): d=-0.45[a,d,f] COPE (Denial): d=0.53[a,d,f]

Author (year); name of app	Trial	Primary outcome measure	Study sample	Intervention group	Control group	Delivery type	Delivery length and support	Within[f] and between[g] effect size (Cohen's *d*)
Grassi et al (2011); *Mobile Stress Management*	RCT	Stress	Female university students	n=not reported; SIT	n=not reported; Control	Stand-alone mobile app on mobile phone	6 videos over 6 days; no support	NA
Watts et al (2013); *Get Happy Program*	Pilot RCT	MDD	Adults from the community	n=15; CBT via mobile app	n=20; CBT via computer	Stand-alone mobile app on mobile phone + iPad	6 modules over 8 weeks; MHP	PHQ-9: *d*=1.56[a,e,f] *d*=-0.14[a,g] *d*=1.69[b,e,f] *d*=-0.28[b,g] BDI-II : *d*=1.90[a,e,f] *d*=-0.11[a,g] *d*=2.11[b,e,f] *d*=-0.48[b,g] K10: *d*=1.93[a,e,f] *d*=0.01[a,g] *d*=1.23[b,e,f] *d*=0.03[b,g]

[a]posttest

[b]follow-up

[c]within immediate coaching session

[d]*P*<.05

[e]*P*<.001

[f]within-group effect size;

[g]between-group effect size

Quality Assessment

The quality of the studies varied but was generally low (see Table 2). Three studies describing 2 apps reported adequate sequence generation [21,22,25], whereas 3 studies [19,20,24] did not outline their sequence generation method. Three studies [21,22,25] reported allocation to conditions by an independent (third) party, whereas 3 other studies [19,20,24] did not provide sufficient information on allocation. Two studies that included diagnostic interviews [21,22] reported using blinded outcome assessors, and 4 studies [19,20,24,25] did not report blinding of assessors or used self-report outcome measures. Two studies [11,23] were not eligible for ratings for sequence generation, allocation concealment, or blinding of outcome assessors due to the pre-post study design. In 6 studies [11,21-25], ITT analyses (completeness of follow-up data) were conducted; 1 of these failed to describe dropout rates [23], and only 1 study [11] described reasons for dropout during the intervention. Two studies [19,20] did not state the nature of the statistical analyses or dropout rate at all. Insufficient information and a high risk of bias of selective outcome reporting was present in 3 studies [19,20,23] and 2 studies [21,22] respectively. Three studies [11,23,25] had a high risk of other sources of bias (eg, absence of a control group, possible treatment infidelity) while for 5 studies [19-22,24] the risk of bias from other sources was unclear (due to significant difference at baseline for stress outcome, unequal number of participants in intervention and control group, and insufficient information). None of the included studies met all 6 quality criteria of the Cochrane tool (see Table 2).

Table 2. Risk of bias assessed by Cochrane Risk of Bias Tool[a].

Trials	Sequence generation	Allocation concealment	Blinding	Incomplete outcome data	Selective outcome reporting	Other sources of bias	Total
Burns et al, 2011	NA	NA	NA	0	0	2	2
Grassi et al, 2011	1	1	1	1	1	1	6
Kauer et al, 2012	0	0	0	1	2	1	4
Reid et al, 2011	0	0	0	1	2	1	4
Rizvi et al, 2011	NA	NA	NA	1	1	2	4
Villani et al, 2011	1	1	1	1	0	1	5
Villani et al, 2012	1	1	1	1	1	1	6
Watts et al, 2013	0	0	1	0	0	2	3

[a]0: low risk of bias; 1: insufficient information; 2: high risk of bias; NA: not applicable.

Effects of the Mental Health Apps

Depression

Four studies describing 3 mobile apps [11,21,22,25] targeted depression. Burns et al [11] found a significant reduction in depression caseness (Mini-International Neuropsychiatric Interview [MINI]: Z=2.15, beta [week]=-.65, *P*=.03), as well as depression and anxiety symptoms at posttest (Patient Health Questionnaire [PHQ-9]: *d*=1.95, *P*<.001; Quick Inventory of Depression Symptoms-Clinician Rated: *d*=2.28, *P*<.001; Generalized Anxiety Disorder-7 item scale: *d*=1.37, *P*<.001) in a pilot test of the guided Mobilyze! app alongside a website and EMA for adults from the general population. The Mobilyze! app will be publicly available for download soon.

