Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 15 (2013), Issue 10   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Viewpoints

Letters to the Editor

Original Paper

Determinants of Successful Weight Loss After Using a Commercial Web-Based Weight Reduction Program for Six Months: Cohort Study

Elisa Postrach[1], MSc; Rosa Aspalter[2], MD; Ulf Elbelt[1,3], MD; Michael Koller[4], PhD; Rita Longin[2], MSc; Jörg-Dieter Schulzke[1], MD, PhD; Luzia Valentini[1], PhD

[1]The Division of Nutritional Medicine, Department of Gastroenterology and Hepatology, Charité- Universitätsmedizin Berlin, Berlin, Germany
[2]KiloCoach e.U., Vienna, Austria
[3]Department of Endocrinology, Diabetes and Nutrition, Charité- Universitätsmedizin Berlin, Berlin, Germany
[4]Centre for Clinical Studies, University Hospital Regensburg, Regensburg, Germany

Corresponding Author:
Luzia Valentini, PhD
The Division of Nutritional Medicine
Department of Gastroenterology and Hepatology
Charité- Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 514 113
Fax: 49 30 450 514 923
Email: luzia.valentini@charite.de

Abstract

Background: The Internet is widely available and commonly used for health information; therefore, Web-based weight loss programs could provide support to large parts of the population in self-guided weight loss. Previous studies showed that Web-based weight loss interventions can be effective, depending on the quality of the program. The most effective program tools are visual progress charts or tools for the self-monitoring of weight, diet, and exercises. KiloCoach, a commercial program currently available in German-speaking countries, incorporates these features. A previous investigation showed that the program effectively supports users in losing weight.

Objective: We investigated weight loss dynamics stratified by weight loss success after 6-month use of KiloCoach. Furthermore, we analyzed possible associations between intensity of program use and weight loss. The results are intended for tailoring user recommendations for weight-loss Internet platforms.

Methods: Datasets of KiloCoach users (January 1, 2008 to December 31, 2011) who actively used the platform for 6 months or more were assigned to this retrospective analysis. Users (N=479) were 42.2% men, mean age of 44.0 years (SD 11.7), with a mean body mass index (BMI) of 31.7 kg/m^2 (SD 3.2). Based on the weight loss achieved after 6 months, 3 success groups were generated. The unsuccessful group lost <5%, the moderate success group lost 5%-9.9%, and the high success group lost ≥10% of their baseline body weight. At baseline, the unsuccessful (n=261, 54.5%), moderate success (n=133, 27.8%), and high success (n=85, 17.8%) groups were similar in age, weight, BMI, and gender distribution.

Results: After 6 months, the unsuccessful group lost 1.2% (SD 2.4), the moderate success group lost 7.4% (SD 1.5), and the high success group lost 14.2% (SD 3.8) of their initial weight ($P<.001$). Multivariate regression showed that early weight loss (weeks 3-4), the total number of dietary protocols, and the total number of weight entries were independent predictors for 6-month weight reduction (all $P<.001$) explaining 52% of the variance in weight reduction. Sensitivity analysis by baseline carried forward method confirmed all independent predictors of 6-month weight loss and reduced the model fit by only 11%. The high success group lost weight faster and maintained weight loss more efficiently than the other groups ($P<.001$). Early weight loss was associated with weight maintenance after 1 year and 2 years (both $P<.001$). Weight dynamics did not differ between men and women over 6 months when adjusted for baseline and usage parameters ($P=.91$). The percentage of male long-term users was unusually high (42.2%).

Conclusions: Our results suggest that early weight loss and close program adherence (ie, 5 dietary protocols per week and weekly entering of current weight), especially in the early phase of program usage, can improve weight loss outcome.

(J Med Internet Res 2013;15(10):e219) doi:10.2196/jmir.2648

KEYWORDS

Internet; weight loss; overweight; obesity; weight reduction program; efficiency; program evaluation; preventive health services; sex

Introduction

In 2008, more than 50% of European men and women were overweight [1]. Obesity was present in 21% of women and 22% of men aged 20 years or older [2]. Between 1980 and 2008, the prevalence of obesity nearly doubled worldwide [1]. During the same period, Internet availability and usage also increased significantly. In the European Union, the percentage of citizens aged between 16 and 74 years with Internet access at home increased from 33% in 2004 to 67% in 2011 [3]. Furthermore, 38% of Europeans searched the Web for health-related information in 2011 [4]. This suggests that weight loss programs delivered via the Internet have the potential to reach and be accepted by large numbers of European citizens. Commercial weight loss platforms are steadily increasing in number and warrant special attention.

Web-based weight loss programs provide a health intervention that is flexible, timesaving, and cost-effective [5]. For weight loss interventions with increased intensity (eg, treatment by a doctor or other professionals), overweight and obese individuals reported an increasing number of obstacles (eg, lack of money or time, disgrace) [6]. Web-based programs overcome traditional access barriers of face-to-face counseling (eg, by protecting user anonymity [7] or by reducing travel times [8]), thus appealing to broad levels of the population [9]. Several Web-based weight loss interventions have been shown to be efficient in supporting weight loss [10-14]. Two systematic reviews with meta-analysis on Web-based weight loss interventions found that Internet-based programs have the potential to achieve weight loss and can result in weight loss outcomes comparable to other weight loss interventions [15,16]. Online tools that visualize goal progress or feedback, such as a body mass index (BMI) calculator or progress graphs, were found to be especially supportive in weight reduction [17].

Since 2005, a commercial online weight loss program that incorporates both feedback and visualizing tools has been available in German-speaking countries (KiloCoach). A previous study showed that KiloCoach users who continuously entered dietary protocols for at least 60 days (n=946) lost 4.1% (SD 5.5) of their baseline weight [18]. Program users who entered protocols for 1 year (n=104) lost 6.4% (SD 7.3) of their baseline body weight [18].

The primary objective of the present work was not to evaluate the overall weight loss efficacy of this program, but to investigate the weight loss dynamics of KiloCoach users who used the program for at least 6 months and to associate final weight loss with the use of different program tools. Based on our results, we aimed at drafting user recommendations on how to use the program more effectively in future.

Methods

Program Description

KiloCoach is available on the Web [19]. The key concept of this commercial program is to induce lifestyle changes that lead to weight loss. Users are encouraged to adapt healthier eating and activity habits by means of self-monitoring combined with tailored feedback and information about health and nutrition.

Self-monitoring includes optional logbook and weight entries in addition to dietary protocols (Multimedia Appendix 1). The most important program tool is the dietary protocol, which is the electronic version of the common written protocols for recording food intake (Figure 1). It provides the electronic facility to quickly record all food items and drinks from a database of approximately 40,000 items. Additionally, physical activity can be recorded daily. Based on the dietary protocol, energy intake and expenditure are calculated, analyzed, and visualized in real time to provide immediate feedback to the user. The logbook is a private blog that offers the opportunity to document special situations or additional anthropometric measurements. Thus, logbook entries are defined as small personal notes entered by the user. A weight entry is a weight recorded by the user. To assist with weight reduction, KiloCoach calculates an upper threshold for daily energy intake (kcal) based on a user's body data and individual weight loss goal, considering that weight loss should not exceed 1 kg per week. Further supportive features are analyzing tools that analyze diet composition or identify food groups that mainly contribute to energy intake, for example (Multimedia Appendix 1). Finally, users can actively participate in the weight loss community's online forums and contact nutrition, sports, coaching, as well as medical experts, if required.

Because KiloCoach is based on a healthy diet, encourages participants to increase physical activity, and anticipates a weight loss speed of 0.5 to 1.0 kg per week, it fulfills the internationally accepted criteria for recommendable weight loss programs [20].

Datasets from KiloCoach users who became members for at least 2 months between January 1, 2008 and December 31, 2010, and who did not report confounding factors to weight loss (diabetes mellitus, hypothyroidism, limited motility, pregnancy, or lactation) were eligible for the present analysis.

Additional inclusion criteria were at least 1 dietary protocol available during program use, no program interruption longer than 3 months, age ≥ 18 years, and BMI between 27 to 39.9

kg/m^2 [10,21]. The BMI cut-off points were chosen to allow the direct comparison of this retrospective analysis with the results of a prospective controlled trial that is currently being carried out in our center (clinicaltrials.gov registration: NCT01634204) and with other studies already published.

All 1123 datasets fulfilling the predefined criteria were extracted by KiloCoach and sent unmodified to the Charité - Universitätsmedizin Berlin. The datasets of users who were still active after 6 months were selected by the Charité - Universitätsmedizin Berlin. This selection was done first because various guidelines on the treatment of overweight and obesity recommend a weight loss duration of 6 months [20,22]. Moreover, a duration of 6 months is used in many international weight loss trials, which allows for comparisons.

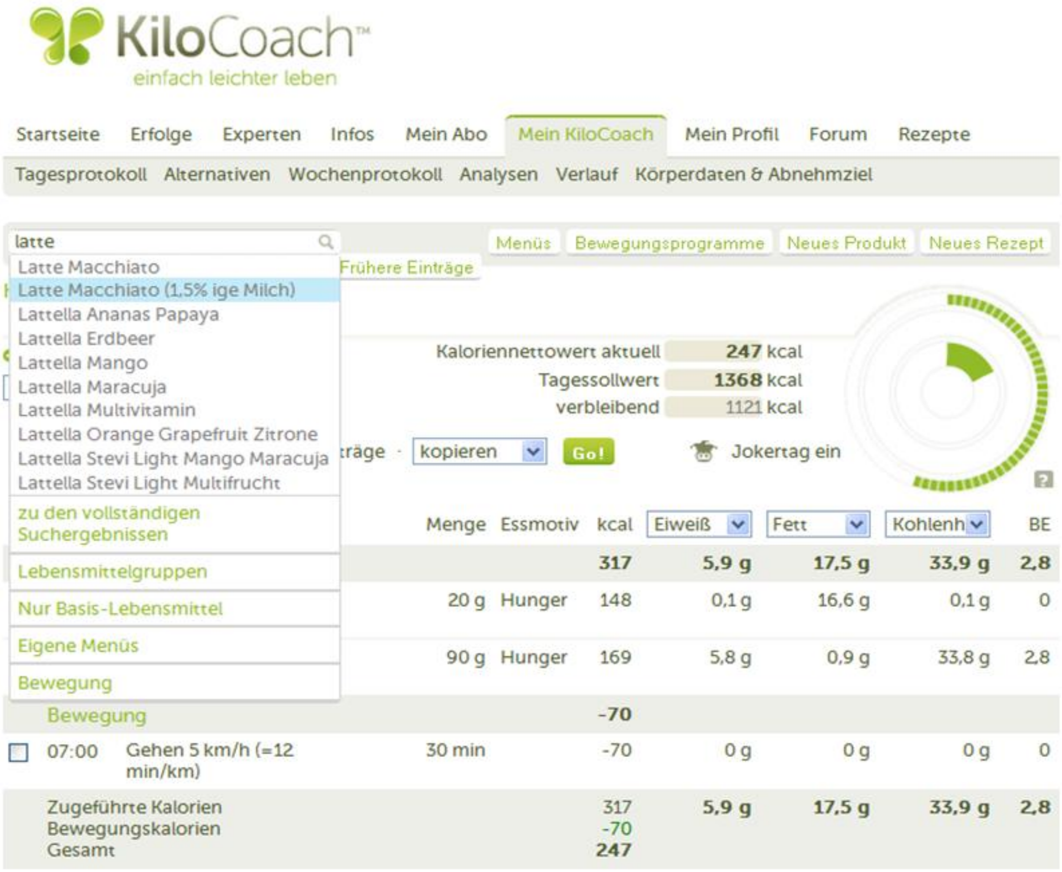
The resulting sample eligible for analysis, referred to as study sample, contained 479 datasets (Figure 2). An observation period from January 1, 2008 to December 31, 2011 was chosen to allow a usage period of at least 12 months for every user.

Each dataset contained self-reported personal data, such as age, sex, height, and body weight. The BMI was calculated from self-reported weight and height. Additional data were duration of membership, number of purchased membership days, and indicators for program usage and compliance expressed as frequencies of dietary protocols, weight entries, logbook entries, and meals per day. As expected for a self-guided program, not every user entered his or her weight at the same time or used the program following the same pattern. Thus, we averaged weight and BMI over periods of 2 weeks for the first 6 months (weeks 1-2, weeks 3-4,..., weeks 25-26). Frequencies of dietary protocols, weight entries, and logbook entries were expressed in absolute numbers over the same time periods.

Users were divided into 3 weight loss success groups, referred to as unsuccessful, moderate success, and high success. These groups were based on an achieved percentage weight loss of <5%, 5%-9.9%, and ≥10% of initial weight, respectively, after 6 months of program usage. This classification was chosen because weight reduction lower than 5% is considered insignificant [22], between 5% and 10% is moderate [23], and above 10% is high [24].

Data entries after the 6-month period (after weeks 25-26) were referred to as follow-up because the weight loss phase changes into weight maintenance after 6 months [20]. For follow-up data, intervals were created for 1, 1.5, and 2 years (weeks 47-57, weeks 72-84, and weeks 93-114, respectively).

Figure 1. Screenshot of a dietary protocol in KiloCoach (see Multimedia Appendix 1 for explanation).

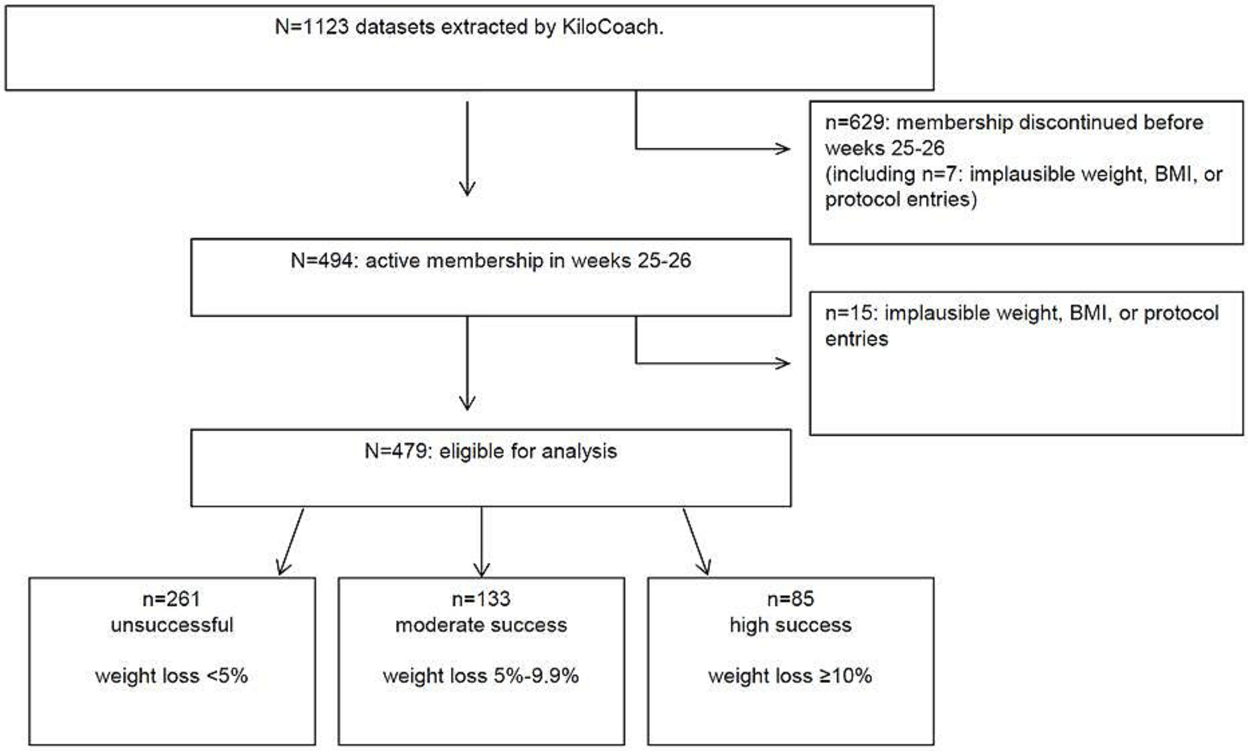


latte

Kaloriennettowert aktuell 247 kcal
Tagessollwert 1368 kcal
verbleibend 1121 kcal

		Menge	Essmotiv	kcal	Eiweiß	Fett	Kohlenh	BE
				317	**5,9 g**	**17,5 g**	**33,9 g**	**2,8**
		20 g	Hunger	148	0,1 g	16,6 g	0,1 g	0
		90 g	Hunger	169	5,8 g	0,9 g	33,8 g	2,8
	Bewegung			**-70**				
☐ 07:00	Gehen 5 km/h (=12 min/km)	30 min		-70	0 g	0 g	0 g	0
	Zugeführte Kalorien Bewegungskalorien Gesamt			317 -70 **247**	**5,9 g**	**17,5 g**	**33,9 g**	**2,8**

Figure 2. Dataset flow through the screening process to success group allocation.



Statistical Analysis

Overview

Statistical analysis was carried out with SPSS version 19 (IBM Corp, Armonk, NY, USA), and *P* values <.05 were considered statistically significant. Descriptive results are given as means and standard deviations (SD) if not indicated otherwise. Missing system values for the numbers of dietary protocols and logbook entries originated from nonuse in the respective time periods and were consequentially replaced by 0 numbers resulting in exact and complete information on program usage.

In the study sample, 39.31% of possible weight entries (2448/6227) were missing because of the self-guided character of the program. For example, users were not aware to actively re-enter weight during periods of weight stabilization, which often exceeded 2 weeks during active weight loss attempts. Overall, 1187 of 6227 weight entries (19.06%) were unavailable between 2 active weight entries. Only 1261 of 6227 (20.25%) were missing after the last weight entry, and 1013 (80.33%) of these belonged to the unsuccessful group, when the last observation carried forward method (LOCF) was used to complement missing values. Consequentially, only 248 weight data were missing after the last weight entry in the moderate success and high success groups (8.75% of moderate and high or 3.98% of total sample). There is no consensus on how to deal with missing values [25]. From a clinical standpoint, LOCF provided the best estimate to complement missing values of self-reported weight in this scenario. Consequentially, we chose LOCF imputation for primary analysis.

Nevertheless, we additionally performed a sensitivity analysis by using the baseline carried forward (BCF) method, a more conservative estimate for completing missing values. In our sample, which is characterized by a high number of in-between missing weight entries, BCF leads to clinically implausible weight in users with significant weight loss. It also results in a worst-case scenario for success group allocation, because all users who did not enter weight in weeks 25-26 were allocated to the unsuccessful group even when a significant weight loss was confirmed by active weight entries 2 weeks before (see subsequent sensitivity analyses also). This negative scenario was considered adequate to test the robustness of LOCF results.

Missing weight data in the follow-up period (after weeks 25-26) were not complemented by LOCF or any other imputation.

Linear Analysis by General Linear Model Repeated Measures and Multivariate Regression Analysis

Changes from baseline in outcome measures were analyzed with a general linear model (GLM) for repeated measures. In a 14-level model time (duration of platform usage), group (success group), and sex were tested and adjusted for baseline values and usage markers (age, BMI, number of protocols, and number of weight entries as covariates). The 13 degrees of freedom (df) contrast describing the difference in trajectories over time among success groups was taken as a primary indication of different weight dynamics dependent on final weight reduction. The 13-df contrast describing the difference in trajectories over time between men and women was taken as a primary indication of different weight dynamics dependent on sex.

We applied multivariate linear regression analyses controlled for collinearity (variance inflation factor <3), autocorrelation (Durbin-Watson statistic=1.18), and outliers (standardized residues <3.2) to evaluate the impact of early weight loss, baseline characteristics, and user behavior on total weight loss at 6 months in the primary analyses and the sensitivity analyses. In the primary analyses, case-wise diagnoses identified 2 participants with extreme weight loss of 18% to 20% (no: 617 and 923), who significantly influenced the results (z scores > 3.1). We kept both participants in the model. Omitting their data would have increased the adjusted multivariate coefficient of determination (adjusted R^2) by 2% without changing the main results of the analyses. Thus, all participants were included in the analyses (N=479).

Post Hoc Analysis

The post hoc analysis of differences among all groups was done with the Kruskal-Wallis test for continuous and discrete variables and the chi-square test (χ^2) for binary variables. Two group comparisons were performed using the Mann-Whitney *U* test; Spearman rank correlation (ρ) was used for bivariate analyses. All post hoc analyses were exploratory; therefore, no Bonferroni adjustment was applied [26]. In some figures, box-whisker plots were used displaying the 25th, 50th, and 75th percentiles in the boxes and the minimum and maximum as whiskers, except for extreme values.

Sensitivity Analyses

To test the robustness of the LOCF imputation, we conducted a sensitivity analysis. First, we used the BCF method to complement all missing weight data within the first 6 months of program usage in long-term users (N=479).

Second, we evaluated separately the subgroup of users who coincidentally entered their body weight at our endpoint for weight loss, weeks 25-26 (N=214). In this group, final weight was self-reported by all users and only in-between missing weight entries had to be completed by LOCF. This sample provides the most probable reproduction of weight dynamics and success group allocation, albeit at the cost of group size and representability.

Results

Kilocoach Users With and Without Weight Entry After Six Months

KiloCoach users who actively used the platform for at least 6 months were referred to as long-term users and were chosen as the study sample (N=479). The study sample accounted for 42.7% of the total KiloCoach population.

Baseline weight, height, and BMI were statistically different between long-term and short-term users, but the numeric difference was insignificant, as shown in Table 1. The study sample had a higher proportion of men and, as expected, significantly more purchased membership days compared to short-term users.

Table 1. General characteristics of long-term users (≥6 months activity), short-term users (<6 months activity), and all users of KiloCoach.

General characteristics	Total population N=1123	Long-term users n=479	Short-term users n=644	P [a]
Sex (male), n (%)	435 (38.7)	202 (42.2)	233 (36.2)	<.001
Age (years), mean (SD)	44.2 (11.8)	44.0 (11.7)	43.8 (11.8)	.20
Initial weight (kg), mean (SD)	92.5 (13.8)	94.4 (14.4)	92.0 (14.3)	<.001
Height (cm), mean (SD)	172 (8.7)	172 (8.8)	171 (8.6)	.04
BMI (kg/m^2), mean (SD)	31.3 (3.1)	31.7 (3.2)	31.0 (3.0)	<.001
Purchased membership days, mean (SD)	275 (236)	413 (264)	172 (143)	<.001

[a] Long-term vs short-term users; Mann-Whitney *U* test used except for sex (chi-square test used).

Primary Analysis of the Dynamics and Predictors of Weight Loss

General Characteristics of the Success Groups

Table 2 shows the baseline characteristics of the KiloCoach study sample. Of the entire sample, 54.5% (261/479) were unsuccessful (lost <5% initial body weight), 27.8% (133/479) were moderately successful (lost 5% to 9.9% initial body weight), and 17.8% (85/479) were highly successful (lost ≥10% initial body weight). Six-month weight reduction significantly differed among the success groups and averaged 5.3% (SD 5.6) in the total group. Users were aged between 18 and 74 years. Sex distribution, age, baseline body weight, height, and baseline BMI did not differ significantly among the success subgroups, although the number of purchased membership days differed significantly among the groups (Table 2).

Table 2. General characteristics of the study sample according to the percentage weight loss achieved after 6 months: unsuccessful (lost <5% of initial body weight), moderate success (lost 5%-9.9% of initial body weight), and high success (lost ≥10% initial body weight).

General characteristics	All (N=479)	Unsuccessful (n=261)	Moderate success (n=133)	High success (n=85)	*P* [a]
Sex (male), n (%)	202 (42.2)	111 (42.5)	49 (36.8)	42 (49.4)	.18
Age (years), mean (SD)	44.0 (11.7)	43.6 (11.6)	46.2 (11.4)	46.2 (12.2)	.08
Initial weight (kg), mean (SD)	94.4 (14.4)	95.2 (14.9)	92.9 (4.9)	94.4 (14.1)	.26
Height (cm), mean (SD)	172 (8.8)	173 (9.3)	172 (8.3)	172 (8.0)	.77
BMI (kg/m^2), mean (SD)	31.7 (3.2)	31.9 (3.4)	31.3 (3.0)	31.6 (3.1)	.28
Purchased membership days, mean (SD)	413 (264)	374 (245)	420 (247)	521 (312)	<.001
6-month weight loss (%), mean (SD)	5.3 (5.6)	1.2 (2.4)	7.4 (1.5)	14.2 (3.8)	<.001

[a] Kruskal-Wallis 1-way ANOVA over the 3 subgroups used except for sex (chi-square test used).

The Dynamics of Weight Loss

The weight loss dynamics showed significant differences among the 3 groups over the weight loss period of 6 months (GLM repeated measures, 13 df contrast, *P*<.001).

The post hoc analysis showed that, despite similar initial weight and BMI, weight reduction after 2 weeks of KiloCoach usage already differed significantly among the 3 groups, with a mean of 0.4% (SD 0.6) in the unsuccessful group, 0.7% (SD 0.7) in the moderate success group, and 0.9% (SD 0.7) in the high success group (*P*<.001) (Figure 3). For all the following time points, weight loss differed significantly among the 3 success groups (*P*<.001). Furthermore, compared to the moderate success group, the high success group achieved significantly more weight reduction from weeks 3-4 onward (mean 3.2%, SD 1.5 vs mean 2.5%, SD 1.5, *P*<.001). The high success group lost weight faster than the other groups. Although both the moderate and the high success groups reached their maximum percentage weight loss, weight loss was significantly more in the high success group (*P*<.001) after 6 months. This finding was in contrast to the unsuccessful group who lost weight only until weeks 11-12. Achieved weight loss remained stable in the unsuccessful group until weeks 15-16, but was followed by weight regain by weeks 25-26. Multimedia Appendix 2 provides a detailed description of the course of weight loss in all groups.

Bivariate Correlations to Identify Indicators for Successful Weight Loss

Baseline Characteristics

Baseline body weight and BMI (rho=–0.027, *P*=.55; rho=–0.034, *P*=.46, respectively) and user's sex (rho=–0.053, *P*=.25) did not correlate with percentage weight reduction after 6 months in the unadjusted bivariate analysis; however, a positive correlation was observed for the user's age (rho=0.138, *P*=.002).

Early Weight Loss

We already observed a significant positive correlation (*P*<.001) with final weight loss in the first 2 weeks of program usage (Figure 4). The correlation became steadily stronger during the following 4 weeks, reaching an association comparable to the 3-month outcome by weeks 5-6 (Figure 4).

Figure 3. Percentage weight loss over 6 months for the unsuccessful (<5% weight loss), moderate success (5%-9.9% weight loss), and high success (≥10% weight loss) groups using KiloCoach. *= start of significant difference among the 3 subgroups (Kruskal-Wallis 1-way ANOVA, *P*<.001).

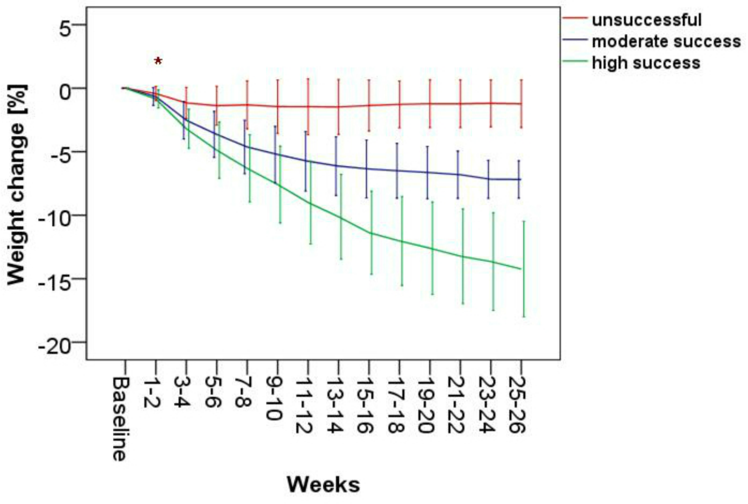
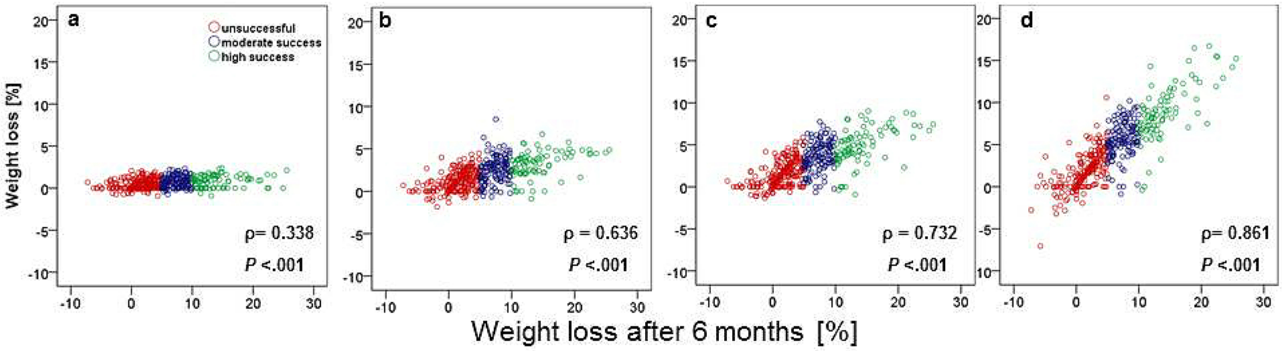


Figure 4. Correlation (Spearman rho, ρ) between early weight loss and weight loss outcome after 6 months during weeks 1-2 (a), weeks 3-4 (b), weeks 5-6 (c), and after 3 months (d) for the unsuccessful (<5% weight loss), moderate success (5%-9.9% weight loss), and high success (≥10% weight loss) groups using KiloCoach.



Program Usage

The weekly number of dietary protocols, weight entries, and logbook entries differed significantly among success groups over 6 months (all $P<.001$). These weekly numbers were significantly lower for the unsuccessful group compared to the moderate success group ($P<.001$; Figure 5), and the moderate success group had significantly less dietary protocols ($P=.005$), weight entries ($P<.001$), and logbook entries ($P=.001$) per week than the high success group. For numeric results, see Multimedia Appendix 2. After 6 months, the 3 program tools with the strongest correlations with percentage weight loss after 6 months were weekly numbers of dietary protocols (rho=0.589), weight entries (rho=0.631), and logbook entries (rho=0.599, all $P<.001$).

Reported energy intake was lowest in the unsuccessful group and highest in the high success group (unsuccessful: mean 1705 kcal/d, SD 607; moderate success: mean 1984 kcal/d, SD 675; high success: mean 2156 kcal/d, SD 740, $P<.001$ among groups) and correlated positively with weight loss after 6 months (rho=0.22, $P<.001$). In-line with this finding, the number of daily meals increased with increasing success (rho=0.247, $P<.001$). Both variables are interpreted as indicators for the accuracy of the dietary protocols rather than as objective measures of dietary intake.

The use of all program features decreased with increasing usage period in all success groups. For example, 70.1% (SD 21.2) of all dietary protocols present after 6 months of KiloCoach use were written within the first 3 months.

Multivariate Analysis

We performed a multivariate analysis to investigate the impact of early weight loss and program usage adjusted for possible confounding factors, such as sex, age, and baseline BMI. Percentage weight loss as a continuous variable after 6 months was chosen as the dependent variable. The total number of logbook entries showed strong collinearity to the total number of weight entries (variance inflation factor=16.1) and had to be removed from the model. The model is summarized in Table 3. The F test disclosed significant associations ($P<.001$) with the adjusted R^2, indicating that 52.4% of the variance in weight loss was explained by the model. Early weight loss by weeks 3-4, the total number of protocols, and the total number of weight entries qualified as independent predictors of 6-month weight reduction. The model summary was $R=0.728$, adjusted $R^2=0.524$, $F_{6,472}=88.6$, $P<.001$. Replacing weight loss by weeks 3-4 with weight loss during the first 2 weeks as dependent variable resulted in a considerable deterioration of the model fit by reducing the adjusted R^2 to 39.7%.

Sensitivity Analyses

Sensitivity analyses were carried out using BCF imputation and subgroup evaluation.

Baseline Carried Forward Imputation

Missing weight data were imputed using baseline weight. Because success groups were based on percentage weight loss after 6 months, group sizes changed using this model. The unsuccessful group increased to 344 users, whereas the moderate and high success groups decreased to 71 and 64 users, respectively. The average weight loss after 6 months was 3.3% (SD 5.4) in the total group (Multimedia Appendix 3).

Results of the primary analysis were confirmed for the weight loss dynamics in the GLM repeated measures (13 df contrast, $P<.001$) (Figure 6) and in the post hoc evaluation because weight loss among the 3 groups differed already significantly in weeks 1-2 ($P<.001$) and weight loss was higher in high compared to moderate success group from weeks 3-4 onward ($P<.001$).

Multivariate analysis confirmed the results of the primary analysis. Early weight loss by weeks 3-4, total number of protocols, and total number of weight entries qualified as independent predictors of 6-month weight reduction (all $P<.001$) in a model summary ($R=0.644$, adjusted $R^2=0.412$, $F_{6,472}=131.9$, $P<.001$).

Figure 5. Use of program tools in the success groups during the first 6 months of program usage. Weekly number of dietary protocols (a) and number of weight entries (b) differed significantly among the unsuccessful (<5% weight loss), moderate success (5%-9.9% weight loss), and high success (≥10% weight loss) groups.

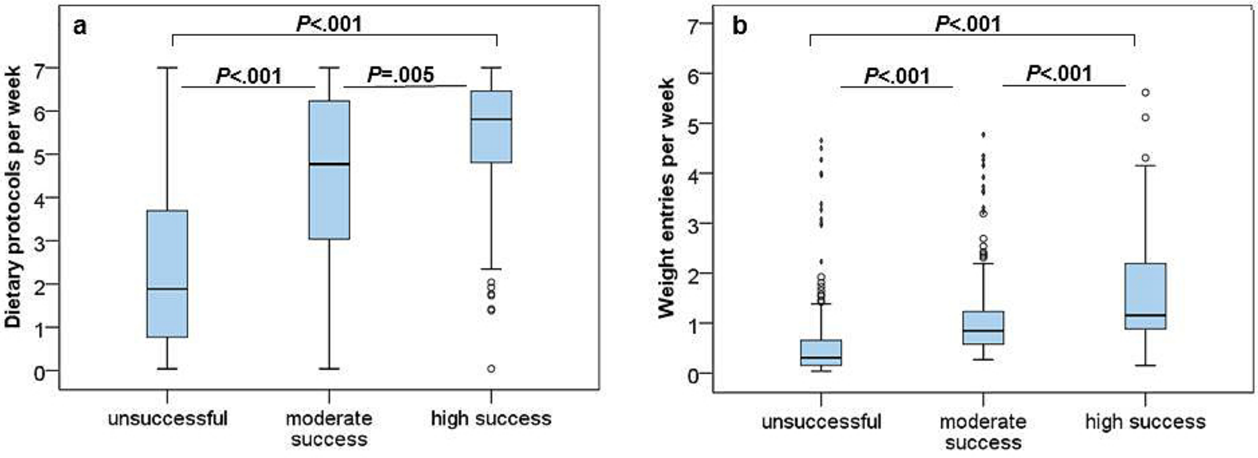
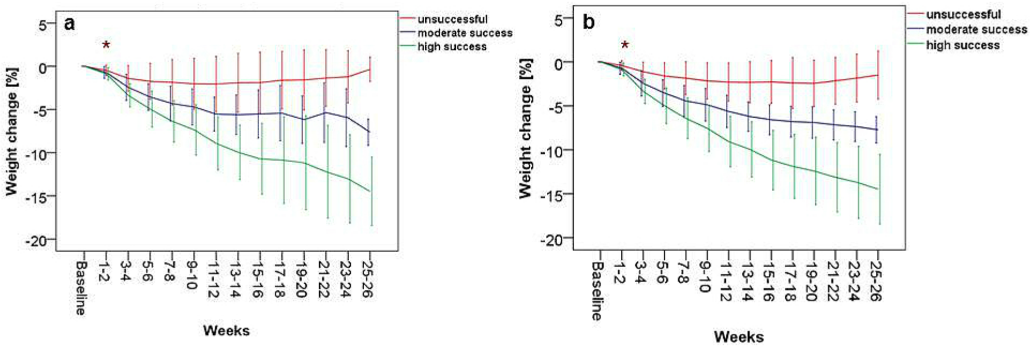


Table 3. Multiple regression analyses including regression coefficients (b) and standardized regression coefficient (ß-weight) for predicting percentage weight reduction after 6 months of program usage.

Multiple regression	b	*t* test	
		ß-weight	*P*
Weight loss week 3-4 (%)	1.579	0.457	<.001
Sex (female= 0, male =1)[a]	–0.015	–0.001	.97
Age (years)	0.008	0.018	.59
Baseline BMI (kg/m^2)	0.060	0.035	.27
Total protocols (n)	0.027	0.286	<.001
Total weight entries (n)	0.040	0.188	.001

[a] Dummy-coded term.

Figure 6. Sensitivity analysis. Development of percentage weight loss in the unsuccessful (<5% weight loss), moderate success (5%-9.9% weight loss), and high success (≥10% weight loss) groups when missing weight data in the study sample were imputed using the baseline carried forward (BCF) method (a) and when only users with active weight entry in weeks 25-26 were included and missing data were imputed using last observation carried forward (LOCF) method (b). * = start of significant difference among the 3 subgroups (Kruskal-Wallis 1-way ANOVA, P<.001).



Subgroup Evaluation of Users With Six-Month Weight Entry

Only users who entered a weight in weeks 25-26 were included in this analysis (n=214). This subgroup was chosen because of its superior data quality. No after last entry weight data had to be imputed in this sample, rendering it most concise for final weight loss and allocation to success groups. The sizes of the success groups changed: unsuccessful (n=79), moderate success (n=71), and high success (n=64). The subgroup analysis resulted in a 6-month weight loss of 7.5% (SD 6.0) (Multimedia Appendix 3).

Results of the primary analysis were confirmed for the weight loss dynamics in the GLM repeated measures model (13 df contrast, $P<.001$) (Figure 6) and in the post hoc analysis. Again, weight loss differed significantly between the 3 success groups after 2 weeks of program usage ($P<.001$).

Multivariate analysis confirmed the results of the primary analysis. Early weight loss by weeks 3-4 ($P<.001$), total number of protocols ($P<.001$), and total number of weight entries ($P=.002$) qualified as independent predictors of 6-month weight reduction. The model summary was $R=0.591$, adjusted $R^2=0.330$, $F_{6,211}=18.8$, $P<.001$.

Gender Differences

The weight dynamics were similar between men and women in the trajectory over time during the weight loss period of 6 months when adjusted for age, BMI, number of protocols, and number of weight entries (GLM repeated measures, 13 df contrast, $P=.91$).

In the post hoc bivariate analysis, age, baseline BMI, and purchased membership were comparable between men and women (Table 4). Percentage 6-month weight loss and assignment to success groups were not statistically different between the sexes (Table 4).

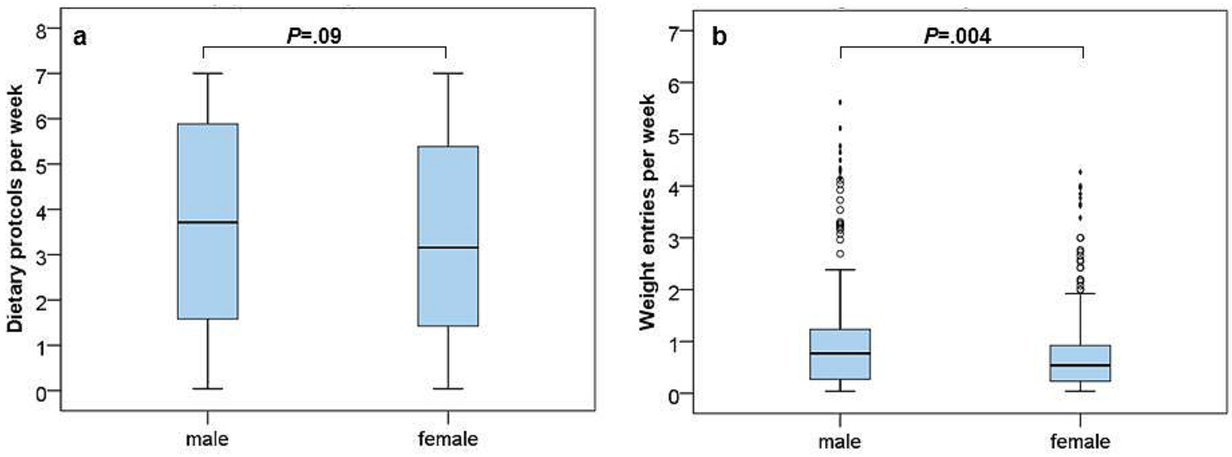
Over 6 months, males entered significantly more weight entries ($P=.004$) and logbook entries per week than females ($P=.004$). Additionally, males tended to write more dietary protocols than females ($P=.09$) although this did not reach statistical significance (Figure 7).

Table 4. Gender differences of the study sample.

Gender difference	Male	Female	P [a]
n	202	277	.001
Age (years), mean (SD)	44.4 (11.2)	45.1 (12.0)	.44
Initial weight (kg), mean (SD)	102.0 (13.3)	88.7 (12.3)	< .001
Baseline BMI (kg/m^2), mean (SD)	31.7 (3.3)	31.6 (3.2)	.99
Purchased membership days, mean (SD)	441 (280)	393 (250)	.13
Group distribution, n (%)			.18
Unsuccessful	111 (55.0)	150 (54.2)	
Moderate success	49 (24.3)	84 (30.3)	
High success	42 (20.8)	43 (15.5)	
6-month weight loss (%), mean (SD)	5.7 (5.7)	4.9 (5.4)	.25

[a] Mann-Whitney *U* test used except for proportions in success groups (chi-square test used).

Figure 7. Gender-specific use of program tools in the first 6 months of program participation. Dietary protocols (a) were used more frequently by men, who also entered significantly more body weights per week (b) than women.



Weight Maintenance

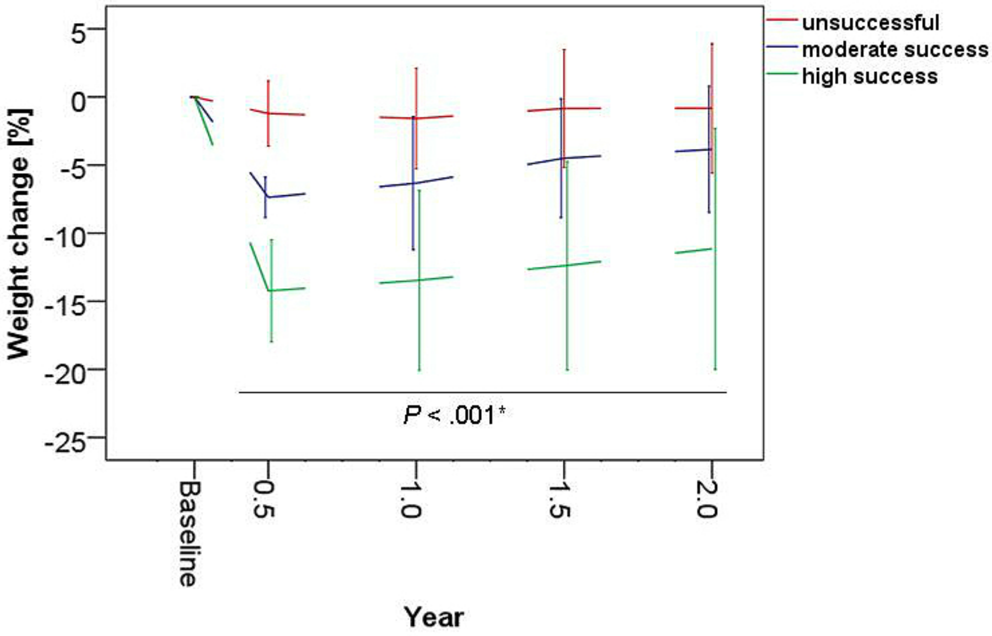
At the end of the follow-up period (ie, the time between weeks 25-26 and 2 years after the first dietary protocol), the unsuccessful group maintained a slight body weight reduction of 0.8% (SD 4.7) compared to baseline weight (Figure 8). Although some weight regain was observed in the moderate success group, the weight loss of 3.9% (SD 4.6) differed significantly from that of the unsuccessful group (*P*=.006) after 2 years. Although the high success group also regained some weight after 2 years compared to their weight at 6 months, this group maintained a clinically significant weight loss of 11.2% (SD 8.9) and differed significantly from the unsuccessful (*P*<.001) and moderate success (*P*=.001) groups. See Multimedia Appendix 2 for a detailed description of the weight maintenance period.

There were still significant positive correlations between early weight loss in weeks 1-2 and weight maintenance up to 2 years (Table 5). Also during follow-up, no significant gender-related maintenance pattern was observed.

Table 5. Spearman correlations (ρ) between early weight loss and weight maintenance in the study sample.

% Weight loss (long term)	% Weight loss (early)					
	Weeks 1-2		Weeks 3-4		After 3 months	
	ρ	*P*	ρ	*P*	ρ	*P*
After 1 year	0.249	<.001	0.402	<.001	0.604	<.001
After 1.5 years	0.237	.007	0.396	<.001	0.547	<.001
After 2 years	0.278	.008	0.447	<.001	0.553	<.001

Figure 8. Development of percentage weight loss in the success groups compared to baseline. Weight maintenance during the follow-up period (1, 1.5, and 2 years after the first dietary protocol) of the unsuccessful (<5% weight loss), moderate success (5%-9.9% weight loss), and high success (≥10% weight loss) groups of the study sample. * = significant difference among the 3 subgroups (Kruskal-Wallis 1-way ANOVA).



Discussion

Principal Findings

The KiloCoach online program aims at supporting self-guided body weight reduction and provides information, tools for self-monitoring and analysis of eating habits, and social support. Our aim was to identify predictors of weight loss effectiveness in long-term users of KiloCoach in a real world setting. *Long term* was defined as program adherence of at least 6 months. This criterion was fulfilled by 44% of KiloCoach users who represent a relevant proportion of the total KiloCoach population.

We showed that, despite comparable baseline characteristics, weight loss dynamics and weight maintenance differed significantly among the 3 success groups of our study sample. Multivariate analysis showed that early weight loss (weeks 3-4), number of dietary protocols, and number of weight entries to be independent predictors for final weight loss after 6 months in the primary analysis confirmed by 2 sensitivity analyses. Moderate and high successors used program tools more frequently than unsuccessful users. The total population and study sample consisted of an unexpectedly high percentage of male users, who demonstrated weight reduction comparable to females and even higher intensity of program usage.

Early Weight Loss

So far, 3 trials have reported that early weight loss is positively related to final weight loss [27-29]. All trials evaluated in-person weight loss programs. In the DiOGenes study, early weight loss after 1 and 3 weeks of dieting predicted weight loss after 8 weeks on a low-calorie diet (800 kcal/d) [29]. Fabricatore et al [27] found that early weight loss after 3 weeks of treatment predicted successful weight loss after 1 year. Elfhag et al [28] reported that weight loss after 5 weeks of treatment best forecasted weight loss after 8 to 10 months of group sessions.