In a randomized controlled trial (RCT) of a guided mobiletype app with EMA conducted by Kauer et al [21] and Reid et al [22], no significant differences were found at posttest and follow-up on outcomes of depression, anxiety, and stress among adolescents from general practice compared to an attention control group (Depression and Anxiety Stress Scale [DASS] anxiety: *d*=0.07, *P*=.76; DASS depression: *d*=0.09, *P*=.69). However, it should be noted that the control group received largely the same intervention as the experimental group, with the exception of two components; ESA training via EMA and minimal feedback reports. Mediator analyses yielded an indirect effect of group on depression via ESA (beta=–0.610, 95% CI –5.596 to –0.003). Significant small to moderate within-group differences over time were found for the intervention group (DASS stress: *d*=0.37 at posttest; *d*=0.59 at follow-up; DASS anxiety: *d*=0.31 at posttest; *d*=0.45 at follow-up; DASS depression: *d*=0.34 at posttest; *d*=0.64 at follow-up) and control group (DASS stress: *d*=0.14 at posttest; *d*=0.41 at follow-up; DASS anxiety: *d*=0.07 at posttest; *d*=0.08 at follow-up; DASS depression: *d*=0.42 at posttest; *d*=0.61 at follow-up). Between-group effect sizes were small and nonsignificant (DASS stress: *d*=0.14 at posttest; *d*=0.22 at follow-up; DASS Anxiety: *d*=0.25 at posttest; *d*=0.07 at follow-up; DASS depression: *d*=0.11 at posttest; *d*=0.09 at follow-up). The mobiletype app is not publicly available for download to date.

Watts et al [25] found a significant reduction over time (*P*<.001) and large effect sizes in a pilot RCT of a partially guided CBT-based program for depression delivered either via a computer or mobile app (Get Happy Program) (PHQ-9: *d*=1.56; Beck Depression Inventory [BDI-II]: *d*=1.90; Kessler10 [K10]: *d*=1.93). No differences between the two groups were found for depression over time with *P*>.05 (PHQ-9: *P*=.34, *d*=–0.14; BDI-II: *P*=.52, *d*=–0.11; K10: *P*=.90, *d*=–0.01). The Get Happy app is not publicly available for download to date.

Anxiety/Stress

Three RCTs describing 1 unguided mobile app (Mobile Stress Management) using SIT [19,20,24] found a significant decrease in state and trait anxiety (State and Trait Anxiety Inventory [STAI]) and a significant increase in active coping skills among oncology nurses [20,24] and female university students [19] compared to a control group. Grassi et al [19] used a simplified version of the Mobile Stress Management app, which was also effective for reducing stress. However, both Villani et al [20] and Grassi et al [19] did not provide statistical results for intervention versus control group comparisons. Villani et al [24] reported significant decreases in state anxiety over time ($F_{1,28}$=71.365, *P*≤.001) and a significant *group x time* interaction effect for state anxiety ($F_{1,28}$=27.476, *P*≤.001). Within-group effect sizes (converted from *t* test statistics) were small for active coping strategies (COPE Inventory [COPE] Active: *d*=0.42), and large for state anxiety (STAI: *d*=0.84) and denial coping strategies (COPE Denial: *d*=1.08). The Mobile Stress Management app is publicly available for download (Italian version only).

Substance Use

A pilot feasibility study aiming to reduce substance use (alcohol, drugs, and tobacco) among adults suffering from borderline personality disorder using a mobile app (DBT Coach [23]) in conjunction with face-to-face DBT therapy, indicated a significant reduction (*P*<.05) within each DBT Coach session in emotional intensity and urge to use substances (*d*=0.52 and *d*=0.29 respectively). Furthermore, a significant reduction (*P*<.05) in symptoms of depression (BDI: *P*=.014, *d*=0.55), global symptom severity (Brief Symptom Inventory: *P*=.021, *d*=0.43), and confidence in participants' ability to use opposite action (ie, emotion regulation) skills (Behavior Confidence Questionnaire: *P*=.008, *d*=0.59) was noted from pre- to post assessment. Multimedia Appendix 5 outlines the ITT

within-group effect sizes. The DBT Coach app is publicly available for download.

Ecological Momentary Assessment

Mixed findings were obtained from the 2 studies using EMA as part of the intervention. In the Burns et al [11] study, promising accuracy rates (60-91%) were achieved in predicting categorical contextual states (eg, location) based upon participant EMA entries. For participant states rated on continuous self-report scales (eg, mood), predictive capability was poor. Notwithstanding these technological outcomes, Reid et al [21] and Kauer et al [22] demonstrated that increased self-monitoring with EMA by participants did lead to increased ESA and thereby reduced depressive symptoms.

Intervention Feasibility and Adherence

Three studies providing usability and feasibility outcomes (eg, acceptability of the technology, perceived usefulness, perceived utility) reported moderate to high rates of mobile phone usage, feasibility, and participant satisfaction with the intervention [11,23,25]. The dropout rate was reported in 4 studies and varied between 12.5% and 34.3% [11,21,22,25]. Reported reasons for dropout, where described, were mostly due to technical problems [11].

Discussion

Principal Results and Comparison With Prior Work

In general, the studies included in this systematic review showed promising results for evidence-based mental health apps in reducing depressive symptoms and caseness, stress, anxiety, and substance use, similar to previous reviews of mHealth [6,7]. However, due to the high risk of bias in some studies, these findings need to be considered with caution pending replication. Due to the absence of a control group in 2 studies [11,23], it was difficult to determine whether the beneficial effects were attributable to the app itself, a function of natural remission or regression to the mean, or in case of the DBT Coach app [23], due to the face-to-face DBT therapy offered to all participants in conjunction with the app. Additionally, a clear conclusion about the efficacy of the DBT Coach for substance use treatment cannot be drawn yet, since—besides the absence of a control group—change in substance use (eg, amount of alcohol units per week) prior to or after treatment was not reported, nor was a distinction made between different types of substance use (alcohol, drugs, nicotine cessation). Furthermore, some studies failed to provide sufficient information regarding dropout rates or did not report the statistical analyses used [19,20].

The mobiletype app was the only intervention that failed to yield any significant direct effect on depression, although a significant indirect effect was found in a reduction of depressive symptoms through the direct effect of increased ESA [21]. Because the attention-placebo control group received almost the same intervention as the experimental group, except for the ESA component, the nonsignificant finding is likely to be the cause of this finding. This study suggests that repeated self-monitoring over time using EMA on a mobile device may increase ESA and thereby reduce depressive symptoms. Evidence supports a similar mechanism underlying improvements in depression with CBT, where one of the most important components of CBT for depression involves rating one's mood and activities in a diary to raise awareness of how activities influence mood states [26]. The development of mobile devices has facilitated the collection of EMA data, thereby providing a portable and convenient delivery mode with which an individual can incorporate EMA and regular mood monitoring in their daily lives and improve ESA as part of treatment for depression. Although EMA shows promising results in predicting categorical contextual states, it needs to be further optimized to be able to accurately predict mood states [11]. Once refined in such a way to maximize accuracy and temporal resolution and minimize bias, EMA holds considerable potential to reveal dynamic interplay between mood, cognition, and behavior, increase participant self-awareness of such processes, and thereby enhance mental health treatment [27]. Together with the use of biomedical and/or activity sensors, timely personalized feedback can be generated to prompt users. mHealth interventions therefore have the potential to improve current depression treatment considerably [10]. In a similar way to guided Internet interventions [28], guided apps might derive larger effect sizes and adherence rates than stand-alone self-help apps, but more research is necessary to elucidate this.