To our knowledge, this is the first time that the effects of early weight loss on later weight loss and weight maintenance have been evaluated for a commercial online weight loss program. We found weight loss in the third and fourth week of platform use was highly predictive for the 6-month outcome and was also significantly associated with weight maintenance up to 2 years. Interestingly and against intuition, early weight loss does not seem to be associated with pretreatment motivation [28], a psychological factor that could not be evaluated in the present study. However, self-motivation quickly fluctuates in relation to lapses and relapses [30]. Users with significant early weight loss were likely to have reinforced their motivation during program use and thereby increased self-efficacy to lose weight. Early weight loss might be a modifiable factor that could be influenced by educational advice supplied by the platform provider.

Program Usage

By means of weight tracking, dietary protocols, and analyzing tools, KiloCoach users are encouraged to self-monitor and change their lifestyle. Previous studies showed that self-monitoring behavior of weight, diet, and activity are cardinal behaviors of successful weight controllers [30-32]. Our results also depict that usage intensity of KiloCoach was associated with higher weight loss. Especially the features of self-monitoring weight and diet were predictive for 6-month weight loss, even when adjusted for early weight loss. These results are in-line with the finding of Krukowski et al [33] that overall online self-monitoring is associated with weight loss outcome after 6 months. They further described that the achievement of weight loss greater than 5% was more likely in users who consistently self-monitored in the early usage phase (ie, within the first 4 weeks).

Food and exercise diaries were previously identified as the most effective program features for weight loss in retrospective studies [12,14] and in 1 prospective randomized trial [34]. Adherence to the old method of paper dietary protocols is a well-known problem; it is a common experience that documenting more than 7 days in succession leads to inadequate results because of decreased interest and boredom of the individuals [35]. Thus, we were surprised to find a high mean number of online dietary protocols over 6 months ranging from mean 2.4 (SD 2.0) per week in the unsuccessful group to mean 5.3 (SD 1.7) per week in the high success group. The mean total number of dietary protocols was 63 and 137, respectively, meaning that users generated dietary protocols on 35% to 75% of their participation days. Computer-based technology seems to facilitate self-monitoring of diet, and further technical advancements of Web-based programs are expected to continuously improve self-monitoring adherence [36,37].

Both energy intake and number of meals were significantly higher in the high success group than in the unsuccessful group, which appears idiosyncratic with regard to successful weight loss. These 2 factors are most likely indicators for the average quality of dietary protocols, which suggests that dietary protocols were filled in more thoroughly by the high success group.

Male Users

It was surprising to observe a percentage of 39% male users in the total KiloCoach population and of 42% in the study sample. Comparing this finding to other non-Web-based weight loss interventions, this proportion is very high. For example, only 3.6% of all users of a special offer to access Weight Watchers were men [38]. In a workplace-based weight loss program, only 6.6% of the enrolled participants [39] and 7.4% of the completers [40] were men. Considering the lower barriers for participation in online weight loss programs, such as anonymity, flexible integration into everyday life, and the self-guided procedures, one could assume that men are particularly attracted by such an intervention. Still, the proportion of male users in KiloCoach is high, even compared with other Web-based weight loss programs. Other programs reported male participation ranging from 14% [12,41] to 26% [14,15].

Currently, we can only speculate about the reasons why KiloCoach is more attractive to men. According to the program owner, the website and the user interface of the program are designed to appeal to men and women. However, men and women seem to prefer different program features. For example, men seem to prefer calculative tools, especially for weight forecasts, whereas women seem to be more interested in nutritional information, including the calculation of one's own recipes, and peer support (personal communication).

Effectiveness

We reported an average weight loss of 5.3% (SD 5.6) of baseline weight after 6 months of KiloCoach usage in our primary analysis. This finding confirms previous results using the same platform (weight loss of 4.4%, SD 5.1 and 6.4%, SD 7.3 in users who followed the program for at least 60 days or still entered dietary protocols after 1 year) [18]. The program shows a satisfying overall result compared to structured, in-person weight loss programs that usually average between 5% and 10% weight loss over 6 months [42,43]. The proportion of our long-term users who lost weight successfully (ie, ≥5% of baseline weight) was 46%. For comparison, Krukowski et al [33] showed that 53% of participants in their online study arm condition achieved a weight loss of 5% or more after 6 months. In another study, a smaller proportion (36%) of study completers achieved this weight loss [44].

A recent Cochrane review showed that structured in-person treatment resulted in a mean weight loss that was only 2.1 kg higher than that achieved by Web-based intervention [15]. When compared with minimal personal interventions (eg, information, standard care), Web-based programs lead to an even higher weight loss of 1.5 kg after 6 months [15]. Thus, Web-based weight loss intervention programs are a feasible and efficient compromise between high and minimal resource interventional programs.

With regard to weight maintenance, the high success group maintained a clinically significant weight loss after 2 years. This finding also points toward a possible advantage over other successful weight loss programs because individuals often fail to maintain their weight loss over longer periods of time [24].

Strengths and Limitations

A strength of our work is that, for the first time, a weight loss platform for the German-speaking countries Germany, Austria, and Switzerland was investigated concerning weight loss dynamics and usage of program tools. Another strength is that all long-term users were evaluated. Thus, the results are representative for this group. Furthermore, we showed with this study that detailed long-term electronic documentation of nutrition and exercise seems to be possible.

The retrospective design might be seen as a disadvantage, but it presented an immediate opportunity to investigate existing data of the KiloCoach database, representing field data from the real world in a population in which scientific evaluation was not prospectively intended. This approach can be advantageous because nutritional behavior is responsive to observation, which might bias the results.

Critics might further note that all information about the users is self-reported. Body weight might be especially underreported, but Harvey-Berino et al [45] found that self-reported weight is highly accurate and weight loss conveyed by self-reporting is comparable to de facto weight loss. Users included in this study had to buy a membership for program usage and had to lose weight on their own; they would not have benefited from program usage if the reported data did not agree with reality.

One limitation of our study is the absence of information about motivation at the beginning of program use or demographic characteristics other than age and sex (eg, ethnicity, marital status, education) that might have influenced weight loss. Problems in identifying predictors have been summarized recently [30].

Meanwhile, a controlled prospective trial has started to further increase the knowledge about efficacy and weight dynamics in KiloCoach users.

Conclusion

KiloCoach is an effective self-help tool to reduce weight that attracts more men than other programs and enables long-term monitoring of dietary intake and physical activity. To achieve the best possible weight loss result, users should closely adhere to the program for frequent and thorough usage of self-monitoring tools. To be more specific, writing dietary protocols 5 days per week and entering body weight on 1 day per week is recommended according to our results. Close program adherence seems to be especially important during the first period of program usage, with the aim to induce early weight loss and improve chances for further success.

Acknowledgments

EP was granted a doctoral scholarship by the Charité - Universitätsmedizin Berlin. The project is supported by the German Diabetes Trust (327/09/12). Monika Schoell is greatly acknowledged for her linguistic support.

Authors' Contributions

Conception and study design: LV, EP, MK; provision of raw data: RA, RL; generation and collection of data: EP; data analysis and interpretation: MK, EP, LV, UE; manuscript drafting and finalization: EP, LV, MK, UE, JDS; manuscript review and revision: EP, LV, MK, UE, JDS, RA, RL. All authors read and approved the final version of this manuscript.

Conflicts of Interest

RA is the owner of the commercial program KiloCoach. She provided the study team with user data and declares that the submitted data have been unselected and unmodified. KiloCoach will not have direct financial benefit from publication. RL is an employee of KiloCoach. KiloCoach did not sponsor the project or members of the project team. The other authors do not have any financial relations to KiloCoach and declare no conflicts of interest.

Multimedia Appendix 1

KiloCoach dietary protocol and analyzing tools.

[PPT File (Microsoft PowerPoint Presentation), 2MB - jmir_v15i10e219_app1.ppt]

Multimedia Appendix 2

Detailed description of weight loss and weight maintenance period and of program usage intensity.

[PDF File (Adobe PDF File), 62KB - jmir_v15i10e219_app2.pdf]

Multimedia Appendix 3

Sensitivity analyses - weight change results.

[PDF File (Adobe PDF File), 55KB - jmir_v15i10e219_app3.pdf]

References

1. World Health Organization. Overweight: Situation and trends URL: http://www.who.int/gho/ncd/risk_factors/overweight_text/en/index.html [accessed 2013-02-28] [WebCite Cache ID 6ElwZjXNS]
2. World Health Organization. Risk factors: Adult risk factors data by WHO region URL: http://apps.who.int/gho/data/view.main.1740 [accessed 2013-02-28] [WebCite Cache ID 6EluRxJCI]
3. Eurostat. 2013 Feb 25. Individuals using the Internet, by place of use URL: http://epp.eurostat.ec.europa.eu/tgm/web/_download/Eurostat_Table_tin00075PDFDesc_22f6b5d8-5c9e-405e-8391-f24473d9e54d.pdf [accessed 2013-03-01] [WebCite Cache ID 6Emxgkxth]
4. Eurostat. 2013 Feb 25. Individuals using the Internet for seeking health-related information URL: http://epp.eurostat.ec.europa.eu/tgm/web/_download/Eurostat_Table_tin00101PDFDesc_7996fb27-6b4d-42b8-9511-68e91bae49df.pdf [accessed 2013-03-01] [WebCite Cache ID 6Emyf6feg]
5. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
6. Ciao AC, Latner JD, Durso LE. Treatment seeking and barriers to weight loss treatments of different intensity levels among obese and overweight individuals. Eat Weight Disord 2012 Mar;17(1):e9-16. [Medline: 22751277]

7. Saperstein SL, Atkinson NL, Gold RS. The impact of Internet use for weight loss. Obes Rev 2007 Sep;8(5):459-465. [doi: 10.1111/j.1467-789X.2007.00374.x] [Medline: 17716303]
8. 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]
9. Krukowski RA, West DS, Harvey-Berino J. Recent advances in internet-delivered, evidence-based weight control programs for adults. J Diabetes Sci Technol 2009 Jan;3(1):184-189 [FREE Full text] [Medline: 20046664]
10. 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]
11. 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]
12. Neve M, Morgan PJ, Collins CE. Weight change in a commercial web-based weight loss program and its association with website use: cohort study. J Med Internet Res 2011;13(4):e83 [FREE Full text] [doi: 10.2196/jmir.1756] [Medline: 21993231]
13. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
14. Johnson F, Wardle J. The association between weight loss and engagement with a web-based food and exercise diary in a commercial weight loss programme: a retrospective analysis. Int J Behav Nutr Phys Act 2011;8:83 [FREE Full text] [doi: 10.1186/1479-5868-8-83] [Medline: 21810222]
15. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675. [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
16. 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]
17. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: the relationship between web features and weight loss. Telemed J E Health 2008 Oct;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]
18. Longin R, Grasse M, Aspalter R, Waldherr K. Effectiveness of the online weight reduction program KiloCoach™ and comparison with other evaluated commercial direct intervention and online programs. Obes Facts 2012;5(3):372-383. [doi: 10.1159/000339726] [Medline: 22722385]
19. KiloCoach. URL: http://kilocoach.com/de-de/ [accessed 2013-09-24] [WebCite Cache ID 6JsE7wVuw]
20. National Institute for Health and Clinical Excellence, National Collaborating Centre for Primary Care. Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children. 2006 Dec. URL: http://www.nice.org.uk/nicemedia/live/11000/38294/38294.pdf [accessed 2013-03-27] [WebCite Cache ID 6FQVXRwbN]
21. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272. [doi: 10.1038/oby.2006.34] [Medline: 16571852]
22. Hauner HBG, Hamann A, Husemann B, Koletzko B, Liebermeister H, Wabitsch M, et al. Prävention und Therapie der Adipositas. 2007. URL: http://www.adipositas-gesellschaft.de/fileadmin/PDF/Leitlinien/Adipositas-Leitlinie-2007.pdf [accessed 2013-03-26] [WebCite Cache ID 6FPR51mWo]
23. Dämon S, Schätzer M, Höfler J, Tomasec G, Hoppichler F. Nutrition and diabetes mellitus: an overview of the current evidence. Wien Med Wochenschr 2011 Jun;161(11-12):282-288. [doi: 10.1007/s10354-011-0888-4] [Medline: 21442215]
24. Wing RR, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Jul;82(1 Suppl):222S-225S [FREE Full text] [Medline: 16002825]
25. Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials. 2011. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf [accessed 2013-08-12] [WebCite Cache ID 6IoZJo4fC]
26. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998 Apr 18;316(7139):1236-1238 [FREE Full text] [Medline: 9553006]
27. Fabricatore AN, Wadden TA, Moore RH, Butryn ML, Heymsfield SB, Nguyen AM. Predictors of attrition and weight loss success: Results from a randomized controlled trial. Behav Res Ther 2009 Aug;47(8):685-691 [FREE Full text] [doi: 10.1016/j.brat.2009.05.004] [Medline: 19497559]
28. Elfhag K, Rössner S. Initial weight loss is the best predictor for success in obesity treatment and sociodemographic liabilities increase risk for drop-out. Patient Educ Couns 2010 Jun;79(3):361-366. [doi: 10.1016/j.pec.2010.02.006] [Medline: 20223613]

29. Handjieva-Darlenska T, Handjiev S, Larsen TM, van Baak MA, Jebb S, Papadaki A, et al. Initial weight loss on an 800-kcal diet as a predictor of weight loss success after 8 weeks: the Diogenes study. Eur J Clin Nutr 2010 Sep;64(9):994-999. [doi: 10.1038/ejcn.2010.110] [Medline: 20588292]
30. Stubbs J, Whybrow S, Teixeira P, Blundell J, Lawton C, Westenhoefer J, et al. Problems in identifying predictors and correlates of weight loss and maintenance: implications for weight control therapies based on behaviour change. Obes Rev 2011 Sep;12(9):688-708. [doi: 10.1111/j.1467-789X.2011.00883.x] [Medline: 21535362]
31. 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]
32. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
33. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170. [doi: 10.1037/a0028135] [Medline: 22545978]
34. Hunter CM, Peterson AL, Alvarez LM, Poston WC, Brundige AR, Haddock CK, et al. Weight management using the internet a randomized controlled trial. Am J Prev Med 2008 Feb;34(2):119-126. [doi: 10.1016/j.amepre.2007.09.026] [Medline: 18201641]
35. Thompson FE, Byers T. Dietary assessment resource manual. J Nutr 1994 Nov;124(11 Suppl):2245S-2317S [FREE Full text] [Medline: 7965210]
36. 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]
37. Lieffers JR, Hanning RM. Dietary assessment and self-monitoring with nutrition applications for mobile devices. Can J Diet Pract Res 2012;73(3):e253-e260. [Medline: 22958633]
38. Mitchell NS, Ellison MC, Hill JO, Tsai AG. Evaluation of the effectiveness of making Weight Watchers available to Tennessee Medicaid (TennCare) recipients. J Gen Intern Med 2013 Jan;28(1):12-17. [doi: 10.1007/s11606-012-2083-8] [Medline: 22618582]
39. Touger-Decker R, Denmark R, Bruno M, O'Sullivan-Maillet J, Lasser N. Workplace weight loss program; comparing live and internet methods. J Occup Environ Med 2010 Nov;52(11):1112-1118. [doi: 10.1097/JOM.0b013e3181f9ee8c] [Medline: 21063189]
40. Bruno M, Touger-Decker R, Byham-Gray L, Denmark R. Workplace weight loss program: impact on quality of life. J Occup Environ Med 2011 Dec;53(12):1396-1403. [doi: 10.1097/JOM.0b013e3182338644] [Medline: 22036946]
41. Jonasson J, Linné Y, Neovius M, Rössner S. An Internet-based weight loss programme -- a feasibility study with preliminary results from 4209 completers. Scand J Public Health 2009 Jan;37(1):75-82. [doi: 10.1177/1403494807085066] [Medline: 18684784]
42. Foster GD, Makris AP, Bailer BA. Behavioral treatment of obesity. Am J Clin Nutr 2005 Jul;82(1 Suppl):230S-235S [FREE Full text] [Medline: 16002827]
43. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
44. Collinson A, Lindley R, Campbell A, Waters I, Lindley T, Wallace A. An evaluation of an Internet-based approach to weight loss with low glycaemic load principles. J Hum Nutr Diet 2011 Apr;24(2):192-195. [doi: 10.1111/j.1365-277X.2010.01138.x] [Medline: 21843154]
45. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a web-based obesity treatment program. Telemed J E Health 2011 Nov;17(9):696-699 [FREE Full text] [doi: 10.1089/tmj.2011.0032] [Medline: 21882997]

Abbreviations

BCF: baseline carried forward
BMI: body mass index
GLM: general linear model
LOCF: last observation carried forward

Edited by G Eysenbach; submitted 03.04.13; peer-reviewed by R Krukowski, M Neve, J Gao; comments to author 15.06.13; revised version received 12.07.13; accepted 29.08.13; published 14.10.13.

Please cite as:
Postrach E, Aspalter R, Elbelt U, Koller M, Longin R, Schulzke JD, Valentini L
Determinants of Successful Weight Loss After Using a Commercial Web-Based Weight Reduction Program for Six Months: Cohort Study
J Med Internet Res 2013;15(10):e219
URL: http://www.jmir.org/2013/10/e219/
doi:10.2196/jmir.2648
PMID:24126250

Original Paper

Adherence to a Web-Based Physical Activity Intervention for Patients With Knee and/or Hip Osteoarthritis: A Mixed Method Study

Daniël Bossen[1], MSc; Michelle Buskermolen[1], BSc; Cindy Veenhof[1], PhD; Dinny de Bakker[1,2], PhD; Joost Dekker[3], PhD

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

[2]Tilburg University, Tranzo, Tilburg, Netherlands

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

Corresponding Author:
Daniël Bossen, MSc
Netherlands Institute for Health Services Research (NIVEL)
Otterstraat 118-124
Utrecht, 1568 BN
Netherlands
Phone: 31 302729649
Fax: 31 302729729
Email: d.bossen@nivel.nl

Abstract

Background: Web-based interventions show promise in promoting a healthy lifestyle, but their effectiveness is hampered by high rates of nonusage. Predictors and reasons for (non)usage are not well known. Identifying which factors are related to usage contributes to the recognition of subgroups who benefit most from Web-based interventions and to the development of new strategies to increase usage.

Objective: The aim of this mixed methods study was to explore patient, intervention, and study characteristics that facilitate or impede usage of a Web-based physical activity intervention for patients with knee and/or hip osteoarthritis.

Methods: This study is part of a randomized controlled trial that investigated the effects of Web-based physical activity intervention. A total of 199 participants between 50-75 years of age with knee and/or hip osteoarthritis were randomly assigned to a Web-based intervention (n=100) or a waiting list (n=99). This mixed methods study used only data from the individuals allocated to the intervention group. Patients were defined as users if they completed at least 6 out of 9 modules. Logistic regression analyses with a stepwise backward selection procedure were executed to build a multivariate prediction usage model. For the qualitative part, semistructured interviews were conducted. Both inductive and deductive analyses were used to identify patterns in reported reasons for nonusage.

Results: Of the 100 participants who received a password and username, 46 completed 6 modules or more. Multivariate regression analyses revealed that higher age (OR 0.94, *P*=.08) and the presence of a comorbidity (OR 0.33, *P*=.02) predicted nonusage. The sensitivity analysis indicated that the model was robust to changes in the usage parameter. Results from the interviews showed that a lack of personal guidance, insufficient motivation, presence of physical problems, and low mood were reasons for nonusage. In addition, the absence of human involvement was viewed as a disadvantage and it negatively impacted program usage. Factors that influenced usage positively were trust in the program, its reliability, functionality of the intervention, social support from family or friends, and commitment to the research team.

Conclusions: In this mixed methods study, we found patient, intervention, and study factors that were important in the usage and nonusage of a Web-based PA intervention for patients with knee and/or hip osteoarthritis. Although the self-guided components offer several advantages, particularly in relation to costs, reach, and access, we found that older patients and participants with a comorbid condition need a more personal approach. For these groups the integration of Web-based interventions in a health care environment seems to be promising.

Trial Registration: The Netherlands National Trial Register (NTR): 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(10):e223) doi:10.2196/jmir.2742

KEYWORDS

adherence; usage; Web-based intervention; mixed method study

Introduction

Osteoarthritis (OA) in the knee or hip is a prevalent musculoskeletal disorder characterized by joint pain, joint stiffness, and functional disability [1]. Regular physical activity (PA) has been recognized as an effective lifestyle strategy in the nonpharmacological management of knee and hip OA [2,3]. Despite recommendations, people with knee or hip OA are less physically active than the general population [4,5].

In an attempt to enhance a physically active lifestyle in patients with knee and/or hip OA, we developed a Web-based PA intervention. The intervention, entitled *Join2move*, is a self-paced 9-week PA program in which the patient’s favorite recreational activity is gradually increased during fixed time periods. In a recent randomized controlled trial (RCT) among 199 participants with knee and/or hip OA [6], *Join2move* was demonstrated to be effective compared to a waiting list control group. Besides enhanced levels of PA, this study showed significant improvements in physical functioning, self-efficacy, pain levels, tiredness, and anxiety in the intervention group.

Unfortunately, substantial rates of nonusage were observed. A considerable proportion of potential users was never exposed to important program content. This is consistent with other studies [7-16]. For example, two studies [15,16] testing a Web-based PA intervention reported that 60% of their diabetes patients accessed the website once a week. The issue of nonusage is described in Eysenbach’s Law of Attrition [17]. According to Eysenbach, characteristics related to the participant, intervention, and study may play a pivotal role in the adoption or rejection of Web-based interventions. Studies have demonstrated that older age groups [10,18-22], people with a healthy lifestyle [10,20], those with social ties [23], higher educated patients [22], and women [22,24] are more likely to adhere to Web-based interventions. In addition to user characteristics, the characteristics of the intervention itself can also influence usage. For instance, self-guided interventions with minimal human “push factors” (eg, online counseling or emails) show higher rates of nonusage than programs with substantial human involvement [17,25,26]. Other intervention characteristics that predict usage are program duration and complexity. Generally, shorter, more concise interventions achieve better usage rates compared with more extensive interventions [27]. Moreover, it is known that study-related factors (eg, attention, commitment, and a belief in the importance of research), especially in RCTs [26], are positively related to usage [18,28].

Although considerable research has been devoted to quantitative predictors of nonusage, little qualitative research has been conducted on the underlying reasons for nonusage. Therefore, we conducted a mixed methods study to gain a deeper understanding of actual usage patterns, possible attrition predictors, and reasons for (non)usage. This is a necessary step toward enhancing program usage and may help us to make the *Join2move* intervention even more effective.

In this study, we utilized a mixed methods design employing both quantitative and qualitative (interviews) methods. By integrating the quantitative and qualitative results, we aimed to identify patient-, intervention-, and study-related characteristics that may facilitate or impede the usage of Web-based intervention for patients with knee and/or hip OA. Since this study was explorative by nature, no a priori hypotheses were formulated.

Methods

Study Design and Participants

Data from this study were retrieved from a randomized controlled trial that aimed at evaluating the effectiveness of the *Join2move* intervention for patients with hip and/or knee OA [6]. In brief, the design of the study was a randomized, nonblinded, controlled, two-arm trial. Ethical approval was obtained from the medical ethics committee of the VU University Medical Center Amsterdam. Enrollment started on January 3, 2011 and ended on November 5, 2011. Sedentary volunteers with knee and/or hip OA were recruited via articles in newspapers and health-related websites. The eligibility criteria for participants were (1) aged 50-75, (2) self-reported OA in knee and/or hip, (3) self-reported inactivity (<30 minutes of moderate PA less than 5 days in a week), (4) no face-to-face consults for OA with a health care provider, other than general practitioner, in the last 6 months, (5) ability to access the Internet weekly, and (6) no contra-indications to exercise without supervision. In total, 199 eligible participants were randomly assigned either to the intervention (n=100) or waiting list control group (n=99). Baseline, 3-month, and 12-month follow-up data were collected via online questionnaires. Primary outcomes were PA, physical functioning, and self-perceived effect. Self-perceived effect was assessed by asking participants about the degree of change since their previous assessment (much worse to much better*)*. Both short-term and long-term results revealed positive effects of *Join2move* with respect to PA, physical functioning, self-perceived effect, and several other secondary outcomes [6].

Intervention

Over the course of 1 year, experts from the Netherlands Institute for Health Services Research developed the *Join2move* intervention. The *Join2move* intervention is based on a previously developed and evaluated behavioral graded activity (BGA) program for patients with knee and/or hip OA [29]. Details of the *Join2move* intervention and the development process are described in another publication [30]. In brief, the *Join2move* intervention is a fully automated Web-based intervention that contains automatic functions (automatic messages on the website and automatic emails) without human support. Screenshots illustrating different stages of the

Join2move program are presented in Multimedia Appendix 1. Participants are presented with the homepage (see Figure 1). *Join2move* is a self-paced 9-week PA program in which the patient's favorite recreational activity is gradually increased in a time-contingent manner (ie, on fixed time points). In the first week, users select a central activity such as cycling or walking and perform a 3-day self-test. Based on the performance from the self-test, a range of goals is automatically generated and presented on the website. In this way, achievable goals are set. Users have the option to choose one of the proposed short-term goals between a lower and upper limit. Depending on the selected goal, 8 tailored modules are generated and presented weekly on the 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 on a 10-point Numerical Rating Scale (0 is no pain, 10 is worst possible pain). Performance is measured by three items: (1) "I completed the module as instructed", (2) "I did more than the instructed module", and (3) "I did less than the instructed module" (due to time constraints, weather conditions, pain in my knee and/or hip, or other physical complaints). Subsequently, tailored to the answers from the evaluation form, automated text-based messages are generated. 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. When users indicated that a module was missed due to pain in knee and hip or other physical complaints, they had the option to repeat the module (a maximum of three times), adapt the intensity of the module, or proceed to the next module. Since personal messages are updated on a weekly basis, users are encouraged to log in once a week. Automated emails are generated if participants do not log on the website for 2 weeks. At the end of the program, the website presents a motivational message to perform regular PA in the future. In total, the program lasted 9 weeks.

Figure 1. Homepage Join2move.

Bewegen helpt!

Gedoseerd bewegen houdt bij mensen met artrose de conditie op peil, spieren sterk, gewrichten soepel en geest gezond. Hierdoor worden dagelijkse activiteiten, zoals wandelen, tuinieren en fietsen gemakkelijker uitvoerbaar.

Bent u tussen de 50 en 70 jaar, heeft u last van artrose in de heup en/of knie en wilt u stapsgewijs meer gaan bewegen? Schrijf u dan in voor het beweegprogramma.

Met vriendelijke groet,
Daniël Bossen

Meneer Verveer (61) - *"Artrose In Beweging was voor mij de stok achter de deur om meer te bewegen. Ik vind het heerlijk om samen met mijn vrouw te wandelen en te genieten van de natuur"*

Nieuw onderzoek februari 2012
In februari start het NIVEL met een nieuw onderzoek naar de effectiviteit van het internet beweegprogramma. U kunt zich nu inschrijven voor deelname! Schrijf u hier in!

Data Collection and Outcomes of the Quantitative Study

Overview

Program usage (ie, the number of completed program modules) was monitored throughout the intervention period. A module consisted of a text-based assignment plus accompanying evaluation form, which was presented on the website for 7 consecutive days. Once a participant had filled out the evaluation form 7 days after receiving the weekly assignment, the module was defined as completed and the user was automatically presented with a new weekly assignment. If a scheduled weekly module was missed, participants had the option to repeat the module, adapt the difficulty, or continue with the next module. In total, 9 weekly modules were available to the participant. This was automatically registered. After some consideration, the research team had decided that completion of at least 6 modules was required to improve PA and other primary effects. Patients were defined as users if they completed at least 6 out of 9 modules. Participants who did not reach this threshold were defined as nonusers. Predictors of usage were collected through online baseline questionnaires and can be categorized as demographic, clinical, or psychological predictors. The potential predictors were not selected on theoretical grounds.

Demographic Predictors

Demographic predictors were gender, education (low: primary and lower vocational education; middle: secondary and middle vocational education; high: higher vocational and university education), and age (years) as demographic predictors.

Clinical Predictors

Clinical predictors in this study were location of OA (knee, hip or both), duration of OA complaints (years), and body mass index (BMI) (weight in kilograms divided by height in meters squared). Pain and fatigue were assessed on a 10-point Numerical Rating Scale (0 is no pain/not tired, 10 is worst possible pain/extremely tired). Self-reported PA was measured by the validated PA Scale for the Elderly (PASE) [31]. 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. Performance of the activities (assigned according to the level of intensity: light, moderate, and strenuous) is 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. Physical functioning was determined by a subscale of the Knee OA Outcome Score (KOOS) [32,33] and the Hip Injury OA Outcome Score (HOOS) [34,35]. The KOOS and HOOS are self-administered questionnaires designed to assess patients' opinions about their knee- and/or hip-related problems. The questionnaires assess 5 indicators on a 5-point Likert scale: pain, symptoms, physical functioning, sport/recreation functioning, and quality of life. The presence of self-reported comorbidity was obtained through a specific list of comorbid diseases. The list described the most prevalent chronic diseases and disorders in The Netherlands [36].

Psychological Predictors

Anxiety and depression were evaluated by a 14-item Hospital Anxiety and Depression scale [37]. Seven items on this questionnaire are related to anxiety and seven are related to depression. A lower score represents less anxiety and depression. Self-efficacy was evaluated by the Arthritis Self-Efficacy Scale for pain and other symptoms [38,39]. We used the subscales self-efficacy for pain and self-efficacy for other symptoms (eg, fatigue, depression). The score ranges from 1-10, where a higher score indicates greater self-efficacy.

Active and passive pain coping were determined by the Pain Coping Inventory questionnaire [40]. This 33-item questionnaire determines active and passive pain-coping strategies. A higher score on the active pain-coping subscale indicate a more adequate pain coping, and a higher score on the passive pain-coping subscale indicates inadequate pain coping. Locus of control, the extent to which one believes that one's health is determined by one's behavior, was examined with the Multidimensional Health Locus of Control Scale (MHLC) [41]. We used two subscales of the MHLC: (1) belief of control by powerful others (6 items) and (2) internal locus of control (6 items). For each subscale, a higher score indicates a greater level of belief in a particular subscale.

Data Collection and Outcomes of the Qualitative Study

One year after being assigned to the program, a subgroup of participants from the intervention group was interviewed. All participants from the intervention group (n=100) were categorized into two groups: (1) users and (2) nonusers. Since the nonuser group showed considerable divergence in extent of program use (0 to 5 modules), we decided to invite more nonusers than users for our interview sample. This was executed by a stratified purposive sampling procedure [42]. After the stratified sampling, participants were contacted by phone, invited to participate, and scheduled for a face-to-face interview until the sampling goal was reached. The goal was to conduct 15 interviews (10 users and 5 nonusers). To reach this sampling goal, 24 participants were invited; 15 agreed to be interviewed and 9 decided not to participate due to a lack of interest. All participants who declined to be interviewed were nonusers. Semistructured interviews were conducted by the same interviewer (MB) in the respondents' homes and lasted approximately 60 minutes. Interviews were digitally audiorecorded with the participants' permission. The interviews were transcribed by means of the program Express Scribe [43]. During the interview process, we used an open-question guide (see Multimedia Appendix 2). This interview guide contained three topics: (1) patient characteristics, (2) intervention characteristics, and (3) study characteristics. The intervention characteristics contained three of the five themes described by Eysenbach's law of attrition [17]: (1) *Relative advantage*, the degree to which the innovation is perceived to be superior to the ideas that it replaces [44], (2) *Complexity*, the degree to which an innovation is perceived as relatively difficult to understand and use [44], and (3) *Compatibility*, the degree to which an innovation is perceived as being consistent with the values, experiences, and needs of potential adopters [44].

Analyses

Quantitative Analyses

Descriptive analyses were performed to describe participant characteristics and program usage. Logistic regression analysis with a stepwise backward selection procedure was used to build the most parsimonious prediction model. Program use (user/nonuser) was employed as a dichotomous dependent variable. Demographic, clinical, and psychological variables were the independent variables. Statistical analyses were conducted in two phases. First, potential predictors of interest were screened by univariate logistic regressions. Second, variables that achieved $P<.20$ were included in a multivariate stepwise regression analysis. Variables with the highest P value were removed one by one, until all remaining variables were $P<.10$. Only the final model was reported. Since this mixed methods study is explorative rather than hypothesis confirming, we decided to use the threshold value of $P<.10$. A sensitivity analysis was conducted to determine the robustness of usage thresholds. The sensitivity analysis was performed by changing the threshold of 6 modules to 5 modules (minus 1) and 7 modules (plus 1); this was subsequently repeated in univariate and multivariate analyses. Model fitting was evaluated with the

Receiver Operating Characteristic (ROC) curve and the Hosmer-Lemeshow test. Statistical analyses were performed using SPSS Statistics 20.0.

Qualitative Analyses

Interviews were analyzed by means of deductive and inductive content analysis [42]. In the deductive approach, a template was created based on three concepts of Eysenbach's law of attrition (relative advantage, complexity, and compatibility) [17]. Guided by these predetermined concepts, text sections were analyzed and coded. In addition to the deductive approach, an inductive method with no predetermined structure was employed. Based on the grounded theory approach [45], recurrent themes from the interview data were identified, coded, labeled, and grouped into broader concepts. While the deductive "top-down" approach tests pre-existing concepts of (non)usage, the inductive "bottom-up" approach starts with patterns observed from the interview data. Data analysis was performed using the software MAXQDA [46] for textual analysis. All interviews were analyzed by the researcher (MB). To assess interrater reliability, a random sample of five interviews was analyzed by a second investigator (DB). Codes were compared and disagreements were resolved by discussion between the 2 researchers. No major differences were found in codes between the 2 researchers.

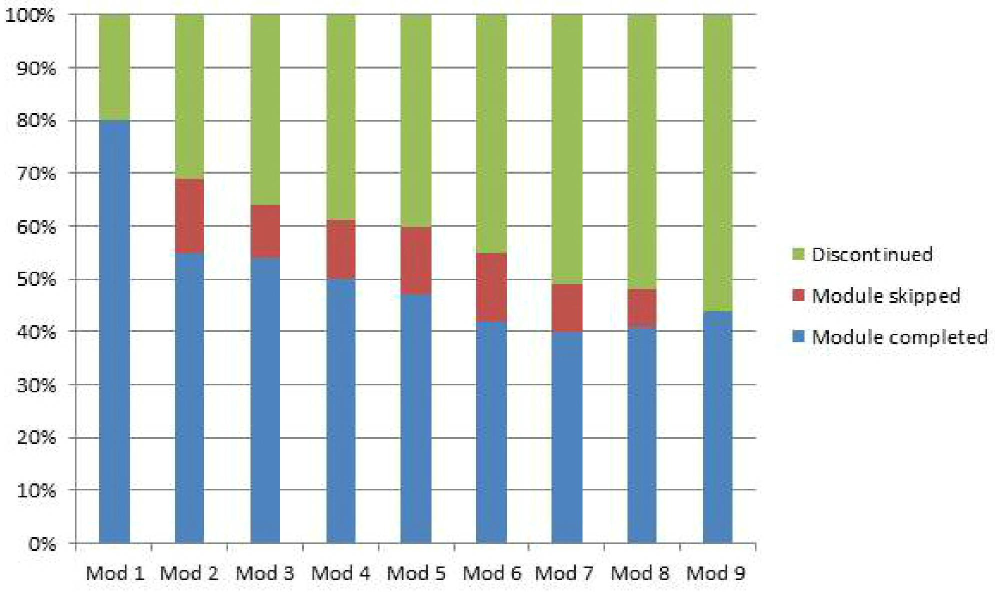
Results

Quantitative Results

Program Completion

Of the 100 participants who received a password and username to enroll, 49 users made a start with the first module and 6 participants never logged in to their personal website. Figure 2 depicts an overview of the module completion rate; 80% of the subjects completed the first module. This percentage declined to 55% during the second module. Approximately 50 of the 100 users completed modules 3, 4, 5, 6, 7, 8, and 9. The average number of modules completed was 5.6 (SD 2.9) out of 9 modules. Since personal messages were updated on a weekly basis, patients had the opportunity to complete a module within 7 days. Overall, 19 of the 100 participants completed all modules of the program, and 46 of the 100 users used at least 6 out of 9 modules. Consequently, 46 participants were defined as users and 54 as nonusers. Users finished a median of 8 (SD 1.1) modules and nonusers a median of 2 (SD 1.5) modules. Adverse events, such as extreme pain and injuries, were not reported during the program.

Figure 2. Program use.



Predictors of Program Usage

presents demographic, clinical, and psychological baseline variables for users and nonusers. Univariate analyses showed that age, BMI, symptoms, and comorbidity reached the threshold of *P*<.20. Based on these variables, three multivariate models were built, which resulted in the most parsimonious predictors including age and comorbidity (Table 2). Higher age (*P*=.08, OR 0.94) and presence of comorbidity (*P*=.02, OR 0.33) were negative predictors for program completion. The sensitivity analysis indicated that the model was robust to changes in the parameter usage. The area under the ROC curve for the model was .68 (95% CI 0.57-0.79). The Hosmer-Lemeshow test of goodness of fit was not statistically significant (*P*=.43), indicating that the data fitted the model well.

Qualitative Results

Overview

The qualitative deductive and inductive analysis resulted in the identification of several reasons for (non)usage. The majority of reasons were found by the deductive analysis. Additionally, the inductive analysis identified a number of personal factors (eg, social environment and emotional factors) relating to (non)usage. Reasons are divided into patient, intervention, and study characteristics and are illustrated by interview quotes. Additional quotes illustrative of each theme are provided in Multimedia Appendix 3.

Patient Characteristics

Interviewees reported that a low mood interfered with their ability to perform modules. One participant summarized this sentiment by saying, “I had a bad year and I was not at ease with myself. I was not in the right mood to exercise. It was all too much” [woman, hip OA, nonuser]. Lack of self-discipline was another identified reason for nonusage. As one man put it “This kind of program does not work for me. I find it difficult to stay motivated all the time. At the beginning I was motivated but then it went downhill quickly. I got lazy and other activities became more important” [man, knee OA, nonuser]. Another reason for discontinuation was the presence of an additional health problems, other than OA. Due to pain and/or other (medical) treatments, it was difficult for interviewees to continue their involvement in the *Join2move* program. In addition, participants who regarded themselves as already physically active found it less necessary to participate. By contrast, patients who felt themselves responsible for their own progress were most likely to use the program. These individuals perceived the program as something that needed to be done, rather than appreciation or enjoyment. Furthermore, those who emphasized the importance of their partner, family, or friends in maintaining the *Join2move* program were mostly adherent. One participant commented: “Regularly, my husband and friends joined me because I told them about the program. This motivated me to continue” [woman, knee OA, user].

Intervention Characteristics

Participants reported that several characteristics of the *Join2move* intervention were identified as a reason for (dis)continuation. Overall, they expressed positive feedback regarding the complexity of the program. Usability problems with respect to the functionality of the website were not reported. The values “trust” and “reliability” were important in the decision to engage the *Join2move* program. To cite one patient: “Join2move is based on an evidence-based theory. This persuaded me to participate and to continue with the program” [man, knee OA, user]. Further, patients consistently reported that the Web-based character of the intervention was an advantage compared with face-to-face treatments. The flexibility of being able to complete modules at one’s own pace without time or travel restrictions was cited as an advantage. On the other hand, the Web-based character also had a downside. Some participants had a strong need for personal guidance. In the words of one participant: “Although it was possible to fill out an evaluation form about pain and performance, sometimes I just needed a personal chat to talk about my progress” [man, knee OA, nonuser]. Moreover, gradually increasing a self-selected activity was not always compatible with expectations. As one participant said: “I expected a package of specific exercises instead” [woman, knee OA, nonuser].

Study Characteristics

Study-related factors were also cited as reasons for remaining or not remaining engaged in the program. Some participants felt under obligation to continue. They described a feeling of commitment to the organizers of the study. “Because I was allocated to the intervention group, I wanted to finish the entire program. Maybe a little old-fashioned but I found it inappropriate to stop halfway” [woman, knee OA, user]. Some participants perceived the questionnaires used as being too long or too difficult. The questionnaire consisted of 17 pages with a total of 171 items. Participants not only lost interest in completing the questionnaires but were also less motivated to continue with the program.

Table 1. Baseline demographic and clinical characteristics.

		Users, n=46	Nonusers, n=54	*P* value
Demographic predictors				
Gender, n (%)				
	Male	17 (37)	23 (43)	.57
	Female	29 (63)	31 (57)	
Age (years), mean (SD)		60 (6.3)	62 (6.5)	.09
Education				
	Lower education	7 (15)	6 (11)	.60
	Middle education	18 (39)	18 (33)	.41
	Higher education	21 (46)	30 (56)	.42
Clinical predictors				
Location OA, n (%)				
	Knee	30 (65)	36 (67)	.89
	Hip	11 (24)	11 (20)	.80
	Both	5 (11)	7 (13)	.64
OA duration (years), mean (SD)		2.8 (1.3)	2.8 (1.1)	.86
BMI (kg/m ^{2}), n (%)				
	Normal weight (<25)	22 (48)	17 (31)	.10
	Overweight (>25)	24 (52)	37 (69)	
Comorbidity, mean (SD)				
	No, n (%)	36 (78)	30 (56)	.02
	Yes, n (%)	10 (22)	24 (44)	
Physical activity		117 (66.1)	130 (65.5)	.29
Pain, 0-10		5.4 (2)	5.4 (2.3)	.92
Fatigue, 0-10		4.7 (2.7)	5.2 (2.8)	.34
Symptoms		56 (15.6)	60 (17.8)	.17
ADL		58.3 (22.3)	55.3 (19.9)	.47
Sport and recreation		58 (22)	55 (19.9)	.47
Quality of life		38.7 (16.9)	42 (17.4)	.32
Psychological predictors, mean (SD)				
Self-efficacy pain		3.4 (0.8)	3.4 (0.9)	.67
Self-efficacy other symptoms		3.5 (0.9)	3.4 (0.9)	.60
Active pain coping		2.0 (0.4)	2.1 (0.4)	.34
Passive pain coping		1.8 (0.4)	1.9 (0.4)	.26
Anxiety		4.7 (3)	4.5 (2.9)	.62
Depression		3.8 (2.9)	3.8 (3)	.88
Internal locus of control		23 (5.4)	23.7 (4.3)	.46
Powerful others locus of control		15.3 (4.4)	15.9 (4.5)	.54

Table 2. Univariate and multivariate analyses for predictors for usage[a].

	B[b]	SE[c]	OR (95% CI)	*P* value
Univariate analyses				
Age, years	−.06	.04	.94 (0.88-1.01)	.09
BMI (normal weight/overweight)	−.69	.42	.50 (0.22-1.13)	.10
Comorbidity (no/yes)	−.93	.44	.39 (0.14-0.84)	.02
Symptoms (0-100)	−.02	.01	.98 (0.96-1.01)	.17
Multivariate analyses				
Age, years	−.07	.04	.94 (0.87-1)	.08
Comorbidity (no/yes)	−1.1	.46	.33 (0.13-0.82)	.02

[a] The reference groups are nonusage, normal weight, and no comorbidity.

[b] B=beta coefficient.

[c] SE=standard error.

Discussion

Principal Findings

The aim of this mixed methods study was to identify patient, intervention, and study characteristics that facilitate or impede the usage of a Web-based intervention for patients with knee and/or hip OA. Results from this study showed that participants with knee and/or hip OA used the *Join2move* program less than intended. Of all participants, 94% started the program, 46% reached the threshold of 6 out of 9 completed modules, and 19% finished all 9 weekly modules. To put these rates into perspective, we refer to Hansen et al [7] who found that merely 7% of inactive participants logged in once to a self-guided Web-based PA intervention, and Irvine et al [8] showed that 46% of the users completed all 12 sessions of a self-guided Web-based PA intervention. In a study among patients with rheumatoid arthritis, Van den Berg et al [47] reported that 86% of the patients assessed a website once per week for the duration of 3 months. When considered in light of these studies, our usage rates can be interpreted as reasonable. However, Web-based interventions differ widely in terms of population, content, setting, and methods of measuring usage. For example, while our study used number of modules completed for measuring usage, the above-mentioned studies used log-in data [7,8] or questionnaires [47] as measures. Further, our intervention was self-directed, while the program by Van den Berg et al [47] contained supervision. These differences may have had a major impact on usage and indicates that direct comparison with other reported Web-based interventions remains difficult. In an effort to overcome this issue, the systematic review by Kelders et al [26] adopted the concept of intended usage. This is a universal measure for adherence, which is defined as the extent to which users should experience the content of the intervention to derive maximum benefit.

Considering the predictors of usage, it appeared from the quantitative analysis that age and comorbidity proved to be significantly related to program usage. Younger participants were more likely to use the intervention modules than older participants. This is in contrast to previous studies that have found correlations between older age and higher usage rates [9,21,22]. This discrepancy in findings can be explained by the fact that the mean age of our study sample was significantly higher (62 years) than the mean age of the other studies (42, 44, and 39 years respectively) [9,21,22]. In fact, the younger participants from our sample should be compared with the older subjects from other studies. This suggests that participants aged roughly 50-60 years are most adherent to Web-based interventions. Apart from this, the presence of an additional medical condition increased the odds of not using *Join2move*. These results were also confirmed in the interviews. Patients mentioned that physical discomfort during PA and specific comorbid-related factors such as pain, medication use, and disease-related constraints hampered their program performance. Another explanation might be that the program was solely focused on OA and no attention was paid to additional diseases. Participants with an additional illness might feel that the *Join2move* program did not suit their needs. Unfortunately, it was not possible to examine the influence of each comorbidity on usage due to the low number of cases per disease category. Further research is required to examine which of the comorbidities is most predictive in relation to (non)usage.

With respect to the intervention, participants indicated that the automatic gradual increase of PA as well as working toward a short-term goal were mechanisms that supported them in completing weekly modules. Compared with face-to-face treatments, the flexibility of completing modules at one’s own pace without time or travel restrictions was cited as a major advantage. However, older patients, those with comorbidity and patients who attach great importance to personal contact indicated that the lack of human involvement was a disadvantage. Furthermore, from the interviews it became clear that those who felt themselves responsible for their own progress were most likely to use the program. This, however, was not confirmed in the quantitative analysis. Although we included questions about responsibility and persistence, the questionnaires were not sensitive enough to confirm the conclusions from the qualitative analysis. This illustrates very well why we have chosen dual data collection. The weakness of questionnaires was compensated by interview data. Other mentioned motivations for (non)usage were trial specific. While questionnaires impede usage, commitment to the research team was described as an important facilitator for usage. We did not find any predictive value for education and gender, in contrast to other studies [22,24].

Limitations

A major weakness is the potential presence of recall bias. In an effort to prevent attention bias during the previously conducted randomized controlled trial, the length of time between program participation and interviews was approximately 12 months. As a consequence, participants may not have accurately remembered the intervention in detail. This may have affected the reliability of our results. Another weakness is that results are limited in their generalizability because participants were mainly older, healthy, and highly educated patients with knee and/or hip OA. Furthermore, the role of motivation as proximate determinant of usage behavior was not investigated in this study. Future research should examine the role of motivation on program usage. A last limitation was that participants were included on the basis of self-reported OA. Diagnosis was not confirmed through clinical tests or x-ray reports due to practical reasons. Although self-reported OA is a common inclusion strategy in the field of osteoarthritis research, it is presumable that we have included false positive OA patients in the study.