Usability, Helpfulness, and Satisfaction

Usability, helpfulness, and satisfaction ratings, where assessed, were moderate to high [11,23,25], indicating that mHealth apps are perceived to be a useful vehicle for enhancing access to evidence-based monitoring and self-help. However, common technical problems (eg, battery failure, connectivity, freezing of app) need to be overcome. Adherence rates (if reported) were high, in line with previous research in mHealth [29], but higher when compared to adherence rates seen with Internet-based interventions [30]. It might be that the method of delivery (mobile phone) and its portability and flexible usage, and/or its delivery of personalized feedback may account for these higher retention rates for mobile apps. However, some of the included studies provided subjects with monetary rewards for participation, which is likely to artificially raise adherence rates as well.

Sustainability of Results

Most studies included only posttest assessment or a short-term follow-up (6 weeks). Although 1 study showed sustainable results at 3-month follow-up [25], sustainability of results over a medium- to long-term timeframe requires further investigation and replication. As such, on the basis of current evidence, sustainability of results cannot yet be determined.

Since mental health apps downloadable for use by the general population are increasing rapidly, despite evidence for their efficacy being largely unknown, the focus of this systematic review was on apps only. We applied very stringent inclusion criteria to ensure that we identified the evidence-based mental health apps that could be downloadable in the future by the general public from app stores, for example, Google Play for Google Android [14] or the Apple App Store [15]. Therefore, several highly sophisticated programs using mobile technology were excluded, such as the myCompass program [31] for depression, anxiety, and stress. The CBT-based myCompass

program is delivered via a website with an Internet-enabled mobile phone component and encourages real-time self-monitoring of moods, mood triggers, and lifestyle behaviors using SMS text messaging and email prompts. Other examples of similar programs include an SMS-based txt2quit intervention [32] and a video-based STUB IT intervention [33], both of which have been shown to be effective for smoking cessation, and an SMS-based intervention [34] to increase medication adherence in individuals with schizophrenia. We were also unable to include the innovative INTREPID research [35], which used virtual reality exposure therapy on mobile phones to reduce anxiety.

There are more than 3000 mental health apps for Android, Apple, and Microsoft freely available to download to date, compared to the 8 evidence-based apps we identified through our systematic review. Only 2 of the apps included in this review are currently available for public download, comprising less than 1% of the commercially available apps. A recently published review on existing (commercial) mHealth apps for the most prevalent health conditions in the Global Burden of Disease list provided by the World Health Organization [36] echoes this finding. The authors concluded that the development of mHealth apps was first and foremost driven by commercial and economic motivations rather than scientific motivations behind research. Although the numerous protocols [9,37] and case studies [38,39] we excluded indicate a nascent field of research, the rapid growth and development of thousands of non-evidence-based mental health technologies has generated the need for independent regulation. This is underlined by the alarming findings from previous research [40-42] indicating that only 13-26% of Web-based or app-based interventions for smoking cessation adhere to treatment guidelines. A recent study on commercial apps using EMA for alcohol use echoes these findings [43]. The US Food and Drug Administration has taken an important step towards the development of quality control guidelines for health apps [44], but there are still major issues and dangers concerning the lack of quality control of commercially available mental health apps. Further research and work must be undertaken to develop, test, and disseminate evidence-based mHealth interventions among the public to ensure optimal public health outcomes.

Limitations

This review has several limitations. First, despite an extensive search, the number of included studies was small, which restricted our interpretations as to whether mHealth apps have an effect on reducing mental health symptoms. Second, the number of participants in the included studies was small. As a result, the studies were probably underpowered to detect the more subtle effects of the interventions. Furthermore, small sample sizes hamper the precision and accuracy of the statistical results and therefore limit our interpretations [45]. Third, the quality of the included studies was low. Historically, low quality trials yield positive results [46]. Due to the small number of studies, we were unable to examine whether significant differences existed between higher- and lower-quality studies. Fourth, there were no studies that examined the long-term efficacy of mental health apps. Therefore, long-term effects remain as yet unknown. Finally, only studies from peer-reviewed, English language journals were included in this review. However, the effect of language bias has been shown to have a minimal impact on the conclusions of systematic reviews [47].

Future Research

There is a very clear need for more research in this area. Trials with an RCT design of high quality to minimize risk of bias are needed to determine the efficacy of mental health apps. Unfortunately, the competitive nature and time-consuming process of grant applications and RCT designs necessary for such high-quality research contrasts sharply with the speed of development in this highly innovative technology. Component testing with small sample sizes may offer one solution to help bridge the gap between academia and real-world applications [48]. Research is particularly weak in the domains of sleep disturbance, anxiety disorders, and smoking cessation and needs further investigation. The cost-effectiveness and cost-utility of mHealth, compared to standard care or Internet-based treatment, requires further examination.

Conclusions

In summary, although a firm conclusion cannot yet be drawn, the current systematic review suggests that mobile apps for mental health have the potential to be effective in reducing depression, anxiety, stress, and possibly substance use for individuals experiencing these symptoms. Given the widespread usage of mobile and smartphones and increasing uptake of tablet devices, mHealth has the potential to increase treatment accessibility globally. The difference in the volume of commercial apps compared to the small number of tested evidence-based apps is striking. It warrants the need for public education and further development and research into evidence-based mental health apps and consideration of industry regulation.

Acknowledgments

This study is funded by Black Dog Institute, University of New South Wales. HC is supported by National Health and Medical Research Council Fellowship 525411.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search string MEDLINE and Embase.

[GIF File, 40KB - jmir_v15i11e247_app1.GIF]

Multimedia Appendix 2

Search string Cochrane Compendex and Inspec.

[GIF File, 52KB - jmir_v15i11e247_app2.GIF]

Multimedia Appendix 3

Search string PsycINFO and PsycTESTS.

[GIF File, 75KB - jmir_v15i11e247_app3.GIF]

Multimedia Appendix 4

Individual author names.

[PDF File (Adobe PDF File), 26KB - jmir_v15i11e247_app4.pdf]

Multimedia Appendix 5

Psychosocial studies of applications on mobile devices (intention-to-treat).

[PDF File (Adobe PDF File), 182KB - jmir_v15i11e247_app5.pdf]

References

1. Source Digit. Global Mobile Penetration Reached 91 Percent in Q3 2012; 6.4 Billion Mobile Subscribers Worldwide. 2012. URL: http://sourcedigit.com/1264-global-mobile-penetration-q3-2012/ [accessed 2013-06-17] [WebCite Cache ID 6HQu0WEDj]
2. CNet. Google ties Apple with 700,000 Android Apps. URL: http://news.cnet.com/8301-1035_3-57542502-94/google-ties-apple-with-700000-android-apps/ [accessed 2013-06-17] [WebCite Cache ID 6HQuFcE3y]
3. Mobi Health News. Report: 13K iPhone consumer health apps in 2012. URL: http://mobihealthnews.com/13368/report-13k-iphone-consumer-health-apps-in-2012/ [accessed 2013-06-17] [WebCite Cache ID 6HQuIsXID]
4. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
5. National Institute for Clinical Excellence (NICE). Depression: management of depression in primary and secondary care (amended). In: NICE Clinical Practice Guideline. London: NICE; 2007:23.
6. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
7. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
8. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
9. Proudfoot J, Nicholas J. Monitoring evaluation in low intensity CBT interventions. In: Bennett-Levy J, Richards DA, Farrand P, Christensen H, Griffiths KM, Kavanagh DJ, et al, editors. Oxford Guide to Low Intensity CBT Interventions. Oxford: Oxford University Press; 2010:97-104.
10. Warmerdam L, Riper H, Klein M, van den Ven P, Rocha A, Ricardo Henriques M, et al. Innovative ICT solutions to improve treatment outcomes for depression: the ICT4Depression project. Stud Health Technol Inform 2012;181:339-343. [Medline: 22954884]
11. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
12. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. Prof Psychol: Res Pract 2011;42(6):505-512.