Future Directions and Implications

In light of rising health care costs and the large population of patients with knee and/or hip OA, *Join2move* is an effective, low-cost, and promising program for improving PA levels in patients with knee and/or hip OA. We believe that the quantitative and qualitative results provide insights that are of relevance to the field of Web-based health education. Future Web-based PA programs should include gradual activity programs with attainable short-term goals. Goal setting, preferably by participants themselves, as well as feedback on performance seem to be powerful tools for increasing the usage of Web-based interventions. Future studies should also pay special attention to older patient groups and patients with a comorbid condition. For these groups a more personal approach is needed. In a further study, we will investigate if guidance by a physical therapist will lead to higher levels of usage. The fact that participants described a feeling of commitment to the organizers of the study may indicate that observed usage patterns cannot be replicated in a real-life setting. Conducting more practically oriented research is an important way to explore usage rates in real-world settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots intervention.

[PDF File (Adobe PDF File), 654KB - jmir_v15i10e223_app1.pdf]

Multimedia Appendix 2

Interview guide.

[PDF File (Adobe PDF File), 9KB - jmir_v15i10e223_app2.pdf]

Multimedia Appendix 3

Interview quotes.

[PDF File (Adobe PDF File), 119KB - jmir_v15i10e223_app3.pdf]

References

1. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med 2000 Oct 17;133(8):635-646. [Medline: 11033593]
2. Macera CA, Hootman JM, Sniezek JE. Major public health benefits of physical activity. Arthritis Rheum 2003 Feb 15;49(1):122-128 [FREE Full text] [doi: 10.1002/art.10907] [Medline: 12579603]
3. Roddy E, Zhang W, Doherty M, Arden NK, Barlow J, Birrell F, et al. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus. Rheumatology (Oxford) 2005 Jan;44(1):67-73 [FREE Full text] [doi: 10.1093/rheumatology/keh399] [Medline: 15353613]
4. 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]
5. Sjöström M, Oja P, Hagströmer M, Smith B, Bauman A. Health-enhancing physical activity across European Union countries: the Eurobarometer study. J Public Health 2006 May 19;14(5):291-300. [doi: 10.1007/s10389-006-0031-y]

6. Bossen D, Veenhof C, van Beek K, Dekker J, de BD. The Effectiveness of a Web-based Physical Activity Intervention in Patients with Knee and/or Hip Osteoarthritis: A Randomized Controlled Trial. 2013 Sep 23 Presented at: Medicine 2.0 Congress; Sept. 23-24, 2013; London, England p. 1-32.
7. 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]
8. 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]
9. 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]
10. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
11. 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]
12. 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]
13. Robroek SJ, Lindeboom DE, Burdorf A. Initial and sustained participation in an internet-delivered long-term worksite health promotion program on physical activity and nutrition. J Med Internet Res 2012;14(2):e43 [FREE Full text] [doi: 10.2196/jmir.1788] [Medline: 22390886]
14. Binks M, van Mierlo T. Utilization patterns and user characteristics of an ad libitum Internet weight loss program. J Med Internet Res 2010;12(1):e9 [FREE Full text] [doi: 10.2196/jmir.1347] [Medline: 20350926]
15. 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]
16. 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]
17. 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]
18. 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]
19. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
20. 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]
21. 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]
22. Van 't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
23. Poirier J, Cobb NK. Social influence as a driver of engagement in a web-based health intervention. J Med Internet Res 2012;14(1):e36 [FREE Full text] [doi: 10.2196/jmir.1957] [Medline: 22356829]
24. Robroek SJ, Brouwer W, Lindeboom D, Oenema A, Burdorf A. Demographic, behavioral, and psychosocial correlates of using the website component of a worksite physical activity and healthy nutrition promotion program: a longitudinal study. J Med Internet Res 2010;12(3):e44 [FREE Full text] [doi: 10.2196/jmir.1402] [Medline: 20921001]
25. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]

26. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
27. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
28. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
29. 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]
30. 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]
31. 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]
32. 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]
33. 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]
34. 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]
35. 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]
36. Dutch Bureau of Statistics (CBS). Den Haag/Heerlen. 2008. Zelfgerapporteerde medische consumptie, gezondheid en leefstijl URL: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=81173ned&D1=2-30&D2=0-13,47-51&D3=0&D4=l&VW=T [accessed 2013-03-20] [WebCite Cache ID 6JEKbyIO3]
37. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
38. 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]
39. Taal E, Rasker JJ, Wiegman O. Group education for rheumatoid arthritis patients. Semin Arthritis Rheum 1997 Jun;26(6):805-816. [Medline: 9213379]
40. 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]
41. 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]
42. Patton MQ. How to use qualitative methods in evaluation. Newbury Park, Calif: Sage Publications; 1987.
43. Express Scribe transcription software. NCH Software. 2011. URL: http://www.nch.com.au/scribe/index.html [accessed 2013-10-01] [WebCite Cache ID 6K1iD8Ssx]
44. Rogers EM. Diffusion of innovations. New York: Free Press; 2003.
45. Strauss AL, Corbin J. Basics of qualitative research: techniques and procedures for developing grounded theory. Thousand Oaks: Sage Publications; 1998.
46. MAXqda software computer program. Berlin: Lucanus GmbH; 2007. URL: http://www.maxqda.com/ [accessed 2013-10-01] [WebCite Cache ID 6K1ivPBtJ]
47. van den Berg MH, Ronday HK, Peeters AJ, le Cessie S, van der Giesen FJ, Breedveld FC, et al. Using internet technology to deliver a home-based physical activity intervention for patients with rheumatoid arthritis: A randomized controlled trial. Arthritis Rheum 2006 Dec 15;55(6):935-945 [FREE Full text] [doi: 10.1002/art.22339] [Medline: 17139640]

Abbreviations

BGA: behavior graded activity
BMI: body mass index
HOOS: Hip Injury OA Outcome Score
KOOS: Knee OA Outcome Score
MHLC: Multidimensional Health Locus of Control Scale

OA: osteoarthritis
OR: odds ratio
PA: physical activity
PASE: PA Scale for the Elderly
RCT: randomized controlled trial
ROC: Receiver Operating Characteristic

Edited by G Eysenbach; submitted 31.05.13; peer-reviewed by L van Genugten, R Kretschmann, R Kok; comments to author 20.06.13; revised version received 31.07.13; accepted 07.09.13; published 16.10.13.

Please cite as:
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
URL: http://www.jmir.org/2013/10/e223/
doi:10.2196/jmir.2742
PMID:24132044

Original Paper

Rethinking the Dose-Response Relationship Between Usage and Outcome in an Online Intervention for Depression: Randomized Controlled Trial

Liesje Donkin[1], BSc, MSc (Hons), MAPS; Ian B Hickie[1], MD, FRANZCP; Helen Christensen[2], PhD; Sharon L Naismith[1], BA (Hons), MClinPsych(Neuro), DPsych (Neuro); Bruce Neal[3], MB ChB, PhD; Nicole L Cockayne[1], BSc (Hons), MPH; Nick Glozier[1,4], MBBS, FRANZCP, PhD

[1]Brain & Mind Research Institute, The University of Sydney, New South Wales, Australia

[2]Black Dog Institute, Prince of Wales Hospital, Hospital Road, Randwick, New South Wales, Australia

[3]The George Institute for Global Health, The University of Sydney, PO Box M201, Missenden Road, New South Wales, Australia

[4]Disciplines of Psychiatry and Sleep Medicine, Sydney Medical School, The University of Sydney, Sydney, Australia

Corresponding Author:
Liesje Donkin, BSc, MSc (Hons), MAPS
Brain & Mind Research Institute, The University of Sydney
100 Mallet Street
Camperdown, NSW
New South Wales, 2050
Australia
Phone: 61 21847886
Fax: 61 96234656
Email: Liesje.Donkin@sydney.edu.au

Abstract

Background: There is now substantial evidence that Web-based interventions can be effective at changing behavior and successfully treating psychological disorders. However, interest in the impact of usage on intervention outcomes has only been developed recently. To date, persistence with or completion of the intervention has been the most commonly reported metric of use, but this does not adequately describe user behavior online. Analysis of alternative measures of usage and their relationship to outcome may help to understand how much of the intervention users may need to obtain a clinically significant benefit from the program.

Objective: The objective of this study was to determine which usage metrics, if any, are associated with outcome in an online depression treatment trial.

Methods: Cardiovascular Risk E-couch Depression Outcome (CREDO) is a randomized controlled trial evaluating an unguided Web-based program (E-couch) based on cognitive behavioral therapy and interpersonal therapy for people with depression and cardiovascular disease. In all, 280 participants in the active arm of the trial commenced the program, delivered in 12 modules containing pages of text and activities. Usage data (eg, number of log-ins, modules completed, time spent online, and activities completed) were captured automatically by the program interface. We estimated the association of these and composite metrics with the outcome of a clinically significant improvement in depression score on the Patient Health Questionnaire (PHQ-9) of ≥5 points.

Results: In all, 214/280 (76.4%) participants provided outcome data at the end of the 12-week period and were included in the analysis. Of these, 94 (43.9%) participants obtained clinically significant improvement. Participants logged into the program an average of 18.7 times (SD 8.3) with most (62.1%, 133/214) completing all 12 modules. Average time spent online per log-in was 17.3 minutes (SD 10.5). Participants completed an average of 9 of 18 activities available within the program. In a multivariate regression model, only the number of activities completed per log-in was associated with a clinically significant outcome (OR 2.82, 95% CI 1.05-7.59). The final model predicted 7.4% of variance in outcome. Curve estimates indicated that significant logarithmic (P=.009) and linear (P=.002) relationships existed between activities completed per log-in and clinically significant change.

Conclusions: Only one objective measure of usage was independently associated with better outcome of a Web-based intervention of known effectiveness. The 4 usage metrics retained in the final step of the regression accounted for little outcome variance. Medium level users appeared to have little additional benefit compared to low users indicating that assumptions of a linear relationship between use and outcome may be too simplistic and further models and variables need to be explored to adequately understand the relationship.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR): ACTRN12610000085077; http://www.anzctr.org.au/ACTRN12610000085077.aspx (Archived by WebCite at http://www.webcitation.org/6K9FQtKBn).

(J Med Internet Res 2013;15(10):e231) doi:10.2196/jmir.2771

KEYWORDS

adherence; Internet; eHealth; depression; patient compliance

Introduction

Web-based interventions for psychological conditions have been found to have a moderate to large effect size [1,2] that is comparable to face-to-face interventions [3-5]. However, in a review of Web-based interventions, the median proportion of users completing all modules in a trial was 56% [6]. Drawing on the medication literature, this level of exposure to an intervention would be considered suboptimal, but no similar models exist for Web-based interventions. Given this, it is unclear how important the degree of program usage is for outcomes in online interventions.

To date, much of the reporting of engagement is of dropout attrition [7] (the proportion of participants that do not complete the trial or provide follow-up data) or of treatment completers or persisters (those that complete the intervention). However, reporting on attrition alone does not adequately describe how the user interacts with the program nor does it inform developers of how much of the intervention needs to be completed in order for participants to obtain a benefit. An alternative way to gain these insights is to assess a measure of usage or of adherence. Usage refers to the level of activity within a program, whereas adherence refers to the degree to which the user's activity within the program matches the pattern of activity that was intended by the program developers. For example, a user who completes all 10 modules in a program will have 100% usage on the modules' metric of usage. However, if these modules were supposed to be completed weekly and the user only completed 6 of these on time, the user was 60% adherent on the modules' adherence metric. Alternatively, if a user completes all 20 of the compulsory activities in a program when scheduled to do so, the user is 100% adherent. However, if the user completes these activities several times, the user's usage statistic may be much higher. These 2 concepts provide a measure of activity within a program, with one focused on general activity (usage) and the other focused on whether this activity matches the developer's expectations (adherence). Therefore, adherence is a specific subset of usage that has timing factors as a component of what is measured. Despite these differences, both of these variables provide important information about program engagement and provide an opportunity for researchers to understand whether it is exposure to program material or adherence that is needed to obtain a clinically significant effect.

Web-based interventions have an advantage over traditional medication trials in terms of measuring usage because there are many objective metrics readily available [8]. Such objective measures include the number of times the participant logs into the program, the number of modules completed, the number of completed activities, and broad patterns of usage, such as time spent online and the repetition of optional activities completed. Assessment may be further refined by composite measures [8], such as time spent per activity or number of modules completed per log-in. Despite the relative ease of capturing these data in online interventions, few studies report these. Even when reported, common practice is to report dropout attrition only or a singular measure of use, which inadequately describes the level of program use in these trials. Thus, little insight is gained about the impact of usage on program outcomes.

Recent articles have begun to explore the relationship between program usage and outcomes [6,8,9]. For example, based on a post hoc median split of website activity (calculated as number of log-ins multiplied by duration in minutes per log-in), high users of an Internet program aimed at smoking cessation were more likely to quit and remain continually abstinent than low users [10]. The same has been found in eating disorders in which increased completion of program components and tasks in online interventions has been found to predict better outcomes [9,11,12]. Likewise, greater improvements in anxiety and depression were seen as individuals worked through increasing numbers of modules on an online cognitive behavior therapy (CBT) program [13]. Finally, better engagement online has been found to positively influence the consumption of fruit and vegetables [8]. Such analyses indicate that the dose of the behavioral intervention appears to influence outcome [10,13-15].

A recent systematic review of Web-based interventions showed that several potential usage metrics (number of log-ins, self-reported activity completions, and time spent online) were not consistently associated with outcome for Web-based intervention for psychological disorders [16]. Only the relationship between proportion of modules completed and outcome appear to be consistent. The assumption behind these approaches is that there is a linear relationship between outcome and exposure to content. However, the relationship between dose-response may not be linear, but rather curvilinear (eg, reaches a saturation point where no further benefit is obtained). Likewise, the association may be modified by sociodemographic factors [17] or psychological traits [18,19].

The inability to consistently detect a dose-response relationship may be influenced by the usage metric utilized. Because most studies only report 1 or 2 such usage metrics and rarely examine the relationship between these metrics, little is known about the relative contribution of the different metrics or the relationship of these to outcomes. Previous attempts have been made to define the measurement of usage, most often in the form of adherence [20,21], by producing combined measures of engagement [8] and to standardize reporting of this [22,23], but variations in reporting continue to exist in the literature. These variations in reporting may be because of a lack of consensus about the relationship between usage and outcome or the best way to measure usage, leaving researchers confused. Given this, this study aims to evaluate the role of several different usage metrics and combinations of these on the outcome of a randomized controlled trial (RCT) of an online depression treatment trial. Furthermore, this study seeks to determine which of these, if any, are more important in predicting and explaining a clinically significant change. It was hypothesized that usage would be associated with outcome and that modules completed would have the strongest relationship with outcome, consistent with the systematic review by Donkin et al [16].

Methods

Overview

Cardiovascular Risk E-couch Depression Outcome (CREDO) is a randomized, double-blind, parallel, attention-controlled, Internet-delivered trial targeting depressive symptoms in those with risk factors for or diagnosis of cardiovascular disease (CVD). The method and primary results of CREDO have been published elsewhere [24,25]. This study is a secondary analysis of the usage of the intervention.

Participants

Trial participants were recruited from the 45 and Up Study [26], a longitudinal study of health and aging in New South Wales, Australia. Potential participants were invited to participate in the CREDO trial if they were aged between 45 and 75 years, provided a valid email address, self-reported significant risk factors for or a history of CVD, and screened positive for at least moderate psychological distress on the Kessler Psychological Distress 10 scale (K10) [27,28] during the 45 and Up Study baseline data collection. Potential participants underwent a further screening process for trial inclusion to ensure a current level of depressive symptoms. Once identified as being suitable for the trial, participants were randomized either into the intervention arm using E-couch, an Internet cognitive behavior therapy (iCBT) intervention, or to HealthWatch, an online attention control. Both E-couch and its predecessor, MoodGYM [29,30], have been shown to be effective in improving symptoms of depression [31]. To determine the effect of E-couch usage on outcome, only those participants in this arm who completed the outcome measure at 3 months were included in this analysis.

Intervention

E-couch is an iCBT program containing psychoeducation about depression with components of CBT, interpersonal psychotherapy (IPT), applied relaxation, and physical activity. In its open access format [32], E-couch allows users to choose which aspects of treatment they wish to engage with in the form of choosing their own toolbox. For the purpose of CREDO, the program was restructured linearly so that it contained 12 modules that required users to work through each module sequentially rather than being able to choose which section they wanted to engage with.

Activities were spread throughout E-couch. The CBT component had 12 activities, IPT had 4 activities, and the exercise component had 2 activities. The relaxation component contained a recording of relaxation exercises, but because this did not require participants to enter anything into the program, it was not included as an activity in this analysis. See Multimedia Appendices 1-4 for screenshots of exercise examples. Users were sent an email when their module opened and a reminder email again 3 to 4 days later if the module had not yet been completed. If they still had not completed the module 1 week after it was opened, they received a reminder phone call prompting their return to the site.

Outcome Measure

The primary outcome measure of the study was the 9-item Primary Health Questionnaire (PHQ-9) [33], a widely used self-report tool designed for the assessment of depressive symptoms in community samples. Items are scored on a scale of 0 to 3 and are provided with a summary score ranging from 0 to 27. The PHQ-9 has shown to have sufficient sensitivity and specificity for major depressive disorder [29,33] and to be an indicator of minimal clinically important change for individuals [34]. For this analysis, the standard definition of a clinically significant improvement of a reduction of 5 points in PHQ-9 score [33] was used as the outcome measure. This was utilized in favor of a continuous measure because it was considered to be the most clinically meaningful.

Usage Metrics

Overview

A number of measures were used to assess usage of the intervention as recorded objectively by the program and did not rely on participant self-report.

Proportion of Modules Completed

The proportion of the 12 possible E-couch modules that the individual completed was recorded. A complete module consisted of the user clicking through each page of the module until they had viewed all pages. No time limit or activity level was required to complete the module other than what was required to click through the module's pages.

Proportion of Activities Completed

Data were captured for each type of activity section (ie, cognitive activities, relationship activities, and physical activities) and overall activity completion. A total of 18 different activities were available for completion in the program and were spread throughout the modules. To complete more activities, users needed to complete more modules. In order for an activity to be counted as completed, the individual had to have engaged with the task in some way (eg, provided text or worked through the activity by clicking on the required sections).

Number of Program Log-Ins

The number of times the participant logged into the program over the course of the 12-week period was recorded. Participants were expected to complete 1 log-in per week; therefore, they were expected to have logged into the program on 12 occasions. All modules could theoretically be completed in 1 log-in at week 12 (a module being made available each week for 12 weeks). Participants were able to log in as many times as they wished per module, allowing this metric to range from 1 to an unknown limit imposed by the study duration and participants' availability.

Total Number of Activities Completed

Users were able to complete each activity as many times as they wanted and were not limited to 18 activities. Given this, the total number of activities completed was collected.

Total Time Spent in the Program

The total time spent logged into the program each week was recorded. The program continued to keep time if the user did not log out; therefore, the time spent on the final page for each log-in was excluded in case of failure to log out. Average time spent online per log-in and total time in the program were used in this analysis. Average time spent online per log-in was capped at 60 minutes to reduce the impact of outliers. This impacted 1 participant; average time spent online per log-in was limited to 60 minutes from 83 minutes. No minimum average time requirement was defined for the program.

Average Number of Activities Completed per Log-In

This was calculated by dividing the total number of activities completed by the total number of log-ins to the program.

Average Number of Minutes per Log-in

This was calculated by dividing the total time in minutes spent in the program by the total number of log-ins to the program.

Average Number of Modules Completed per Log-in

This was calculated by dividing the number of modules completed by number of times that they logged in to the program.

Combined Modules-Activities Measure

An aggregated measure was calculated by adding the number of modules completed (range 0-12) with the number of compulsory activities completed (range 0-18) to give a total range of 0 to 30.

Data Analysis

Data analyses were completed using SPSS version 20.0 (IBM Corp, Armonk, NY, USA). Data were examined for normality and where the assumptions of normality were not met, nonparametric tests were utilized. Chi-square (χ^2), independent samples *t* tests, and Mann-Whitney *U* tests were used to determine if there were any differences between those who persisted with the study (ie, provided postintervention outcome data at week 12) and those who did not. Univariate associations of demographic variables with outcome and usage were evaluated using Spearman rank correlation (ρ) and chi-square tests. Similarly, Spearman rank correlations, independent samples *t* tests, Mann-Whitney *U* tests, and chi-square tests were used to examine the relationship between usage variables and clinically significant improvement.

A binary logistic regression model using the enter method was then completed to assess the ability of the usage variables to predict clinically significant improvement. Demographic and usage variables were included in the regression model if there was *P* value of $P<.20$ for its association with the outcome. Autocorrelations between usage variables were assessed before modeling. Where significant autocorrelations were found (considered to be a correlation of $r>0.80$), the variables were identified as a priori (ie, being more relevant) and were entered into the model.

Ethical Approval

Written informed consent was obtained from all the participants and ethics approval for the 45 and Up Study was provided by the University of New South Wales Human Research Ethics Committee. Ethics approval for the CREDO trial was obtained from the University of Sydney Human Research Ethics Committee.

Results

Overview

Of the 562 participants who provided consent and met trial criteria, 280 (49.8%) were randomized into the E-couch iCBT program. Of these, 214 (76.4%) persisted with the study and provided postintervention outcome data. There were no significant differences between persisters and those who did not provide outcome data in age, sex, country of birth, marital status, or baseline depression severity (Table 1). However, those who spoke English at home were more likely to persist with the trial (OR 2.91, 95% CI 1.51-7.38). As indicated in Table 1, significant differences existed between persisters and nonpersisters on all 3 basic usage metrics.

Table 1. Association of demographics, baseline depression score, and basic usage metrics with study persistence.

Variable	Persisters (n=214)	Nonpersisters (n=66)	OR (95% CI)	*t* (df)	*P* value
Age (years), mean (SD)	57.39 (6.5)	57.68 (7.1)		0.31 (278)	.75
Depression score (PHQ-9), mean (SD)	12.13 (3.5)	11.27 (3.0)		–1.81 (278)	.07
Sex (female), n (%)	133 (62.1)	40 (60.6)	1.07 (0.61-1.88)		
Speak English at home, n (%)	203 (94.9)	57 (86.4)	2.91 (1.15-7.38)		
Marital status (partnered), n (%)	151 (70.6)	53 (80.3)	0.59 (0.30-1.15)		
Highest qualification (postschool), n (%)	157 (73.4)	47 (71.2)	1.11 (0.60-2.06)		
Born in Australia, n (%)	158 (73.8)	53 (80.3)	0.69 (0.35-1.36)		
Number of modules completed (range 0-12), mean (SD)	12 (3.8)	3.21 (2.1)		–21.05 (196)	<.001
Number of activities completed (range 0-44), mean (SD)	11.9 (9.1)	1.6 (3.0)		–14.18 (276)	<.001
Number of log-ins (range 1-65), mean (SD)	18.72 (8.3)	7.18 (4.7)		–14.25 (196)	<.001

Intervention Usage of Those Who Persisted With the Study

Of the 214 participants, 62.1% (133/214) completed all 12 modules and 79% (169/214) completed 10 modules or more. In all, 2 participants (1%) did not complete any modules, but did complete the assessments. Participants completed 9 of 18 (50%) available activities in the program on average; the number of activities completed ranged from 0 to 18. Several participants completed the same activity on multiple occasions, as allowed by the program, with the total number of activities completed ranging from 0 to 44 (mean 11.9, SD 9.1). Participants logged into the program an average of 18.7 times (SD 8.3, range 1-65). The mean total time spent in the program was 318.3 minutes (SD 204.3, range 24.7-1221.7).

On average, participants completed 0.5 activities per log-in (range 0-7.57) and 0.6 modules (range 0-2.0) per each log-in. The average amount of time spent online per log-in was 17.3 minutes (SD 10.5), whereas the average time to complete a module was 33.19 minutes (SD 23.18). For the combined usage measure of activities and modules, the mean score was 19.47 (SD 7.49).

Demographic Factors, Usage, and Outcome

Overview

In all, 94 (43.9%) of participants obtained clinically significant improvement during the study. There was no difference between those who obtained clinically significant improvement and those who did not in age, level of education, country of birth, language spoken at home, marital status, or baseline depression score.

Older age was associated with greater time spent online (ρ=0.27, *P*<.001), more log-ins (ρ=0.19, *P*=.01), and total number of activities completed (ρ=0.16, *P*=.02). Men completed more modules ($\chi^2_1 = 5.0$, *P*=.03) than women did. There were no sex differences in the other basic usage measures or obtaining a clinically significant outcome ($\chi^2_1 = 1.6$, *P*=.69). A significant correlation was found between baseline depression severity and number of modules completed (ρ=–0.141, *P*=.04) with people who were more depressed completing fewer modules. However, there was no significant association between baseline depression severity and other usage metrics. No relationships were found between the interactions of age, sex, and baseline depression severity with usage on outcome.

Usage Factors Associated With Clinically Significant Improvement in PHQ-9 Score

Associations between basic and composite usage measures and clinically significant improvement were examined (Table 2). Of the basic usage measures, there was no significant difference between those who obtained clinically significant change and those who did not in the number of modules completed, the number of log-ins to the program, or the proportion of the 18 potential activities completed. However, a significant difference was found in the total number of minutes spent in the program between those who obtained clinically significant change on the PHQ-9 (mean 351.1, SD 206.4) and those who did not (mean 292.6, SD 199.8; t_{212}=2.09, *P*=.04). Likewise, a significant difference was found in the number of activities completed between those who obtained change (mean 13.5, SD 9.5) and those who did not (mean 10.7, SD 8.7; t_{212}=2.33, *P*=.02).

Of the composite measures, a significant difference was found between those who obtained clinically significant change and those who did not in average number of activities completed per log-in (mean difference 0.20, range 0.07-0.33; t_{212}=3.02, *P*=.01) and average time spent online per log-in (mean difference 3.26 minutes, range 0.88-5.63; t_{212}=2.71, *P*=.01). No other composite usage metrics were associated with significant outcomes.

A total of 214 cases were analyzed in a binary logistic regression model using the backwards likelihood ratio method. In total, 6 usage variables were associated with outcome in the univariate analyses at the *P*<.20 level as prespecified for inclusion in the model, but because of high autocorrelations, total number of activities completed and proportion of activities completed were removed. The combined measures were retained because they were considered a priori to provide a better reflection of use over time rather than the simple quantification of use.

The remaining usage measures (average minutes per log-in, average number of activities completed per log-in, and total time spent in the program) were entered into the model with sex and age. In the parsimonious model, total time spent online in the program, time spent online per log-in, activities completed per log-in, and the combined modules-activities measure remained in the model. Of these, only the number of activities completed per log-in was statistically associated with a clinically significant improvement (see Table 3.).

According to the Nagelkerke R^2 statistic, the variance in the outcome predicted by this model was 7.4%. The likelihood of the model predicting whether or not the individual would obtain clinically significant change or not was 61.2%. A further regression was modeled using the variables excluded based on autocorrelations as sensitivity analysis. This yielded similar results with only activities completed per log-in being found to contribute significantly to the final model.

To examine the linearity of the relationship between usage and outcome, the linear model of clinically significant change and significant usage metrics included in the linear regression were compared with logarithmic and quadratic curve estimation. Significant curve estimations were found for the 4 usage variables included in the final step of the analysis, except for the combined activities-modules metric (see Table 4), although they did not significantly outperform the linear model in any case.

Table 2. Univariate associations of usage metrics of E-couch with clinically significant change in depression in CREDO.

Variable	Overall sample, mean (SD)	Clinically significant improvement, mean (SD)[a]		Difference (95% CI)	*P* value	Effect size (Cohen's *d*)
		Yes (n=94)	No (n=120)			
Proportion of 12 modules completed	0.9 (0.3)	0.9 (0.3)	0.9 (0.3)	0.03 (–0.05, 0.11)	.42[c]	0.11
Proportion of all 18 activities completed	0.5 (0.3)	0.6 (0.3)	0.5 (0.3)	0.07 (–0.20, 2.70)	.09[b]	0.23
Total number of activities completed (range 0-44)	11.9 (9.1)	13.6 (9.5)	10.8 (8.7)	2.89 (0.44, 5.34)	.02[c]	0.32
Number of program log-ins (range 1-65)	18.7 (8.3)	18.9 (7.2)	18.6 (9.1)	0.23 (–2.04, 2.49)	.84[b]	0.03
Total number of minutes spent online in program	318.3 (204.3)	351.1 (206.4)	292.6 (199.8)	58.42 (3.34, 113.47)	.04[b]	0.29
Average number of activities completed per log-in (range 0-2.4)	0.6 (0.5)	0.8 (0.4)	0.6 (0.5)	0.20 (0.07, 0.33)	.04[b]	0.46
Number of modules completed per log-in (range 0-2.0)	0.6 (0.3)	0.6 (0.3)	0.6 (0.2)	–0.01 (–0.08, 0.06)	.73[b]	0.06
Average number of minutes online per log-in (range 1.6-63.8)	17.3 (8.9)	19.1 (10.4)	15.9 (7.2)	3.26 (0.88, 5.63)	.01[b]	0.37
Average number of minutes online per module (range 5.6-165.8)	33.2 (23.8)	35.3 (22.9)	31.5 (24.5)	3.80 (2.69, 10.30)	.25[b]	0.16
Combined measured (range 0.0-30.0)	19.5 (7.5)	20.4 (7.0)	18.8 (7.8)	1.595 (–0.43, 3.62)	.12[b]	0.21

[a] Defined as a reduction of 5 points or more on the PHQ-9.

[b] *t* test analysis.

[c] Mann-Whitney *U* test.

[d] Combined measure of number of compulsory activities completed (of possible 18) and number of modules completed (of possible 12) with scores ranging from 0-30.

Table 3. Final step in the binary logistic regression model using the enter ratio and adjusted for age and gender examining the relationship of usage measures to obtaining clinically significant change.

Usage variable	B	SE	Wald χ^2	*P*	Exp(B)	95% CI
Time spent online in minutes	0.00	0.00	0.88	.35	1.00	1.00-1.00
Time spent online per log-in	0.01	0.03	0.08	.78	1.01	0.96-1.06
Activities completed per log-in	1.04	0.51	4.21	.04	2.82	1.05-7.59
Combined modules and activities measure	–0.04	0.03	1.24	.27	0.96	0.90-1.03
Constant	–0.86	0.30	8.13	<.001	0.42	

Table 4. Comparison of linear, logarithmic, and quadratic models for usage variables included in the linear regression.

Usage variable and model	Unstandardized coefficients		Standardized coefficients	*F* (df)	*P*	Adjusted R^2
	B	SE	β			
Activities completed per log-in						
Linear	0.217	0.070	0.208	9.61 (1,212)	.002	0.039
Logarithmic	0.105	0.040	0.178	6.98 (1,212)	.009	0.027
Quadratic	0.023	0.101	0.044	4.81 (2,211)	.009	0.035
Total time spent online						
Linear	0.000	0.000	0.142	4.38 (1,212)	.04	0.016
Logarithmic	0.114	0.047	0.163	5.79 (2,212)	.02	0.022
Quadratic	0.001	0.000	0.418	3.23 (2,211)	.04	0.020
Total time spent online per log-in						
Linear	0.010	0.004	0.183	7.33 (1,212)	.007	0.029
Logarithmic	0.146	0.064	0.155	5.21 (1,212)	.02	0.019
Quadratic	0.001	0.011	0.022	4.04 (2,211)	.02	0.028
Combined activities-modules metric						
Linear	0.007	0.005	0.106	1.22 (1,212)	.12	0.007
Logarithmic	0.084	0.055	0.104	1.15 (1,212)	.28	0.001
Quadratic	0.011	0.019	0.169	1.23 (1,211)	.30	0.002

Usage Groups and Outcome for Persisters

Patterns of usage were also explored by trichotomizing usage metrics using tertiles of low, medium, and high users. When exploring this categorization against obtaining clinically significant change, significant relationships were found between outcome and time spent online (χ^2_2 =6.6, *P*=.04), time spent online per log-in (χ^2_2 =6.8, *P*=.03), and activities completed per log-in (χ^2_2 =6.7, *P*=.04). In the time spent online variable, significantly more high users obtained clinically significant change than low users (high users obtaining change = 53.5%, low users= 32.4%, *P*=.01), in time spent online per log-in, more high users obtained change than medium users (high users obtaining change = 54.9%, medium users= 33.3%, *P*=.01), and in activities completed per log-in, significantly more high users obtained change than low users (high users obtaining change = 56.3%, low users = 36.6%, *P*=.02) or medium users (medium users= 38.9%, *P*=.04). See Figures 1-3 for graphical representations of these findings.

Figure 1. The difference in percentage of participants achieving clinically significant change across usage groups relating to total time spent online in the program.

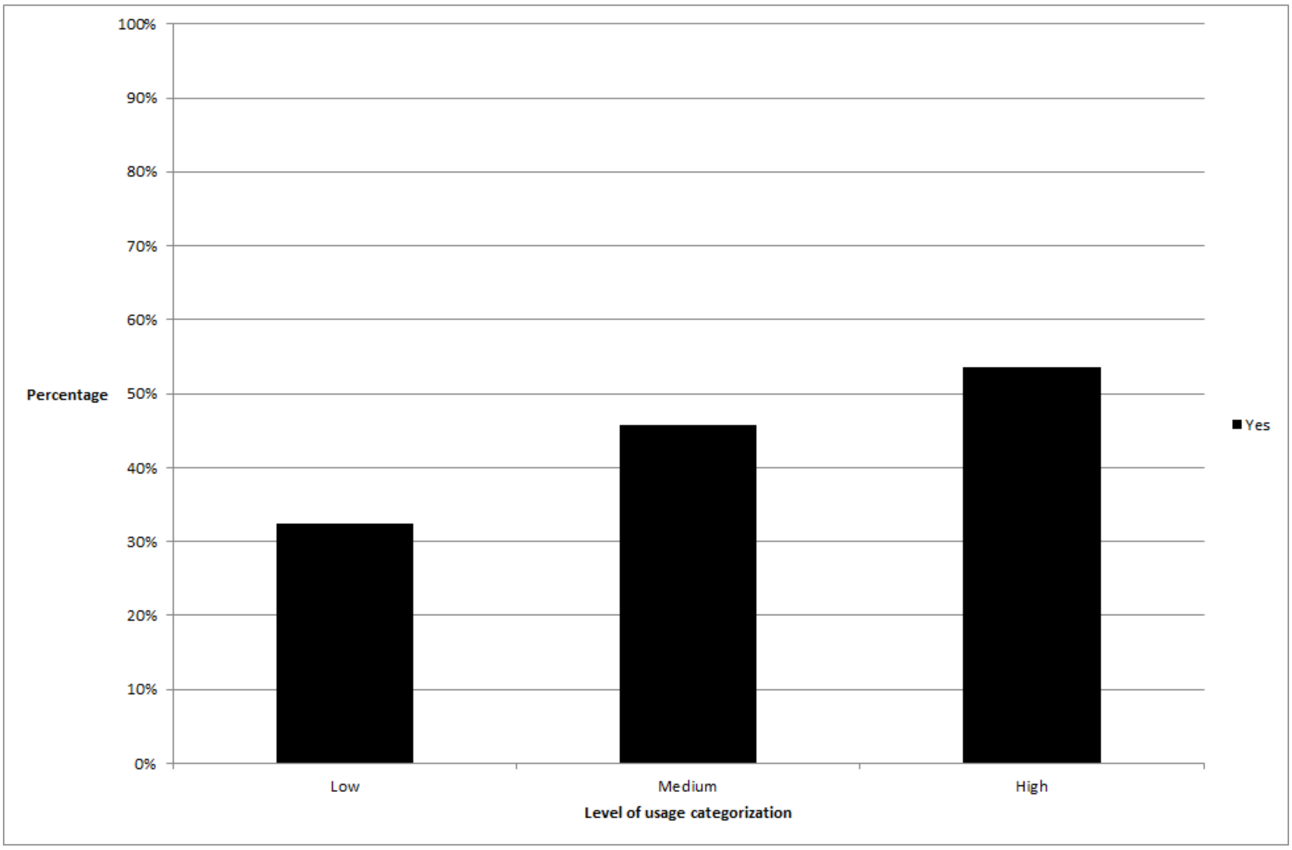


Figure 2. The difference in percentage of participants achieving clinically significant change across usage groups relating to time spent online per log-in.

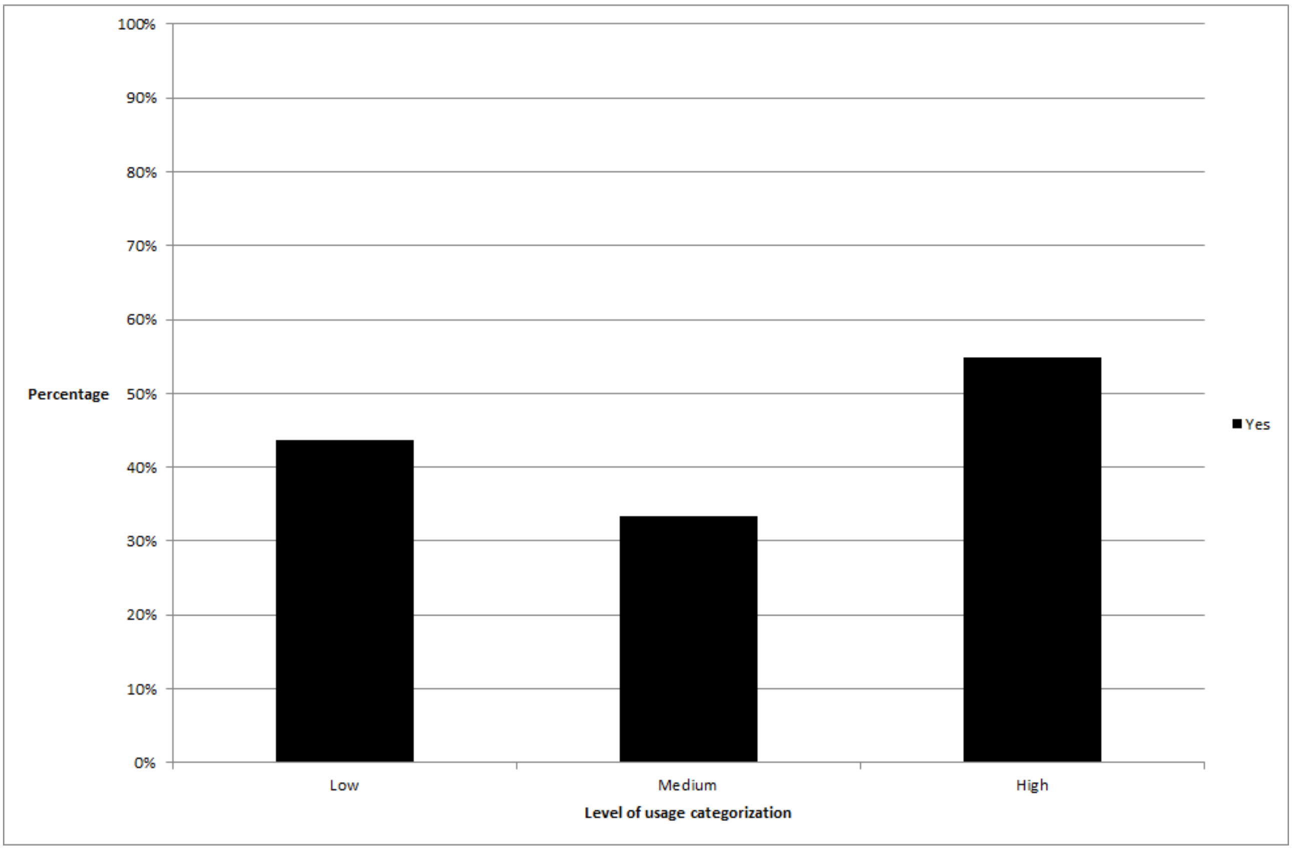
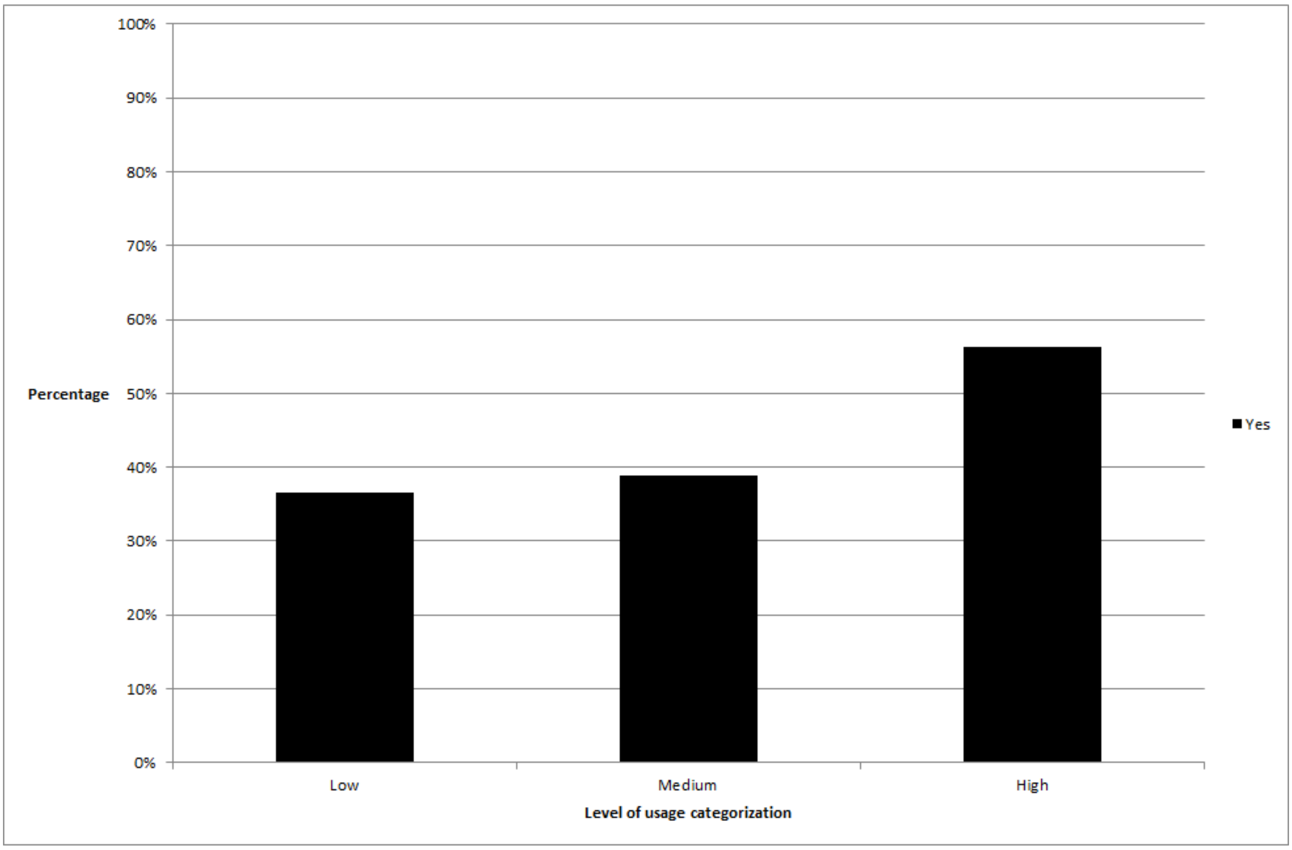


Figure 3. The difference in percentage of participants achieving clinically significant change across usage groups relating to activities completed per log-in.



Sensitivity Analysis

A sensitivity analysis was completed using the continuous variable of PHQ-9 change score. This allowed for slightly increased power and the inclusion of more variables into the analysis. However, when a regression analysis was completed, activities completed per log-in remained the only significant predictor of outcome. Curve estimation did not meet significance for activities completed per log-in (logarithmic curve estimation: *P*=.06; quadratic curve estimation: *P*=.09) and time spent online per log-in (logarithmic: *P*=.09). Similar to clinically significant change, tertile splits found a significant difference between high and low/medium level users for time spent online per log-in and for activities completed per log-in.

Discussion

Principal Findings

Current schools of thought regarding the role of usage in Web-based interventions tend to draw on literature from other treatment paradigms with the assumption that benefit obtained from the program is proportionate to the level of program use. This study found relationships between only a few objectively measured usage variables and clinically significant improvement in participants who persisted with the iCBT program. Of the 4 usage variables that were included in the linear regression model, only the number of activities completed per log-in independently predicted outcome in the final regression model. Further, the significance of nonlinear models for several of the usage variables indicates that the relationship between use and outcome may not be as simple as a linear relationship. Instead, it supports a view that the benefits of use may occur after following a high level of activity during each engagement with the program, not necessarily as a result of ongoing longer-term engagement with the program, and that the number of modules completed in the program is a poor indicator of benefit obtained. Likewise, the analysis of levels of usage indicates that although high program users generally do better, medium users do not necessarily benefit more than low users.

Across analyses, it appears that those participants who were more actively engaged in the program (completing more activities each time they logged in and spending more time in the program with each log-in) were most likely to benefit from the program. These findings are not dissimilar to face-to-face CBT, in which it is the ongoing completion of homework activities across therapy sessions that best predicts outcome [35]. Thus, this indicates that users who are more actively engaged in their treatment may do better.

The activities completed per log-in metric accounted for very little of the variance in the outcome. The inability of some other metrics to predict outcome may reflect limited variability, particularly concerning modules completed, a metric that has been associated with outcome previously [16]. Within this study, the level of module completion (all modules completed by 62.1% of participants) and number of log-ins (a mean of 18 when 12 would have been required to complete the program) indicates a relatively high usage. The other metrics of usage showed greater variability and were more likely to be associated with outcome, with the variable with the greatest variance (activities completed per log-in) being the most strongly related to outcome.

As expected, participants who did not persist with the trial provided outcome data with lower levels of usage. This is likely to reflect early dropout and not being exposed to the content; therefore, they were unable to be more adherent to the program overall. People who were older appeared to adhere more, consistent with other studies [36]. However, the longer time spent online in the older age groups may not actually be a good measure of use because it may reflect less familiarity with using the computer or a slower cognitive processing speed rather than indicating something that may influence outcome. It was also found that men completed more modules than women did, somewhat contrary to previous studies [36,37], although a recent systematic review failed to consistently find a relationship between demographic variables and usage [37] indicating that other factors, such as patient beliefs and personal motivations, may also influence usage [38]. Despite sex and age being found to be associated with the number of activities completed, no interaction was found between these demographics, usage, and outcome. Likewise, no interaction was found between baseline depression severity, number of modules completed, and outcome. Therefore, this supports the view that program usage has a greater impact on outcome than demographics.

The lack of any strong predictive relationship between the usage metrics identified here and outcome may challenge the traditional view of a dose-response relationship relating to outcome. The high rates of usage in this study may have meant that many participants reached a dose-response plateau where they had been exposed to an appropriate level of the program and were unlikely to obtain further benefits from additional exposure. This model indicates that patients may reach therapy saturation at certain levels of use and would likely obtain the effects of the program early on. If this were the case, we would expect that outcome gains would be obtained with medium usage and then be maintained as patients persisted with the intervention. Such an effect has been seen in longer Web-based interventions [15]; for example, Christensen et al [39] found no further improvement in symptoms between 4 and 5 modules. Conversely, of those usage metrics that were associated with outcome here, medium users appeared to derive minimal if any benefit compared to low users and it was the high users that benefited, implying a difference between modules delivered and adherence.

The high rates of use may reflect this analysis only selecting participants that persisted with the study, whereas other usage-outcome association studies have utilized the last observation carried forward (LOCF) technique. Although retaining only those who had persisted may have biased the analysis to more adherent people, using the LOCF technique conflates the measurement of persistence (the number of people who complete the program) which, in turn, leads to these participants also appearing to have poor usage (because only a small proportion of the program completed). Additionally, LOCF may also underestimate the overall effectiveness of the intervention because some studies have reported that people who notice an improvement in their symptoms drop out [37,40], but a LOCF approach would assume no improvement. Additionally, the analysis could have included more complex and potentially more accurate methods for handling missing data, such as multiple imputations or mixed models with maximum likelihood estimation. Given that missing outcome data was likely to occur in cases with low usage rates (because of not completing the program and not providing outcome data), and that the aim of the analysis was to explore this relationship, complete case analysis was preferred.

The univariate associations between usage and outcome found in this analysis are consistent with our recent review which indicated that, of the online intervention studies which reported usage, most (31/33 studies) found a positive relationship between usage variables (34/37 variables) and outcome [16]. However, when further analysis is completed, such as within the present paper, the ability of these variables to predict improvement in the form of clinically significant change is limited. As such, these findings coupled with the curve estimations in this analysis may challenge the perception of the linearity of the relationship between usage and outcome. This implies that it is not the exposure to the material alone that improves outcome, as evidenced by the lack of association between basic usage metrics such as modules completed, but the gradual exposure to and active engagement with the material over time, as evidenced by the strong relationship of composite variables. Given this, we can conclude that concentrated use of the program (eg, completing multiple modules per log-in) or passive exposure to material (as measured by modules completed) may not be as useful as regular shorter periods of use with higher levels of activity in each of these log-ins.

Future Direction

A number of recommendations based on these findings can be made. The finding that those who completed a high number of activities per log-in achieved a greater benefit than those who undertook few, and that a medium activity per log-in count conferred no more benefit than a low activity per log-in count indicates that maximizing usage behavior online may improve outcomes. One way of doing this is ensuring that more activities are included with each module, thereby encouraging users to be more engaged with their treatment. These activities could include the use of activities related to the therapeutic modality of the program or multiple choice quizzes to assess learning with reference back to sections containing material related to incorrect answers.

Programs that limit program exposure at each log-in to allow adequate time for learning, the completion of activities, and skills implementation to occur may also be beneficial as well as incorporating a “hook” to encourage users to return the following week [38]. This directive and potentially restrictive nature of the intervention needs to be balanced with user perception of freedom within the program to encourage ongoing engagement [41].

Providing education and setting early expectations about what users need to do to achieve benefit from the program (ie, being more active while online) may be a helpful approach to improving program outcomes. This would include emphasizing that users are more likely to obtain benefit if they are more active and complete activities when they become available.

Likewise, encouraging users to complete the activities on multiple occasions, particularly when waiting for the next module to become available, may also improve outcomes. However, definitive conclusion about the usefulness of these strategies is beyond the scope of this analysis and would benefit from further research.

A further opportunity is to measure usage and program benefits throughout the course of the intervention to determine at what point users reach their therapy saturation and obtain little if any program gains after this time point. Conversely, such monitoring may indicate that a certain level of usage is required to obtain a benefit. Ensuring users are actively engaging in the program is likely to require frequent monitoring, which in itself can influence the outcome of the intervention, and the use of measures with good test-retest properties. Future programs utilizing this design will also need to consider the burden of intensive monitoring on users and the potential for this to increase the propensity to drop out from the study. Electronic measurement automation may provide a way to reduce monitoring burden.

Previous research has found that although usage of program components was related to early improvements, it was the completion of homework exercises that was correlated with long-term improvements [42]. Given this, developing and using a measure designed to capture real world implementation of online learning and completion of offline homework activities may be key in better understanding how program use may impact outcome. This may be as simple as asking users if they have completed their homework tasks or providing details of how they have implemented the previous module's learnings, much like feedback occurs in current psychotherapy. However, reporting is likely to be prone to self-report bias and may only provide a crude estimate until more sophisticated tools are developed. Standardizing the assessment of usage across trials and programs would be of huge benefit in understanding these processes [21] and suggests a role for multinational nongovernment organizations and developers groups, such as the International Society for Research on Internet Interventions, in this process. However, until a consensus about the best way to measure and define usage is reached, it is difficult to implement standardization throughout trials.

Limitations

As mentioned previously, the inclusion of only those participants who provided outcome data is likely to have limited the generalizability of the results. However, it is unclear what effect that this may have had on the analysis and how to best manage this. Using the LOCF may actually fail to give an accurate picture of the progress of these participants. This is largely because the LOCF method assumes no progress in this group. Although research has indicated that obtaining benefits may be key in helping people to persist with an intervention [38], it may also contribute to users ceasing to use the intervention because they believe that they no longer require assistance. Given this, generalizing assumptions about usage to those who fail to persist with interventions should be done with caution. Additionally, usage research may benefit from using more sophisticated analytic approaches (eg, latent class modeling [43,44] or gaussian mixture models [45]) to assess whether there are groups who are more likely to respond and if these groups differ in usage. Growth modeling of outcomes and usage would be better able to analyze the usage-outcome association.

In addition, the specific inclusion criteria of this trial and the unique nature of each intervention likely limits the generalizability of the results of this analysis to other populations. However, recent findings have indicated that this may be less of a factor than initially thought, and that the sampling bias related to trials may not actually limit the ability of trial results to be generalized based on demographic factors [46].

Finally, despite that the data were drawn from a RCT, this substudy was observational in nature and no manipulation of variables related to usage occurred. Given this, the ability to imply causation is limited. Future research needs to explore the manipulation of these variables, such as controlling number of log-ins to determine if unlimited access affects outcomes, exploring usage and outcome in single log-in sites, or limiting the amount of activities or modules that can be completed per log-in to future test the hypotheses drawn from the findings of this study.

Conclusions

Future research would benefit from exploration of the relationship between usage metrics and outcome to further investigate the nature of this relationship. Although this analysis found only 1 metric was predictive of outcome, this finding is limited by the context of this study. Future research needs to continue to explore this research in trials and naturalistic implementation of Web-based interventions to determine if this is the case.

Acknowledgments

This CREDO research trial is supported by the Cardiovascular Disease and Depression Strategic Research Program (Award Reference No. G08S 4048) funded by the National Heart Foundation of Australia and beyondblue: the national depression initiative. The 45 and Up Study is managed by The Sax Institute in collaboration with major partner Cancer Council New South Wales; and partners the National Heart Foundation of Australia (NSW Division); NSW Health; beyondblue: the national depression initiative; Department of Family and Community Services (FACS) NSW; and UnitingCare Ageing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot showing an example of psychoeducation about the link between thoughts and moods.

[PNG File, 248KB - jmir_v15i10e231_app1.png]

Multimedia Appendix 2

Screenshot showing psychoeducation teaching the user about cognitive distortions.

[PNG File, 111KB - jmir_v15i10e231_app2.png]

Multimedia Appendix 3

Screenshot showing an example of an activity that teaches users about cognitive restructuring.

[PNG File, 118KB - jmir_v15i10e231_app3.png]

Multimedia Appendix 4

Screenshot showing an example of an activity from the interpersonal therapy component of E-couch.

[PNG File, 137KB - jmir_v15i10e231_app4.png]

References

1. Gainsbury S, Blaszczynski A. A systematic review of Internet-based therapy for the treatment of addictions. Clin Psychol Rev 2011 Apr;31(3):490-498. [doi: 10.1016/j.cpr.2010.11.007] [Medline: 21146272]
2. 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]
3. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
4. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
5. McCabe M, Price E. Internet-based psychological and oral medical treatment compared to psychological treatment alone for ED. J Sex Med 2008 Oct;5(10):2338-2346. [doi: 10.1111/j.1743-6109.2008.00885.x] [Medline: 18554261]
6. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
7. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
8. Couper M, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
9. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008 Apr;76(2):341-346. [doi: 10.1037/0022-006X.76.2.341] [Medline: 18377129]
10. 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]
11. Carrard I, Fernandez-Aranda F, Lam T, Nevonen L, Liwowsky I, Volkart AC, et al. Evaluation of a guided internet self-treatment programme for bulimia nervosa in several European countries. Eur Eat Disord Rev 2010 Sep 21:138. [doi: 10.1002/erv.1043] [Medline: 20859989]
12. Troop N, Schmidt U, Tiller J, Todd G, Keilen M, Treasure J. Compliance with a self-care manual for bulimia nervosa: predictors and outcome. Br J Clin Psychol 1996 Sep;35 (Pt 3):435-438. [Medline: 8889084]
13. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
14. An LC, Perry CL, Lein EB, Klatt C, Farley DM, Bliss RL, et al. Strategies for increasing adherence to an online smoking cessation intervention for college students. Nicotine Tob Res 2006 Dec;8 Suppl 1:S7-12. [Medline: 17491165]

15. Wantland D, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
16. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
17. 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]
18. Wagner B, Brand J, Schulz W, Knaevelsrud C. Online working alliance predicts treatment outcome for posttraumatic stress symptoms in Arab war-traumatized patients. Depress Anxiety 2012 Jul;29(7):646-651. [doi: 10.1002/da.21962] [Medline: 22678971]
19. Blankers M, Koeter MW, Schippers GM. Baseline predictors of treatment outcome in Internet-based alcohol interventions: a recursive partitioning analysis alongside a randomized trial. BMC Public Health 2013;13:455 [FREE Full text] [doi: 10.1186/1471-2458-13-455] [Medline: 23651767]
20. Danaher B, 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]
21. Kelders S, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
22. 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]
23. 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;40(2):82-97. [doi: 10.1080/16506073.2011.573807]
24. Cockayne NL, Glozier N, Naismith SL, Christensen H, Neal B, Hickie IB. Internet-based treatment for older adults with depression and co-morbid cardiovascular disease: protocol for a randomised, double-blind, placebo controlled trial. BMC Psychiatry 2011;11:10 [FREE Full text] [doi: 10.1186/1471-244X-11-10] [Medline: 21232162]
25. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
26. 45Up Study Collaborators, Banks E, Redman S, Jorm L, Armstrong B, Bauman A, et al. Cohort profile: the 45 and up study. Int J Epidemiol 2008 Oct;37(5):941-947 [FREE Full text] [doi: 10.1093/ije/dym184] [Medline: 17881411]
27. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health 2001 Dec;25(6):494-497. [Medline: 11824981]
28. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [Medline: 12214795]
29. Mackinnon A, Griffiths KM, Christensen H. Comparative randomised trial of online cognitive-behavioural therapy and an information website for depression: 12-month outcomes. Br J Psychiatry 2008 Feb;192(2):130-134 [FREE Full text] [doi: 10.1192/bjp.bp.106.032078] [Medline: 18245031]
30. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
31. Griffiths K, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
32. E-couch. URL: https://ecouch.anu.edu.au/welcome [accessed 2013-10-05] [WebCite Cache ID 6K9Gum7wd]
33. 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]
34. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [Medline: 15550799]
35. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: an updated meta-analysis. Cognit Ther Res 2010 Oct;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]
36. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
37. Neil A, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]

38. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
39. 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]
40. Postel M, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
41. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
42. de Graaf LE, Huibers MJ, Riper H, Gerhards SA, Arntz A. Use and acceptability of unsupported online computerized cognitive behavioral therapy for depression and associations with clinical outcome. J Affect Disord 2009 Aug;116(3):227-231. [doi: 10.1016/j.jad.2008.12.009] [Medline: 19167094]
43. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: A review. Stat Methods Med Res 2012 Apr 19:epub (forthcoming). [doi: 10.1177/0962280212445839] [Medline: 22517270]
44. Spycher BD, Minder CE, Kuehni CE. Multivariate modelling of responses to conditional items: New possibilities for latent class analysis. Stat Med 2009 Jun 30;28(14):1927-1939. [doi: 10.1002/sim.3550] [Medline: 19205074]
45. Shiyko MP, Ram N, Grimm KJ. Growth mixture models. In: Hoyle RH, editor. Handbook of Structural Equation Modeling. New York: The Guilford Press; 2012.
46. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Sampling bias in an internet treatment trial for depression. Transl Psychiatry 2012;2:e174 [FREE Full text] [doi: 10.1038/tp.2012.100] [Medline: 23092978]

Abbreviations

CBT: cognitive behavior therapy
CREDO: Cardiovascular Risk E-couch Depression Outcome
CVD: cardiovascular disease
iCBT: Internet cognitive behavior therapy
IPT: interpersonal psychotherapy
LOCF: last observation carried forward
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 13.06.13; peer-reviewed by S Kelders, R Crutzen; comments to author 01.07.13; revised version received 30.09.13; accepted 02.10.13; published 17.10.13.

Please cite as:
Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, Glozier N
Rethinking the Dose-Response Relationship Between Usage and Outcome in an Online Intervention for Depression: Randomized Controlled Trial
J Med Internet Res 2013;15(10):e231
URL: http://www.jmir.org/2013/10/e231/
doi:10.2196/jmir.2771
PMID:24135213

Review

Assessing the Internal and External Validity of Mobile Health Physical Activity Promotion Interventions: A Systematic Literature Review Using the RE-AIM Framework

Kacie CA Blackman[1], PhD; Jamie Zoellner[1,2], PhD, RD; Leanna M Berrey[1], BS; Ramine Alexander[1], BS; Jason Fanning[3], MS; Jennie L Hill[1,2], PhD; Paul A Estabrooks[1,2,4], PhD

[1]Department of Human Nutrition, Foods, and Exercise, Virginia Tech, Blacksburg, VA, United States
[2]Fralin Translational Obesity Research Center, Virginia Tech, Blacksburg, VA, United States
[3]Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[4]Department of Family Medicine, Virginia Tech Carilion Medical School, Roanoke, VA, United States

Corresponding Author:
Kacie CA Blackman, PhD
Department of Human Nutrition, Foods, and Exercise
Virginia Tech
Integrated Life Sciences Building
1981 Kraft Drive
Blacksburg, VA, 24060
United States
Phone: 1 323 442 ext 7250
Fax: 1 323 442 8201
Email: kaciea1@vt.edu

Abstract

Background: Mobile health (mHealth) interventions are effective in promoting physical activity (PA); however, the degree to which external validity indicators are reported is unclear.

Objective: The purpose of this systematic review was to use the RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) framework to determine the extent to which mHealth intervention research for promoting PA reports on factors that inform generalizability across settings and populations and to provide recommendations for investigators planning to conduct this type of research.

Methods: Twenty articles reflecting 15 trials published between 2000 and 2012 were identified through a systematic review process (ie, queries of three online databases and reference lists of eligible articles) and met inclusion criteria (ie, implementation of mobile technologies, target physical activity, and provide original data). Two researchers coded each article using a validated RE-AIM data extraction tool (reach, efficacy/effectiveness, adoption, implementation, maintenance). Two members of the study team independently abstracted information from each article (inter-rater reliability >90%) and group meetings were used to gain consensus on discrepancies.

Results: The majority of studies were randomized controlled trials (n=14). The average reporting across RE-AIM indicators varied by dimension (reach=53.3%, 2.67/5; effectiveness/efficacy=60.0%, 2.4/4; adoption=11.1%, 0.7/6; implementation=24.4%, 0.7/3; maintenance=0%, 0/3). While most studies described changes in the primary outcome (effectiveness), few addressed the representativeness of participants (reach) or settings (adoption) and few reported on issues related to maintenance and degree of implementation fidelity.

Conclusions: This review suggests that more focus is needed on research designs that highlight and report on both internal and external validity indicators. Specific recommendations are provided to encourage future mHealth interventionists and investigators to report on representativeness, settings, delivery agents for planned interventions, the extent to which protocol is delivered as intended, and maintenance of effects at the individual or organizational level.

(J Med Internet Res 2013;15(10):e224) doi:10.2196/jmir.2745

KEYWORDS

physical activity; mobile technology; review; generalizability

Introduction

The numerous health benefits of physical activity (PA) are well known, but still it is estimated that roughly 31% of the world's adult population (28% men, 34% women) is classified as insufficiently active [1]. Likewise, it is a concern in the United States where only 6-11% of children [2] and 8.2% of adults meet the national PA guidelines based on objective PA assessments [3]. Given these low PA rates, there is a need for increased attention to the development of effective and scalable PA promotion interventions that can reach a large number of people at a low cost [4].

One such approach is the use of mobile technology, since ownership is on the rise in adults and children [5,6]. By 2012, it was estimated that there were 7 billion mobile-connected devices across the globe and the number of mobile devices outnumbered the human population [7]. In the United States, according to a 2012 nationally representative survey, more than 88% of American adults own mobile phones, which is an 11% increase from 2011 [8]. Fifty-three percent of American mobile phone owners own a smartphone [8]. Furthermore, roughly 75% of 12-17 year olds own mobile phones and this is a drastic surge (ie, up 30%) from 2004 [6].

This growth in mobile technology ownership has led to the development of a number of mobile health (mHealth) intervention reviews [9-12]. Specifically, related to PA, mHealth interventions that deliver information and behavioral strategies through short message service (SMS) via mobile phones have been developed to increase PA [13-17]. In addition, ecological momentary interventions through palmtop computers and mobile phones [14] can enhance interventions and aid in improving health outcomes. The potential utility of mHealth interventions to promote PA is also evident in the large number of commercially available fitness applications that promote behavioral tracking (eg, Nike+Running, Runtastic), link to external technology devices (eg, Fitbit), or directly encourage different intensities of PA (eg, Zombies, Run!). A recent meta-analysis on the use of mobile devices [10] and text messaging review [9] for PA promotion summarized the literature in this area and concluded that interventions delivered through this modality were effective for increasing PA. Similarly, a review of Internet-based PA interventions concluded that interactive technology interventions were effective for PA promotion [18]. However, a recent Cochrane review of mHealth interventions for preventive health care suggested that the availability of studies using randomized controlled trials was insufficient to determine if these approaches could influence PA or other health behaviors [12].

Despite the popularity of commercially available health-related applications, there is little evidence that mobile phone-based interventions with demonstrated efficacy have been translated beyond the research setting and been broadly adopted [19]. Some potential reasons for the lack of translation of these interventions into more widespread use are that the scientific approach typically emphasizes high internal validity at the expense of external validity [20] and that the traditional research pace impedes the flow of disseminating relevant findings [21]. To date, reviews of mHealth interventions have evaluated the quality of studies through the lens of internal validity and emphasized improved reporting on potential confounding factors [22]. As a result, the conclusions are largely limited to factors related to intervention efficacy and the extent to which these mHealth interventions report on or achieve external validity to different settings and populations is unclear [13-17,23]. This issue was recently underscored by the publication of the CONSORT-EHEALTH reporting standards [24]. The standards included eight highly recommended and four essential categories of reporting, which highlight the need for additional attention to external validity. Briefly, the four essential categories include (1) reporting on the context within which participants accessed the intervention, (2) the delivery mode, features, and functionality of the intervention, (3) the use of prompts to interact with the intervention, and (4) any co-interventions that may occur.

To improve the reporting across behavioral interventions, Glasgow and colleagues developed the RE-AIM (reach, effectiveness, adoption, implementation, maintenance) framework to evaluate the degree to which behavioral interventions, including those targeting PA, report on internal and external validity factors [25]. The framework specifies standards related to the reporting of "Reach" into the target population and representativeness of the study sample; "Efficacy/effectiveness" of the intervention on the primary outcome tested under either optimal or real-world conditions, quality of life, and avoidance of unintended or negative consequences; "Adoption" rates of organizations and staff that would ultimately use the intervention and the characteristics of those organizations and staff; the degree to which the intervention is "Implemented" as intended; and the "Maintenance" of effects at the individual level and sustainability of the intervention at an organizational or delivery level (RE-AIM) [26]. The RE-AIM framework has demonstrated utility in summarizing reports of internal and external validity factors across numerous bodies of literature (eg, weight loss maintenance, health literacy, tobacco use, and PA interventions for older adults and for breast cancer survivors) [27-37]. Collectively, these previous reviews have provided recommendations and future directions to enhance the likelihood of research to practice. Many of these recommendations align with those proposed in the CONSORT-EHEALTH standards [24]. In particular, the context within which participants access mHealth interventions is documented within the adoption (ie, description of intervention location and staff) and each of the other three essential standards are captured within an assessment of the implementation dimension (ie, cost, intervention description including frequency, type, and duration of contacts). The primary purpose of this systematic review is to determine the degree to which studies testing mHealth interventions to promote PA report on factors that inform generalizability across settings and populations. Recommendations to improve the likelihood of broad dissemination of effective mHealth interventions are also provided based on the literature [24,27-38].

Methods

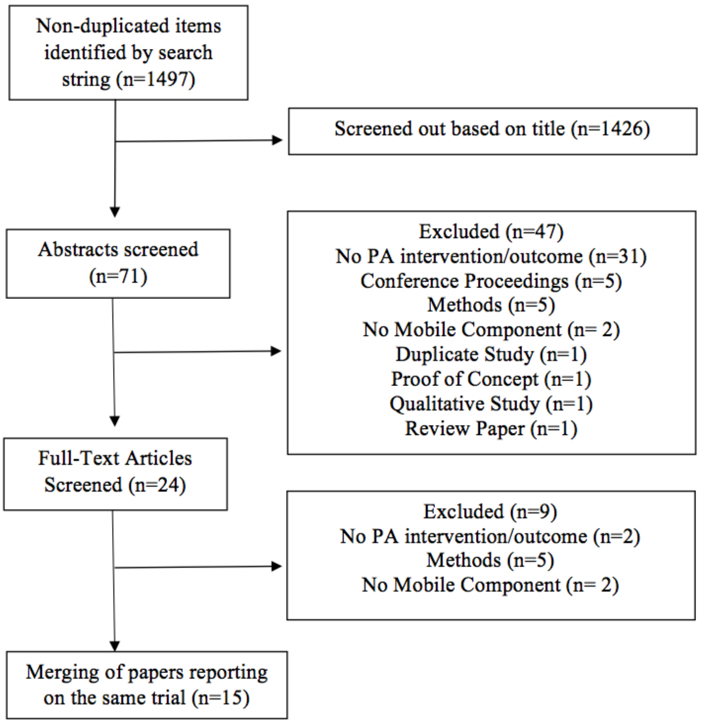
Selection of Studies for Review

We replicated the search strategy used in a recently published meta-analysis publication that focused solely on effectiveness of mHealth interventions for PA promotion at the individual level [10]. Our literature search was conducted between August 2011 and July 2012 and included articles published between 2000 and 2012 that met the inclusion criteria indicated in Table 1. Review articles, observational (eg, cross-sectional, descriptive) commentaries, methodological articles, and articles not explicitly related to PA were excluded. Implementation of mobile technologies included data collection or conveyance of intervention information via SMS or native mobile device software or hardware. The search strategies to identify eligible articles included queries using three online databases (PsycINFO, PubMed, and Scopus) and a hand search of reference lists for articles that met inclusion criteria. The search terms included mobile phone, cell phone, PDA, SMS, or text messaging combined with PA or exercise [10]. In addition to comprehensively evaluating the reporting of RE-AIM criteria on a single trial, data was extracted from companion articles (eg, qualitative/quantitative methods measuring implementation) of studies that met inclusion criteria. Figure 1 outlines the identification of the 20 articles representing 15 trials that were included in this systematic review.

Table 1. Inclusion criteria.

Data type	Inclusion criteria
Participants	Any age
Language	English
Study design	Experimental and quasi-experimental
Control condition	Any comparator including active control, inactive control, or participants as their own control (ie, pre- and post-measures)
Intervention	Implementation of mobile technologies
Measurement	Assesses physical activity directly among participants
Primary outcome	Physical activity
Type of data	Original, quantitative outcome data

Figure 1. Flow diagram of study selection.



Assessment of Reporting Comprehensiveness Across RE-AIM Dimensions

Comprehensiveness of reporting was determined using a previously developed 21-item validated data extraction tool that included both internal and external validity indicators based on the RE-AIM framework [27,28,31-33]. Table 2 includes details on each of the indicators assessed across the RE-AIM framework. In addition, we examined the degree to which reporting across the implementation dimension of RE-AIM addressed essential CONSORT-EHEALTH standards in terms of application costs, intervention features, theoretical backgrounds, prompts, and co-interventions [24]. Finally, due to the emerging nature of this body of literature, we also documented whether studies reported on adaptations that were made across intervention testing [34]. This was useful in order to determine the extent to which intervention fidelity was maintained during intervention implementation while allowing adaptations to the intervention to be made by delivery agents/organizations/systems to improve feasibility/acceptability/utility [39].

Coding Protocol and Scoring

All studies were coded independently by two members of the research team with the exception of the first three studies which were coded by five members of the research team to promote familiarity with the data extraction tool. For each of the 21 items, coders indicated whether or not the indicator was reported (ie, yes or no), and subsequently extracted specific data. After independently coding, the Kappa statistic [41] was calculated to evaluate inter-rater reliability. The average Kappa statistic for consistency of coding was 0.90, indicating strong inter-rater reliability. For the differences that did arise, researchers met to discuss articles, resolve uncertainty, and gain consensus in the coding by revisiting the specific article.

To calculate the proportion reporting for each item, the number of “yes” codes was summed across the 15 studies and then divided by 15. Then the resulting number became the proportion reporting for that particular item. An overall comprehensiveness of reporting score for each article was calculated based on the number of reported indicators (possible score 0-21). Comprehensiveness of reporting score categories have been published in a past RE-AIM review [28], with articles scoring 15-21, 8-14, and less than 8, considered as high, moderate, and low quality reporting, respectively.

Table 2. RE-AIM internal and external validity indicators.

RE-AIM dimension	Indicator	Description	Importance
Reach			
	Individual level	The number, proportion, and representativeness of participants.	
	Method to identify target population	Describe the process by which the target population was identified for participation in the intervention.	Helps investigators develop an approach to determining who may be suitable for the intervention. Examples include using an electronic medical record query or mass media approaches [20].
	Inclusion criteria	Explicit statement of characteristics of the target population that were used to determine if a potential participant was eligible to participate.	Inclusion criteria should be as inclusive as possible to improve the external validity of findings [40].
	Exclusion criteria	Explicit statement of characteristics that would prevent a potential participant from being eligible to participate.	Exclusion criteria should be considered carefully to prevent potential harm to prospective participants, but should also avoid excluding individuals based on criteria that could be related to SES (eg, ability to travel to intervention site), comorbidities, or other factors that could influence an externally valid depiction of intervention effects [40].
	Participation rate	Sample size divided by the target population denominator.	Provides information on the acceptability of the study and interventions from the perspective of the target population [26].
	Representativeness	Explicit statement of characteristics of the study participants in comparison to the target population.	Identifies disparities in participation and informs the degree to which the study results are generalizable to the target population [26].
Efficacy/effectiveness			
	Individual level	The measure of the primary outcome, quality of life, and on avoiding unintended negative consequences.	
	Measures/results for at least 1 follow-up	The study variable(s) are measured at a time point after baseline.	To evaluate whether the intervention outcomes were statistically significant or changed (positively/negatively) [26].
	Intent-to-treat analysis utilized	Analyzing participants in trials in the groups to which they were randomized, regardless of whether they received or adhered to the allocated intervention.	Reduces bias from omitting individuals who were lost to follow-up and improves generalizability [63].
	Quality-of-life (QOL) or potential negative outcomes	QOL: Includes a measure of quality of life with some latitude for coding articles that refer to well-being or satisfaction with life. Negative outcomes: To evaluate unanticipated consequences and results that may be a product of the intervention and may have caused unintended harm.	Provide a metric to compare across interventions with different behavioral targets and provides a better sense of the impact that the intervention on the participants' perceptions of health [26]. Allows for the weight of the harms and benefits of an intervention [26].
	Percent attrition	The proportion that was lost to follow-up or dropped out of the intervention.	High attrition lowers statistical power and treatment-correlated attrition of participants from conditions threatens internal validity [42].
Adoption			
	Organizational level (setting and staff)	The number, proportion, and characteristics of adopting organizations and staff.	
	Description of intervention location	The explicit statement of characteristics of the location of the intervention.	Provides an understanding of resources needed for future researchers [26].
	Description of staff who delivered intervention	The explicit statement of characteristics of the staff who delivered the intervention.	Provides information on the characteristics may be needed to deliver an intervention and assist with retention of participants [35].

RE-AIM dimension	Indicator	Description	Importance
	Method to identify staff who delivered intervention (target delivery agent)	Describe the process by which the staff was identified for participation in the study.	Helps investigators develop an approach to identify and engage staff that may be suitable for intervention delivery [35].
	Level of expertise of delivery agent	Training or educational background in of those delivering the intervention.	Allows for the assessment of generalizability of those delivering an intervention to typical practice settings delivery [35].
	Inclusion/exclusion criteria of delivery agent or setting	The explicit statement of characteristics of the setting/agent that were used to determine if a potential setting/agent is eligible to participate.	Inclusion criteria should be as inclusive as possible to improve the external validity of findings. Exclusion criteria should not systematically remove potential settings or staff that typical in the practice domain [20].
	Adoption rate of delivery agent or setting	The number of participating delivery settings or agents divided by the number of eligible and approached delivery settings or agents.	Provides information on the acceptability of the study and interventions from the perspective of the setting and staff that will ultimately be responsible for intervention delivery [26].
Implementation			
	Organizational level	The degree to which the intervention is delivered as intended.	
	Intervention duration and frequency	Duration: length the intervention over days, weeks, and months as well as the length of each intervention contact. Frequency: number of contacts with participants	Useful for replication and comparison of resources needed to resources available in a practice setting [26].
	Extent protocol delivered as intended (%)	Description of fidelity to the intervention protocol.	This provides insight into the feasibility of delivering all components of an intervention at the pre-determined date and time [26].
	Measures of cost of implementation	The ongoing cost (eg, money, time) of delivery across all levels of the intervention.	This is helpful for future researchers to be able to determine if conducting a specific intervention has economically feasible delivery [35].
Maintenance			
	Individual and organization level	The measure of behavior at the individual level and sustainability of the intervention at an organizational level.	
	Assessed outcomes ≥ 6 months post intervention	Description of follow-up outcome measures of individuals available at some duration after intervention termination.	Provides information on the maintenance of intervention outcomes over time [26].
	Indicators of program level maintenance	Description of program continuation after completion of the research study.	Provides information on whether the intervention can be integrated into an existing system/organization [26].
	Measures of cost of maintenance	The ongoing cost of maintaining delivery across all levels of the intervention.	Sustainability costs provides information for practice settings to determine the resources needed for long-term intervention delivery [28].

Results

Study Characteristics

All trials were published after 2006 and 13 were conducted in Western countries. Six studies were conducted in the United States [43-52], three trials were conducted in the United Kingdom [53-55], three trials were conducted in Australia [56-59], and one each was conducted in Hong Kong [60], New Zealand [61], and Taiwan [62]. The majority of studies were randomized controlled trials and one was a quasi-experimental trial without a control group [47-49]. Most studies intervened at the individual level, two intervened at the group level [43-45,58,59], and one targeted both levels [52]. The length of the studies ranged from 2 weeks to 2 years, with an average of 19 weeks. The most commonly reported intervention length was 12 weeks.

Five studies measured PA only through self-report [43-45,50,52,56,57], four used objective measures [46,52,55,62], and three used both self-report and objective measures [47-49,53,61]. Of the seven studies that objectively measured PA, half of the studies used a pedometer [47-49,52,60,61]. Each of the following objective PA measures were collected once: both biaxial and triaxial accelerometers [58,59], uniaxial accelerometers [53], biaxial accelerometers [51], the walking distance of the incremental shuttle walking test [62], and a mobile sensing platform [46].

In addition to PA, the majority of studies (n=11) reported on other outcomes. Eight studies reported on body mass index (BMI) [43-49,53,54,58-60,62]; two on BMI-z scores [58,59,61]; five studies reported on physiological outcomes [43-45,51,53,61,62]; four studies on psychological outcomes [47-49,53,58-60]; three studies on weight [43-45,53,54]; two each on sedentary activity/screen time [52,58,59]; diet [43-45,58,59]; and percent body fat [51,60]; and one on each on sugar sweetened beverages intake [50]; upper body muscular endurance and core abdominal isometric muscular endurance [58,59]; waist-to-hip ratio [54]; and waist circumference [43-45].

The types of mobile devices used were similar across studies. Nearly all studies (n=13) used mobile phones while two used personal digital assistants [43-45,53]. Most frequently, mobile technology was implemented as a way to monitor outcomes via self-report [43-45,51-53] or data from an external pedometer/accelerometer was manually entered on the mobile phone [50,53]. Additionally, mobile technology was used to provide prompts [47-49,54,61] to encourage behavior change [55,56] and provide health promotion information sent through SMS [58,59]. Furthermore, in two studies, mobile technology was used as an interactive mobile application [46,57], in one study to deliver an exercise program [62] and in another study as a mobile PA diary [47-49]. Table 3 shows the overall quality of RE-AIM reporting across the 21-item validated extraction tool, which is displayed as the proportion reporting.

Reach

Reach was the second most reported dimension at 53.3% (2.67/5). Approximately half of all studies reported on four of the five items (method used to identify target population, inclusion and exclusion criteria, and participation rate). The least reported component was representativeness, with only four studies reporting [43-45,47-49,56,62]. None of the studies reported on characteristics of dropouts. All trials reported on sample size, which ranged from 17-210 participants with a median of 78. The participation rate ranged from 48-91 with a median of 51. The methods that were used to identify the target population included utilizing existing databases [43-45,54,56], regional diabetes services [61], recruitment coordinators [51], listservs [52], and an invitation letter from stakeholders [55]. In addition to English speaking, the most common inclusion criteria were PA requirement [46,50,53,54,56,58,59], weight requirement [43-45,53], or required possession/access to a mobile phone [46-49,54,56]. The majority of studies described recruitment strategies (n=11), which happened through various ways. Eleven studies focused on child participants, while four focused on adult participants [55,58,59]. Approaches for recruiting children included sending recruitment letters home [52,63], giving a presentation at school assemblies [55], teacher referrals [58,59], and using university listservs [52]. The majority of studies that enrolled adult participants were recruited mainly through local or mass media. Local mass media strategies included distributing flyers [43-45], using voicemail announcement systems [43-45], using mailing lists [43-45], posting ads on city buses [43-45], placing newspaper announcements [47-49], email [54,57], and using local mass media outlets [50,52]. Other studies' recruitment methods comprised obtaining names and contact information from pulmonary rehabilitation coordinators [51], contacting individuals on a weight research registry via letter or email [43-45], targeting previous trial participants [43-45], and using a market research recruitment agency [46,53].

Efficacy/Effectiveness

Efficacy/effectiveness was the most reported dimension at 60.0% (2.4/4). All studies reported on measures or results for at least one follow-up. Approximately three quarters of the studies reported on percent attrition, which ranged from 0-53%. Four studies reported on intent-to-treat analysis [51,55,58-60], six stated present-at-follow-up analyses were used [50,52,54,55,57,62], and one used present-at-follow-up and intention-to-treat analyses [43-45]. Of the two studies that reported a high attrition rate (ie, 25% or higher) [52,56], only one [56] used intent-to-treat analysis.

The majority of studies (n=12) reported whether the trial was an efficacy or effectiveness trial. Of these studies, eight were efficacy trials [47-52,55-57,62] and four were effectiveness trials [46,57-59,62]. A little over 30% of studies reported quality of life or potential negative outcomes and found that mobile PA interventions generally improved quality and did not have any significant negative outcomes [43-45,51,53,61,62]. In terms of PA outcomes for the 14 controlled trials, six studies found that the intervention group had significant differences compared to the control group, four studies had mixed results, and four had nonsignificant differences between groups. In the only quasi-experimental, single group study included in this study, the post-assessments of PA were significantly higher than pre-assessments [47-49]. Only one study assessed cost effectiveness, which indicated that cost per participant associated with a mobile phone-based exercise program was $580 and coaching was added at a low incremental cost of US$80 [51]. Moreover, there were no significant differences in PA outcomes (ie, six-minute walk distance) between these two groups (ie, mobile-coached versus mobile self-monitored) in the study [51].

Adoption

The average proportion reporting on Adoption items was 11% (0.7/6). Level of expertise of delivery agent was the most reported adoption component (n=5). The descriptions of staff level of expertise included a nutritionist [43-45], a master's level prepared exercise physiologist [43-45], a research assistant [56], a behavioral counselor [56], a nurse [51], and a psychologist [52]. No studies reported on method to identify staff who delivered the intervention, description of staff who delivered the intervention, inclusion/exclusion criteria of delivery agent, or adoption rate of delivery agent.

Setting-level reporting was similar to staff-level reporting. Only five studies specified the intervention location: a school [58,59,62], a research center physiologist, and an outpatient setting from four regional adolescent diabetes services [61]. Last, only two studies described the intervention location [58,59,62], two studies noted inclusion/exclusion criteria of setting [58,59,62], and one indicated adoption rate of setting [58,59].

Implementation

The average proportion reporting on Implementation indicators was 24% (0.7/3). Intervention duration and frequency were the most frequently reported items (n=6) [43-45,50-52,57-59]. Few studies reported on measures of cost of implementation (n=3) [51,58,59,62] or the degree to which the intervention protocol was carried out as intended (n=2) [53,58,59]. Though it appeared that no study charged participants for the applications, none explicitly stated this. Further, no study reported on adaptations made to intervention during the study. More than half (n=8) of the studies had a theoretical basis [43-46,56,58,59,62], with social cognitive theory used most frequently (n=3) [50,56,58,59]. Almost all studies (n=13) stated the degree to which participants received prompts, co-interventions, and other intervention components, including methods such as self-monitoring of outcomes through mobile technology (eg, mobile phone or PDA), class attendance, application usage, or the completion of intervention.

Maintenance

Maintenance was the dimension that was reported least among the RE-AIM dimensions, with no items (0%, 0/3) reported. The reporting on indicators of individual-level or program-level maintenance were not reported in any trial.

Comprehensiveness of Reporting on RE-AIM Criteria

The average comprehensiveness of reporting score was 6.9 out of a possible 21-item reporting coding sheet and scores ranged from 3-13. None of the studies were categorized as high reporting quality, six studies were moderate (range 8-11) [43-45,51,52,56,58,59,62], and nine studies were of low reporting quality [46-50,53-55,57,60,61].

Table 3. Proportion of mobile health interventions reporting RE-AIM dimensions and components (n=15).

RE-AIM Dimensions	RE-AIM Components	Proportion Reporting[a], %
Reach		
	Method to identify target population	60.0
	Inclusion criteria	80.0
	Exclusion criteria	60.0
	Participation rate	46.7
	Representativeness	26.7
	Average across Reach Components	53.3
Efficacy/effectiveness		
	Measures/results for at least one follow-up	100.0
	Intent to treat analysis utilized	33.3
	Quality-of-life or potential negative outcomes	33.3
	Percent attrition	73.3
	Average across Efficacy/Effectiveness Components	60.0
Adoption		
	Description of intervention location	13.0
	Description of staff who delivered intervention	0.0
	Method to identify staff who delivered intervention (target delivery agent)	0.0
	Level of expertise of delivery agent	33.3
	Inclusion/exclusion criteria of delivery agent or setting	13.3
	Adoption rate of delivery agent or setting	6.7
	Average across Adoption Components	11.1
Implementation		
	Intervention duration and frequency	40.0
	Extent protocol delivered as intended (%)	13.3
	Measures of cost of implementation	20.0
	Average across Implementation Components	24.4
Maintenance		
	Assessed outcomes ≥ 6 months post intervention	0.0
	Indicators of program level maintenance	0.0
	Measures of cost of maintenance	0.0
	Average across Maintenance Components	0.0

[a] Based on denominator of 15 intervention trials, reported across 20 articles.

Discussion

Principal Findings

Our review highlighted a recent increase in studies conducted to determine the efficacy or effectiveness of mHealth interventions for the promotion of PA. We identified gaps across and within each of the RE-AIM dimensions, potentially as a result of the relative early stages of this area of research. We also understand that there is a need to advance research by utilizing innovative, flexible, and rapid research designs and “rapid-learning research systems” where researchers, funders, health systems, practitioners, and community partners collaborate [21]; however, the lack of internal and external validity reporting identified indicated that few innovative designs are currently being used in this area of investigation.

Still, the comprehensiveness of reporting on RE-AIM criteria across these mHealth articles was relatively low with a number of gaps in reporting on both internal (eg, extent that the protocol was delivered as intended) and external validity factors (eg, description of intervention location and staff). At the individual level (ie, reach, efficacy/effectiveness, and maintenance), the reporting on issues related to reach and maintenance are particularly problematic. At the organizational or delivery level (ie, adoption, implementation, maintenance), there are large gaps in reporting across each of the dimensions. These gaps extend to the reporting across the four CONSORT-EHEALTH standards of access as well as the degree to which intervention features and functionality were addressed. Based on our findings, the results reported on mHealth PA interventions, from both an internal and external validity perspective, should be considered with caution.

Consistent with past research, this body of literature does not typically describe the target population or give indications as to the degree to which the study samples are representative of a larger population [27,28,31,33,35,37,40,42]. Thus, inferences cannot be made regarding who may be likely to benefit from these interventions based on different demographic, economic, or behavioral factors. Similarly, it is unclear which subgroups of the population may be more or less likely to engage in mHealth PA interventions. This is especially important to document given that those studies that did examine the representativeness of the study sample to the target population found that nonparticipants were less educated [43-45,56] and, if they did engage in the study, had greater difficulty in operating technology [47-49,62]. Additionally, almost all of the studies used convenience sampling procedures rather than recruiting from a known target population denominator. It is vital to recruit larger numbers of subgroups of the target population so that individuals that could most benefit from the intervention are actually receiving it. This information, across studies, is critical to ensure that interventions are designed to address broad access to the intervention and the needs of subgroups of a target population that suffer from health disparities (eg, lower education levels).

Similar to other areas of research, efficacy or effectiveness based upon changes to the PA and percent attrition were reported consistently across the majority of studies while the maintenance of those changes were not [27-29,31,32,36,37]. However, the generally positive effects found across studies may be overestimated based on the degree to which attrition was considered in follow-up analyses. That is, only one in every three studies reported using intent-to-treat analyses, with the majority limiting the description of study results to those who were present for follow-up assessments. Given the recidivism related to physical inactivity, it may be surprising that no study examined the maintenance of PA change at least 6 months past completion of the intervention. On one hand, the area is relatively new and it appears that researchers have emphasized determining the degree to which these interventions can initiate change. On the other hand, mHealth interventions may be less likely to encourage PA change maintenance due to advancements in newer technology that could make current interventions obsolete, the potential of technical problems that may reduce motivation, or simply decreased participant engagement over time. Until maintenance is documented in mHealth PA intervention studies, it is left to researcher and participant speculations on how well these interventions can contribute to maintained PA change over a long period of time.

Organizational or delivery level facets of RE-AIM have consistently been underreported across behavior change intervention studies; yet, studies on mHealth PA interventions appear to be even less likely to report on organizational adoption, implementation, and maintenance [29,30,37-39]. To be balanced, the majority of the studies included were reported as efficacy trials and some adoption information like inclusion and exclusion criteria of the staff and locations of intervention delivery may not be relevant. Similarly, efficacy and effectiveness trials do not typically have a goal to achieve and track maintained delivery of an intervention beyond the life of a grant. However, to allow for replication and determination of generalizability, even within highly controlled efficacy trials, it is necessary to provide the description of (1) the intervention costs and location, (2) the characteristics of the intervention and those who delivered it, (3) the degree to which the intervention was delivered as intended, and (4) if any adaptations were made to the intervention during the study period. An additional potential critique of this literature is the tendency for participants to use non-assigned treatments [22] that may contribute to the intervention's effectiveness. However, this body of literature included reports of co-interventions that, to some degree, address this issue. Still, no articles reported explicit tracking of non-assigned treatments so that possible contributions to effectiveness could be determined [22].

Understanding costs across RE-AIM dimensions is also key for dissemination [38]. In addition to implementation costs, other costs may be accrued both by organizations implementing these interventions as well as by the participants using them. For example, tracking of costs related to recruitment, equipment, technology (eg, batteries/chargers, mobile phone, service plan), and maintenance (eg, continue program once study period or funding is over) can improve the information necessary for decision making. From the perspective of the CONSORT-EHEALTH standards, future costs to the consumer should be considered in relation to the expected reach and effectiveness of mHealth PA interventions [24]. Without information on adoption and implementation, it is difficult to know the resources needed to successfully implement mHealth interventions in diverse locations or with staff of different levels of expertise.

Based upon the growth of research in the area of mHealth PA interventions and the review of this literature to date, there are a number of ways to improve the assessment and reporting on individual and organizational level factors that will improve our understanding of both the internal and external validity of this work. In Table 4, we provide a number of recommendations across RE-AIM dimensions specific to mHealth PA intervention research. In addition to these recommendations, the use of mixed method approaches that blend qualitative and quantitative data collected from participants and from those who implement the intervention could add depth to the data collected in mHealth PA studies and improve subsequent replication and implementation efforts [34]. Further, from a translational science perspective, tracking mHealth intervention costs across RE-AIM dimensions can inform adoption and delivery decisions within community and clinical practice settings.

Table 4. Recommendations.

RE-AIM component	Recommendations for reporting on future mHealth PA studies
Reach	
	Report on characteristics (eg, demographics, behavioral outcomes) of nonparticipants and compare them to participants to understand the representativeness of the study sample. If not possible for Institutional Review Board reasons to compare nonparticipants directly, participants can be compared to the general local population.
	Indicate exclusion criteria so that it is clear as to why certain individuals were not eligible for participation.
	Report on inclusion criteria (eg, computer/Internet literacy [24]) so that investigators can understand why specific individuals were selected.
	Describe recruitment methods and adaptations to recruitment methods so that future researchers will know the best ways to recruit for mHealth PA interventions.
	Recruit participants from a known denominator that are representative of the target population.
	Calculate the participation rate based upon a known denominator: # eligible approached and agreed to participate/total # eligible and approached.
	Describe how participants accessed the application, and cost to access application [24].
Effectiveness	
	Use intention-to-treat methods.
	Assess potential negative outcomes of the intervention and quality of life before and after the intervention.
	Indicate subgroup effects, especially those related to health equity issues.
Adoption	
	Report on characteristics of the location where the intervention is delivered and the staff who deliver the intervention and describe reasons for selection of this location and staff.
	If applicable, explicitly state inclusion/exclusion criteria of participating staff.
	If delivery locations or staff volunteer or are recruited for the study, calculate participation rate of settings/staff based on the number who volunteer divided by the number who were invited.
	Describe the level of human involvement required for the trial compared to the level of human involvement for a routine application [24].
	Describe the level of prompts/reminders required for the trial compared to the level of prompts/reminders for a routine application [24].
	Describe any interventions (including training sessions/support) that are implemented in addition to the targeted mHealth intervention [24].
Implementation	
	Report on intervention content, duration, and frequency of in-person and virtual sessions (eg, SMS, applications).
	Provide information intervention costs (eg, price of mobile technology, mobile phone data plan, time it takes to implement each session).
	Indicate percent delivered as intended (eg, text messages sent/unsent/received/not received; any application functioning problems or other technology problems).
	Reports of engagement should use standard or harmonized reporting methods (eg, number of sessions, number of bug fixes).
	Describe adaptations made to the intervention during implementation.
Maintenance	
	Include an assessment of maintenance of PA change 6 months after the completion of the intervention.
	Provide a description of how the intervention could be sustained or, if applicable, provide data on the degree to which the intervention is sustained over time.
	Report on strategies included during intervention design related to technical staff and potential participants to produce interventions that are functional and persuasive for a long period of time.