13. Ehrenreich B, Righter B, Rocke DA, Dixon L, Himelhoch S. Are mobile phones and handheld computers being used to enhance delivery of psychiatric treatment? A systematic review. J Nerv Ment Dis 2011 Nov;199(11):886-891. [doi: 10.1097/NMD.0b013e3182349e90] [Medline: 22048142]
14. Google. Google play URL: https://play.google.com/store [accessed 2013-06-17] [WebCite Cache ID 6HQuU4TgM]
15. Apple. iTunes URL: http://www.apple.com/itunes/ [accessed 2013-06-17] [WebCite Cache ID 6HQudENrw]
16. 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]
17. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J: L. Erlbaum Associates; 1988.
18. Rosnow RL. Effect sizes for experimenting psychologists. Can J Exp Psychol 2003 Sep;57(3):221-237. [Medline: 14596479]
19. Grassi A, Gaggioli A, Riva G. New technologies to manage exam anxiety. Stud Health Technol Inform 2011;167:57-62. [Medline: 21685642]
20. Villani D, Grassi A, Cognetta C, Cipresso P, Toniolo D, Riva G. The effects of a mobile stress management protocol on nurses working with cancer patients: a preliminary controlled study. Stud Health Technol Inform 2012;173:524-528. [Medline: 22357050]
21. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
22. Reid SC, Kauer SD, Hearps SJC, Crooke AHD, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011;12:131. [Medline: 22123031]
23. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
24. Villani D, Grassi A, Cognetta C, Toniolo D, Cipresso P, Riva G. Self-help stress management training through mobile phones: An experience with oncology nurses. Psychol Serv 2013 Aug;10(3):315-322. [doi: 10.1037/a0026459] [Medline: 23937091]
25. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
26. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression. New York: Guilford Press; 1979.
27. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
28. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
29. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-84; discussion 485. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
30. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
31. Black Dog Institute. MyCompass. URL: http://www.blackdoginstitute.org.au/docs/mycompassbackgroundinfo.pdf [accessed 2013-06-24] [WebCite Cache ID 6HcqhVbSf]
32. Milne K, Bowler S, Li J, Salmon P. Evaluation of the first year of the Txt 2 Quit service: 17 June 2008–16 June 2009. Wellington, New Zealand: The Quit Group; 2009. URL: http://www.quit.org.nz/file/research/FINAL%202008-09%20Txt2Quit%20evaluation%20report%2020090731.pdf [accessed 2013-06-18] [WebCite Cache ID 6HSGpLKIk]
33. Whittaker R, Maddison R, McRobbie H, Bullen C, Denny S, Dorey E, et al. A multimedia mobile phone-based youth smoking cessation intervention: findings from content development and piloting studies. J Med Internet Res 2008;10(5):e49 [FREE Full text] [doi: 10.2196/jmir.1007] [Medline: 19033148]
34. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
35. Riva G, Gorini A, Gaggioli A. The Intrepid project - biosensor-enhanced virtual therapy for the treatment of generalized anxiety disorders. Stud Health Technol Inform 2009;142:271-276. [Medline: 19377166]
36. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]

37. Bockting CL, Kok GD, van der Kamp L, Smit F, van Valen E, Schoevers R, et al. Disrupting the rhythm of depression using Mobile Cognitive Therapy for recurrent depression: randomized controlled trial design and protocol. BMC Psychiatry 2011;11:12 [FREE Full text] [doi: 10.1186/1471-244X-11-12] [Medline: 21235774]
38. Ly KH, Carlbring P, Andersson G. Behavioral activation-based guided self-help treatment administered through a smartphone application: study protocol for a randomized controlled trial. Trials 2012;13:62 [FREE Full text] [doi: 10.1186/1745-6215-13-62] [Medline: 22607302]
39. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
40. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
41. Bock B, Graham A, Sciamanna C, Krishnamoorthy J, Whiteley J, Carmona-Barros R, et al. Smoking cessation treatment on the Internet: content, quality, and usability. Nicotine Tob Res 2004 Apr;6(2):207-219. [doi: 10.1080/14622200410001676332] [Medline: 15203794]
42. Bock BC, Graham AL, Whiteley JA, Stoddard JL. A review of web-assisted tobacco interventions (WATIs). J Med Internet Res 2008;10(5):e39 [FREE Full text] [doi: 10.2196/jmir.989] [Medline: 19000979]
43. Cohn AM, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
44. U.S. Food and Drug Administration. Draft guidance for industry and food and drug administration staff- mobile medical applications. URL: http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm263280.htm [accessed 2013-09-05] [WebCite Cache ID 6JNnp2sdo]
45. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
46. Cuijpers P, Andersson G, Donker T, van Straten A. Psychological treatment of depression: results of a series of meta-analyses. Nord J Psychiatry 2011 Dec;65(6):354-364. [doi: 10.3109/08039488.2011.596570] [Medline: 21770842]
47. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. Clin Orthop Relat Res 2007 Feb;455:23-29. [doi: 10.1097/BLO.0b013e31802c9098] [Medline: 17279036]
48. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523. [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]