Limitations

Our review includes some limitations. First, our conclusions and recommendations are based on the degree to which these studies reported on specific RE-AIM dimensions. It is possible that some of these data have been collected, but not reported. To address this, we included all available articles on any given trial. Still, investigator plans and data for maintenance/sustainability or designing for dissemination may exist but go unreported; however, a transparent reporting of any existing plans would provide additional important context for any intervention study. In addition, a lack of reporting on an outcome cannot be equated to a lack of an intervention's ability to achieve that outcome (eg, lack of reporting on maintenance cannot be equated to a lack of maintenance). Second, because mHealth PA interventions are relatively novel and this is an emergent research area, the goal of the studies included within this review may have been to establish internal validity (eg, effectiveness of study outcomes), and therefore we must be cautious of being overly critical of these studies relative to their reporting of organizational adoption or maintenance factors.

Conclusions

There is an emergent body of literature reporting on mHealth PA interventions. On average, the studies provide initial evidence that these interventions may have promise in helping participants initiate PA. However, few studies report on key internal (eg, delivery as intended) or external (eg, descriptions of participants, settings, and delivery staff) factors. As a result, the degree to which these findings are robust and generalizable cannot be determined. Improved reporting across RE-AIM dimensions and the use of intention-to-treat, tracking of costs, and mixed methods approaches are recommended to ensure mHealth PA interventions are developed that can be broadly applicable across target populations, intervention delivery locations, and staff of differing levels of expertise.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2008. Physical inactivity - a global public health problem URL: http://www.who.int/dietphysicalactivity/factsheet_inactivity/en [WebCite Cache ID 6CoJleLjJ]
2. Trojano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
3. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]
4. 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]
5. Smith A. Pew Internet & American Life Project. 2010 Jul 07. Mobile Access 2010 URL: http://pewinternet.org/Reports/2010/Mobile-Access-2010.aspx [WebCite Cache ID 6CoN4Wmtq]
6. Lenhart A, Ling R, Campbell S, Purcell K. Pew Internet & American Life Project. 2010 Apr 20. Teens and mobile phones URL: http://pewinternet.org/Reports/2010/Teens-and-Mobile-Phones.aspx [accessed 2012-12-11] [WebCite Cache ID 6Cos8OAac]
7. Cisco Systems Inc. 2012. Cisco Visual networking index: Global mobile data forecast, 2011-2016 URL: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html [WebCite Cache ID 6CooPIV6t]
8. Smith A. Pew Internet & American Life Project. 2012 Jun 26. 17% of cell phone owners do most of their online browsing on their phone, rather than a computer or other device URL: http://pewinternet.org/Reports/2012/Cell-Internet-Use-2012.aspx [WebCite Cache ID 6Cos49n7k]
9. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
10. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
11. 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]
12. Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. Cochrane Database Syst Rev 2012;12:CD007457. [doi: 10.1002/14651858.CD007457.pub2] [Medline: 23235643]
13. 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]
14. 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]
15. 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]

16. Lau PWC, Lau EY, Wong DP, 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]
17. Militello LK, Kelly SA, Melnyk BM. Systematic review of text-messaging interventions to promote healthy behaviors in pediatric and adolescent populations: implications for clinical practice and research. Worldviews Evid Based Nurs 2012 Apr;9(2):66-77. [doi: 10.1111/j.1741-6787.2011.00239.x] [Medline: 22268959]
18. 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]
19. Dolan B. 2010. Smartphones to drive mHealth to 500 million users URL: http://mobihealthnews.com/9523/smartphones-to-drive-mhealth-to-500-million-users/ [WebCite Cache ID 6CotJPpNY]
20. Estabrooks PE, Gyurcsik NC. Evaluating the impact of behavioral interventions that target physical activity: issues of generalizability and public health. Psychology of Sport and Exercise 2003 Jan;4(1):41-55. [doi: 10.1016/s1469-0292(02)00016-x]
21. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
22. Danaher BG, Lichtenstein E, McKay HG, Seeley JR. Use of non-assigned smoking cessation programs among participants of a Web-based randomized controlled trial. J Med Internet Res 2009;11(2):e26 [FREE Full text] [doi: 10.2196/jmir.1172] [Medline: 19632976]
23. Fischer S, Stewart TE, Mehta S, Wax R, Lapinsky SE. Handheld computing in medicine. J Am Med Inform Assoc 2003 Apr;10(2):139-149 [FREE Full text] [Medline: 12595403]
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. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? Ann Behav Med 2004 Feb;27(1):3-12. [doi: 10.1207/s15324796abm2701_2] [Medline: 14979858]
26. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]
27. Akers JD, Estabrooks PA, Davy BM. Translational research: bridging the gap between long-term weight loss maintenance research and practice. J Am Diet Assoc 2010 Oct;110(10):1511-22, 1522.e1 [FREE Full text] [doi: 10.1016/j.jada.2010.07.005] [Medline: 20869490]
28. Allen K, Zoellner J, Motley M, Estabrooks PA. Understanding the internal and external validity of health literacy interventions: a systematic literature review using the RE-AIM framework. J Health Commun 2011;16 Suppl 3:55-72. [doi: 10.1080/10810730.2011.604381] [Medline: 21951243]
29. Antikainen I, Ellis R. A RE-AIM evaluation of theory-based physical activity interventions. J Sport Exerc Psychol 2011 Apr;33(2):198-214. [Medline: 21558580]
30. Bull SS, Gillette C, Glasgow RE, Estabrooks P. Work site health promotion research: to what extent can we generalize the results and what is needed to translate research to practice? Health Educ Behav 2003 Oct;30(5):537-549. [Medline: 14582596]
31. Dzewaltowski DA, Estabrooks PA, Klesges LM, Bull S, Glasgow RE. Behavior change intervention research in community settings: how generalizable are the results? Health Promot Int 2004 Jun;19(2):235-245 [FREE Full text] [Medline: 15128715]
32. Estabrooks P, Dzewaltowski DA, Glasgow RE, Klesges LM. Reporting of validity from school health promotion studies published in 12 leading journals, 1996-2000. J Sch Health 2003 Jan;73(1):21-28. [Medline: 12621720]
33. Glasgow RE, Bull SS, Gillette C, Klesges LM, Dzewaltowski DA. Behavior change intervention research in healthcare settings: a review of recent reports with emphasis on external validity. Am J Prev Med 2002 Jul;23(1):62-69. [Medline: 12093425]
34. Kessler RS, Purcell EP, Glasgow RE, Klesges LM, Benkeser RM, Peek CJ. What does it mean to "employ" the RE-AIM model? Eval Health Prof 2013 Mar;36(1):44-66. [doi: 10.1177/0163278712446066] [Medline: 22615498]
35. Klesges LM, Dzewaltowski DA, Glasgow RE. Review of external validity reporting in childhood obesity prevention research. Am J Prev Med 2008 Mar;34(3):216-223. [doi: 10.1016/j.amepre.2007.11.019] [Medline: 18312810]
36. McMahon S, Fleury J. External validity of physical activity interventions for community-dwelling older adults with fall risk: a quantitative systematic literature review. J Adv Nurs 2012 Oct;68(10):2140-2154. [doi: 10.1111/j.1365-2648.2012.05974.x] [Medline: 22416905]
37. White SM, McAuley E, Estabrooks PA, Courneya KS. Translating physical activity interventions for breast cancer survivors into practice: an evaluation of randomized controlled trials. Ann Behav Med 2009 Feb;37(1):10-19. [doi: 10.1007/s12160-009-9084-9] [Medline: 19255819]
38. Estabrooks PA, Allen KC. Updating, employing, and adapting: a commentary on What does it mean to "employ" the RE-AIM model. Eval Health Prof 2013 Mar;36(1):67-72. [doi: 10.1177/0163278712460546] [Medline: 23045308]

39. Aarons GA, Green AE, Palinkas LA, Self-Brown S, Whitaker DJ, Lutzker JR, et al. Dynamic adaptation process to implement an evidence-based child maltreatment intervention. Implement Sci 2012;7:32 [FREE Full text] [doi: 10.1186/1748-5908-7-32] [Medline: 22512914]
40. National Institutes of Health. 2012. NIH clinical research trials and you: The basics URL: http://www.nih.gov/health/clinicaltrials/basics.htm [WebCite Cache ID 6IpjfsNhO]
41. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
42. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton-Mifflin; 2002.
43. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
44. Burke LE, Styn MA, Glanz K, Ewing LJ, Elci OU, Conroy MB, et al. SMART trial: A randomized clinical trial of self-monitoring in behavioral weight management-design and baseline findings. Contemp Clin Trials 2009 Nov;30(6):540-551 [FREE Full text] [doi: 10.1016/j.cct.2009.07.003] [Medline: 19665588]
45. Conroy MB, Yang K, Elci OU, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574. [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]
46. Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich J, LeGrand L, et al. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. : UbiComp ’08; 2008 Presented at: Proceedings of the 10th International Conference on Ubiquitous Computing; September 21-24; Seoul, Korea p. 54-63.
47. Fukuoka Y, Lindgren T, Jong S. Qualitative exploration of the acceptability of a mobile phone and pedometer-based physical activity program in a diverse sample of sedentary women. Public Health Nurs 2012;29(3):232-240. [doi: 10.1111/j.1525-1446.2011.00997.x] [Medline: 22512424]
48. Fukuoka Y, Kamitani E, Dracup K, Jong SS. New insights into compliance with a mobile phone diary and pedometer use in sedentary women. J Phys Act Health 2011 Mar;8(3):398-403. [Medline: 21487139]
49. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]
50. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]
51. Nguyen HQ, Gill DP, Wolpin S, Steele BG, Benditt JO. Pilot study of a cell phone-based exercise persistence intervention post-rehabilitation for COPD. Int J Chron Obstruct Pulmon Dis 2009;4:301-313 [FREE Full text] [Medline: 19750190]
52. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]
53. 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]
54. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]
55. Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]
56. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
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. Lubans DR, Morgan PJ, Okely AD, Dewar D, Collins CE, Batterham M, et al. Preventing Obesity Among Adolescent Girls: One-Year Outcomes of the Nutrition and Enjoyable Activity for Teen Girls (NEAT Girls) Cluster Randomized Controlled Trial. Arch Pediatr Adolesc Med 2012 Sep 1;166(9):821-827. [doi: 10.1001/archpediatrics.2012.41] [Medline: 22566517]
59. Lubans DR, Morgan PJ, Dewar D, Collins CE, Plotnikoff RC, Okely AD, et al. The Nutrition and Enjoyable Activity for Teen Girls (NEAT girls) randomized controlled trial for adolescent girls from disadvantaged secondary schools: rationale, study protocol, and baseline results. BMC Public Health 2010;10:652 [FREE Full text] [doi: 10.1186/1471-2458-10-652] [Medline: 21029467]

60. Cheung PPY, Chow BC, Parfitt G. Using environmental stimuli in physical activity intervention for school teachers: A pilot study. Int Electron J Health Educ 2008;11:47-56.
61. Newton KH, Wiltshire EJ, Elley CR. Pedometers and text messaging to increase physical activity: randomized controlled trial of adolescents with type 1 diabetes. Diabetes Care 2009 May;32(5):813-815 [FREE Full text] [doi: 10.2337/dc08-1974] [Medline: 19228863]
62. Liu WT, Wang CH, Lin HC, Lin SM, Lee KY, Lo YL, et al. Efficacy of a cell phone-based exercise programme for COPD. Eur Respir J 2008 Sep;32(3):651-659 [FREE Full text] [doi: 10.1183/09031936.00104407] [Medline: 18508824]
63. The Cochrane Collaboration. 2012. Intention to treat issues URL: http://www.cochrane-net.org/openlearning/html/mod14-4.htm [WebCite Cache ID 6Ipjp6IRU]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth
mHealth: mobile health
PA: physical activity
RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance framework

Edited by G Eysenbach; submitted 01.06.13; peer-reviewed by R Glasgow, D Esteves; comments to author 25.06.13; revised version received 13.08.13; accepted 15.09.13; published 04.10.13.

Please cite as:
Blackman KCA, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, Estabrooks PA
Assessing the Internal and External Validity of Mobile Health Physical Activity Promotion Interventions: A Systematic Literature Review Using the RE-AIM Framework
J Med Internet Res 2013;15(10):e224
URL: http://www.jmir.org/2013/10/e224/
doi:10.2196/jmir.2745
PMID:24095951

Viewpoint

Considerations for Community-Based mHealth Initiatives: Insights From Three Beacon Communities

Nebeyou A Abebe[1*], MA, PMP; Korey L Capozza[2*], MPH; Terrisca R Des Jardins[3*], MHSA; David A Kulick[1*], MPH; Alison L Rein[4*], MS; Abigail A Schachter[4*], BA; Scott A Turske[3*], BA

[1]Louisiana Public Health Institute, Crescent City Beacon Community, New Orleans, LA, United States
[2]HealthInsight, Utah Beacon Community, Salt Lake City, UT, United States
[3]Southeast Michigan Beacon Community, Detroit, MI, United States
[4]AcademyHealth, Washington, DC, United States
[*]all authors contributed equally

Corresponding Author:
Alison L Rein, MS
AcademyHealth
1150 17th St NW
Suite 600
Washington, DC, 20036
United States
Phone: 1 202 292 6700 ext 733
Fax: 1 202 292 6833
Email: alison.rein@academyhealth.org

Abstract

Mobile health (mHealth) is gaining widespread attention for its potential to engage patients in their health and health care in their daily lives. Emerging evidence suggests that mHealth interventions can be used effectively to support behavior change, but numerous challenges remain when implementing these programs at the community level. This paper provides an overview of considerations when implementing community-based mHealth initiatives, based on the experiences of three Beacon Communities across the United States that have launched text messaging (short message service, SMS) pilot programs aimed at diabetes risk reduction and disease management. The paper addresses lessons learned and suggests strategies to overcome challenges related to developing text message content, conducting marketing and outreach, enrolling participants, engaging providers, evaluating program effectiveness, and sustaining and scaling the programs.

(J Med Internet Res 2013;15(10):e221) doi:10.2196/jmir.2803

KEYWORDS

mHealth; mobile health; mobile phone; Type 2 diabetes mellitus; text messaging; short message service (SMS); risk reduction; self management

Introduction

Mobile health (mHealth), defined as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices,” is increasingly used to engage patients in their health and care [1]. Cell phone use is widespread across socioeconomic, racial, ethnic, and age groups; 91% of Americans use cell phones, and 80% of cell phone users engage in text messaging (short message service, SMS) [2]. Additionally, Hispanic and black Americans—who are often medically underserved—are more likely to use text messaging than white Americans (85% vs 80% vs 79%, respectively) [3]. The near-ubiquity of cell phones, and their use for texting, demonstrates the potential of text message-based mHealth programs to reach traditionally underserved and hard-to-reach populations.

A growing body of evidence supports the feasibility of using text messaging and other mHealth applications for health promotion [4-6], behavior change (eg, smoking cessation) [7-10], chronic disease management [11], medication adherence [12,13], prenatal care [14,15], weight loss [16,17], and physical activity [18-20]. These programs target health behaviors by providing information and services—including medical appointment and medication reminders, self-tracking tools, educational resources, lab and clinical results delivery, etc—via timely and often personalized messages [5]. Such mHealth services confer advantages over traditional informational public health campaigns by providing a medium for broader audience engagement and direct interaction.

While using mobile technology as a source of health information is a relatively new concept, recent studies suggest that patients are generally open to receiving health-related text and electronic messages [21,22]. Furthermore, 31% of cell phone owners report using their phone to look for health information in 2012, compared to only 17% in 2010 [23]. This growing appetite for receiving and seeking health information via mobile technology presents new opportunities to engage patients outside of traditional care settings, even those who do not regularly seek health care services.

Recognizing this potential, and intrigued by the opportunity to help manage and possibly prevent chronic disease, several communities across the United States receiving federal funding through the Beacon Community Cooperative Agreement Program have deployed mHealth programs. Full results of the Beacon Communities' evaluation efforts are still pending, but early findings suggest a promising impact of mHealth on behavior change. However, these community-based pilots encountered a number of challenges in the design (eg, developing content; conducting outreach), execution (eg, engaging patients and providers), evaluation, and sustainability of their mHealth programs. The paper also describes lessons learned and offers strategies and promising practices to address these challenges (Table 1).

Table 1. Challenges, considerations, and lessons learned for developing community-based text-messaging programs.

Domain	Challenges / Considerations	Lessons Learned
Developing message content		
	Technical constraints (160 character limit)	Use evidence when developing messages
	Cultural competency / language availability	Tailor messages based on participant data
	Clinical validity / guidelines	Customize messages to local population to enhance user experience
	Literacy level testing	
	Tailoring to local population and individual users	
	Framing messages based on behavior change theories	
Outreach and marketing		
	Selecting media and outreach channels	Engage non-traditional partners
	Events vs marketing	Design a flexible outreach plan
	Targeted vs mass-media marketing	Enrollment proposition is more compelling in health-related contexts (eg, health fair vs sporting event)
Enrolling participants		
	Enrollment method impacts enrollment numbers and engagement	Offer multiple enrollment method options
	Limited technological proficiency and access	High-touch, in-person recruitment is key, but is labor intensive and costly
	Costs of participating/texting	
Engaging providers		
	Lack of payment under fee-for-service	Credibility drives adoption
	Lack of interoperability/data sharing between mHealth platforms and EHRs	Integrate mHealth into other interventions/initiatives
	Multiple and competing priorities	
Evaluating impact		
	Lack of robust mHealth evaluation methodologies	Plan evaluation strategy, identify data sources and outcome metrics from outset of project
	Limited funding	Minimize biases
	Accelerated timeline	Consider level of rigor needed and budget or other resource limitations
	Rigor / quality vs speed / cost	
	Biases (attrition, sampling, non-response)	
Sustaining and scaling		
	Sustaining programs after grant funding ends	Leverage community partnerships for financial and in-kind support
	Lack of provider reimbursement for mHealth	Partner with health plans
	Securing partnerships and resources	Incorporate mHealth into other payment reform strategies

About the Programs

The 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act authorized the Office of the National Coordinator for Health Information Technology to create the Beacon Community Cooperative Agreement Program, which granted 17 diverse communities across the United States $12-15 million each over 3 years to build and strengthen their health information technology (health IT) infrastructures and to test innovative technologies to improve care quality and population health and reduce costs [24]. Several of the Beacon Communities launched community-based mHealth programs as part of their health IT-enabled quality improvement efforts.

The Crescent City and Southeast Michigan Beacon Communities, located in and around New Orleans, Louisiana, and Detroit, Michigan, respectively, were two of three Beacon Communities (the Greater Cincinnati Beacon Community in Ohio was the third) to pilot txt4health**,** a text message–based health information service that aimed to help adults ages 18 and up to understand their risk of developing type 2 diabetes and steps they can take to reduce that risk. txt4health targeted highly vulnerable, at-risk populations in these communities, many of whom were overweight or obese, low-income, and/or uninsured. Participants enrolled in txt4health by texting the word “health” to 300400 or online via the txt4health website. Upon enrollment, participants completed a diabetes risk assessment, the results of which were used to place them in a risk category and to tailor subsequent text messages. Participants received 4-7 messages per week, including general educational messages, diet and exercise tips, health reminders, and information about local health care providers and resources. Participants could also set and track progress toward weight loss and exercise goals by responding to text message prompts.

The Utah Beacon Community, located in and around Salt Lake City, Utah, launched Care4Life**,** a two-way text messaging program designed to enhance self-management among adults aged 18 and up with type 2 diabetes. Care4Life participants received 1-7 messages per day over a period of 6 months. Like txt4health, Care4Life messages included general diabetes education, health improvement suggestions, and encouragement toward self-entered weight loss and exercise goals. In addition, Care4Life included robust coaching and interactive support functionality; participants could elect to receive medication, blood sugar testing, blood pressure monitoring, and clinical appointment reminders, as well as requests to report back medication adherence, blood sugar readings, physical activity, and weight. As with txt4health, Care4Life enrollees could join via text message or online; however, participants who enrolled by text received one-way educational messages only—unless they signed up for more protocols at a later date—whereas those who enrolled online received the full suite of two-way message options. Participants could also track their data via a Care4Life Web portal. See Table 2 for a comparison of txt4health and Care4Life.

Table 2. txt4health and Care4Life program characteristics.

	txt4health	Care4Life
Beacon Communities implementing	Crescent City and Southeast Michigan	Utah
Target population	Adults age 18+ at risk for type 2 diabetes	Adults aged 18+ diagnosed with type 2 diabetes with HbA1c>8[a]
Message types	Diabetes risk assessment	Diabetes education and health improvement
	General diabetes education	Medication, glucose testing, blood pressure monitoring, and clinical appointment reminders
	Diet and exercise tips	Encouragement toward self-entered weight loss and exercise goals
	Health reminders	Requests to report back blood sugar readings, medication adherence, exercise and weight loss goals
	Ability to set and track personal weight loss and exercise goals	
	Information about how to find local providers and resources	
Program duration	14 weeks	26 weeks
Message frequency	4-7 per week	7-49 per week

[a] HbA1c=hemoglobin A1c, a measure of blood sugar control.

Program Planning: Developing Message Content

For the Crescent City and Southeast Michigan Beacon Communities, developing txt4health text message content involved more than simply adhering to character limits and considering literacy levels. The messages needed to contain clinically valid health information, presented in a tailored way that would appeal to the target population and ultimately promote behavior change.

To ensure clinical validity, the txt4health messages were developed by an advisory group that included members of the Crescent City, Southeast Michigan, and Cincinnati Beacon teams, as well as experts from the txt4health mHealth vendor (Voxiva), the Centers for Disease Control and Prevention, and the American Diabetes Association. The clinical content was based on evidence-based guidelines, including those endorsed by the American Diabetes Association (ADA) and the National Diabetes Education Program (NDEP); for instance, the risk assessment was used to assign txt4health participants into risk categories developed by the ADA, in order to tailor the text messages to the appropriate risk level. To vet their messages further, the Crescent City Beacon conducted focus groups and in-depth interviews with local providers and community leaders, as well as workgroups with representatives from consumer organizations including the Juvenile Diabetes Research Foundation.

The advisory group also drew from relevant research to formulate messages based on theories of behavior change, especially the Health Belief Model, a conceptual model describing factors that influence whether people engage in health behaviors such as preventive care or adherence to treatment regimens [25]. According to the model, the likelihood of engaging in a health behavior is influenced by one’s perceived susceptibility to a particular disease or condition, the perceived seriousness or severity of that condition, the perceived benefits of the behavior, and the perceived barriers to engaging in that behavior [25]. The communities deploying txt4health observed a lack of perceived risk of developing diabetes among potential enrollees, either in the form of real or perceived apathy toward the risk factors of diabetes, or in a lack of understanding of how risk factors affect the onset of diabetes. To address these perceptions, the advisory group crafted educational messages and recommendations based on the Health Belief Model, emphasizing participants' susceptibility to diabetes (based on their risk assessment); the potentially severe consequences of developing or failing to control diabetes; the simple steps that can be taken to reduce one's risk; and the short- and long-term benefits of those steps. See Figure 1 for sample messages that address each dimension of the HBM.

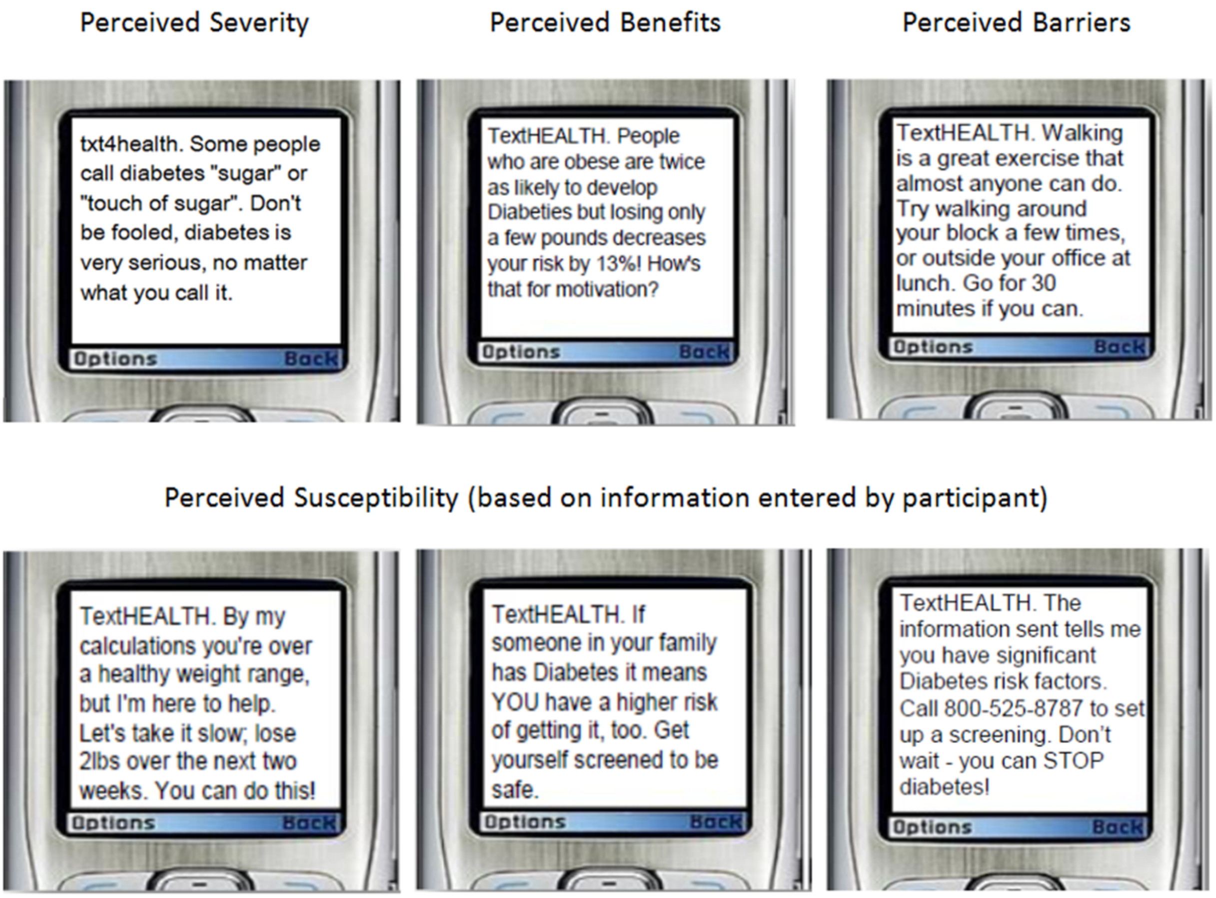
To appeal to txt4health's broad target population, which included many people who did not regularly access health care, the Crescent City and Southeast Michigan Beacon teams also endeavored to the craft messages in an approachable, encouraging, and friendly "voice". Furthermore, the messages aimed to be culturally competent, reflecting an understanding of local interpretations of disease and the colloquial language used to describe it. For example, in Southeast Michigan diabetes is often referred as "sugar". By incorporating this term into the text messages, the txt4health team hoped participants would perceive the program as more accessible and relatable and would thus engage with it more actively. They also subjected the messages to literacy level testing to ensure that they were straightforward, easily understandable, and did not contain medical jargon.

The Crescent City and Southeast Michigan teams also created community-specific messages to reflect the local context. About once per week, both communities sent messages via txt4health to notify participants about upcoming local events such as health fairs. Additionally, recognizing that personal safety could be a concern in Detroit and the surrounding cities, the Southeast Michigan team included alternatives to outdoor activities (eg, renting an exercise video or exercising while watching TV) in their txt4health messages.

In contrast to the experience with txt4health, the Care4Life messages used in Utah were previously developed by a diabetes education expert, based on the ADA clinical guidelines, the NDEP, and principles of the American Association of Diabetes Educators AADE7 Self-Care Behaviors. The Utah Beacon team decided to use these pre-developed messages because they had already undergone a rigorous vetting process and were based on established clinical guidelines—an important feature given that the intervention was embedded in a clinical environment [26]. To preserve the validity of the messages, the text was not modified or customized for the Utah Beacon population; however, participants could select the types and frequency of messages received (eg, medication reminders, weight loss tracking, blood sugar testing requests), and the messages addressed participants by name. See Figure 2 for sample Care4Life messages.

Research suggests that framing health-related messages to target beliefs, perceptions, and subjective norms can influence their impact on attitudes and intentions and ultimately encourage behavior change [27-31]. While the txt4health teams used the Health Belief Model to target perceptions and beliefs, future iterations of both txt4health and Care4Life could evaluate how further tailoring and framing the messages based on user demographics or health risk assessment data could enhance user experience and the programs' impact on targeted health behaviors.

Figure 1. Sample txt4health messages and relevant Health Belief Model dimension.



Perceived Severity

txt4health. Some people call diabetes "sugar" or "touch of sugar". Don't be fooled, diabetes is very serious, no matter what you call it.

Perceived Benefits

TextHEALTH. People who are obese are twice as likely to develop Diabeties but losing only a few pounds decreases your risk by 13%! How's that for motivation?

Perceived Barriers

TextHEALTH. Walking is a great exercise that almost anyone can do. Try walking around your block a few times, or outside your office at lunch. Go for 30 minutes if you can.

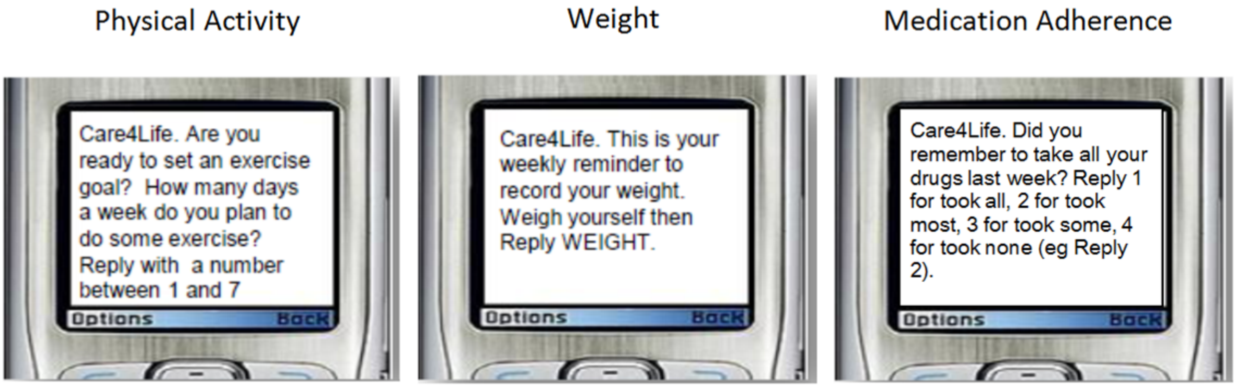
Perceived Susceptibility (based on information entered by participant)

TextHEALTH. By my calculations you're over a healthy weight range, but I'm here to help. Let's take it slow; lose 2lbs over the next two weeks. You can do this!

TextHEALTH. If someone in your family has Diabetes it means YOU have a higher risk of getting it, too. Get yourself screened to be safe.

TextHEALTH. The information sent tells me you have significant Diabetes risk factors. Call 800-525-8787 to set up a screening. Don't wait - you can STOP diabetes!

Figure 2. Sample Care4Life messages.



Physical Activity

Care4Life. Are you ready to set an exercise goal? How many days a week do you plan to do some exercise? Reply with a number between 1 and 7

Weight

Care4Life. This is your weekly reminder to record your weight. Weigh yourself then Reply WEIGHT.

Medication Adherence

Care4Life. Did you remember to take all your drugs last week? Reply 1 for took all, 2 for took most, 3 for took some, 4 for took none (eg Reply 2).

Outreach and Marketing

Overview

All three Beacon Communities engaged in marketing and outreach in order to drive interest and enrollment in their mHealth programs. This engagement involved experimentation with different partners, media, and outreach strategies (Table 3), as generating the desired levels of enrollment proved challenging.

Outreach efforts in the Utah Beacon Community targeted patients treated at the 19 primary care clinics participating in a community-wide quality improvement effort around diabetes care. The outreach process was dictated by what clinic staff were willing and able to take on in terms of workload and provision of access to patient data. Since the pilot was conducted as a randomized controlled trial, outreach was also limited by human subjects research protocols. Clinic staff queried the clinics' electronic health record (EHR) systems to identify patients with type 2 diabetes who might benefit from Care4Life and mailed invitations that directed them to sign up online.

However, finding that many patients had difficulty with online enrollment, about 4 months into the recruitment efforts the Beacon team shifted to a more hands-on, in-person approach where staff stationed in clinics assisted patients with enrollment and offered training on how to navigate the Care4life program.

In contrast to Care4Life, txt4health targeted populations at risk for or with undiagnosed diabetes, whether or not they sought or accessed care. Therefore, while the Crescent City and Southeast Michigan Beacons supplied marketing materials to promote txt4health in primary care clinics, they also undertook comprehensive, multipronged marketing campaigns via a wide variety of outreach channels and settings. These channels included mass-media marketing through traditional vehicles such as television and radio, as well as social media and online advertising.

The Crescent City and Southeast Michigan Beacons were challenged in reaching a typically hard-to-reach target audience. With the support and resources from a leading national advertising and marketing firm, the Southeast Michigan Beacon drew from third-party syndicated research to identify areas in the Detroit area likely to have a high density of diabetic and pre-diabetic individuals and targeted outreach activity in these areas. They advertised txt4health in public transit, bus shelters, laundromats, barbershops, salons, and other settings frequented by the target audience, which allowed them to optimize outreach while conserving limited resources.

In Crescent City, the Beacon Community Advisory Group was instrumental in devising the txt4health outreach strategy, conducting focus groups and key informant interviews with consumers, providers, and other community members to better understand how to reach the target audience. The Crescent City team worked closely to leverage the advisory group members' communication channels (eg, health fairs, church meetings, retail stores) for outreach efforts. For example, a large retail partner allowed txt4health staff to directly engage customers through in-store activations, and pharmacists working at this retail chain promoted txt4health directly to customers at risk for diabetes. Additionally, Crescent City engaged non-traditional partners such as faith-based organizations and university student groups to host enrollment contests and events. Crescent City relied less heavily on online marketing than Southeast Michigan, instead asking community partners to include txt4health information on their websites.

Having a flexible outreach plan allowed all three communities to change their tactics based on their experiences with various strategies. They also found that community partnerships, health fairs, and public events represented key opportunities to engage potential txt4health and Care4Life participants. In Crescent City and Southeast Michigan, in addition to driving enrollment, these partnerships and events also helped garner support and goodwill toward txt4health and the broader Beacon Community initiative.

Table 3. mHealth outreach and marketing channels and tactics.

Outreach Channel	Tactic	Beacon Communities Using		
		Crescent City	Southeast Michigan	Utah
Mass media	Television and radio public service announcements, (paid) radio advertising, "flash mobs," online advertising, transit media (bus exteriors and interiors), in-place media (targeted signage), and earned media.	X	X	
Electronic marketing	Facebook, online advertising.	X	X	
Events	Community events, entertainment venues, health fairs, screenings.	X	X	X
Community partner marketing	Community partner events, websites, and newsletters.	X	X	X
Beacon interventions	Integrating mHealth program into other Beacon interventions (eg, Patient Health Navigator program; Emergency Department Diabetic Patient Identification program; diabetes quality improvement initiative).		X	X
Primary care practices	Directly involving primary care providers in promoting the program to patients and visitors	X	X	X
	Displaying marketing materials in exam and waiting rooms (eg, table tents, "prescription-like" tear-pads, posters).	X	X	X
Mass mailings	Using clinic data to identify patients likely to benefit from the program and mailing them an invitation with instructions on how to enroll.			X

Enrolling Participants

As noted above, these outreach and marketing campaigns were designed to drive enrollment in txt4health and Care4Life. Participants could enroll in one of three ways (Table 4). The Beacon Communities found that the method used to enroll interested participants had an important influence on the total number of enrollees, as well as their subsequent level of engagement with the program.

In the case of txt4health, enrolling via text allowed potential participants to immediately "opt-in" to the service and proceed to the health risk assessment and subsequent messaging curriculum. Participants with a computer and Internet access could enroll online via the txt4health website and confirm their enrollment by responding to a confirmation text message triggered by data entered in the online form. Like text-based enrollment, this method required proactive participation on the part of the user, which translated to a higher likelihood of program "activation" upon receipt of the confirmation text message.

Third-party enrollment was used at health fairs and community events where Beacon staff promoted txt4health. At these events, potential enrollees provided their name, cell phone number, and ZIP code on a roster that included a consent waiver. After the event, Beacon staff would enter this information into the online enrollment form, triggering confirmation text messages for participants to respond to in order to activate their enrollment. The Beacon teams observed a significant drop-off in txt4health participant activation among those who were enrolled by a third party, which they attribute to many potential factors, including the time lag between initial sign-up and confirmation of participation in the program; the provision of incorrect or incomplete contact information on the roster; and/or the lack of direct personal participation in the enrollment process.

For Care4Life, the mode of enrollment had implications for program experience once enrolled. Those who enrolled by text message received one-way educational messages only but could add more message types via the Web portal or via text message at a later date. Those who enrolled online could receive the full suite of two-way message options including reminders, coaching, and requests for feedback. Additional information was required for online enrollment, including a series of health-related questions designed to set personalized reminder and coaching messages. This full enrollment process was more time consuming for patients to complete via text message than via a Web-based form.

To address this challenge, Beacon staff in participating clinics were equipped with computers and made available to help patients enroll in Care4Life online. Using this method, they signed up more than 400 patients in the program over a 6-month period. One clinic in Utah adopted a more aggressive approach, incorporating the Care4Life outreach and enrollment process directly into existing workflows. During regularly scheduled outreach calls to diabetic patients, medical assistants scheduled the quarterly recommended office visit and then signed interested patients up for Care4Life during the same call. Using this method, the clinic enrolled more than 40 patients in a period of 2 weeks, which represented a significant boost in enrollment.

Based on their experiences, these three Beacon Communities identified some key barriers to enrollment. First, although the txt4health and Care4Life programs were free, messaging rates applied for participants without unlimited texting plans, which proved cost-prohibitive for many potential participants who did not want to use up their limited messages. The txt4health teams identified this issue prior to the program launch, since at that time many people in the target population used government-issued cell phones whose service plans did not include unlimited texting. Unfortunately, it was not possible to assess how many potential participants were deterred from enrolling due to cost.

Limited technology proficiency and/or access to computers or the Internet presented additional barriers to online and text-based enrollment. During a follow-up telephone survey of 104 patients invited to participate in Care4Life, 35% reported limited or no access to a computer and 38% reported having trouble using a Web browser. Certain patient demographic characteristics were associated with lower technological proficiency; for instance, while older patients (age 50+) represented the majority of txt4health and Care4Life users, this group tended to struggle with using texting more than their younger counterparts. Many Care4Life participants also had very basic or older phones with more outdated features (eg, numerical keys rather than keyboard) that made texting more difficult, especially for older users. To address this challenge, Beacon staff in Utah and New Orleans trained Care4Life and txt4health participants, respectively, to send and receive text messages during in-person enrollment events.

All three Beacon Communities found that potential participants often needed the assurance of in-person interaction and personal relationship (eg, with a trusted provider) to get involved with these new, novel programs. While surveys indicated that the traditional marketing tactics such as advertisements and brochures increased community awareness of the programs, direct in-person engagement drove enrollment to a much greater extent. People were much more receptive to the txt4health and Care4Life "enrollment proposition" when they were open to or seeking health information. Whether at community events or in the clinic, additional staff support was critical to deliver the high-touch, one-on-one personal interactions and drive enrollment in the programs. However, this made the process much more labor-intensive and costly than anticipated. Other communities launching similar mHealth programs may learn from this experience by anticipating and budgeting for additional staff time and resources to support in-person enrollment efforts.

Table 4. Enrollment methods for txt4health and Care4Life.

Enrollment method	txt4health	Care4Life
Text	Text the word "health" to 300400	Text a unique, clinic-specific enrollment code to 300400
Online	Enter cell phone number and ZIP code in online enrollment form, then respond to confirmation text message triggered by the form	Complete online enrollment form (up to 26 questions), then respond to confirmation text message triggered by the form
Third party	Allow third party to complete online enrollment on behalf of user. Enrollee must respond to confirmation text message triggered by the online form.	Allow third party to complete online enrollment on behalf of user. Enrollee must respond to confirmation text message triggered by the online form.

Engaging Providers

Based on the experiences of these three Beacon Communities, integrating mHealth programs in the clinical setting has several advantages. As anticipated, the Beacons found that introducing the programs in the health-related or clinical settings (rather than via mass media or community events) led to higher rates of adoption and greater credibility among participants. Unfortunately, several barriers may constrain further integration of mHealth into the clinical setting.

Two primary barriers to engaging clinicians and care teams in mHealth are related to the predominant fee-for-service payment structure. First, the current structure rewards patient volume, which limits appointment times—typically to 8-10 minutes. Second, providers are reimbursed only for specific activities, which do not generally include discussion or promotion of mHealth programs [32]. As such, while all three Beacon Communities had initially hoped to integrate their mHealth programs into the primary care workflow, provider time constraints and the lack of reimbursement for helping with txt4health and Care4Life enrollment were perceived as hindering these efforts.

Despite these challenges, the Beacon Communities were able to involve care team members in promoting their mHealth programs. Practice coordinators in Southeast Michigan incorporated txt4health enrollment into patient check-in and check-out processes; additionally, diabetic patients participating in the Patient Health Navigator care management program, and those who were identified by the Emergency Department Diabetic Patient Identification program, were encouraged to enroll in txt4health.

The Utah Beacon team also interfaced with primary care clinics because they anticipated benefits to integrating the Care4Life pilot into patients' existing care settings and because providers expressed interest in self-management support for patients outside of the clinic. Since the Utah Beacon offered pay-for-performance incentives to a subset of clinics based on diabetes care quality and outcomes, medical assistants were able to enroll patients in Care4Life as a strategy to reach hemoglobin A1c (HbA1c, a measure of blood sugar control) targets. However, in the absence of such payment incentives, the clinics may not have chosen this approach, as it required medical assistants to devote time they otherwise would have spent on patient care.

Another barrier to integration of mHealth into the care setting is that few existing mHealth tools have the capability to interface with the data management systems or EHRs used by clinicians to record patient encounter information [32]. Therefore, information collected as part of mHealth programs (via text or otherwise) like Care4Life or txt4health is unlikely to be recorded or viewed by providers in their primary documentation and clinical decision support systems, and therefore unlikely to be used to inform decisions at the point of care.

In cases where mHealth programs are able to send data to EHRs, providers have expressed concern about how to handle those data, raising questions of information accuracy and medical liability [33]. Given the variety and volume of data that providers are already struggling to process and manage as they adopt EHRs and other health IT tools, it is unclear whether or to what extent additional patient-generated information from mHealth programs will be integrated into the care process directly through these enabling technologies. Thus, while new reimbursement structures may facilitate integration of mHealth into the primary care workflow, further advances in device interoperability and data integration will also be necessary to achieve this objective.

Evaluating Impact

The Beacon Communities are pioneers in deploying mHealth programs to achieve the triple aim of reduced costs, improved population health, and higher-quality care, and many stakeholders are anxious to see what impact these and similar initiatives have had. All three Beacon Communities are engaged in evaluation efforts, the results of which will be disseminated separately. While initial results are promising in terms of user satisfaction and self-reported behavior change, assessment of these mHealth interventions has proved particularly challenging. The domain of mHealth interventions is new and rapidly evolving, and standardized and robust evaluation methodologies are not yet widely available [32,34]. As a result, much of the existing literature focuses on the feasibility of deploying mHealth programs, rather than their impact on health outcomes; the little evidence available on the impact of mHealth is highly variable and often context specific [32,34].

As it happened, the Beacon Communities were offered the opportunity to launch the txt4health and Care4Life pilot programs more than 1 year into these 3-year efforts. Thus, while the Beacon grants initially allocated funding for robust program evaluations, the post-hoc funding re-allocated to txt4health and Care4Life included relatively few resources specifically for evaluation. As a result, the Beacon teams needed to take a pragmatic approach to evaluation and, in some cases, secure funding from other sources (eg, community partners).

The Beacons also faced an accelerated timeline, needing to complete the entire pilot (including program planning, development, implementation, deployment, evaluation, and close out) in less than 2 years. Given the aforementioned resource constraints and the relative dearth of evidence available at the time of launch regarding best practices for mHealth deployment, the Beacons worked to balance the desire for rigorous evaluations with the need to rapidly roll out the programs. In addition to the roll-out processes described above, the evaluation teams were charged with designing evaluation plans, obtaining Institutional Review Board (IRB) approval, recruiting active users, administering surveys, and collecting and analyzing the resulting data.

In the context of these constraints, the Beacons took different approaches to evaluating their specific programmatic objectives (Table 5). Primary data sources included txt4health and Care4Life system usage data, EHR data, and multimodal surveys offered online and via text, mail, or phone. From these sources, the evaluation teams gathered data to inform multiple outcomes of interest, including enrollment numbers, user demographics, user engagement (eg, number/frequency of texts responded to, duration in the program), clinical outcomes (eg, change in HbA1c), self-reported behavior (eg, medication adherence), patient activation, and user satisfaction. Patient activation is assessed with the Patient Activation Measure (PAM), a valid, highly reliable scale that reflects a developmental model of patient and consumer activation [35].

During data collection, the Beacon teams encountered limitations including attrition bias, sampling bias, non-response to surveys, and incomplete EHR data due to inconsistent primary care follow-up. For example, in Utah, the Care4Life team had planned to evaluate objective change in HbA1c using data from participating provider EHRs. However, these data were missing for many patients who did not come in for regular follow-up appointments as the Utah Beacon team had assumed they would; this greatly reduced the sample size available for analysis. In Crescent City, the txt4health team chose landline random digit dialing (RDD) as a survey method since it was less expensive than more robust methodologies (eg, mobile phone RDD, which requires additional screening to ensure numbers in the sample are active and local). Unfortunately, landline RDD may have introduced selection bias by oversampling populations that are more likely to use landlines (eg, older individuals) and undersampling those more likely to use mobile phones. Since the survey was intended to evaluate a mobile phone–based service, this bias may have important consequences for the validity of the results.

An additional challenge was the difficulty of isolating the impact of mHealth programs on health behaviors and outcomes. As was the case with many mHealth programs, txt4health and Care4Life were implemented in “real-world” settings rather than controlled research settings, in the midst of multiple Beacon Community initiatives aimed at improving diabetes care and outcomes. This context makes it difficult to control for external factors and tease out the impact of—or attribute observed outcomes to—the specific mHealth intervention.

Despite these limitations, useful insights may be gleaned from the available data sources and analyses. In addition to the initial results of self-reported behavior change, patient activation, and user satisfaction, correlations between particular demographic characteristics and enrollment and program usage data may reveal important information. For instance, these data may help determine which people are most likely to enroll in, engage with, and benefit from these programs; how and why they choose to use the programs; whether one-way or two-way messaging is more effective in driving behavior change; and whether certain characteristics correlate with higher likelihood of dropping out of the program [36].

In some cases, the limitations and biases associated with mHealth data sources and evaluation methodologies can be addressed, but generally at a cost. Those engaged in mHealth evaluation efforts must consider the costs and benefits, as well as the anticipated value and intended use of evaluation results. For example, if the evaluation is to be used to assess clinical impact, or to justify further significant resource expenditure to sustain and/or spread an mHealth program, then the value of anticipated outcomes may be worth the costs of rigor. Alternatively, programs that primarily focus on health education and public awareness may only need rigor sufficient to prove the value of the program to community partners and other local supporters and thus may be able to use lower-cost methodologies. Regardless, those embarking on mHealth interventions should carefully consider their evaluation and research aims from the outset, as well as the resources they have at their disposal to achieve their desired outcomes [32].

Table 5. Beacon Community evaluation strategies for txt4health and Care4Life.

mHealth Program (Beacon Community)	Program Component Being Evaluated	Outcome(s) of Interest	Evaluation Method(s)
Txt4health (Crescent City)			
	Social marketing campaign	Awareness of and support for txt4health	Cross-sectional pre- and post- campaign surveys (online and landline Random Digit Dialing)
	User engagement		
		# of users enrolled % of users completing diabetes risk assessment Frequency of setting/achieving physical activity and weight loss goals	Descriptive analysis of system-level usage data
		Patient Activation Measure (PAM) score	Patient Activation Measure (PAM) tool
	User satisfaction		
		User demographics User perceptions of txt4health usability Impact on user behavior User satisfaction	Multimodal survey (telephone, online, or mail)
Txt4health (Southeast Michigan)			
	User engagement		
		# of users enrolled % of users completing diabetes risk assessment Frequency of setting/achieving physical activity and weight loss goals	Descriptive analysis of system-level usage data
		Patient Activation Measure (PAM) score	Patient Activation Measure (PAM) tool
	User satisfaction		
		User demographics User perceptions of txt4health usability Impact on user behavior User satisfaction	Multimodal survey (telephone, online, or mail)
Care4Life (Utah)			
	Clinical outcomes	Change in HbA1c	Electronic health record review
	User engagement		
		Duration in the program # of text message replies with the program Frequency of messages elected to receive	Descriptive analysis of system-level usage data
	User satisfaction	User satisfaction	5-question text message-based survey at 90 days Client Satisfaction Questionnaire (CSQ-8) at 180 days

Scaling and Sustaining mHealth Initiatives

As with the other interventions they implemented, the Beacon Communities launched these mHealth pilots in hopes that—if demonstrated to be effective at achieving their objectives—they would be sustained and eventually scaled to other populations and/or communities. However, developing long-term plans to sustain and scale these programs has proven challenging, and the future of the programs in some communities remains uncertain.

Only Crescent City is planning to expand txt4health statewide throughout Louisiana; in the meantime, participants in the New

Orleans area can still enroll in the program. The Utah Beacon continues to recruit and enroll patients in Care4Life through the end of the Beacon Program in September 2013, but there are currently no plans to sustain or scale it beyond that time. In Southeast Michigan, all marketing and enrollment for txt4health concluded following the end of the year-long pilot. Despite the diverse and uncertain futures of these programs, the Beacon Communities identified several approaches that could facilitate the long-term sustainability of community-based mHealth programs: incorporating mHealth programs into the care setting, engaging payers, and leveraging community resources and organizations to reach target constituents.

As noted previously, incorporating mHealth into the clinical setting facilitates enrollment and thus, represents a promising sustainability strategy. However, changes to the payment structure that reward improvements in patient health are necessary to make mHealth and other health IT strategies sustainable. By participating in new payment models such as accountable care organizations and/or partnering with health plans, providers and staff could be reimbursed for activities that promote patient self-care including explaining mHealth programs, helping patients enroll, and reviewing patient data submitted via mHealth programs. The Beacons are also integrating txt4health into other proven diabetes prevention programs that are reimbursable activities, such as a YMCA diabetes prevention program in Crescent City and an Emergency Department Diabetic Patient Identification intervention in Southeast Michigan. This strategy supports sustainability efforts while enhancing diabetes prevention offerings to at-risk populations.

While these types of payment and care delivery reform efforts are becoming widespread, numerous barriers still hinder efforts to engage payers to support mHealth. For instance, in Michigan, one challenge stemmed from the numerous requirements and approvals required for marketing communications directed toward Medicaid beneficiaries. These requirements clearly applied to txt4health and required the Southeast Michigan Beacon Community to secure state approvals before payers could promote txt4health to Medicaid-insured individuals.

In all three Beacon Communities, building relationships with community organizations and leveraging local resources was critical to the success of their mHealth interventions. Building trust and social capital through these partnerships provided outreach channels to the organizations' constituents, as well as sources of in-kind support (eg, creative development, sponsorship of campaign events), campaign design input, and in some cases, financial support.

For example, the Crescent City Beacon Community advisory group included many traditional public health partners (eg, state and local health departments, health associations) and several non-traditional private sector partners (eg, large health plans, employers, faith-based organizations, fraternity/sorority groups). These organizations were selected based on their reach and influence among targeted communities, and capacity to provide financial and in-kind support for public health initiatives. Maintaining effective communication and allowing advisory group members to help shape the program from project inception was a key to success. Establishing trust among key stakeholders who buy into a mutually beneficial concept can facilitate scaling and sustainability of grassroots activities and pilot programs.

Conclusions

In the midst of widespread attention to mHealth as the "next big thing" in health care, the collective experience of these three communities provides insights into the practical challenges of implementing mHealth programs in the community setting. Beacon Communities encountered a number of barriers at each stage, including issues related to developing tailored, culturally competent messages; designing comprehensive outreach strategies; enrolling participants; engaging providers in mHealth programs; evaluating mHealth programs; and sustaining and scaling pilots. Ideally, others with an interest in implementing community-based mHealth programs will be able to apply these lessons learned to help anticipate and overcome potential challenges in their own initiatives.

The Beacon experiences also yielded important insights into what works. These factors were critical to the success of their mHealth programs and should be considered by other communities:

1. Identify community partners that are willing to engage with and support the program, and leverage their resources and community presence to design the program strategy and reach the target audience.
2. To the extent possible, design outreach, enrollment, and message content around the needs and perspectives of end users to increase program enrollment and engagement.
3. Anticipate that traditional marketing tactics may be insufficient to drive enrollment, and plan and budget additional staff time and resources for in-person engagement with the target audience to help drive enrollment.
4. To the extent feasible, bring care providers into the process—even if it means developing work-around solutions—to help them understand and promote mHealth as a tool to enhance patient care.
5. When planning the evaluation strategy, decide at the outset which aspects of the program will be critical to measure and which will not, to determine what stakeholders "need to know" versus what would be "nice to know".
6. Last, share lessons learned with others to allow them to benefit from your experience.

A certain level of readiness is necessary for both providers and patients to begin to use cell phones as sources of and channels for sharing health information. This readiness may take time to develop. Much like the adoption curve for other technologies (eg, automated teller machines, online retail transactions), the use of mHealth may require time for market adoption and product improvement [36,37]. Working toward a culture of greater patient engagement in health and care will also further the potential impact of mHealth. And, as noted previously, framing and tailoring mHealth messages to target health beliefs and perceptions may enhance their impact on behavior change.

A primary take-away from the Beacon Community experiences with txt4health and Care4Life is that mHealth technology itself is not a "silver bullet". As is increasingly evident in the adoption of many other health IT tools, the full value of mHealth will be realized only when attitudes, behaviors, and health care delivery also change.

Acknowledgments

Support for the development of this paper was provided through the Office of the National Coordinator for Health Information Technology's Beacon Community Cooperative Agreement Program. The authors would like to acknowledge the following for their work on the txt4health and Care4Life campaigns: the Utah Beacon Community Team; the Southeast Michigan Beacon Community Team; the Crescent City Beacon Community Team (including Anjum Khurshid, Snigdha Mukherjee, and Lisanne Brown); Ann Albright, CDC, Diabetes Translation; Judith McDivitt, CDC, National Diabetes Education Program; the New Orleans Txt4Health Community Advisory Board; Pamela Johnson and Martin Salvi, Voxiva. The authors would also like to thank Margo Edmunds for her early review and feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Voxiva, txt4health, and Care4Life.

[PDF File (Adobe PDF File), 204KB - jmir_v15i10e221_app1.pdf]

References

1. World Health Organization. mHealth: New horizons for health through mobile technologies. Global Observatory for eHealth series 2011;3:1-102 [FREE Full text]
2. Smith A. Pew Internet and American Life Project. 2013 Jun 05. Smartphone Ownership 2013 URL: http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013.aspx [accessed 2013-08-08] [WebCite Cache ID 6Ij4yKJaQ]
3. Duggan M, Rainie L. Pew Internet and American Life Project. Washington, DC: Pew Internet and American Life Project; 2012 Nov 25. Cell phone activities 2012 URL: http://pewinternet.org/Reports/2012/Cell-Activities.aspx [accessed 2013-06-20] [WebCite Cache ID 6KCPNJfBR]
4. 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]
5. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Apr;32(1):56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
6. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of Text Messaging-based Interventions for Health Promotion: A Meta-Analysis. Social Science & Medicine 2013;97:41-48 [FREE Full text] [doi: 10.1016/j.socscimed.2013.08.003]
7. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
8. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
9. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
10. 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]
11. Yun TJ, Arriaga RI. A text message a day keeps the pulmonologist away. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2013 Presented at: CHI '13: 2013 ACM SIGCHI Conference on Human Factors in Computing Systems; 2013; Paris, France p. 1769-1778 URL: http://www.cc.gatech.edu/~arriaga/YunArriagaCHI13.pdf
12. Petrie KJ, Perry K, Broadbent E, Weinman J. A text message programme designed to modify patients' illness and treatment beliefs improves self-reported adherence to asthma preventer medication. Br J Health Psychol 2012 Feb;17(1):74-84. [doi: 10.1111/j.2044-8287.2011.02033.x] [Medline: 22107110]
13. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: A pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]

14. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
15. Hoff A, Nunez-Alvarez A, Martínez KM, Lacoursiere DY. CSU San Marcos National Latino Research Center. Maternal & newborn health: Text4baby San Diego evaluation overview: October 2011-October 2012 URL: http://www.csusm.edu/nlrc/documents/report_archives/Text4Baby_SanDiego_Evaluation_Overview.pdf [accessed 2013-06-20] [WebCite Cache ID 6HWbfAq9l]
16. 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]
17. Donaldson EL, Fallows S, Morris M. A text message based weight management intervention for overweight adults. J Hum Nutr Diet 2013 Jun 6:n/a. [doi: 10.1111/jhn.12096] [Medline: 23738786]
18. 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]
19. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
20. Williams AD. Use of a text messaging program to promote adherence to daily physical activity guidelines: a review of the literature. Bariatric Nursing and Surgical Patient Care 2012 Mar;7(1):13-16. [doi: 10.1089/bar.2012.9999]
21. Snuggs S, McRobbie H, Myers K, Schmocker F, Goddard J, Hajek P. Using text messaging to prevent relapse to smoking: intervention development, practicability and client reactions. Addiction 2012 Dec;107(Suppl 2):39-44. [doi: 10.1111/j.1360-0443.2012.04084.x] [Medline: 23121358]
22. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009 Mar;15(1):17-25 [FREE Full text] [doi: 10.1177/1460458208099865] [Medline: 19218309]
23. Fox S, Duggan M. Pew Internet and American Life Project. Washington, DC; 2012 Nov 08. Mobile Health 2012 URL: http://www.pewinternet.org/Reports/2012/Mobile-Health.aspx [accessed 2013-06-20] [WebCite Cache ID 6KCPVlEll]
24. McKethan A, Brammer C, Fatemi P, Kim M, Kirtane J, Kunzman J, et al. An early status report on the Beacon Communities' plans for transformation via health information technology. Health Aff (Millwood) 2011 Apr;30(4):782-788. [doi: 10.1377/hlthaff.2011.0166] [Medline: 21471501]
25. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]
26. Care4Life. Voxiva URL: https://www.voxiva.com/index.php/products/diabetes [accessed 2013-08-16] [WebCite Cache ID 6IvI0cI7w]
27. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
28. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
29. Patrick H, Williams GC. Self-determination theory: its application to health behavior and complementarity with motivational interviewing. Int J Behav Nutr Phys Act 2012;9:18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]
30. Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]
31. Shaw RJ, Bosworth HB, Hess JC, Silva SG, Lipkus IM, Davis LL, et al. Development of a theoretically driven mHealth text messaging application for sustaining recent weight loss. JMIR Mhealth Uhealth 2013 May 07;1(1):e5. [doi: 10.2196/mhealth.2343]
32. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
33. Jackson S. FierceMobileHealthcare. 2011 Jul 29. Mobile technology brings equal parts risk, reward to medical liability URL: http://www.fiercemobilehealthcare.com/story/mobile-technology-brings-equal-parts-risk-reward-medical-liability/2011-07-29 [accessed 2013-06-20] [WebCite Cache ID 6HWHyWxwJ]
34. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
35. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
36. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
37. Vodafone mHealth Solutions. Evaluating mHealth adoption barriers: Human behaviour URL: http://mhealth.vodafone.com/global/health_debate/insights_guides/human_behaviour/index.jsp [accessed 2013-06-21] [WebCite Cache ID 6HXldKmqP]

Abbreviations

ADA: American Diabetes Association
EHR: electronic health record
IRB: Institutional Review Board
NDEP: National Diabetes Education Program
RDD: random digit dialing

Edited by G Eysenbach; submitted 28.06.13; peer-reviewed by R Whittaker, R Furberg, M McKinley; comments to author 24.07.13; revised version received 30.08.13; accepted 06.09.13; published 15.10.13.

Please cite as:
Abebe NA, Capozza KL, Des Jardins TR, Kulick DA, Rein AL, Schachter AA, Turske SA
Considerations for Community-Based mHealth Initiatives: Insights From Three Beacon Communities
J Med Internet Res 2013;15(10):e221
URL: http://www.jmir.org/2013/10/e221/
doi:10.2196/jmir.2803
PMID:24128406

Review

Evidence of Effectiveness of Health Care Professionals Using Handheld Computers: A Scoping Review of Systematic Reviews

Sharon Mickan[1*], BOccThy, MA, PhD; Julie K Tilson[2*], DPT, MS; Helen Atherton[3*], MSc, MPH, PhD; Nia Wyn Roberts[3*], BA(Hons), MSc Econ; Carl Heneghan[1*], BM, BCh, MA (Hons) Oxon, MRCGP, DPhil

[1]Centre for Evidence Based Medicine, Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
[2]Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, United States
[3]Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Sharon Mickan, BOccThy, MA, PhD
Centre for Evidence Based Medicine
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865289352
Fax: 44 01865289352
Email: sharon.mickan@phc.ox.ac.uk

Abstract

Background: Handheld computers and mobile devices provide instant access to vast amounts and types of useful information for health care professionals. Their reduced size and increased processing speed has led to rapid adoption in health care. Thus, it is important to identify whether handheld computers are actually effective in clinical practice.

Objective: A scoping review of systematic reviews was designed to provide a quick overview of the documented evidence of effectiveness for health care professionals using handheld computers in their clinical work.

Methods: A detailed search, sensitive for systematic reviews was applied for Cochrane, Medline, EMBASE, PsycINFO, Allied and Complementary Medicine Database (AMED), Global Health, and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases. All outcomes that demonstrated effectiveness in clinical practice were included. Classroom learning and patient use of handheld computers were excluded. Quality was assessed using the Assessment of Multiple Systematic Reviews (AMSTAR) tool. A previously published conceptual framework was used as the basis for dual data extraction. Reported outcomes were summarized according to the primary function of the handheld computer.

Results: Five systematic reviews met the inclusion and quality criteria. Together, they reviewed 138 unique primary studies. Most reviewed descriptive intervention studies, where physicians, pharmacists, or medical students used personal digital assistants. Effectiveness was demonstrated across four distinct functions of handheld computers: patient documentation, patient care, information seeking, and professional work patterns. Within each of these functions, a range of positive outcomes were reported using both objective and self-report measures. The use of handheld computers improved patient documentation through more complete recording, fewer documentation errors, and increased efficiency. Handheld computers provided easy access to clinical decision support systems and patient management systems, which improved decision making for patient care. Handheld computers saved time and gave earlier access to new information. There were also reports that handheld computers enhanced work patterns and efficiency.

Conclusions: This scoping review summarizes the secondary evidence for effectiveness of handheld computers and mhealth. It provides a snapshot of effective use by health care professionals across four key functions. We identified evidence to suggest that handheld computers provide easy and timely access to information and enable accurate and complete documentation. Further, they can give health care professionals instant access to evidence-based decision support and patient management systems to improve clinical decision making. Finally, there is evidence that handheld computers allow health professionals to be more efficient in their work practices. It is anticipated that this evidence will guide clinicians and managers in implementing handheld computers in clinical practice and in designing future research.

(J Med Internet Res 2013;15(10):e212) doi:10.2196/jmir.2530

KEYWORDS

handheld computers; mobile devices; mhealth; PDA; information seeking behavior; evidence-based practice; delivery of health care; clinical practice; health technology adoption; diffusion of innovation; systematic review; evidence synthesis; documentation

Introduction

Handheld computing devices are changing health care delivery. Clinicians now have instant access to vast amounts of information, including x-ray results, laboratory tests, databases of primary and pre-appraised research, clinical practice guidelines, and drug reference guides. The evolution of handheld computers—smaller, more versatile, and capable of Internet connectivity—has prompted increasing usage by health care professionals. In 2003, 40% of physicians were reported to own a PDA (personal digital assistant) [1], and by 2011, over 68% of doctors in the United Kingdom were reported to own a smartphone [2]. Among medical students, 70% reported owning a smartphone in 2006 [3], increasing to 79% in 2011 [2]. There are many examples of handheld computer use in health care, including electronic prescribing, patient diagnosis and advice, patient review, practice management, reminder notifications, and eLearning.

Given the fast pace of technological innovation, the use of handheld computers has preceded definitive research about clear benefits. To date, most research has evaluated patterns of usage and adoption [4]. However, it is important to understand whether handheld computers are effective and in what settings they demonstrate improved patient care or lead to efficiencies in health care delivery. Syntheses of research evidence offer a high quality and practical way to review the existing research base. This review will scope the evidence of effectiveness across all aspects of health care practice by reviewing systematic reviews, to identify documented positive outcomes.

Methods

Inclusion Criteria

This review included systematic reviews published between 1992 and 2012, of all quantitative study designs, that described effective use of handheld computers by health care professionals. We defined handheld computers, consistent with the MeSH (Medical Subject Headings) term, as small, portable, and fitting in the hand. We were particularly interested in commercially available tools that health care professionals could carry with them in clinical environments. Outcomes were not pre-specified, and all aspects of demonstrated effectiveness in clinical practice were included.

Exclusion Criteria

Systematic reviews were excluded when the focus was on the patients' use of handheld computers, when students were learning in a classroom, and when only laptop computers were included. Systematic reviews were also excluded when they only described patterns of usage and when they focused on evaluating electronic medical records as stand-alone systems.

Search

The following databases were searched on June 7, 2012, and December 11, 2012: Cochrane Database of Systematic Reviews (CDSR) and Database of Abstracts of Reviews of Effectiveness (DARE), Medline, EMBASE, PsycINFO, Allied and Complementary Medicine Database (AMED), Global Health, and Cumulative Index to Nursing and Allied Health Literature (CINAHL). Free-text terms and subject headings to describe handheld computers and health professionals were used as a basis of the search strategy, and these terms covered both older and newer devices (see Multimedia Appendix 1 for search strategy). Sensitive search filters developed by the Health Information Research Unit at McMaster University, Hamilton, Canada, were applied to focus the search on systematic reviews.

Assessment of Quality

All relevant systematic reviews were independently appraised by 2 authors using the Assessment of Multiple Systematic Reviews (AMSTAR) tool [5]. This 11-item evaluation tool assesses methodological quality, presentation, and the risk of bias in systematic reviews. Systematic reviews that did not report a comprehensive search strategy or scored less than 5 out of a possible total of 11 items were excluded.

Data Extraction

Data from included reviews were extracted independently by 2 authors to record the population studied, purpose of the review, search time frame, number and design of included studies, types of handheld computer included, and outcomes reported. A conceptual framework proposed by Free et al [6] was used to create a standardized template for data extraction. Several additional categories were created using an iterative process that involved fitting the data to amended versions of the original framework. This piloting and iterative refinement was carried out by SM and HA and continued until agreement was reached on the most appropriate categories for the data. A new template, which was used to extract objective and self-reported outcomes, summarized them according to the primary function for which the handheld computer was being used (eg, information seeking, patient data collection).

Data Synthesis

It was expected that high levels of data heterogeneity would preclude statistical synthesis. A narrative approach was planned to summarize the evidence for effectiveness of handheld computers to support clinical practice, with respect to the primary function of the handheld computer. This involved presenting the results of each review using summary text, according to the relevant categories as determined at the data extraction stage.

Results

Overview

The initial search identified 506 systematic reviews. Of these, 21 were read for inclusion and assessed for quality using the AMSTAR checklist. Five systematic reviews met the inclusion and quality criteria (Figure 1). Included reviews scored between 5 and 8 of 11 possible points on AMSTAR (Table 1).

Table 2 describes the characteristics of the five included systematic reviews. Physicians, pharmacists, and medical students were the most common populations studied. One hundred and thirty eight unique primary studies contributed to these reviews and were published between 1995 and 2008. Of these 138 primary studies, seven were included in three of the included reviews and 19 in two of the included reviews. The lack of overlap of primary studies across these five reviews highlights the inherent heterogeneity of the field and is also reflective of the differing research questions each review addressed in relation to handheld effectiveness. Most were descriptive intervention studies, and only 8 randomized controlled trials (RCTs) were identified. All studies described handheld computers as PDAs with some having Internet connectivity and others not.

Effectiveness could be categorized across four distinct functions of handheld computers, and all five reviews identified evidence for each of the four functions (Table 3): (1) patient documentation, (2) patient care, (3) information seeking, and (4) professional work patterns. Within each function, a range of positive outcomes were reported using both objective and self-report measures.

Table 1. Quality evaluation of included studies.

Quality criteria	Lindquist et al, 2009	Prgomet et al, 2009	Fox et al, 2007	Kho et al, 2006	Lu et al, 2005
1. Was an a priori design provided?	1	1	1	1	1
2. Was there duplicate study selection and data extraction?	1	1	0	1	0
3. Was a comprehensive literature search performed?	1	1	1	1	1
4. Was the status of publication (ie, grey literature) used as an inclusion criterion?	1	1	0	1	1
5. Was a list of studies (included and excluded) provided?	1	1	0	0	1
6. Were the characteristics of the included studies provided?	1	1	1	0	1
7. Was the scientific quality of the included studies assessed and documented?	0	1	1	0	0
8. Was the scientific quality of the included studies used appropriately in formulating a conclusion?	0	0	1	0	0
9. Were the methods used to combine the findings of studies appropriate?	1	1	1	1	1
10. Was the likelihood of publication bias assessed?	0	0	0	0	0
11. Was the conflict of interest stated?	1	0	0	0	0
Total Score	8	8	6	5	6

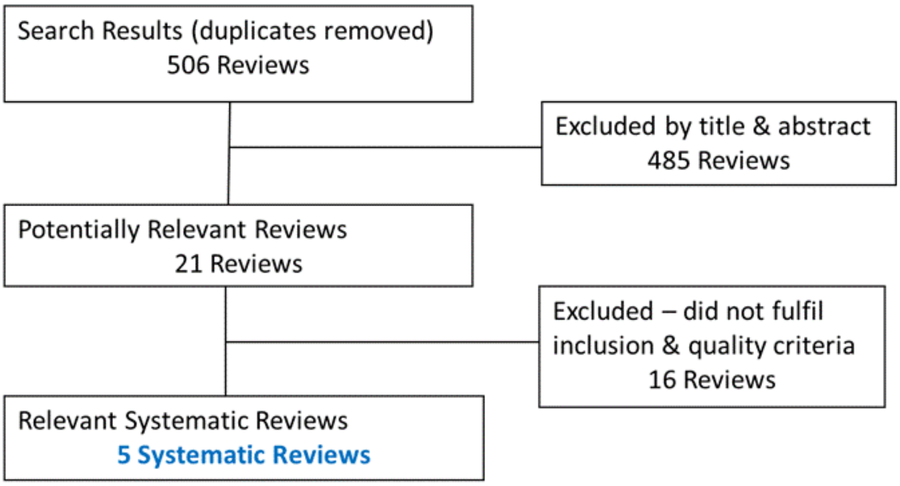
Table 2. Descriptive characteristics of included systematic reviews.

Authors	Title	Population studied	Search time frame	Number of studies	Results
Lindquist et al, 2009	The use of the PDA among personnel and students in health care: a review	Health care professionals and students	1999-2008	48 (6 RCTs)	PDAs are used in patient care by both professionals and students with varied frequency. Their use may improve decision making, reduce number of medical errors and enhance learning.
Prgomet et al, 2009	The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review	Physicians	2000-2006	13 (2 RCTs)	Handheld devices demonstrate greatest benefits in contexts where time is a critical factor; when connecting spatially distributed workers; for overcoming inadequate numbers of computers; and when data access/entry is required at point of care.
Fox et al, 2007	Use of PDAs for documentation of pharmacists' interventions: a literature review	Pharmacists	2001-2006	12	The use of PDAs may increase the frequency and number of interventions documented by pharmacists.
Kho et al, 2006	Use of handheld computers in medical education	Medical staff and students	1995-2004	67	Handheld computers are an important and evolving part of the medical trainee's resources in medical education and patient care.
Lu et al, 2005	A review and framework of handheld computer adoption in health care	Health care professionals	1998-2004	31	Most care providers found PDAs to be functional and useful for documentation and for access to medical references and patient data.

Table 3. Summary of handheld computer functions and reported effective outcomes.

Function of handheld computer	Description of function	Evidence of effectiveness (as documented in included studies)
Patient documentation	Electronic collection and documentation of patient data	More interventions recorded
		Improved documentation rates
		More accurate and detailed description of clinical findings
		More accurate diagnostic coding
		More frequent documentation of side effects
		Increased rate of electronic prescribing
		Fewer errors in discharge lists
		Improved patient identification
		Less information lost
Patient care	Access to electronic evidence based decision support systems, pharmaceutical information, transmission of investigatory images, and access to patient management systems	Reduced prescription error rates
		Fewer unsafe drug treatment decisions
		More changes in drug prescriptions
		Increased self-reported drug knowledge
		Reduced antibiotic prescription
		Decreased average length of stay
		Improved practice efficiency
		Improved diagnosis and patient care
		Shorter intervention times
		More consistent care, according to patient preferences
Information seeking	Looking for information about patients, drugs, guidelines, references, at point of care	Saves time
		More frequent access to electronic resources
		Informs patient education about medication use
		Earlier learning about new developments
Professional work patterns	Integration of handheld computers into work flows to improve efficiency and communication	Integrates well into clinical workflow
		Saves time when retrieving drug information
		Perceived efficient decision making
		Saves time in ward rounds accessing, retrieving, recording data
		More time for direct patient care
		Quicker response times and less failures to respond than with mobile phones and pagers

Figure 1. Criteria flowchart.



Patient Documentation Outcomes

Handheld computers improved patient documentation through more complete records with fewer documentation errors and improved ease and efficiency of documentation. Pharmacists reported improved documentation rates, through recording more interventions and completing more fields [7]. Documentation using PDAs was rated significantly better than paper for detailed description of clinical findings and correct progress assessment [8]. More accurate diagnostic coding and more frequent documentation of side effects were reported [3,8]. The introduction of PDAs significantly increased the average rate of electronic prescribing from 52% to 64% (*P*=.03) [8]. Documentation with PDAs resulted in significantly fewer discrepancies in recording of neonatal patient weight in intensive care (4.4% vs 14.4% [OR 0.29, CI 0.15-0.56]) [8]. When PDAs were used to create discharge order lists, documentation errors were reduced from 22% to 8% (*P*<.05), compared to transcription from paper [1]. An electronic barcode system for identification of patients requiring blood transfusion in the hospital setting was used successfully on a PDA. There were no incidents of blood transfusion to wrong patients or wrong labeling with 41,000 samples over 3 years [9].

Patient Care Outcomes

Improved decision making using handheld and patient management systems was a key benefit. The inclusion of specific intervention rules on handheld devices significantly reduced prescription error rates (0.23 vs 0.45; *P*<.05) [8]. Physicians using a PDA-based CDSS for prescription of nonsteroidal anti-inflammatory drugs made fewer unsafe treatment decisions [9]. Physicians reported that using a drug database developed for a PDA improved their practice efficiency, increased self-reported drug knowledge, and improved patient care [1]. There were twice as many changes in patient management when using electronic resources rather than paper resources (30% vs 18% [OR 2.00, CI 1.11-3.60]), particularly changes in drug prescription (22% vs 13% [OR 1.84, CI 0.95-3.59]) [8].

Physicians reported using PDAs loaded with locally developed guidelines and site-specific laboratory data on average 4 times per day, primarily to access laboratory data. During this 6-month prospective study, use of the PDA led to a significant decrease in antibiotics used from 1925 to 1606 daily doses per 1000 patient days (*P*=.04) and decreased the average length of patient stay by 1 day, from 7.2 to 6.2 bed days (*P*=.02) [8].

Family physicians reported that use of a PDA-based software application for cardiac stress-testing improved diagnosis and care for patients with chest pain [9]. Wireless transmission of investigatory images from PDAs to cardiologists resulted in timely and appropriate ambulance redirection and shorter intervention times [8]. Within this study, the image quality from PDAs was rated as suitable for diagnosis in all cases and identical to reference reports in most cases. Evidence-based guidelines for screening were reported as being fast and easy to use at the point of care [9]. Nurses reported that using a patient management system on a PDA made nursing care more consistent with patient preferences and improved patients' preference achievement [9]. A patient management system available via PDAs in intensive care was described as convenient and functional, especially for patients who had long stays in hospital [9].

Information Seeking Outcomes

Handheld computers have demonstrated effectiveness for supporting health care professionals' information seeking needs. Where PDAs were used for self-directed learning, medical students perceived time savings of around 1 min/encounter [3]; 83% reported being better able to inform patients about medication use when looking at drug reference data [3]. When health care professionals were provided with a PDA with headlines about new books, guidelines, reviews, and medical literature, they reported learning about new developments sooner than without it [9]. Physicians accessed electronic resources via a PDA more often than paper resources (181 vs 131 episodes [OR 1.99, CI 1.41-2.80]), but average time spent in accessing them was similar (9.3 and 9.4 seconds) [8].

Professional Work Pattern Outcomes

Handheld computers can enhance efficiency and improve patterns of work. When a PDA was used for documentation of clinical pharmacy services, 75% of users across several sites indicated that it integrated well into clinical workflow [7]. Physicians reported that PDAs enabled them to save time when retrieving information from a drug database [1]. Use of a PDA led to perceptions of more efficient decision making for patient care [9]. Physicians who utilized PDAs reported improved efficiency of their daily rounds through spending less time accessing, retrieving, and recording data, therefore freeing more time for direct patient care [1]. Median encounter time for each patient was significantly shorter when physicians used PDAs (227 vs 301 seconds) compared to paper [8]. When PDAs were compared to a mobile phone/pager for call outs, they led to shorter response times with fewer failures to respond [8,9].

Discussion

Principal Results

This scoping review has documented the evidence of effectiveness of handheld computers for health care professionals in four functions: patient documentation, patient care, information seeking, and professional work patterns. Across these functions, PDAs appear to provide health care professionals with timely and easy access to relevant information, facilitate accurate and complete documentation, coordinate information at the point of care, and support efficient work flows.

It is important to recognize that the pace of change of technology is faster than the rate of research production in this area. While the included systematic reviews in this review focused on evaluating the effectiveness of PDAs as handheld computers, current practice reflects the widespread use of smartphones, which were only introduced to the market in 2007 [2]. However, technology has changed steadily over time. For example, later PDAs had Internet connectivity and could run specific applications. It is therefore anticipated that the evidence for effectiveness identified in this review will, for the most part, hold true for smartphones. While the devices used may evolve quickly over time, the behaviors and actions of the clinicians using them change at a much slower rate. We can also expect that as hardware and software continue to develop, there will be enhanced and additional benefits. In future updates of this review, we would expect to see systematic reviews of smartphone use. Further, as more patients have smartphones, there are new opportunities for direct communication with health care professionals and for improved self-monitoring and disease prevention. Already, there are many available apps for patient use to enhance wellness through promoting diet and exercise and limiting smoking and alcohol use [10].

Two reviews included in this study superficially addressed issues of cost avoidance and cost savings [1,7]. While savings are likely to be of interest to managers and policymakers, there is need for better understanding of real costs. Medical students and junior doctors have expressed concern about the high costs of smartphones and medical apps [2]. Certainly, widespread implementation of continually evolving handheld computer technology in health care organizations demands economic analyses.

Similarly, the views of health care professionals need to be carefully evaluated in relation to barriers and facilitators of handheld computer use. While positive perceptions about efficiency have been documented, concerns have been raised about lack of user-friendly interfaces [7], encryption of patient data [8], and the practical issues of data crashing [7] and hardware breakage [11]. More recently, doctors have raised concerns about the impact of using smartphones in a clinical environment on the doctor-patient relationship and uncertainty about patients' perceptions and expectations [2].

This study identifies five systematic reviews that provide evidence of the effective use of handheld computers by health care professionals, as a snapshot of current research evidence. It is anticipated that this will provide direction for clinicians and managers who may be implementing handheld computers in clinical practice and for designing future research. The clinicians of 2012 used smartphones and tablets rather than PDAs, but the lessons to be learned from the use of PDAs should not be discounted; technology has become more sophisticated but facilitates similar actions.

Study Limitations

In order to quickly summarize the research evidence within this fairly young academic field, this scoping review included only the evidence of effectiveness reported within included systematic reviews. Reviews of systematic reviews provide a succinct overview of the field, with a focus on studies representing the highest quality of evidence synthesis [12]. While this methodology is especially useful where there is heterogeneity of study design and outcomes, it has limitations. Being removed from the primary data by two levels poses difficulties in synthesizing the resulting data. Although primary studies were checked when deemed necessary, this paper summarized only the data provided by the systematic review authors.

We recognize that neither this scoping review nor any of the included systematic reviews were able to statistically pool effectiveness data. We also acknowledge the high potential for bias associated with the predominance of low quality primary studies. It is also likely that key benefits may have been missed because of measurement limitations in primary studies. Further, it lacks a balanced evaluation of effectiveness for and against each of the functions identified. The evidence reported in this review is also subject to a significant time lag in research production. Over time, systematic reviews of smartphone use will begin to proliferate, but at the time of our search, we did not identify any that met our inclusion criteria.

Areas for Future Research

While this review has presented evidence for better access to patient results and reductions in adverse events and hospital length of stay, there is a need to replicate and better understand these benefits. Effectiveness of handheld computers also needs to be reviewed from the perspectives of patients, health care students, and health care organizations. There is also a need to understand the mechanisms by which handheld computers support clinical practice, and this may require complementary qualitative and mixed methods studies.

Conclusions

There is emerging evidence of effectiveness for the use of handheld computers by health care professionals across a variety of functions that support clinical practice. Handheld computers appear to provide easy and timely access to information and to support more accurate and complete documentation. They can also provide access to evidence-based decision support and patient management systems that improve clinical decision making for patient care. Finally, there is evidence that handheld computers allow health professionals to be more efficient in their work practices, thereby allowing more time for patient contact. This evidence may guide clinicians, managers, and researchers in incorporating the growing number of ever more sophisticated devices into routine clinical practice and future research. We should utilize it in assessing whether emerging devices are living up to their hype.

Acknowledgments

This review was completed without external funding. Special thanks to Katie Marshall, research assistant, for her contribution to data extraction.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Complete search strategy.

[PDF File (Adobe PDF File), 59KB - jmir_v15i10e212_app1.pdf]

References

1. Lu YC, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [doi: 10.1016/j.ijmedinf.2005.03.001] [Medline: 15893264]
2. Payne KB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
3. Kho A, Henderson LE, Dressler DD, Kripalani S. Use of handheld computers in medical education. A systematic review. J Gen Intern Med 2006 May;21(5):531-537 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00444.x] [Medline: 16704405]
4. Garritty C, El Emam K. Who's using PDAs? Estimates of PDA use by health care providers: a systematic review of surveys. J Med Internet Res 2006;8(2):e7 [FREE Full text] [doi: 10.2196/jmir.8.2.e7] [Medline: 16867970]
5. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
6. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
7. Fox BI, Felkey BG, Berger BA, Krueger KP, Rainer RK. Use of personal digital assistants for documentation of pharmacists' interventions: a literature review. Am J Health Syst Pharm 2007 Jul 15;64(14):1516-1525. [doi: 10.2146/ajhp060152] [Medline: 17617503]
8. 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]

9. Lindquist AM, Johansson PE, Petersson GI, Saveman BI, Nilsson GC. The use of the Personal Digital Assistant (PDA) among personnel and students in health care: a review. J Med Internet Res 2008;10(4):e31 [FREE Full text] [doi: 10.2196/jmir.1038] [Medline: 18957381]
10. 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]
11. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
12. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]

Abbreviations

AMSTAR: Assessment of Multiple Systematic Reviews
CDSS: computer-based clinical decision support systems
OR: odds ratio
PDA: personal digital assistant
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 16.01.13; peer-reviewed by H Oh, H Wharrad; comments to author 14.02.13; revised version received 22.02.13; accepted 11.05.13; published 28.10.13.

Please cite as:
Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C
Evidence of Effectiveness of Health Care Professionals Using Handheld Computers: A Scoping Review of Systematic Reviews
J Med Internet Res 2013;15(10):e212
URL: http://www.jmir.org/2013/10/e212/
doi:10.2196/jmir.2530
PMID:24165786

Original Paper

Interpretive Flexibility in Mobile Health: Lessons From a Government-Sponsored Home Care Program

Jeppe Agger Nielsen[1], PhD; Lars Mathiassen[2], PhD

[1]Aalborg University, Department of Political Science, Center for Organization, Management & Administration, Aalborg, Denmark

[2]J Mack Robinson College of Business, Center for Process Innovation, Georgia State University, Atlanta, GA, United States

Corresponding Author:
Jeppe Agger Nielsen, PhD
Aalborg University
Department of Political Science, Center for Organization, Management & Administration
Fibigerstraede 1
Aalborg, 9220
Denmark
Phone: 45 99408212
Fax: 45 98155346
Email: agger@dps.aau.dk

Abstract

Background: Mobile technologies have emerged as important tools that health care personnel can use to gain easy access to client data anywhere. This is particularly useful for nurses and care workers in home health care as they provide services to clients in many different settings. Although a growing body of evidence supports the use of mobile technologies, the diverse implications of mobile health have yet to be fully documented.

Objective: Our objective was to examine a large-scale government-sponsored mobile health implementation program in the Danish home care sector and to understand how the technology was used differently across home care agencies.

Methods: We chose to perform a longitudinal case study with embedded units of analysis. We included multiple data sources, such as written materials, a survey to managers across all 98 Danish municipalities, and semistructured interviews with managers, care workers, and nurses in three selected home care agencies. We used process models of change to help analyze the overall implementation process from a longitudinal perspective and to identify antecedent conditions, key events, and practical outcomes.

Results: Strong collaboration between major stakeholders in the Danish home care sector (government bodies, vendors, consultants, interest organizations, and managers) helped initiate and energize the change process, and government funding supported quick and widespread technology adoption. However, although supported by the same government-sponsored program, mobile technology proved to have considerable interpretive flexibility with variation in perceived nature of technology, technology strategy, and technology use between agencies. What was first seen as a very promising innovation across the Danish home care sector subsequently became the topic of debate as technology use arrangements ran counter to existing norms and values in individual agencies.

Conclusions: Government-sponsored programs can have both positive and negative results, and managers need to be aware of this and the interpretive flexibility of mobile technology. Mobile technology implementation is a complex process that is best studied by combining organization-level analysis with features of the wider sociopolitical and interorganizational environment.

(J Med Internet Res 2013;15(10):e236) doi:10.2196/jmir.2816

KEYWORDS

home health care; mobile health; mobile technology; implementation process; government sponsorship; case study

Introduction

Health care information technology (HIT) has the potential to produce increased quality and efficiency of service delivery [1]. However, HIT implementation is not a straightforward process. It is often as messy as it is exciting, and at times, it may turn into a battlefield where progress occurs through a combination of both “muddling through” and rational decision making [2,3]. Accordingly, when managers implement mobile health, they will likely face both opportunities and challenges.

Mobile health, or mHealth, involves “emerging mobile communications and network technologies for healthcare systems” [4]. The hardware includes laptops, personal digital assistants (PDAs), and smartphones, with more advanced devices integrating and combining functionality [5]. Mobile devices are increasingly involved in many aspects of health care delivery [5-8] because they offer great benefits compared to using personal computers—most importantly the capacity to access information and complete various functions in real time at the point of care [9-11]. Despite these obvious advantages, studies have raised issues related to implementing mobile health care [12-19].

Some researchers have found that the uptake of mobile health systems is more limited than what one might expect from the optimistic tone in the field [12]. Others have highlighted privacy concerns [13], end-user resistance to change [14], lack of adequate training and management support [15], and technical issues suggesting more attention should be given to the overall architecture of the mobile health system and to user interfaces [16]. Another study illustrated how users found laptop computers easier, faster, and more satisfying to use than handheld computers in the data recording process [17]. It has also been demonstrated that mobile devices provide a reservoir of bacteria known to cause infections within the hospital area [8]. Although these studies have enriched our understanding of the impact of mobile health systems, they most often report from pilot projects or from very restricted contexts.

Against this backdrop, the objective of this study was to contribute to mHealth research by examining a large-scale mHealth implementation project in the Danish home care sector. In this context, mobile technology has spread quickly since the mid-2000s, and today most home care agencies have invested in PDAs or smartphones for their health care personnel. Drawing on multiple sources of data covering the period 1998-2008, we demonstrate how mobile technology implementation offered new opportunities and challenges as key stakeholders debated visions for use of mobile technology within the Danish home care sector and transformed health care practices in individual home care agencies (HCAs).

Mobile health dates to the 1990s. The PDA was introduced by Apple in 1993 and became a household product by the end of the 1990s [20]. Parallel to this progress, mobile devices started being used in health care settings in a number of western countries [11]. For instance, in Danish home care, pilots were initiated in 1998 as some agencies tested mobile devices among care workers and nurses. Today, PDAs and smartphones are widely used by health care professionals in most OECD countries [20], and expectations of the transformative potential of mHealth are massive [21] as mobile technologies represent promising new ways in assisting health care professionals as they access, manage, and share critical information at the point of care.

The fact that mobile devices are used by an increasing number of primary staff in health care has attracted considerable research attention. Early studies of mHealth focused on its potential benefits, opportunities, and barriers [22,23]. Comprehensive literature reviews found that mobile devices were widely used by health care professionals and that their use was expected to increase significantly in the years to come [6,20]. Rothschild et al examined US doctors’ use of mobile technology, concluding that doctors found them “to improve patient care and be valuable in learning of recent alerts and warnings” (p. 619 [10]). Overall, mHealth has been shown to improve communication among health care staff [5].

More recently, investigations have presented a set of factors that help explain the adoption of mobile health care systems [7,24-28]. Park and Chen emphasized perceived usefulness and perceived ease of use as key factors for both physicians and nurses in their use of smartphone technology [24]. Similarly, Zhang et al concluded that nurses’ view of usefulness is the main factor in the adoption of mobile technology [25]. Some Scandinavian studies have focused on mobile technology in home care [29]. In a study of Finnish home care agencies, Vuokko illustrated how the introduction of mobile technology impacts home care work and creates concerns among staff related to issues of control, surveillance, and distrust of the management while at the same time, they see benefits in terms of better coordination and documentation [29]. Finally, researchers have started to investigate factors that impact mHealth adoption, usage, and channel preferences from a client perspective [21]. In this line of research, it is demonstrated how mHealth applications can empower clients to track and manage their own health [30].

We go beyond this literature by trying to understand the processes through which mHealth is adopted and implemented into health delivery practices [31-33]. Process-oriented approaches can reveal important lessons on how to manage new technologies, and they have previously been used with success to address the complexities involved in HIT implementation [1,34,35]. These studies suggest technologies have interpretive flexibility [36] as various stakeholders construct the meaning of the technology differently. We assume such an approach may prove useful in exploring mHealth by emphasizing differences in how stakeholders perceive the nature of technology, technology strategy, and technology use [37].

Conceptualizing process “as a sequence of events that describes how things change over time” [33], we emphasize that change may very well unfold differently depending on the context in which organizations are embedded. Following Newman and Robey [38] and Langley [39], we distinguish between antecedent conditions, key events, and practical outcomes and use temporal bracketing to highlight the important phases through which the process unfolds. Process data, in particular from longitudinal studies, are indeed complex. It is therefore important to adopt analytical approaches that can help manage this complexity and bring forward valuable insights and lessons. Contextual considerations are also highly relevant in process studies [32,40] as they can help us understand how organizational implementation of HIT, such as mHealth, is shaped by the wider sociopolitical and interorganizational environment [41-43]. This adds further to the number of variables studied, suggesting a case study design with in-depth examination of how the context and the contents of HIT implementation are formed over time [44,45].

Methods

The research is organized as a longitudinal case study with embedded units of analysis of the implementation of mobile technology into Danish home care agencies. We examined how mobile technology implementation unfolded in three specific home care agencies and complemented this organization-level analysis with a perspective of the broader home care sector by including an examination of how government, IT vendors, and interest groups were involved in shaping the implementation process. We followed the implementation process over a 10-year period from 1998 when the first initiatives were taken to 2008 when technologies were in use in the majority of Danish municipalities.

The Danish health care sector is organized into 5 regions with responsibility for hospitals and 98 municipalities which, according to the Danish Social Services Act, are responsible for home care to help the elderly and disabled cope with everyday life [46]. Even though clients can choose private providers and have the costs compensated by government (the so-called Free Choice model), home care services are predominately delivered by the public sector [47]. Home care services are long-term or temporary. Long-term home care is provided free of charge while citizens may be asked to subsidize the costs of temporary home care depending on level of income [48].

There are approximately 700,000 people over the age of 67 in Denmark. Of these, more than 160,000 (2011) receive long-term home care and a further 8000 people receive temporary home care. Home care involves daily-living assistance in clients' own homes including a wide range of nursing and care services such as coordination with post-acute care, assistance with medication, personal hygiene and care, cleaning, shopping, and preparation of meals. In Denmark, the average duration of hospitalization has been remarkably reduced, which means that the home care sector has experienced increasingly complex tasks. Internationally, Denmark is rated as a leader in the area of home care services with the most far-reaching public-financed services for the elderly. Approximately 70,000 full-time care workers and 6000 nurses are employed in the sector [49].