Abbreviations

BA: behavioral activation
BDI: Beck Depression Inventory
BDI-II: Beck Depression Inventory-Second edition
CBT: cognitive behavior therapy
COPE: COPE Inventory
CT: cognitive therapy
DASS: Depression, Anxiety and Stress Scale
DBT: dialectical behavior therapy
EMA: ecological momentary assessment
ESA: emotional self-awareness
ITT: intention-to-treat
K10: Kessler Psychological Distress Scale-10 item scale
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatric Interview
PHQ: Patient Health Questionnaire
PHQ-9: Patient Health Questionnaire-9 item scale
RCT: randomized controlled trial
SIT: stress inoculation training
STAI: State-Trait Anxiety Inventory

Edited by P Carlbring; submitted 25.06.13; peer-reviewed by J Ruwaard, O Kristjansdottir; comments to author 21.08.13; revised version received 17.09.13; accepted 18.09.13; published 15.11.13.

Please cite as:
Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H
Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review
J Med Internet Res 2013;15(11):e247
URL: http://www.jmir.org/2013/11/e247/
doi:10.2196/jmir.2791
PMID:24240579

Corrigenda and Addenda

Correction: Systematic Reviews and Meta-Analyses of Home Telemonitoring Interventions for Patients With Chronic Diseases: A Critical Assessment of Their Methodological Quality

Spyros Kitsiou[1], PhD; Guy Paré[1*], PhD; Mirou Jaana[2,3*], PhD

[1]Information Technology in Health Care, HEC Montreal, Montreal, QC, Canada

[2]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[3]School of Business, Lebanese American University, Beirut, Lebanon

[*]these authors contributed equally

Corresponding Author:
Spyros Kitsiou, PhD
Information Technology in Health Care
HEC Montreal
3000, chemin de la Côte-Sainte-Catherine
Montreal, QC, H3T 2A7
Canada
Phone: 1 514 340 6000 ext 2653
Fax: 1 514 340 6132
Email: spyros.kitsiou@hec.ca

Related Article:

Correction of: http://www.jmir.org/2013/7/e150/

(J Med Internet Res 2013;15(11):e253) doi:10.2196/jmir.3081

In "Systematic Reviews and Meta-Analyses of Home Telemonitoring Interventions for Patients With Chronic Diseases: A Critical Assessment of Their Methodological Quality" (J Med Internet Res 2013;15(7):e150), there was an error in Table 6. After the second row of the subheading Clarke 2011 under the column “Low interval (95% CI)”, a 0 above the value 52 was inadvertently deleted in the copyediting process of the manuscript. As a result, all the values in the subsequent rows of “Low interval (95% CI)” in Table 6 were shifted upward, for example, the 52 in the third row belongs to the fourth row of the subheading Clarke 2011 under the column “Low interval (95% CI)”. In the last row of Table 6, the low interval should have been 0 and the high interval should have been 71 instead of being blank cells. These errors have been corrected in the online version of the paper on the JMIR website on November 4, 2013, together with publishing this correction notice. A correction notice has been sent to PubMed and the correct full-text has been resubmitted to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 04.11.13; ###Reviewer names will be inserted here### published 07.11.13.