We collected primary data between June 2007 and May 2008, beginning with fly-on-the-wall observations as the first author followed nurses and care workers “at work with PDAs” in a specific home care agency. Although the lessons learned from this pilot are not directly part of the data sources in this article, these initial observations provided valuable insights and improved our understanding of the research context and key stakeholders involved. For this study, we used multiple data sources, including written materials, a survey to managers across all 98 Danish municipalities, combined with semistructured interviews with managers, care workers, and nurses in three selected home care agencies, as summarized in Table 1. (See Multimedia Appendix 1 for more information about the three cases.)

Interviews were tape-recorded and transcribed. Using all our data, we constructed a chronology of the major events that took place during 1998-2008 [39,50]. Also, antecedent conditions and outcomes were identified [38,40] as summarized in Figure 1. To improve reliability, the analysis was presented to key informants in the home care sector and consequently revised [51].

Below, we present the results of our analyses in two steps. First, we provide an overview of the implementation process in the home care sector as summarized in Figure 1. Next, we provide detailed insights into the implementation of mobile technology in three selected home care agencies (municipalities).

Table 1. Data sources.

Data sources	Description
Survey to home care managers in all 98 Danish municipalities to track the adoption rate of mobile technology	Managers were interviewed over the phone in June 2007, which enabled a 100% response rate. We used a structured interview guide and inquired about: How many home care agencies are using mobile technology? When did they start using the technology? Which groups of employees make use of mobile technology? We also inquired about their motives for adopting or rejecting the technology, sources of inspiration, and the importance of governmental subsidies.
Written materials and interviews with key stakeholders in the home care sector	We reviewed government, consultants, and vendor websites for available written materials as a way to further our understanding of how mobile technology implementation was shaped in interplay with the broader context. We interviewed representatives from Local Government Denmark (LGDK, a major interest group for municipalities), the Ministry of Social Affairs, and Ministry of Finance as they were leading in the discussion on innovating home care by using mobile technology. The interviews were conducted in May 2008 and lasted on average about 1 hour.
Three cases of mobile technology implementation within specific home care agencies (HCA 1, HCA 2, & HCA 3).	The three selected home care agencies had each used mobile technology for some time (respectively 5, 2, and 2 years), which allowed us to achieve comprehensive insight into the implementation process and how the technology was used in day-to-day practices. We conducted semistructured interviews in each case with key stakeholders: managers and employees using mobile technology (nurses and care workers). In total, 10 managers or project managers (respectively 4, 4, and 2 in each case) and 24 employees (respectively 7, 8, and 9 in each case) were interviewed. We organized a protocol to structure the interview process and personalized it for specific stakeholder groups. For instance, the protocol for managers included questions that permitted the managers to express how they perceived the nature of mobile technology, the implementation strategy, and mobile technology in use, but also who they saw as the major sources of inspiration and their collaboration with IT vendors and other stakeholders. We conducted a survey of care workers across the three selected home care agencies (N=315, response rate 63%), particularly focusing on care workers' perceptions towards the mobile technology and how they use the technology in daily practice. Written documents (eg, project descriptions, minutes from meetings, and evaluations) were collected in each case. While interviews enlighten the more informal processes and struggles surrounding mobile technology implementation, written documents identified the formal motives behind mobile technology implementation.

Figure 1. Implementation of mobile health in the Danish home care sector.

- Mobile technlogies (eg, PDAs) devolped, tested and used in other settings.
- Electronic Care Records implemented in home care agencies.
- Government reforms shaped the need for technological support of work processes in home care.

- A few IT vendors developed quite simple mobile health solutions based on offline technology.
- A few home care agencies tested mobile devices among care workers and nurses.
- Mobile health was presented as a interesting innovation, but early experiments were hampered by technical constraints.

- Large-scale pilot (CareMobile) involved strong collaboration between key stakeholders in home care and provided a sound business case for mobile technology use.
- IT providers developed advanced mobile health systems based on online technology.
- The usage of mobile technology expanded with more experiments in home care agencies.

- Mobile technology attracted increasing political attention and support.
- Government sponsoring supported quick and widespread adoption.
- A more fragmented and somewhat critical view of mobile technology evolved.

- Mobile technology in use in the majority of home care agencies.
- Ongoing debate over mobile technologies role in improving home care.
- Tension between a managerial and professional logic.

Results

Process Analysis

Antecedents

Danish home care has been through a series of New Public Management-inspired [52] reforms since the mid-1990s aiming to improve accountability and efficiency [47]. These reforms shaped the need for technological support of work in home care, for example, to meet requirements for transparency and more accurate documentation. The first important review of IT usage in home care was carried out in 1994. The study found that IT systems were used on a very small scale, and roughly 90% of all administrative tasks were handled manually [53].Yet, the possibility of using new technology in home care was boosted in the mid-1990s as the Common Language reform established standards and data models, which became common to all IT vendors providing electronic care records and mHealth systems in home care agencies [54]. Accordingly, home care agencies increased their IT usage during the 1990s; for instance, the majority implemented electronic care records. These systems are tied to the electronic medical record systems in hospitals and include a comprehensive database with client information. Based on these databases, mHealth systems were developed giving health care personal access to client information at the point of care. Since the first tests in the late 1990s, more sophisticated devices were implemented as the mobile network increasingly involved online solutions as a replacement for offline solutions. Whereas offline solutions imply that care workers download and upload client data to mobile devices at the office, online solutions afford access to and the update of centrally stored client data in real-time at the point of care.

Phase 1 (1998-2001)

Initial experiments with mobile technology occurred in the late 1990s and have expanded significantly since. The first home care agency (Municipality of Græsted-Gilleleje) started in 1998 as a group of nurses and care workers tested handheld devices. IT vendors and consultants played a decisive role in this early phase. They developed technological solutions based on offline connectivity and worked closely with early adopter organizations by actively engaging in the implementation of pilots. Mobile technology was promoted by IT vendors and managers as an interesting innovation and a fresh way to modernize public home care, but despite optimistic announcements, projects were hampered by technical difficulties, and initial projects were suspended after pilots.

Phase 2 (2002-2005)

Whereas IT vendors and consultants inspired home care agencies early on in the change process, government bodies started to play a more vital role in the next phase. IT vendors started to develop more advanced mHealth systems based on online technology, but it was a large-scale sector-wide pilot, CareMobile, that was launched in 2002 and reported in 2005 that positioned mHealth on the wider political agenda [55]. The CareMobile project was managed in collaboration between the Ministry of Finance, Ministry of Social Affairs, and LGDK, plus it included several IT vendors and consultancy companies as well as six pilot municipalities. CareMobile offered a sound business case for mobile technology adoption, and the final evaluation highlighted that the technology was mature and that investment could be gained in 1 year, as meeting activities and duplicate data entry could be avoided [55]. It was estimated that administrative tasks in home care could be reduced by more than 3000 full-time positions if mobile technology were adopted by all municipalities [56,57]. Accordingly, the usage of mobile technology expanded in this phase with more experiments in home care agencies. mHealth was put on the political agenda, most prominently through the CareMobile project. Furthermore, CareMobile served as an important activity in building legitimacy for mobile technology adoption and use.

Phase 3 (2006-2007)

While mobile technology until 2005 was reserved for an exclusive group of home care agencies, the following years resulted in widespread adoption. In 2007, 76 of 98 municipalities (78%) had adopted mobile technology and another 13 (13%) expected to implement in 2008 (see Multimedia Appendix 1). In this phase, mHealth gradually attracted more political consideration and support. In the ICT strategy for the social sector, mHealth was presented as a high priority area [58]. Key ministers, including the Prime Minister, emphasized on several occasions the benefits of mobile technology [59]. Finally in 2006, the adoption of mobile technology started to accelerate when the government decided to support implementation with approximately €45 million. The positive results from the CareMobile project provided the rationale for allocating government subsidies to mHealth [60]. Subsequently, 66% of all municipalities responded that government funding had decisive importance for their adoption of mobile technology (Table 2). In this sense, the sponsoring activity was very effective. Whereas the earlier phases were characterized by a high degree of consensus (“mobile health is good”), this phase provided a more fragmentary view of mHealth. Consistent with the rapid dissemination of the technology, the powerful DaneAge Association and the trade unions articulated rather critical views emphasizing that mHealth was associated with unnecessary organizational control [61].

Table 2. The importance of government funding for mobile technology adoption (%) (source: Survey to home care managers in all 98 municipalities).

	Yes	No	No answer	Sum	N
Has government funding been of decisive importance for mobile technology adoption?	66	30	4	100	76[a]

[a] The number of municipalities that had adopted mobile technology at the time of the survey (2007).

Outcome

In 2008 when our study ended, the vast majority of home care agencies had adopted mobile technology, and ministries, consultants, and IT vendors still supported the use of mobile technology. Yet, the many and varied experiences of transforming the new technology into new practices continued to influence the ongoing debate over mHealth's role in improving Danish home care. More critical reports appeared describing control and monitoring issues in the practical use of mobile technology and illustrating a more general tension between management and professionalism in home care. While the introduction of the technology at the outset appeared remarkably promising, many projects were hampered by technical difficulties that also likely contributed to an increased resistance from health care personal.

Case Analyses

Summary

To provide detailed insights into the changes that resulted in individual home care agencies, we trace the implementation of mobile technology in three case settings. These cases highlight the interpretive flexibility [37] of mobile technology as summarized in Table 3. Multimedia Appendix 1 includes data tables about mobile technology implementation and use in these three settings.

Home Care Agency 1

This agency implemented mobile technology (Nokia Communicator was the chosen hardware) in 2005 based on online technology. The new system was implemented with Zealand Care as the vendor and responsible for training sessions. Government funding was not of decisive importance for initial investment by HCA 1 in mobile technology, but the CareMobile initiative served as a major source of inspiration. Mobile technology was considered a useful coordination platform for sharing information, for reducing administrative tasks, and especially for decreasing meetings in home care. A manager explained their technology strategy:

> *We could see an advantage as each nurse and care worker had a cell phone at the point-of-care, and we would like to cut down on the time we spent at meetings. In fact, it was a demand from our politicians in the City Council that if we adopted more technology, we had to cut down our meeting activity.*

Working practices changed substantially as strategies for mobile technology were implemented and transformed into daily work (technology in use). Mobile devices were used for documentation (especially registration of time and services provided), for internal communication between managers, nurses, and care workers and to access information at the point of care. The morning joint meetings at central offices were eliminated as health care personnel instead based their working day on information exchange through mobile communication. However, not all face-to-face meetings were canceled, and joint coordination meetings were held in the afternoon twice a week. Contrary to the initial purpose, mobile technology was not used for filling out records. Instead, the health care personal preferred desktop computers with larger screens and larger keyboards for this specific activity.

Care workers received mobile technology with skepticism, particularly due to the canceling of the morning meeting. In contrast, managers viewed mobile technology in positive terms and had a clear vision of what benefits mobile technology could bring to the organization. Although the care workers' perceptions became more positive over time, mixed attitudes towards mobile technology were still apparent after 2 years of use. It was not the technology itself that created debate (it was perceived as easy to use) but the reduction in meeting activity that proved most controversial as it was considered a reduction in knowledge sharing and collegial relationships. One care worker expressed: "I miss the morning session. I do miss the social and collegial contacts." Overall, mobile technology implementation in HCA 1 demonstrates how the new technology impacted day-to-day working routines (eg, communication and access to information) and conflicted with perceived advantages of established practices (eg, morning meetings).

Home Care Agency 2

This agency also implemented an mHealth system with online connectivity in 2005. The system was provided by Ramböll, and PDAs were the selected hardware including telephone features. Government funding was not of decisive importance for the decision by HCA 2 but sponsoring did help roll out the technology throughout the organization. HCA 2 took a different approach to mHealth compared to HCA 1. The technology strategy was not so much to support a cost-saving agenda, but more to promote a modern image of home care. A home care manager stated: "It was very much related to status...to give our staff advanced mobile technology will certainly raise the status...In many home care agencies it's about efficiency, saving and control...this has not been the case here."

HCA 2 did not mandate detailed time registration and abolish joint morning meeting (as in HCA 1). Instead, it was up to each home care unit whether they wanted to use the technology for these purposes. This more lenient implementation strategy appeared to influence the care workers' interpretation of municipal control as being weak compared to HCA 1 and HCA 3 (Multimedia Appendix 1). Although this agency continued to print schedules rather than access them on the PDA, most health care personnel used the PDA in communication with managers and colleagues internally and with hospitals or general practitioners externally. Many workers remained ambivalent towards mobile technology. On one side, they saw the ability to gain information at the point of care and the telephone-and-text-message features as positive. One worker stated: "I think it is a major advance that we can now order medicines online."

On the other hand, technical difficulties were a major source of frustration. Many care workers found it difficult to fill out records on the PDA, and some care workers were skeptical towards the utility of mobile technology. As one worker noted: "It cannot be of any good, except that by using the cell phone I can call the client if (s)he cannot hear the doorbell."

Home Care Agency 3

This agency was actively engaged in designing an mHealth system as early as 2002 and, as a result, selected PDAs with offline connectivity. Computer Sciences Corporation (CSC) was the chosen provider. HCA 3 had registered services and working hours since the mid-1990s, and mobile technology was perceived by managers as a feasible way to facilitate and advance these activities. The key technology strategy was to improve "uniform level of service and contribute to documentation and transparency in the management of home care" [62].

Accordingly, care workers used mobile technology to access client information, look up schedules, and register working hours and services—as expected by managers that praised the new technology. However, sometimes care workers used the technology differently than planned by performing time registration at the end of their work day rather than "on the go". Overall, the health care personnel perceived the monitoring and careful documentation of home care services as an unpopular system of control. The offline connectivity also proved controversial over time, and the personnel started to request more contemporary mobile devices with telephone features included. In response, HCA 3 established a pilot in 2007 based on government sponsorship. The agency decided, however, not to adopt online technology as an evaluation concluded that there were too many technical difficulties [63].

Table 3. Interpretive flexibility in mobile health across three cases.

	HCA 1	HCA 2	HCA 3
Nature of technology	Nokia Communicator with online connectivity and Zealand as provider.	PDAs with online connectivity and Ramböll as provider.	PDAs with offline connectivity and CSC as provider.
	Managers in support of mobile health.	Managers in support of mobile health.	Managers in support of mobile health.
	The reduction in meeting activity proved controversial for care workers, and many experienced difficulties filling out records.	Many care workers experienced technical difficulties, and some remained skeptical towards mobile technology.	Mixed attitude among care workers: many experienced increased control based on detailed time registration, and offline connectivity proved controversial.
Technology strategy	Mobile technology as coordination platform to share information and reduce meeting activities.	Mobile technology as communication medium to improve relationships and status of home care.	Mobile technology as management tool to improve documentation and transparency of service delivery.
Technology use	Use of mobile technology to support coordination by documenting and sharing information about activities. Joint morning meetings were abolished, as mobile technology afforded information exchanges.	Use of mobile technology to support communication with managers and colleagues internally and with hospitals or general practitioners externally.	Use of mobile technology to support management of resources by recording information about working hours and service delivery.

Discussion

Principal Findings

Our process analysis reveals how government sponsoring and collaboration between key stakeholders in the home care sector shaped the widespread adoption of mobile technology, whereas the case analyses demonstrate how mobile technology had interpretive flexibility with considerable variation in perceived nature of technology, technology strategy, and technology in use between the observed agencies. In the following section, we discuss these findings in detail.

First, strong collaboration between key stakeholders in the home care sector (eg, government bodies, vendors, consultants, interest organizations, and managers) helped initiate and energize the change process. The evidence suggests that mobile technology implementation in the Danish case was not only shaped within specific home care agencies, but also energized in a broader sociopolitical and interorganizational context [43]. Powerful stakeholders in the home care sector justified mobile technology usage; IT vendors and consultants inspired home care agencies, especially early on in the change process; and government accelerated the process by financing a pilot and sponsoring implementation across individual agencies. Indeed, the evidence suggests the widespread and relatively fast diffusion of mobile technology across Danish home care agencies was not only facilitated by, but to a large extent dependent on, these broader initiatives.

Second, the sector-wide pilot (CareMobile) served as an important arena for legitimizing use of mobile technology in home care agencies. As the CareMobile project unfolded from 2002 to 2005, it created a generic technology strategy and practical guidelines for using mobile technology to modernize home care. CareMobile involved interplay between key stakeholders across the home care sector that inspired design of possible technology in use scenarios and exchange of key lessons from mHealth. The pilot provided a sound business case for mHealth implementation and served as a significant inspiration for home care agencies. Moreover, CareMobile functioned as an important justification of subsequent government funding. In this sense, CareMobile represents an interesting example of how large-scale pilots can effectively influence widespread HIT implementation.

Third, government funding represented a double-edged sword. As evident in the process analysis, government sponsorship in 2006 facilitated swift diffusion of mobile technology within the sector and helped individual home care agencies engage in acquisition and implementation activities. The funding proved very successful and supported overall government strategies in the area [58]. Thus, this study corroborates findings from other studies [43] suggesting that public financing is an effective tool in supporting and diffusing new information technologies. However, the hasty diffusion accelerated by government sponsorship did not progress as a straightforward and unproblematic process. In the wake of the widespread diffusion, it became clear that mobile technology implementation was hampered by technical difficulties and associated with a system of organizational control. It is also likely that government funding for some agencies was the key driver behind the project (see Table 2) rather than an organizational need to use mobile technology. As mentioned in our interview with the representative from LGDK: “When you throw bags of money in front of the managers in home care, of course they attempt to grasp them.” Government funding was, in effect, an offer home care agencies could hardly refuse.

Fourth, mobile technology had interpretive flexibility and created considerable variation in how stakeholders perceived the nature of the technology, technology strategy, and technology use across home care agencies. Although the three home care agencies we observed all implemented mHealth systems, their approaches took quite different forms as the technology was transformed to fit the local context of each home care agency. They chose different vendors, the technology was perceived differently, they drew on different technology strategies, and they adopted different technology in use arrangements**.** For instance, whereas HCA 1 focused on mobile technology as a coordination platform, HCA 2 implemented mobile technology as a communication medium, and HCA 3 introduced mobile technology as a management tool. In this sense, we see how home care organizations are interpretive systems [64] and how mobile technology has considerable interpretative flexibility [36,37].

Fifth, while mobile technology in some instances was easily integrated with existing work practices, the integration was challenging in other instances. Mobile technology benefitted the health care personal in several ways. In line with other studies [5,10,11,16,26], our analyses demonstrate how health care personnel appreciated the ability to access client information at the point of care. At the same time, our study reveals mixed attitudes towards mobile technology among care workers. These different perceptions made mobile technology implementation more difficult than expected. For example, in HCA 1, the canceling of the joint morning meeting proved controversial as it conflicted with professional values, and in HCA 3, detailed time registrations were interpreted by care workers as a system of control. Accordingly, while the hopes of the transformative potential of mHealth are high, our research supports a more balanced view recognizing the challenges and difficulties in implementing mHealth systems [13-19].

Conclusions

mHealth implementation does not appear to be a straightforward process with a clear beginning and end. Instead, our analyses demonstrate how mobile technology implementation was an interactive and muddled process that, like other aspects of contemporary health care organization, happened “in many places at once” [65]. Accordingly, we suggest that HIT implementation (such as mHealth) is most readily studied by combining organization-level analysis with features of the wider sociopolitical and interorganizational environment [42,43]. By conducting the overall process analysis, we were able to understand the political dynamics of mobile technology implementation. By moving to the organizational aspect in specific case studies, other aspects of the implementation process were given and demonstrated the interpretive flexibility of mHealth.

Our study provides valuable insights to decision makers and health care organizations as they engage in mHealth. The key lessons from the Danish case include the following: (1) participation and collaboration between a variety of stakeholder proved useful in supporting the implementation process, (2) government-sponsored programs can serve as double-edged swords, (3) managers need to be aware of the interpretive flexibility of mobile technology, and (4) mobile technology may in some areas collide with professional values and norms while in other areas being easily integrated in day-to-day working practices.

Although the mHealth initiative in Danish home care is an attractive subject for analysis as it represents a large-scale innovation project and includes comprehensive empirical data, it still represents only one example of mHealth implementation. As a result, some caution is required in generalizing our findings. Yet, there is a steadily increasing practice and research interest in how mHealth can be applied to improve health care delivery. While literature on mHealth often has relied on variance models such as the Technology Acceptance Model [66], we hope our study may encourage other scholars to include process and longitudinal investigations to more fully understand and draw lessons from mobile health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables.

[PDF File (Adobe PDF File), 290KB - jmir_v15i10e236_app1.pdf]

References

1. Singh R, Mathiassen L, Stachura ME, Astapova EV. Sustainable rural telehealth innovation: a public health case study. Health Serv Res 2010 Aug;45(4):985-1004 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01116.x] [Medline: 20459449]
2. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [Medline: 11734382]
3. Currie, WL, Guah MW. Conflicting institutional logics: a national programme for IT in the organisational field of healthcare. Journal of Information Technology 2007;22:235-247.
4. Istepanian R, Laxminarayan S, Pattichis CS (eds). M-Health: Emerging Mobile Health Systems. NY. USA: Springer; 2006.
5. 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]
6. Lu YC, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [doi: 10.1016/j.ijmedinf.2005.03.001] [Medline: 15893264]
7. Wu JH, Wang SC, Lin LM. Mobile computing acceptance factors in the healthcare industry: a structural equation model. Int J Med Inform 2007 Jan;76(1):66-77. [doi: 10.1016/j.ijmedinf.2006.06.006] [Medline: 16901749]
8. Brady RR, Chitnis S, Stewart RW, Graham C, Yalamarthi S, Morris K. NHS connecting for health: healthcare professionals, mobile technology, and infection control. Telemed J E Health 2012 May;18(4):289-291. [doi: 10.1089/tmj.2011.0147] [Medline: 22428552]
9. Chau S, Turner P. Utilisation of mobile handheld devices for care management at an Australian aged care facility. Electronic Commerce Research and Applications 2006;5(4):305-312.
10. Rothschild JM, Fang E, Liu V, Litvak I, Yoon C, Bates DW. Use and perceived benefits of handheld computer-based clinical references. J Am Med Inform Assoc 2006;13(6):619-626 [FREE Full text] [doi: 10.1197/jamia.M2103] [Medline: 16929041]
11. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
12. Vimarlund V, Olve NG. Economic analyses for ICT in elderly healthcare: questions and challenges. Health Informatics Journal 2005;11(4):309-321.
13. Avancha S, Baxi A, Kotz D. Privacy in mobile technology for personal healthcare. ACM Computer Survey 2012;45(1):1-54.
14. Tooey MJ, Mayo A. Handheld technologies in a clinical setting: state of the technology and resources. AACN Clin Issues 2003 Aug;14(3):342-349. [Medline: 12909802]
15. Rosenthal K. "Touch" vs. "tech": valuing nursing-specific PDA software. Nurs Manage 2003 Jul;34(7):58-60. [Medline: 12843721]
16. Ehrler F, Wipfli R, Teodoro D, Sarrey E, Walesa M, Lovis C. Challenges in the Implementation of a Mobile Application in Clinical Practice: Case Study in the Context of an Application that Manages the Daily Interventions of Nurses. JMIR Mhealth Uhealth 2013 Jun 12;1(1):e7. [doi: 10.2196/mhealth.2344]
17. Haller G, Haller DM, Courvoisier DS, Lovis C. Handheld vs. laptop computers for electronic data collection in clinical research: a crossover randomized trial. J Am Med Inform Assoc 2009;16(5):651-659 [FREE Full text] [doi: 10.1197/jamia.M3041] [Medline: 19567799]
18. Wu R, Rossos P, Quan S, Reeves S, Lo V, Wong B, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. J Med Internet Res 2011;13(3):e59 [FREE Full text] [doi: 10.2196/jmir.1655] [Medline: 21875849]
19. Solvoll T, Scholl J, Hartvigsen G. Physicians interrupted by mobile devices in hospitals: understanding the interaction between devices, roles, and duties. J Med Internet Res 2013;15(3):e56 [FREE Full text] [doi: 10.2196/jmir.2473] [Medline: 23470528]
20. Garritty C, El Emam K. Who's using PDAs? Estimates of PDA use by health care providers: a systematic review of surveys. J Med Internet Res 2006;8(2):e7 [FREE Full text] [doi: 10.2196/jmir.8.2.e7] [Medline: 16867970]
21. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
22. Hameed K. The application of mobile Computing and technology to health care services. Telematics and Informatics 2003;20(2):99-106.
23. Shipman JP, Morton AC. The new black bag: PDAs, health care and library services. Reference Services Review 2001;29(3):229-237.
24. Park Y, Chens JV. Acceptance and adoption of the innovative use of smartphone. Industrial Management & Data Systems 2007;107(9):1349-1365.
25. Zhang H, Cocosila M, Archer N. Factors of adoption of mobile information technology by homecare nurses: a technology acceptance model 2 approach. Comput Inform Nurs 2010;28(1):49-56. [doi: 10.1097/NCN.0b013e3181c0474a] [Medline: 19940621]
26. Vishwanath A, Brodsky L, Shaha S. Physician adoption of personal digital assistants (PDA): testing its determinants within a structural equation model. J Health Commun 2009;14(1):77-95. [doi: 10.1080/10810730802592270] [Medline: 19180373]
27. Hong HS, Kim IK, Lee SH, Kim HS. Adoption of a PDA-based home hospice care system for cancer patients. Comput Inform Nurs 2009;27(6):365-371. [doi: 10.1097/NCN.0b013e3181bcab43] [Medline: 19901573]

28. Hung MC, Jen WY. The adoption of mobile health management services. An empirical study. Journal of Medical System 2012;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2]
29. Vuokko R. Experiences from an implementation project - time management and control in home care. In: Proceedings of the 37th Hawaii International Conference on System Sciences. 2004 Presented at: Hawaii International Conference on System Sciences; 5-8 January 2004; Hawaii.
30. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2009;14(11):1-24.
31. Mohr LB. Explaining Organizational Behavior. In: Explaining Organizational Behavior. San Francisco: Jossey-Bass; 1982.
32. Pettigrew AM. Longitudinal field research on change: Theory and practice. Organization Science 1990;1(3):267-292.
33. Van de Ven AH. Suggestions for studying strategy process: a research note. Strategic Management Journal 1992;13:169-188.
34. Davidson E, Chiasson M. Contextual influences on technology use mediation: A comparative analysis of electronic medical record systems. European Journal of Information Systems 2005;4(1):6-18.
35. Cho S, Mathiassen L, Nilsson A. contextual dynamics during health information systems implementation: An event-based actor-network approach. European Journal of Information Systems 2008;17:614-630.
36. Pinch TJ, Bijker WE. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. Social Studies of Science 1984;14(1):399-441.
37. Orlikowski WJ, Gash DC. Technological frames: Making sense of information technology in organizations. ACM Transactions on Information Systems 1994;12(2):174-207.
38. Newman M, Robey D. A social process model of user-analyst relationships. MIS Quarterly 1992;16(2):249-266.
39. Langley A. Strategies for theorizing from process data. The Academy of Management Review 1999;24(4):691-710.
40. Pettigrew AM. What is a processual analysis? Scandinavian Journal of Management 1997;13(4):337-348.
41. Currie WL. Contextualising the IT artefact: Towards a wider research agenda for IS using institutional theory. Information, Technology & People 2009;22(1):63-77.
42. DiMaggio P, Powell W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review 1983;48(2):147-160.
43. King JL, Gurbaxani V, Kraemer KL, McFarlan FW, Raman KS, Yap CS. Institutional factors in information technology Iinnovation. Information Systems Research 1994;5(2):139-169.
44. Yin RK. Enhancing the quality of case studies in health services research. Health Services Research 1999;34(5):1209-1224.
45. Yin RK. Case study research: Design and methods. Fourth edition. Thousand Oaks, CA: Sage; 2009.
46. Jensen PH, Lolle H. The fragmented welfare state: Explaining local variations in services for older people. Journal of Social Policy 2013;42(2):349-370.
47. Rostgaard T. Care as you like it: The construction of a consumer approach in home care in Denmark. Nordic Journal of Social Research 2011;Special issue:1-16.
48. Jarden ME, Jarden JO. Social and health-care policy for the elderly in Denmark, global action on aging. 2002. URL: http://www.globalaging.org/elderrights/world/densocialhealthcare.htm [accessed 2013-10-09] [WebCite Cache ID 6HoKxhXFT]
49. Nielsen JA, Andersen JG. Hjemmehjælp. Mellem myter og virkelighed. Odense, Denmark: Syddansk Universitetsforlag; 2006.
50. Lyytinen K, Newman M. Explaining information systems change: a punctuated socio-technical change model. European Journal of Information Systems 2008;17:589-613.
51. Miles MB, Huberman AM. Data management and analysis methods, in Handbook of Qualitative Research, Denzin NK and Lincoln YS (eds.). Thousand Oaks: Sage Publications; 1994.
52. Hood C. A public management for all seasons? Public Administration 1991;69(1):3-19.
53. Lützhöft R. Ældresektoren. Behov for sammenhængende IT-løsninger. Copenhagen, Denmark: Kommunernes Landsforening, Kontoret for Teknologi; 1994.
54. Hansen MB, Vedung E. Fælles Sprog i ældreplejens organisering. Evaluering af et standardiseret kategorisystem. Odense, Denmark: Syddansk Universitetsforlag; 2005.
55. Ministry of Social Affairs. CareMobil. Introduktion til mobile it-løsninger i ældreplejen. Copenhagen, Denmark; 2005. URL: http://www.sm.dk/data/Lists/Publikationer/Attachments/180/caremobil%20%20lang%20version.pdf [accessed 2013-10-09] [WebCite Cache ID 6KFORrOZx]
56. Ramböll Management. Omsorgsteknologi kan give mere tid til pleje i ældresektoren. Copenhagen, Denmark; 2007. URL: http://ida.dk/sites/prod.ida.dk/files/Omsorgsteknologi_Ramboll_analyse_260407.pdf [accessed 2013-10-09] [WebCite Cache ID 6KFOhMBne]
57. Ministry of Finance. Budgetredegørelse 2004. Copenhagen, Denmark; 2004. URL: http://www.fm.dk/Publikationer/2004/Budgetredegoerelse%202004.aspx?mode=full [accessed 2013-10-09] [WebCite Cache ID 6KFP2J0Jh]
58. Ministry of Social Affairs. Strategi for digitalisering på det sociale område 2005-2008. Copenhagen, Denmark; 2006. URL: http://www.sm.dk/data/Lists/Publikationer/Attachments/112/strategi%20for%20digitalisering%20p%C3%A5%20det%20sociale%20omr%C3%A5de.pdf [accessed 2013-10-09] [WebCite Cache ID 6KFPGTc3t]

59. Politiken. Kvalitetsreform – med mennesker i centrum. Copenhagen, Denmark; 2006 Nov 16. URL: http://www.stm.dk/index.dsp?page=8156&action=page_overview_search&l1_valg=1800&l2_valg=-1 [accessed 2013-10-09]
60. Ministry of Social Affairs. Vejledning til Pulje til Bedre og Mere Fleksibel Hjemmehjælp. Copenhagen, Denmark: Ministry of Social Affairs; 2006.
61. FOA. Registrering og dokumentation i hjemmeplejen - indtryk fra syv kommuner. Copenhagen, Denmark; 2008. URL: http://www.foa.dk/~/media/_Files/Fildeling/undersgelser/Registrering%20og%20dokumentation%20i%20aeldreplejen%20pdf.ashx [accessed 2013-10-09] [WebCite Cache ID 6KFQf7i1r]
62. Lütshöft R. Projektbeskrivelse. Indførelse af Københavns Omsorgs System (KOS) i Sundhedsforvaltningen. Copenhagen, Denmark: Copenhagen Municipality; 2005.
63. Devoteam Consulting. Evaluering af ny mobil teknologi i Københavns Kommune. Copenhagen, Denmark: Devoteam Consulting; 2007.
64. Daft RL, Weick KE. Toward a model of organizations as interpretive systems. Academy of Management Review 1984;9:284-295.
65. Czarniawska B. Shadowing: And other Techniques for doing Fieldwork in Modern Societies. Copenhagen: Liber and Copenhagen Business School Press; 2007.
66. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340.

Abbreviations

CSC: Computer Sciences Corporation
HCA: home care agency
HIT: health care information technology
LGDK: Local Government Denmark
OECD: Organisation for Economic Co-operation and Development
PDA: personal digital assistant

Edited by G Eysenbach; submitted 02.07.13; peer-reviewed by M Kelley, S Maingot; comments to author 04.08.13; revised version received 13.08.13; accepted 17.09.13; published 30.10.13.

Please cite as:
Nielsen JA, Mathiassen L
Interpretive Flexibility in Mobile Health: Lessons From a Government-Sponsored Home Care Program
J Med Internet Res 2013;15(10):e236
URL: http://www.jmir.org/2013/10/e236/
doi:10.2196/jmir.2816
PMID:24172852

Original Paper

Internet and Social Media For Health-Related Information and Communication in Health Care: Preferences of the Dutch General Population

Tom H Van de Belt[1], MSc; Lucien JLPG Engelen[1]; Sivera AA Berben[2], RN, PhD; Steven Teerenstra[3], PhD; Melvin Samsom[4], MD, PhD; Lisette Schoonhoven[5,6], RN, PhD

[1]Radboud REshape and Innovation Center, Radboud University Medical Center, Nijmegen, Netherlands
[2]Regional Emergency Healthcare Network, Radboud University Medical Center, Nijmegen, Netherlands
[3]Department for Health Evidence, section Biostatistics, Radboud University Medical Center, Nijmegen, Netherlands
[4]Executive Board, Radboud University Medical Center, Nijmegen, Netherlands
[5]Scientific Institute for Quality of Healthcare, Radboud University Medical Center, Nijmegen, Netherlands
[6]Faculty of Health Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Tom H Van de Belt, MSc
Radboud REshape and Innovation Center
Radboud University Medical Center
Route 911
Reinier Postlaan 4
Nijmegen, 6525GC
Netherlands
Phone: 31 24 7440080
Fax: 31 24 7440080
Email: tom.vandebelt@radboudumc.nl

Abstract

Background: Health care is increasingly featured by the use of Web 2.0 communication and collaborative technologies that are reshaping the way patients and professionals interact. These technologies or tools can be used for a variety of purposes: to instantly debate issues, discover news, analyze research, network with peers, crowd-source information, seek support, and provide advice. Not all tools are implemented successfully; in many cases, the nonusage attrition rates are high. Little is known about the preferences of the Dutch general population regarding the use of the Internet and social media in health care.

Objective: To determine the preferences of the general population in the Netherlands regarding the use of the Internet and social media in health care.

Methods: A cross-sectional survey was disseminated via a popular Dutch online social network. Respondents were asked where they searched for health-related information, how they qualified the value of different sources, and their preferences regarding online communication with health care providers. Results were weighed for the Dutch population based on gender, age, and level of education using official statistics. Numbers and percentages or means and standard deviations were presented for different subgroups. One-way ANOVA was used to test for statistical differences.

Results: The survey was completed by 635 respondents. The Internet was found to be the number one source for health-related information (82.7%), closely followed by information provided by health care professionals (71.1%). Approximately one-third (32.3%) of the Dutch population search for ratings of health care providers. The most popular information topics were side effects of medication (62.5%) and symptoms (59.7%). Approximately one-quarter of the Dutch population prefer to communicate with a health care provider via social media (25.4%), and 21.2% would like to communicate via a webcam.

Conclusions: The Internet is the main source of health-related information for the Dutch population. One in 4 persons wants to communicate with their physician via social media channels and it is expected that this number will further increase. Health care providers should explore new ways of communicating online and should facilitate ways for patients to connect with them. Future research should aim at comparing different patient groups and diseases, describing best practices, and determining cost-effectiveness.

(J Med Internet Res 2013;15(10):e220) doi:10.2196/jmir.2607

KEYWORDS

social media; patient participation; consumer health information; empowerment; Health 2.0

Introduction

Health care is increasingly featured by the use of Web 2.0 communication and collaborative technologies that are reshaping the way patients and professionals interact [1]. This process, in which Web 2.0 tools are used in health care, is part of Health 2.0 (also known as Medicine 2.0) [2], an important fundament of which is the use of social media [3]. Kaplan and Haenlein [4] define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content." Well-known examples are YouTube, Facebook, and Twitter. These can be used for a variety of purposes: to instantly debate issues, discover news, analyze research, network with peers, crowd-source information, seek support, and provide advice [5]. Research shows that larger health care organizations, such as hospitals, are increasingly using social media [6,7]. In many cases, the ultimate goal is to make health care better or more cost-efficient [8].

Since the arrival of social media interventions for health-related purposes, it has become clear that not all these interventions are actually successful. Although no studies exist that have investigated this problem for social media, eHealth literature, which overlaps with social media because both involve technology, could provide some insight into this problem. It is known that interventions are often not successful and/or the attrition rates may be high [9-11]. Several explanations for unsuccessful use have been described: (1) technology features (eg, imperfections of the technology), (2) inadequate reimbursement or legislation issues, (3) poor coordination and introduction of tools, and (4) personal characteristics of the intended use [9,12]. Personal characteristics seem to be particularly relevant because they concern the end-users of the tool. Examples of such characteristics, which are known to significantly influence use, are negative attitude toward technology, the extent to which a person feels he has the skills and expertise to be a competent caregiver, and age [9]. Therefore, determining the preferences or needs of potential users of tools is an important step in implementation [12-14]. Although studies have assessed patients' preferences regarding the Internet in health care (eg, the preferred language on websites [15], the preferences of a Web-based intervention [16], preferences regarding social media and asthma patients [17], or the needs of elderly patients regarding eHealth [18]), less is known about the preferences or needs of consumers or the general public, especially regarding social media. A survey showed that 32% of all respondents (US adults) had used social media for health care purposes at one time or another [19]. Further insights, however, are lacking. Questions that arise in this context are: Where do people obtain online health-related information? Where do they connect with peers? Are they willing to ask their doctor questions using a webcam? And are there differences between different groups of the population (eg, by gender, age, or education)?

For that reason, we sought to determine the preferences of the general population in the Netherlands regarding the use of the Internet and social media in health care, by using an online survey that was disseminated via an online social network.

Methods

Design, Setting, and Population

A cross-sectional survey was disseminated via a popular Dutch online social network. Hyves was selected as the social network for dissemination of the survey. This social network has long been the most popular Dutch online social network, with 9.7 million members of all ages [20,21], comprising more than half of the Dutch population [22]. Hyves can be used to create a personal profile and connect with friends. Furthermore, users can like pages or create groups. Between October 4 and November 4, 2011, Hyves members aged at least 15 years were randomly invited through Hyves' internal message system. There were no restrictions regarding sex, race, or income. The messages contained a description of the project (in Dutch) and a link to the survey.

Questionnaire Development and Content

Overview

A first draft of the questionnaire was created by TB and subsequently discussed with LE and LS. This version was shared with 3 experts: a social media expert, a researcher (SB), and an epidemiologist. After discussion, consensus was reached and the survey was finalized and uploaded to the online system. The questionnaire consisted of 17 multiple-choice questions divided over 3 sections: (1) sociodemographic, (2) health-related information and Internet, and (3) respondents' preferences regarding communication in health care. All questions were written in Dutch. The final survey (English version) is available in Multimedia Appendix 1.

Sociodemographic Section

The sociodemographic section contained questions about age, gender, and level of education.

Health-Related Information and Internet

In the health-related information and Internet section, respondents were asked where they searched for health-related information and how they qualified the value of different sources. The topics were:

1. Sources of health-related information;
2. Type of online information that is searched for;
3. Frequency of health-related searches; and
4. Perceived reliability of different sources.

Respondents' Preferences Regarding Communication in Health Care

In the preferences section, preferences regarding communication in health care were acknowledged.

Response

A total of 4232 people selected the link to the online survey, of which 679 filled out the survey. After excluding incomplete surveys or surveys completed by respondents under 15 years (n=44), 635 cases were analyzed. The mean response time was 6.13 minutes (SD 2.95).

Statistical Analysis

The data were downloaded from the online system and analyzed in SPSS version 20 (IBM Corp, Armonk, NY, USA). We used descriptive statistics to examine the proportions for different age, gender, and education groups. Proportions for age were summarized in 6 age groups: 15-24, 25-34, 35-44, 45-54, 55-64, and 65 or older.

Answers regarding health-related information and the Internet as well as preferences of communication in health care were extrapolated to the Dutch population based on gender, age, and level of education. We decided to create 2 age groups based on different generations described in the literature [23]. The first group consisted of people aged 15-34 years. This group has been described as Generation Y and consists of people who grew up with the Internet. The second age group consisted of persons aged 35 years or older, including the Generation X and the so-called baby-boomers. Two levels of education were recognized. The first group consisted of people with no education or lower education, whereas the second group consisted of moderately or highly educated people.

For each stratum (combination of gender, age, and educational level), the response within the survey was estimated. The response of the stratum was then weighted by the relative frequency of that stratum within the Dutch population of 2011, acquired via Statistics Netherlands (Centraal Bureau voor de Statistiek, CBS) [22]. CBS is a Dutch governmental institution and part of the Dutch Ministry of Economic Affairs that is responsible for gathering and publishing official statistics about the Netherlands. CBS statistics are only published if they are valid and if the overall quality can be guaranteed. The following example shows how we weighed data: if the percentage of young males and old males saying yes was 40% and 60%, respectively, then this would result in a mean of 50% in our sample. Given that young and old males (from CBS statistics) form 0.3 and 0.7 of the Dutch male population, respectively, the percentage of males who would say yes in the Dutch population was estimated to be (0.3×40%) + (0.7×60%) = 54%.

We present numbers and percentages or means and standard deviations. To properly test differences between groups in the response (eg, male vs female) extrapolated to the Dutch population, we needed to take into account that (1) the precision of the estimated response percentages in strata is determined by the size of the strata in the survey, and (2) these response percentages are weighted by the relative frequency of those strata in the Dutch population. To accomplish this, we used the SPSS procedure 1-way ANOVA to (1) estimate the response percentages with their corresponding precision from the survey, and (2) perform the weighting by specifying the relative frequencies in contrast tests. Because the size of the strata was reasonably large (>25) and the response within strata was not close to zero or 100%, the ANOVA means and standard errors were considered a good approximation of the response percentages of the strata. *P* values <.05 were considered statistically significant.

Results

Sociodemographic

In total, 635 respondents completed the survey, consisting of 95 (15.0%) men and 540 (85.0%) women. Table 1 shows the age distribution for all respondents in 10-year age ranges. In all, 181 respondents (28.5%) had no education or low education and 454 (71.5%) were moderately or highly educated.

Sources of Health-Related Information

Table 2 shows the popularity of different sources of health-related information estimated for the Dutch population. Internet and physicians were found to be the most popular sources (82.7% and 71.1%, respectively). Family and friends were mentioned by 20.5% of the Dutch population. People aged ≤ 34 years consulted their family and friends significantly more often than people older than 34 years (38.1% vs 13.5%, 1-way ANOVA, contrast test t_{627}=3.52, *P*<.001). Higher educated people also consulted their family and friends more often (12.5% for lower educated people vs 24.7% for higher educated persons, 1-way ANOVA, contrast test t_{627}=–2.05, *P*=.04). Patient information leaflets or books were the least popular information source (14.6%).

Type of Online Information Searched For

The most popular information topics that were searched online (Table 3) were side effects of medication and symptoms (62.5% and 59.7%, respectively). People aged 35 years or older searched significantly more often for information on side effects than people younger than 35 years (68.7% vs 46.8%, 1-way ANOVA, contrast test t_{627}=–2.63, *P*=.01). People younger than 35 years searched more often for symptoms than persons aged 35 or older (76.1% vs 53.2%, 1-way ANOVA, contrast test t_{627}=2.65, *P*=.01). Furthermore, women indicated that they searched more often for information on diagnoses than men (58.8% vs 31.5%, 1-way ANOVA, contrast test t_{627}=–4.13, *P*<.001).

Table 1. Survey respondents (N=635).

Subgroup	n (%)
Gender	
Male	95 (15.0)
Female	540 (85.0)
Age	
15 - 24	74 (11.7)
25 - 34	90 (14.2)
35 - 44	144 (22.7)
45 - 54	172 (27.1)
55 - 64	129 (20.3)
65 or older	26 (4.1)
Education	
No/lower education	181 (28.5)
Moderate or high education	454 (71.5)

Table 2. Sources for health-related information.

Subgroup	Total, %[a]	Group 1, %[b]	Group 2, %[b]	t_{627}	*P*
Gender					
Internet	82.7	82.8	82.6	0.11	.91
Physician	71.1	66.2	74.9	–1.45	.15
Family/friends	20.5	19.7	21.2	–0.29	.78
Patient information (leaflets, books)	14.6	11.6	17.5	–1.19	.23
Age					
Internet	82.7	87.4	74.0	0.99	.32
Physician	71.1	63.8	74.0	–1.30	.19
Family/friends	20.5	38.1	13.5	3.52	<.001
Patient information (leaflets, books)	14.6	15.0	14.5	0.04	.97
Education					
Internet	82.7	78.0	85.2	–1.25	.21
Physician	71.1	69.8	71.8	–0.34	.74
Family/friends	20.5	12.5	24.7	–2.05	.04
Patient information (leaflets, books)	14.6	9.3	17.5	–1.43	.15

[a] Estimations for Dutch population (%) based on the study sample of 635 respondents. Note that these estimates are weighted sums of the cell response percentages; therefore, n’s cannot be provided (see Methods) for these percentages.

[b] For gender, group 1=male, group 2=female; for age, group 1=age ≤34 years, group 2=age>34 years; for education, group 1=no or low education, group 2=moderate or high education.

Table 3. Type of health-related information searched for online.

Subgroup	Total, %[a]	Group 1, %[b]	Group 2, %[b]	t_{627}	*P*
Gender					
Side effects medication	62.5	58.2	66.9	–1.17	.24
Symptoms	59.7	58.7	60.5	–0.13	.90
Diagnoses	45.6	31.5	58.8	–4.13	<.001
Patients’ experiences	41.7	37.1	46.0	–1.27	.20
Health care insurance	41.6	38.0	44.9	–0.99	.32
Therapy	39.3	34.6	43.6	–1.35	.18
My hospital	35.4	38.9	32.1	1.2	.23
Ratings of health care providers	32.3	36.6	28.3	1.4	.16
Health problems	14.8	13.8	15.7	–0.20	.83
Manufacturers of medication	8.9	10.7	7.2	0.98	.33
Second opinion	6.8	5.8	7.8	–0.65	.52
Age					
Side effects medication	62.5	46.8	68.7	–2.63	.01
Symptoms	59.7	76.1	53.2	2.65	.01
Diagnoses	45.6	44.4	46.1	–0.21	.83
Patients’ experiences	41.7	35.0	44.3	–1.1	.27
Health care insurance	41.6	34.6	44.3	–1.11	.27
Therapy	39.3	30.3	42.8	–1.43	.15
My hospital	35.4	32.2	36.7	–0.55	.58
Ratings of health care providers	32.3	30.2	33.2	–0.37	.71
Health problems	14.8	10.6	16.4	–0.95	.34
Manufacturers of medication	8.9	6.2	9.7	–0.65	.52
Second opinion	6.8	10.0	5.6	1.01	.31
Education					
Side effects medication	62.5	58.4	64.7	–0.86	.39
Symptoms	59.7	55.2	62.0	–0.88	.38
Diagnoses	45.6	40.9	48.1	–0.98	.33
Patients’ experiences	41.7	36.4	44.6	–1.08	.28
Health care insurance	41.6	31.7	46.9	–1.94	.05
Therapy	39.3	33.6	42.3	–1.08	.28
My hospital	35.4	29.0	38.8	–1.34	.18
Ratings of health care providers	32.3	31.1	31.9	0.17	.86
Health problems	14.8	12.8	15.9	–0.54	.59
Manufacturers of medication	8.9	5.6	10.7	–1.25	.21
Second opinion	6.8	4.3	8.2	–1.03	.30

[a] Estimations for Dutch population (%) based on the study sample of 635 respondents. Note that these estimates are weighted sums of the cell response percentages; therefore, n’s cannot be provided (see Methods) for these percentages.