Please cite as:
Kitsiou S, Paré G, Jaana M
Correction: Systematic Reviews and Meta-Analyses of Home Telemonitoring Interventions for Patients With Chronic Diseases: A Critical Assessment of Their Methodological Quality
J Med Internet Res 2013;15(11):e253
URL: http://www.jmir.org/2013/11/e253/
doi:10.2196/jmir.3081
PMID:24191346

Corrigenda and Addenda

Metadata Correction: Design and Evaluation of a Simulation for Pediatric Dentistry in Virtual Worlds

Lazaros Papadopoulos[1], BCompSci, MS Medical Informatics, DDS; Afroditi-Evaggelia Pentzou[2], BCompSci; Konstantinos Louloudiadis[3]; Thrasyvoulos-Konstantinos Tsiatsos[2]

[1]Laboratory of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]Multimedia Lab (Division of Technology-Enhanced Learning), Department of Informatics of the Faculty of Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece

[3]Division of Preventive Dentistry, Periodontology and Biology of Implants, School of Dentistry, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:
Lazaros Papadopoulos, BCompSci, MS Medical Informatics, DDS
Laboratory of Medical Informatics
Medical School
Aristotle University of Thessaloniki
Panepistimioupoli
Thessaloniki, 54124
Greece
Phone: 30 2310999272
Fax: 30 2310999263
Email: lazapap@hotmail.gr

Related Article:

Correction of: http://www.jmir.org/2013/10/e240/

(J Med Internet Res 2013;15(11):e268) doi:10.2196/jmir.3134

"Design and Evaluation of a Simulation for Pediatric Dentistry in Virtual Worlds" (J Med Internet Res 2013;15(10):e240) was published on Oct 29th in issue 10 (with correct citation information in the table of contents) but the article itself originally displayed incorrect "please cite as" citation information, displaying issue 11 instead of issue 10. This error was corrected in the online version of the paper on the JMIR website on November 26, 2013, together with publishing this correction notice. As a result, the URL for this paper has changed from http://www.jmir.org/2013/11/e240/ to http://www.jmir.org/2013/10/e240/. No content changes were made.

A correction notice has been sent to PubMed. This was done before submission to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 25.11.13; this is a non–peer-reviewed article;accepted 25.11.13; published 26.11.13.

Please cite as:
Papadopoulos L, Pentzou AE, Louloudiadis K, Tsiatsos TK
Metadata Correction: Design and Evaluation of a Simulation for Pediatric Dentistry in Virtual Worlds
J Med Internet Res 2013;15(11):e268
URL: http://www.jmir.org/2013/11/e268/
doi:10.2196/jmir.3134
PMID:24284162

Corrigenda and Addenda

Correction: Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study

Andrey Zheluk[1], BAppSc, GradDip(History), MBA, GradDip(IT); Casey Quinn[2], BCom, Mphil, PhD; Daniel Hercz[3], BSc, MSc Biostat; James A Gillespie[1], BA (Hons), PhD

[1]Menzies Centre for Health Policy, The University of Sydney, University of Sydney NSW, Australia

[2]PRMA Consulting Ltd, New York, NY, United States

[3]University of Sydney, University of Sydney NSW, Australia

Corresponding Author:
Andrey Zheluk, BAppSc, GradDip(History), MBA, GradDip(IT)
Menzies Centre for Health Policy
The University of Sydney
D02 Victor Coppleson Building
University of Sydney NSW, 2006
Australia
Phone: 61 2 9351 2818
Fax: 61 2 9351 5204
Email: andreyzheluk@gmail.com

Related Article:

Correction of: http://www.jmir.org/2013/11/e256/

(J Med Internet Res 2013;15(11):e267) doi:10.2196/jmir.3112

The order of references in "Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study" (J Med Internet Res 2013 (Nov 12); 15(11):e256) was incorrect. References 60-95 were out of order, while references 1-59 were correct and remain unchanged. Specifically, reference 60 should have been deleted, and references 61-76 shifted up one to become references 60-75. Reference 78 is now 76, while reference 77 remains unchanged. References 79-87 are now shifted up one to become references 78-86, while references 88-95 remain unchanged. Reference 87 is replaced with a new reference (Jolliffe).

This error has been corrected in the online version of the paper on the JMIR website on November 26, 2013, together with publishing this correction notice. A correction notice has been sent to PubMed. This was done before submission to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 22.11.13; this is a non–peer-reviewed article;accepted 22.11.13; published 26.11.13.

Please cite as:
Zheluk A, Quinn C, Hercz D, Gillespie JA
Correction: Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study
J Med Internet Res 2013;15(11):e267
URL: http://www.jmir.org/2013/11/e267/
doi:10.2196/jmir.3112
PMID:24284136

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org