[b] For gender, group 1=male, group 2=female; for age, group 1=age ≤34 years, group 2=age>34 years; for education, group 1=no or low education, group 2=moderate or high education.

Frequency of Health-Related Searches

We determined the frequency of online health-related searches extrapolated to the Dutch population. In all, 92.0% indicated that they searched for health-related information at least once a year and 24.4% searched for health-related information at least every month.

Table 4 shows the search behavior of Dutch people before consulting a physician (eg, general practitioner or specialist). In all, 42.3% indicated that they sometimes searched online for health-related information and 18.4% indicated that they never searched online for information before visiting a physician. Table 4 also shows the search behavior after visiting a physician (general practitioner or specialist). In all, 44.4% indicated that they sometimes searched online for health-related information after visiting their physician and 17.0% indicated that they never searched online for information after having visited their physician.

Perceived Reliability of Sources and Other Preferences

Table 5 shows the perceived reliability of sources of health-related information. On a scale from 1 (very unreliable) to 10 (very reliable), people rated their physician and their personal opinion as most reliable (7.3 and 7.5, respectively). Internet and family/friends scored 6.0 and 5.9 on the scale of reliability, respectively. The least reliable source is information retrieved via social media: 3.8 of 10. Family/friends were found to be more reliable by younger persons than older ones (6.7 vs 5.6, 1-way ANOVA, contrast test t_{627}=3.29, *P*=.001). Furthermore, higher educated people rated their personal opinion as more reliable than lower educated persons did (7.7 vs 7.0, 1-way ANOVA, contrast test t_{627}=–2.35, *P*=.02).

Respondents' Preferences Regarding Communication in Health Care

Table 6 shows to which extent Dutch people would like to communicate using social media or webcams. In all, 25.4% prefer to communicate with their health care provider via social media. Furthermore, 21.2% would like to communicate with their health care providers via a webcam. No statistical differences were found between subgroups.

Table 4. Online searches for health-related information before and after visiting physician (general practitioner or specialist).

Moment of search (before/after)	Total[a]	Gender[a]				Age[a]				Education[a]			
		Male	Female	t_{627}	*P*	≤ 34	> 34	t_{627}	*P*	No/low	Mod/high	t_{627}	*P*
Search before, %				–1.69	.09			3.34	.001			1.65	.10
Very often	4.3	4.1	4.5			11.0	1.7			2.9	5.1		
Often	18.3	18.4	18.2			29.0	14.0			25.4	14.5		
Sometimes	42.3	34.8	49.3			39.9	43.3			50.4	37.9		
Rarely	16.7	18.5	15.0			9.3	19.6			8.7	21.0		
Never	18.4	24.1	13.1			10.9	21.4			12.6	21.6		
Search after, %				–3.52	<.001			–0.88	.38			0.48	.63
Very often	2.5	1.5	3.5			3.9	2.0			2.7	2.4		
Often	14.8	10.9	18.5			12.8	15.6			10.8	17.0		
Sometimes	44.4	36.6	51.8			36.6	47.5			53.3	39.6		
Rarely	21.2	29.5	13.4			27.3	18.7			22.1	20.7		
Never	17.0	21.5	12.8			19.4	16.1			11.1	20.2		

[a] Estimations for Dutch population (%) based on the study sample of 635 respondents. Note that these estimates are weighted sums of the cell response percentages; therefore, n's cannot be provided (see Methods) for these percentages.

Table 5. Perceived reliability of sources for health-related information.

Subgroup	Total, mean[a]	Group 1, mean[b]	Group 2, mean[b]	t_{627}	*P*
Gender					
Self	7.5	7.4	7.5	–0.42	.67
Physician	7.3	7.4	7.2	0.87	.38
Internet	6.0	6.0	6.0	0.18	.86
Friends/family	5.9	6.0	5.8	0.98	.33
Social media	3.8	3.7	3.8	0.04	.97
Age					
Self	7.5	7.8	7.3	1.47	.14
Physician	7.3	7.6	7.2	1.12	.26
Internet	6.0	6.2	5.9	0.97	.33
Friends/family	5.9	6.7	5.6	3.29	.001
Social media	3.8	3.2	4.0	–2.03	.04
Education					
Self	7.5	7.0	7.7	–2.35	.02
Physician	7.3	7.1	7.4	–0.77	.44
Internet	6.0	6.0	6.0	–0.31	.76
Friends/family	5.9	6.1	5.8	1.07	.29
Social media	3.8	3.6	3.8	–0.69	.49

[a] Estimations for Dutch population (on a scale from 1 to 10; 1=very unreliable, 10=very reliable), based on the study sample of 635 respondents. Note that these estimates are weighted sums of the cell response numbers; therefore, n’s cannot be provided (see Methods) for these scores.

[b] For gender, group 1=male, group 2=female; for age, group 1=age ≤34 years, group 2=age>34 years; for education, group 1=no or low education, group 2=moderate or high education.

Table 6. Preferences for communication in health care.

Subgroup	Total, %[a]	Group 1, %[b]	Group 2, %[b]	*t* (df)	*P*
Gender					
Would like to ask questions to health care provider via social media[a]	25.4	27.4	23.7	0.64 (573)	.52
Would like to communicate with health care provider via webcam[c]	21.2	25.2	17.5	1.41 (563)	.16
Age					
Would like to ask questions to health care provider via social media[a]	25.4	19.6	27.8	–1.04 (573)	.30
Would like to communicate with health care provider via webcam[c]	21.2	11.7	25.0	–1.84 (563)	.07
Education					
Would like to ask questions to health care provider via social media[a]	25.4	23.6	26.6	–0.43 (573)	.67
Would like to communicate with health care provider via webcam[c]	21.2	18.0	22.9	–0.76 (563)	.45

[a] Estimations for the Dutch population (%) based on survey sample of 581 (54 respondents excluded because they selected no opinion). Note that these estimates are weighted sums of the cell response percentages; therefore, n's cannot be provided (see Methods) for these percentages.

[b] For gender, group 1=male, group 2=female; for age, group 1=age ≤34 years, group 2=age>34 years; for education, group 1=no or low education, group 2=moderate or high education.

[c] Estimations for the Dutch population based on survey sample of 571 (64 respondents excluded because they selected no opinion). Note that these estimates are weighted sums of the cell response percentages; therefore, n's cannot be provided (see Methods) for these percentages.

Discussion

Principal Findings

As far as we are aware, ours is the first study to investigate online search behavior and preferences regarding the use of social media in health care in the Netherlands. Making use of official statistics, survey results for 635 respondents were successfully extrapolated to the general Dutch population.

The Internet was found to be the number one source for health-related information (82.7%), closely followed by information provided by health care professionals (71.1%). For all groups, the least frequently used source of information was hard copy information, such as leaflets/books. This is higher than AlGhamdi et al [24] found in a survey that included the same age population. They showed that 58.4% of all respondents searched online for health-related information and that health care professionals were the primary source of health-related information. Our findings correspond with a study performed in Brazil, which found that the Internet was the primary source of health-related information for 86% of all respondents [25]. Similar results were also found in a study involving patients suffering from a chronic disease. Approximately 90% of the respondents that searched for additional disease-related information indicated that they used the Internet [26]. However, the same study showed that 55% of all respondents used information leaflets as a source of information versus 14.5% in the present study. This difference can be explained by differences in the study population: our study included any individual instead of patients with a chronic condition only. Another explanation could be that there are differences in broadband penetration between the 2 countries (United States 56.1% vs Netherlands 92.9%) [27]. Health care providers should recognize that a large majority of the Dutch population use online sources for health-related information. Therefore, they should focus on providing high-quality patient information via online channels.

The Dutch population searches online for several health-related topics. In all, 9 of 10 persons indicated that they searched for health-related information at least once a year and 1 in 4 searched for health-related information at least every month. Three topics that were most frequently mentioned (>45.6%) are side effects of medication, symptoms, and diagnoses. People aged 35 years or older searched more often for side effects of medication than their younger counterparts did. This is probably because of a higher consumption of medication by older generations.

Approximately one-third (32.3%) of the Dutch population searches for ratings of health care providers. This is slightly more than was found in a recent report about online health in the United States [28]. This report showed that 10% to 20% of the US population searches for physician ratings, reviews, and rankings. We foresee that more people will search for ratings in the near future, as a rapid rise of health care-related rating websites created by the government, patients' organizations, and other parties can be witnessed [29]. An example of such a rating site is Zorgkaart Nederland [30], a website containing a database with information about all health care providers in the Netherlands. Anyone can rate their health care provider and add their comments or experiences. Currently, it contains information about 112,832 health care providers. The observation that an increasing number of people share their experiences online is supported by our finding that the Dutch population rates their own opinion as important. Interestingly, patients' ratings are significantly associated with official patient surveys about the quality of care [31]. This may be an important finding for future researchers and/or governmental parties (eg, health care inspection) because it could help them in determining high-quality care providers, but also in detecting harmful or unwanted situations.

Approximately 1 in 4 persons would like to use social media to consult their physician and 1 in 5 persons would like to communicate with their physician using a webcam. With the growing number of mobile devices, such as smartphones and tablets, we expect the numbers of people wanting to communicate via social media channels or via webcams to increase as well particularly because usability issues for mobile devices are becoming less relevant [32] and there are tools available that use safe connections that protect data and respect the privacy of users, such as Facetalk [33]. Therefore, future researchers should focus on describing best practices for online patient-physician communication and determine the cost-effectiveness. It would also be interesting to study the extent to which face-to-face technology and social media support patient empowerment, which is a term used to describe the process in which consumers are taking an active role in their care process and where the traditional doctor-patient relationship is disappearing [34].

Limitations

Our study has some limitations that need to be discussed. Although using a social network was helpful in reaching a large group of people very quickly and at relatively low cost, there are some relevant downsides. The online system that sent invitations to Hyves' members randomly did not allow us to register the number of invitations sent. Furthermore, we were not able to distinguish between people who had actually seen the request but had refused to fill in the survey or people who had not seen the request at all (eg, invitation ended up in spam or junkmail folder). As a result, it was impossible to determine exact response percentages. Although we know that people of all genders, ages, and education levels were active on Hyves at the time of the study and that we corrected for overrepresented or underrepresented groups by using official statistics, it is important to consider that all respondents were recruited via an online social network. As a result, we may have missed a specific subgroup of the Dutch population consisting of people without access to the Internet. However, we believe this group to be small because 92.9% of the Dutch population has Internet access [27]. In relation to the survey, it is important to consider that it did not include questions about diseases and use of medication by respondents, which made it impossible to distinguish between ill and healthy respondents. Realizing that ill patients may have other preferences, future surveys should include questions on this matter. Because the present survey was focused on types of information (eg, social media, Internet, books) future studies should aim to further specify this. For example, they should study which types of social media are used, which search engines are used to search for information, and how consumers rate the reliability of different social media networks or websites.

Conclusion

The Internet is the main source of health-related information for the Dutch population. One in 4 persons would communicate with their physician via social media channels and it is expected that this number will further increase. Therefore, health care providers should explore new ways of communicating online and should facilitate ways for patients to connect with them. Future research should aim at comparing different patient groups and diseases, describing best practices, and determining cost-effectiveness.

Acknowledgments

We would like to acknowledge Mr Marc de Vries and Mrs Saskia Wijsman for their enthusiasm and support in disseminating the survey via the Hyves online social media network. Furthermore, we would like to thank all respondents for their benevolence. The study was funded by the Radboud University Medical Center, Nijmegen, the Netherlands.

Authors' Contributions

All authors contributed to the study protocol. Data were analyzed by TB and ST and interpreted by LE, SB, and LS. TB wrote the first version of the manuscript, which was reviewed by SB, LS, LE, and MS at various stages until its final version. All authors read and approved the final version submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey (English version).

[PDF File (Adobe PDF File), 306KB - jmir_v15i10e220_app1.pdf]

References

1. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
2. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
3. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
4. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010;53. [doi: 10.1016/j.bushor.2009.09.003]
5. McCracken A. #ClinMed: What have social media got to do with medicine? Clin Med 2012 Aug;12(4):310-311. [Medline: 22930871]
6. Van de Belt TH, Berben SA, Samsom M, Engelen LJ, Schoonhoven L. Use of social media by Western European hospitals: longitudinal study. J Med Internet Res 2012;14(3):e61 [FREE Full text] [doi: 10.2196/jmir.1992] [Medline: 22549016]
7. Hawker MD. Social networking in the National Health Service in England: a quantitative analysis of the online identities of 152 primary care trusts. Stud Health Technol Inform 2010;160(Pt 1):356-360. [Medline: 20841708]
8. The Change Foundation. 2011 Jun 01. Using social media to improve healthcare quality URL: http://www.changefoundation.ca/library/using-social-media-to-improve-healthcare-quality-part-1/ [accessed 2013-09-27] [WebCite Cache ID 6Jww3BxJC]
9. 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]
10. 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]
11. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, et al. Why is it difficult to implement e-health initiatives? A qualitative study. Implement Sci 2011;6:6 [FREE Full text] [doi: 10.1186/1748-5908-6-6] [Medline: 21244714]
12. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
13. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernandez ME. Planning Health Promotion Programs: An Intervention Mapping Approach, 3rd Edition. San Francisco: Jossey-Bass; 2011.
14. Grol R, Wensing M, Eccles M. Improving Patient Care: The Implementation of Change in Clinical Practice. Edinburgh: Elsevier Butterworth Heinemann; 2005.
15. Singh PM, Wight CA, Sercinoglu O, Wilson DC, Boytsov A, Raizada MN. Language preferences on websites and in Google searches for human health and food information. J Med Internet Res 2007;9(2):e18 [FREE Full text] [doi: 10.2196/jmir.9.2.e18] [Medline: 17613488]
16. Vandelanotte C, Duncan MJ, Plotnikoff RC, Mummery WK. Do participants' preferences for mode of delivery (text, video, or both) influence the effectiveness of a Web-based physical activity intervention? J Med Internet Res 2012;14(1):e37 [FREE Full text] [doi: 10.2196/jmir.1998] [Medline: 22377834]
17. Baptist AP, Thompson M, Grossman KS, Mohammed L, Sy A, Sanders GM. Social media, text messaging, and email-preferences of asthma patients between 12 and 40 years old. J Asthma 2011 Oct;48(8):824-830. [doi: 10.3109/02770903.2011.608460] [Medline: 21864099]
18. Stroetmann VN, Hüsing T, Kubitschke L, Stroetmann KA. The attitudes, expectations and needs of elderly people in relation to e-health applications: results from a European survey. J Telemed Telecare 2002;8 Suppl 2:82-84. [doi: 10.1258/135763302320302154] [Medline: 12217149]
19. PwC Health Research Institute. 2011 Nov. Top health industry issues of 2012 URL: http://www.pwc.com/us/en/health-industries/publications/top-health-industry-issues-of-2012.jhtml [accessed 2013-03-10] [WebCite Cache ID 6F0r9VMQZ]
20. Hyves. 2012. URL: http://www.hyves.nl/about/press/ [accessed 2012-02-24] [WebCite Cache ID 65glvugo0]
21. STIR. 2011. URL: http://www.stir.nl/resultaten/archief/2011/ [accessed 2013-06-18] [WebCite Cache ID 6HSrn3y1Q]
22. Statistics Netherlands. URL: http://www.cbs.nl/en-GB/menu/home/default.htm?Languageswitch=on [accessed 2013-03-10] [WebCite Cache ID 6F0t4LxEl]
23. Piper LE. Generation Y in healthcare: leading millennials in an era of reform. Front Health Serv Manage 2012;29(1):16-28. [Medline: 23050334]
24. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]
25. Moretti FA, Oliveira VE, Silva EM. Access to health information on the internet: a public health issue? Rev Assoc Med Bras 2012;58(6):650-658 [FREE Full text] [Medline: 23250092]
26. Greaves RA, Wilkinson LF, Orchard T. Accessing information on IBD: a survey of sources and reliability of internet resources. GUT 2012;61:396 [FREE Full text] [doi: 10.1136/gutjnl-2012-302514d.242]
27. Internet World Stats. 2012 Jun 30. Europe Internet usage URL: http://www.internetworldstats.com/stats4.htm#europe [accessed 2013-09-02] [WebCite Cache ID 6JKHUq4GA]

28. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2013-03-10] [WebCite Cache ID 6F0tn2EL1]
29. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
30. Zorgkaart Nederland. URL: http://www.zorgkaartnederland.nl/ [accessed 2013-03-10] [WebCite Cache ID 6F0typXiK]
31. 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]
32. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
33. Facetalk. URL: http://www.facetalk.nl/ [accessed 2013-03-10] [WebCite Cache ID 6F17j484d]
34. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2013 May 28. [doi: 10.1111/hex.12089] [Medline: 23711169]

Abbreviations

CBS: Centraal Bureau voor de Statistiek (Statistics Netherlands)

Edited by G Eysenbach; submitted 10.03.13; peer-reviewed by F Grajales III, L Alpay; comments to author 23.04.13; revised version received 27.06.13; accepted 29.08.13; published 02.10.13.

Please cite as:
Van de Belt TH, Engelen LJLPG, Berben SAA, Teerenstra S, Samsom M, Schoonhoven L
Internet and Social Media For Health-Related Information and Communication in Health Care: Preferences of the Dutch General Population
J Med Internet Res 2013;15(10):e220
URL: http://www.jmir.org/2013/10/e220/
doi:10.2196/jmir.2607
PMID:24088272

Original Paper

A Web-Based Tool to Support Shared Decision Making for People With a Psychotic Disorder: Randomized Controlled Trial and Process Evaluation

Lian van der Krieke[1], MSc, MA; Ando C Emerencia[2], MSc; Nynke Boonstra[3], PhD; Lex Wunderink[1,3], PhD; Peter de Jonge[1], PhD; Sjoerd Sytema[1], PhD

[1]University of Groningen, University Medical Center, University Center for Psychiatry, Groningen, Netherlands

[2]University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science, Groningen, Netherlands

[3]Friesland Mental Health Care Service, Leeuwarden, Netherlands

Corresponding Author:
Lian van der Krieke, MSc, MA
University of Groningen
University Medical Center
University Center for Psychiatry
Hanzeplein 1
Groningen, 9700 RB
Netherlands
Phone: 31 503612108
Fax: 31 503619722
Email: j.a.j.van.der.krieke@umcg.nl

Abstract

Background: Mental health policy makers encourage the development of electronic decision aids to increase patient participation in medical decision making. Evidence is needed to determine whether these decision aids are helpful in clinical practice and whether they lead to increased patient involvement and better outcomes.

Objective: This study reports the outcome of a randomized controlled trial and process evaluation of a Web-based intervention to facilitate shared decision making for people with psychotic disorders.

Methods: The study was carried out in a Dutch mental health institution. Patients were recruited from 2 outpatient teams for patients with psychosis (N=250). Patients in the intervention condition (n=124) were provided an account to access a Web-based information and decision tool aimed to support patients in acquiring an overview of their needs and appropriate treatment options provided by their mental health care organization. Patients were given the opportunity to use the Web-based tool either on their own (at their home computer or at a computer of the service) or with the support of an assistant. Patients in the control group received care as usual (n=126). Half of the patients in the sample were patients experiencing a first episode of psychosis; the other half were patients with a chronic psychosis. Primary outcome was patient-perceived involvement in medical decision making, measured with the Combined Outcome Measure for Risk Communication and Treatment Decision-making Effectiveness (COMRADE). Process evaluation consisted of questionnaire-based surveys, open interviews, and researcher observation.

Results: In all, 73 patients completed the follow-up measurement and were included in the final analysis (response rate 29.2%). More than one-third (48/124, 38.7%) of the patients who were provided access to the Web-based decision aid used it, and most used its full functionality. No differences were found between the intervention and control conditions on perceived involvement in medical decision making (COMRADE satisfaction with communication: $F_{1,68}$=0.422, P=.52; COMRADE confidence in decision: $F_{1,67}$=0.086, P=.77). In addition, results of the process evaluation suggest that the intervention did not optimally fit in with routine practice of the participating teams.

Conclusions: The development of electronic decision aids to facilitate shared medical decision making is encouraged and many people with a psychotic disorder can work with them. This holds for both first-episode patients and long-term care patients, although the latter group might need more assistance. However, results of this paper could not support the assumption that the use of electronic decision aids increases patient involvement in medical decision making. This may be because of weak implementation of the study protocol and a low response rate.

Trial Registration: Dutch Trial Register (NTR) trial number: 10340; http://www.trialregister.nl/trialreg/admin/rctsearch.asp?Term=10340 (Archived by WebCite at http://www.webcitation.org/6Jj5umAeS).

(J Med Internet Res 2013;15(10):e216) doi:10.2196/jmir.2851

KEYWORDS

psychotic disorders, schizophrenia; computers; computer-assisted decision making; shared decision making; feasibility studies, randomized clinical trial

Introduction

Shared decision making in mental health care has been dubbed an ethical imperative [1]. Since the rise of recovery-oriented medicine, patients have been acknowledged as experiential experts and equal partners in communication with clinicians. Research has shown that people with severe and persistent mental disorders are no exception. People with psychotic disorders are able and willing to participate in medical decision making [2,3]. However, the desire for participation is greater than the amount of participation they actually experience [4,5]. A range of obstacles hamper successful implementation. Most clinicians believe in the benefits of shared decision making, but time constraints and a large number of clinical responsibilities prevent them from practicing it [6,7]. Moreover, patients may not be used to actively participating in medical decision making and they can lack access to medical information that is easily intelligible [8].

Drake and Deegan [9] stressed the need for decision aids and support centers to ensure the development of an infrastructure that facilitates the practice of shared decision making. Several initiatives have been developed in this area. For instance, in Germany, Hamann et al [3] investigated the effectiveness of a shared decision-making intervention with a printed decision aid for inpatients with schizophrenia. They found that patients using the decision aid had better knowledge about their disease and had a higher perceived involvement in medical decisions compared to a control group that received care as usual [3]. Recently, a special case was made for electronic decision aids [10] because they have various advantages over paper-based decision aids, such as presenting personalized information based on smart algorithms. So far, 3 electronic decision aids have been developed and investigated to support shared decision making in the treatment planning for people with severe mental disorders, but the results are inconsistent [11,12]. A pilot study by Deegan et al [11] showed that outpatients were able to work with a Web-based program to support shared decision making in psychopharmacological consultation. Patients used the program on computers at the clinic where experiential experts were available for assistance. Two small-scale randomized clinical trials were conducted [12,13]. The first trial showed that patients were able to electronically design their own care plan, but there was no difference between intervention and control groups in satisfaction with the care planning process, which was the primary outcome [12]. The second trial reported that a Web-based support system encouraging patients to discuss their current status and treatment with their clinician resulted in patients being more verbally active during health visits [13]. More evidence is needed to determine whether electronic decision aids are helpful in clinical practice and can lead to increased patient involvement and better outcomes. In addition, more information is needed about what proportion of patients are willing and able to work with Web-based decision aids and in what form (with or without assistance, using their own computer or a clinic computer). This paper reports on a randomized controlled trial and process evaluation of a Web-based intervention to facilitate shared decision making, with or without assistance, for people with psychotic disorders. Our aim was to investigate this intervention in a naturalistic setting, meaning that all eligible patients were included to be able to determine how many of them would actually use the decision aid.

Methods

Ethical Considerations

Informed consent was obtained by research nurses. Patients were provided with an information brochure and they received a phone number and email address of a research assistant who they could contact for further information. The national Dutch medical ethical committee for mental health care (Medisch-ethische Toetsingscommissie instellingen Geestelijke Gezondheidszorg; METiGG) assessed the study protocol and judged that the study could be conducted without the committee’s approval. The trial was registered at the Dutch Trial register (NTR trial number: 10340).

Setting and Participants

The study was carried out in a Dutch mental health institution (Friesland Mental Health Care Service, city of Leeuwarden) with a catchment area of approximately 650,000 inhabitants. Data were collected from June 2011 to July 2012. The trial was completed when all patients provided their last measurement. Patients were recruited from 2 outpatient teams for psychosis: the early intervention for psychosis team (a multidisciplinary team for the treatment of patients with a first episode of psychosis) and a rehabilitation team (a multidisciplinary team for patients with chronic schizophrenia). We used broad inclusion criteria. Participants had to meet criteria for a nonaffective psychosis (brief psychotic disorder, schizophreniform disorder, schizoaffective disorder, schizophrenia, or psychotic disorder not otherwise specified) as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition, Text Revision) (*DSM-IV-TR*), be between age 21 and 65 years, and be fluent in Dutch. Participating professionals were all clinicians involved in the care for those patients describe previously (psychiatrists, community psychiatric nurses, psychologists). Internet or computer literacy was not part of the inclusion criteria.

To calculate the sample size, we used the SPSS SamplePower software program (IBM Corp, Armonk, NY, USA). Given an alpha of .05, a power of .80, and an effect size of .50 (based on results of a comparable study [3]), we needed n=64 per group. Because we expected a considerable amount of dropout (50%) and we wanted to investigate what proportion of patients in the participating teams would use the Web-based decision aid, we decided to include all eligible patients treated by the participating teams.

Study Design

We conducted an open-label, 2-group, parallel, randomized controlled trial with approximately the same number of patients in each group. Patients were allocated to either an intervention group that was offered a Web-based tool to support shared decision making or a control group that received care as usual. Randomization of patients was conducted by using the online Research Randomizer [14]. We used block randomization in blocks of 8 (numbers 1 to 4 were considered intervention condition; 5 to 8 control condition). A research assistant located at the mental health institution participating in the study created a spreadsheet file listing all participants in ascending order by research number. Another research assistant located at our research center added the randomization conditions to the spreadsheet, assigning participants to the interventions.

Treatment Conditions

Control Condition

Patients in the control condition received care as usual, as described in the local disease management program for the treatment of people with psychosis. Treatment modules were initially chosen by a clinician in accordance with a treatment path that a patient entered based on the staging of the disorder (first episode or stabilizing/rehabilitation phase), clinician-rated scores on the Health of the Nation Outcome Scale (HoNOS), and patient-rated scores on the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS-P). During a treatment plan meeting, clinicians informed patients about the indicated treatment modules and also discussed alternatives. A final decision was made in a process of shared decision making (which was not further specified in the disease management program).

Intervention Condition

Patients in the intervention condition received care as described in the local disease management program for the treatment of people with psychosis plus they were offered the opportunity to make use of the Web-based information and decision tool (see Multimedia Appendix 1). This tool is meant to support patients in acquiring an overview of their care needs and of the treatment modules provided by their mental health care organization. The tool functions as a website consisting of 3 webpages and a home page. The home page briefly explains the aim and procedure of the website. The first webpage presents a questionnaire about care needs based on items of the CANSAS-P (see Figure 1). The second webpage offers a digital catalog with descriptions of treatment modules dynamically linked to the outcomes of the questionnaire in the first webpage (see Figure 2). For instance, a reported need for more information about symptoms and medication use was linked to information in a module about psychoeducation, whereas a reported need on items about living a meaningful life and doubts about the future was linked to a module about loss and longing.

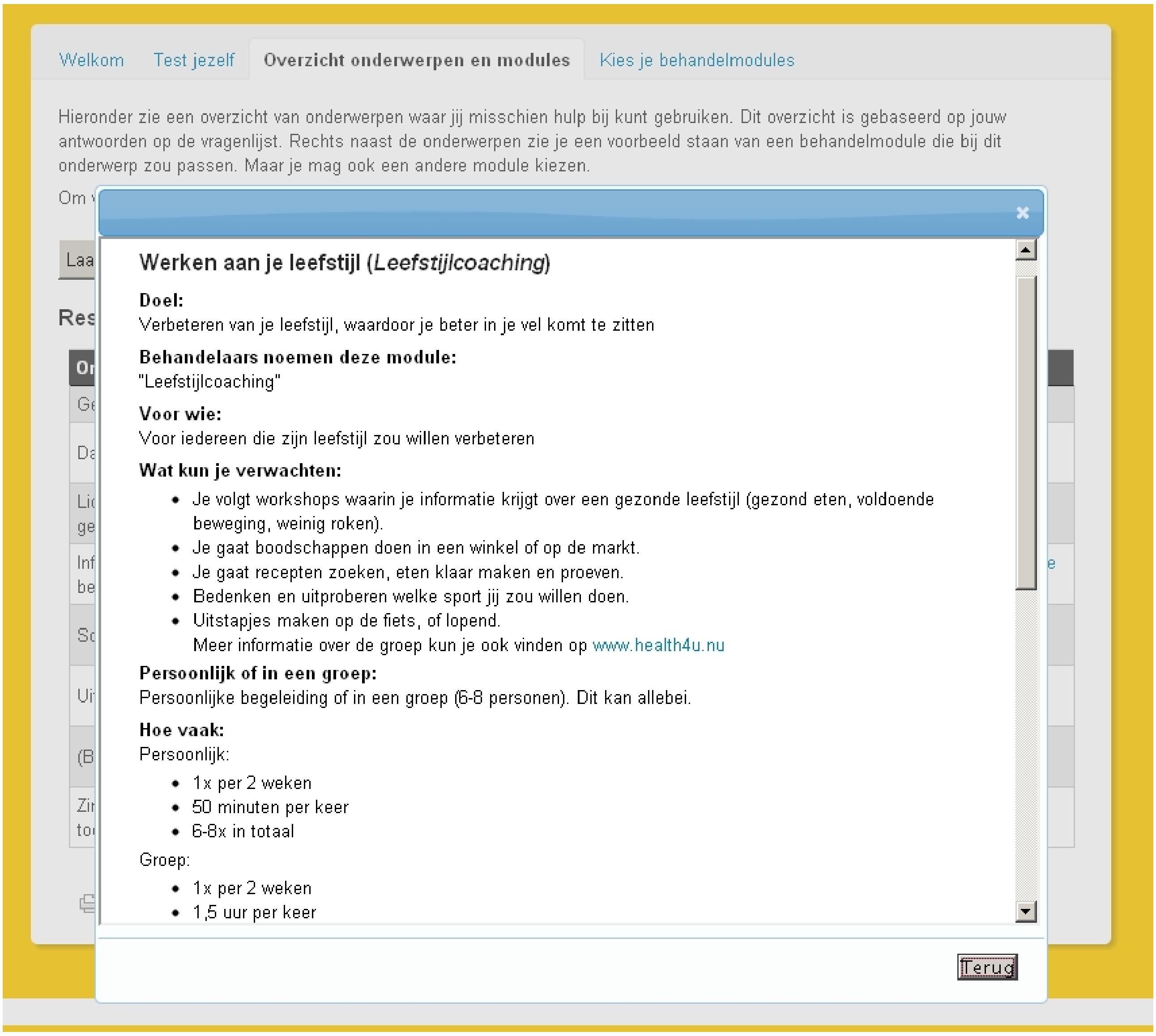
In addition to this selection of modules, patients also had the opportunity to view all available treatment modules irrespective of the questionnaire outcomes. The information about the available modules in the catalog included an overview of its content and duration; a description of problems/symptoms the treatment module is usually indicated for; names, functions, and pictures of clinicians involved; a short story by a patient who tells his/her experience with the treatment module (see Figure 3); and, if available, a brief interview with a clinician who tells about his/her experience with the treatment module (advantages, disadvantages, motivation to provide the treatment, etc). The third webpage presents a list of all treatment modules in a checkbox format. The content and design of this Web-based tool was based on an earlier usability study and needs assessment [15]. During the development process, the content of the tool was validated by clinicians and patients. This content was frozen during the trial.

Patients using the Web-based tool were asked to look through the treatment modules and choose the modules of their preference by ticking the appropriate checkboxes. Patients could print the checkbox form and take it with them to their treatment plan evaluation session to discuss with their clinician.

Patients were informed about the Web-based decision aid by research nurses during a biyearly appointment for Routine Outcome Monitoring (ROM), and they were offered an information brochure. Patients were given the opportunity to use the decision aid either on their own (at their home computer, or at a computer of the service) or with support of an assistant. Furthermore, an assistant was available by phone for help on 3 days each week. Patients received a log-in account by email or on paper from an assistant. No further instructions were given about the optimal timing of frequency regarding the use of the decision aid.

Figure 1. Screenshot of the first webpage with a questionnaire (in Dutch) about care needs.

Figure 2. Screenshot of the second webpage including a digital catalog with descriptions of treatment modules.



Hieronder zie een overzicht van onderwerpen waar jij misschien hulp bij kunt gebruiken. Dit overzicht is gebaseerd op jouw antwoorden op de vragenlijst. Rechts naast de onderwerpen zie je een voorbeeld staan van een behandelmodule die bij dit onderwerp zou passen. Maar je mag ook een andere module kiezen.

Werken aan je leefstijl (*Leefstijlcoaching*)

Doel:
Verbeteren van je leefstijl, waardoor je beter in je vel komt te zitten

Behandelaars noemen deze module:
"Leefstijlcoaching"

Voor wie:
Voor iedereen die zijn leefstijl zou willen verbeteren

Wat kun je verwachten:

- Je volgt workshops waarin je informatie krijgt over een gezonde leefstijl (gezond eten, voldoende beweging, weinig roken).
- Je gaat boodschappen doen in een winkel of op de markt.
- Je gaat recepten zoeken, eten klaar maken en proeven.
- Bedenken en uitproberen welke sport jij zou willen doen.
- Uitstapjes maken op de fiets, of lopend.
  Meer informatie over de groep kun je ook vinden op www.health4u.nu

Persoonlijk of in een groep:
Persoonlijke begeleiding of in een groep (6-8 personen). Dit kan allebei.

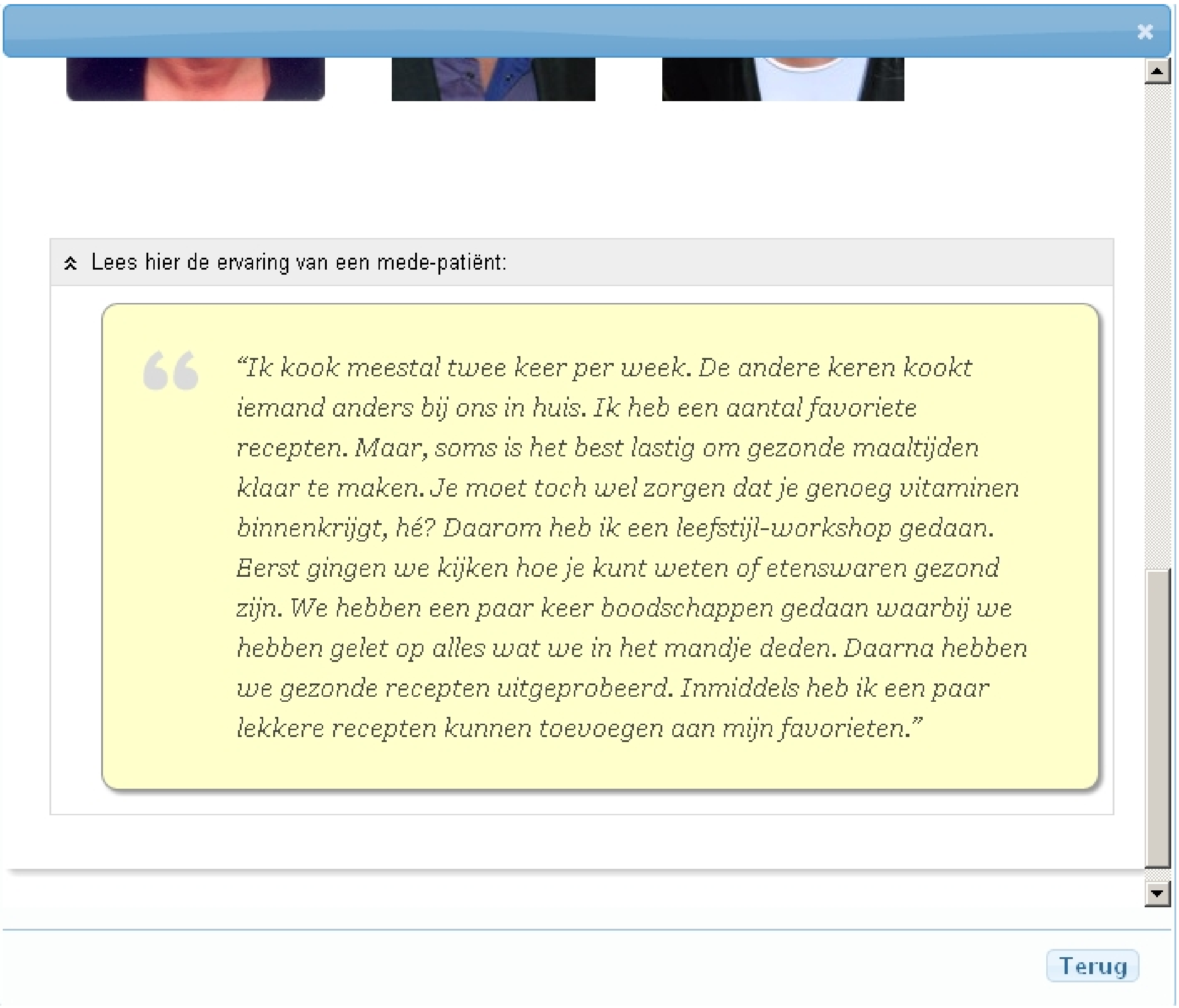
Hoe vaak:
Persoonlijk:

- 1x per 2 weken
- 50 minuten per keer
- 6-8x in totaal

Groep:

- 1x per 2 weken
- 1,5 uur per keer

Figure 3. Screenshot of the second webpage with a short patient story.



Lees hier de ervaring van een mede-patiënt:

“Ik kook meestal twee keer per week. De andere keren kookt iemand anders bij ons in huis. Ik heb een aantal favoriete recepten. Maar, soms is het best lastig om gezonde maaltijden klaar te maken. Je moet toch wel zorgen dat je genoeg vitaminen binnenkrijgt, hé? Daarom heb ik een leefstijl-workshop gedaan. Eerst gingen we kijken hoe je kunt weten of etenswaren gezond zijn. We hebben een paar keer boodschappen gedaan waarbij we hebben gelet op alles wat we in het mandje deden. Daarna hebben we gezonde recepten uitgeprobeerd. Inmiddels heb ik een paar lekkere recepten kunnen toevoegen aan mijn favorieten.”

Procedure

After randomization, baseline measurement took place during a biyearly face-to-face ROM session for all participating patients. Participating clinicians were asked to complete an attitude questionnaire around the same time. Up to 6 weeks after the ROM session, patients in the intervention condition had the opportunity to make use of the Web-based tool. Approximately 6 weeks after ROM, a meeting was planned between the patient and a key clinician in which ROM results were evaluated and a new treatment plan was created or an existing one was adjusted. Patients were sent a final questionnaire by mail. Upon returning the questionnaire to our research center, they received a gift certificate worth €7.50. We deviated from the procedure described in the original research protocol in 1 important aspect: we conducted 1 follow-up measurement instead of 2 because a second follow-up meeting appeared to be not feasible within the time limits.

Measures

Baseline

Self-reported quality of life was measured with the Manchester Short Assessment of Quality of Life (MANSA) [16]. Patients rate their satisfaction with life on different life domains, in 16 items on a 7-point Likert scale, ranging from very dissatisfied to very satisfied. Higher scores indicate a better quality of life.

Psychosocial functioning was measured with the HoNOS [17]. Clinicians rate patients on 12 domains on a 5-point severity scale ranging from no problem to severe or very severe problem. Lower scores indicate a better psychosocial functioning.

Symptom severity was measured with the Positive and Negative Syndrome Scale (PANSS) [18]. Clinicians rate patients during an interview on 7 items about positive symptoms, 7 items about negative symptoms, and 16 items about general psychopathology on a 7-point Likert scale ranging from absent to extreme. Lower scores indicate less symptom severity.

Patients’ preference to participate in medical decision making was measured by the decision-making preference subscale of the Autonomy Preference Index (API) [19]. Patients rate their preference on a 6-item scale in which item scores range from completely disagree (score 0) to completely agree (score 100). A higher score indicates more preference for autonomy.

Outcome

The primary outcome measure was patient-perceived involvement in medical decisions measured with the patient-rated Combined Outcome Measure for Risk Communication and Treatment Decision-making Effectiveness (COMRADE) [20]. The COMRADE consists of 2 subscales, satisfaction with communication and confidence in decision, comprising 20 items in total and scored on a 5-point scale. Higher scores indicate higher perceived involvement.

We used the patient-rated Client Satisfaction Questionnaire (CSQ) [21] as a secondary outcome measurement. The CSQ used in this study consists of 9 items, scored on a 4-point scale. Higher scores indicate higher satisfaction. For the intervention group, we added 6 questions about satisfaction with the Web-based decision tool.

Analysis

Descriptive statistics were used to investigate client characteristics. Baseline measures of both conditions were compared using unpaired *t* tests or chi-square tests. Difference between the intervention and the control condition on the primary outcome measure was examined using a general linear model with adjustments for patient age and partner status (having a partner yes/no).

Process Evaluation

The intervention described previously can be considered a complex intervention because it consists of several components (use of new technology, implementation in regular care, evaluation) and is highly dependent on the context in which it is delivered. Complex interventions are interventions that contain various interacting components of which the whole is more than the sum of its parts [22,23]. For these interventions, a randomized controlled trial needs to be supplemented by a process evaluation to evaluate their effect. Process evaluations explore implementation issues and contextual factors within the trial. They help to distinguish between ineffective interventions (failure of intervention) and badly delivered interventions (implementation failure) [22].

The process evaluation of this study consisted of

1. Open interviews with a sample of 15 patients who did and did not receive the allocated intervention. An interview guide was created in accordance with the guidelines provided by Hennink et al [24]. A verbatim transcript was created for each interview. Coding and analysis was performed with the ATLAS.ti software package.
2. Researcher observation of clinicians discussing implementation of the intervention during clinical meetings, which were recorded in a notebook by a research assistant. Themes of interest were identified by the research team and further discussed with the clinical teams when necessary.
3. A questionnaire-based survey among clinicians consisting of 3 parts: (1) investigating their attitude toward shared decision making and the use of a Web-based decision aid (based on Punter [25] and Holmes-Rovner et al [26] with internal consistency alpha =.85); (2) examining potential hampering factors for shared decision making (based on Charles et al [27]); and (3) exploring to what extent clinicians considered patients to be capable and interested in shared decision making (based on Hamann et al [3]).

This process evaluation provided data to shed light on how well the intervention was implemented, to what extent the trial outcomes were related to the quality of the implementation and the setting in which it was implemented, and what processes might have mediated these relations.

Results

Process Evaluation

In the process evaluation, we collected data to answer 5 questions about potential problems related to implementation and context.

The first question was: Could the outcomes be affected by a negative attitude of clinicians toward shared decision making or the Web-based decision aid? In a questionnaire-based survey, clinicians' attitudes were investigated. On a 5-point Likert scale ranging from completely disagree to completely agree, clinicians agreed or completely agreed with 4 statements about shared decision making in general, and 9 statements about the use of a decision aid in decision-making processes. The mean total score on this scale was 3.52 (SD 0.49), meaning that most clinicians showed a positive attitude toward shared decision making and the use of decision aids. Table 1 shows to what extent clinicians agreed or disagreed with the statements.

The second question was: Do clinicians think there are too many hampering factors to realize a process of shared decision making? In addition, 18 clinicians reported that in processes of shared decision making, the following factors were often or almost always experienced as hampering decision making: patients receive contradictory advice from multiple clinicians (12/18, 67%), patients have difficulty accepting their diagnosis (12/18, 67%), and patients are indecisive (10/17, 59%). The following factors were reported as never or sometimes hampering: patients want to participate to a greater degree than the clinician prefers (15/18, 83%), patients have other interfering health problems (15/18, 83%), lack of time (14/18, 78%), cultural differences (14/18, 78%), patients bring in too much information to discuss (13/18, 72%), patients ask for a treatment that is not evidence-based (12/17, 71%), clinician has too little information to make a decision (12/17, 71%), patients do not understand the information (12/18, 67%), patients are too anxious or worried to listen to what the clinician has to say to them (11/18, 61%), and patients refuse treatment that could benefit them (10/18, 56%).

The third question was: Could the outcomes be affected by the clinicians' judgment about patients' capabilities and interests? Clinicians were asked to what extent they considered patients to be capable and interested in shared decision making. Of the 128 patient observations, clinicians rated most patients as being able to understand the arguments presented, being capable of making reasonable decisions, and being interested in the topics discussed as well as in participating in medical decision making. Patients who were rated by their clinicians as not capable of making decisions (score 1-3) had a significantly lower score than patients rated as capable of making decisions on both subscales of the COMRADE (COMRADE satisfaction with communication: t_{48}=–3.857, P<.001; COMRADE confidence in decision: t_{47}=–2.368, P=.02. This means that patients who perceived their involvement in medical decision making to be low were judged by clinicians to be less capable of participating in decision making.

The fourth question was: Could any problems be observed with fulfillment of the study protocol? Through researcher observation, several recurring themes were identified during clinical meetings in which the trial was discussed. Case managers sometimes were hesitant and felt troubled to invite intervention patients to make use of the decision tool. First, they were doubtful whether patients were able to handle either the computer program or participation in a research trial. Second, they were not sure that patients would benefit from the decision aid because not all treatment options included in the decision aid were actually offered by their organization (eg, music therapy was listed among the treatment options, but no music therapy was currently offered because of absence of a music therapist). In addition, various clinicians reported that they were unsure when to discuss outcomes of the decision aid with their patients because not all conducted a formal treatment evaluation session with their patients following their ROM assessment. Some only discussed ROM results within the clinical team and not directly with patients.

The fifth question was: Did patients experience any problems with the intervention that was not covered in the satisfaction questionnaire? Open interviews among patients who chose to use or not use the website provided some additional details on the process. First, all patients were initially informed about the decision aid by an information booklet and in a meeting with a research nurse, but most of them received additional explanation from their case manager. Some framed the decision aid predominantly within a research context ("by using the decision aid, you contribute to research"), whereas others described it as an attempt to improve services ("using the decision aid might help you reflect on the treatment you want"). This might have affected patients' expectations of the intervention. Moreover, interviews revealed discrepancies between the policy of the local disease management program and patients' experiences in clinical practice. Most of the interviewed patients could not remember their ROM results being discussed with them and some could not remember whether a treatment plan was created.

Allocation and Reception of Intervention

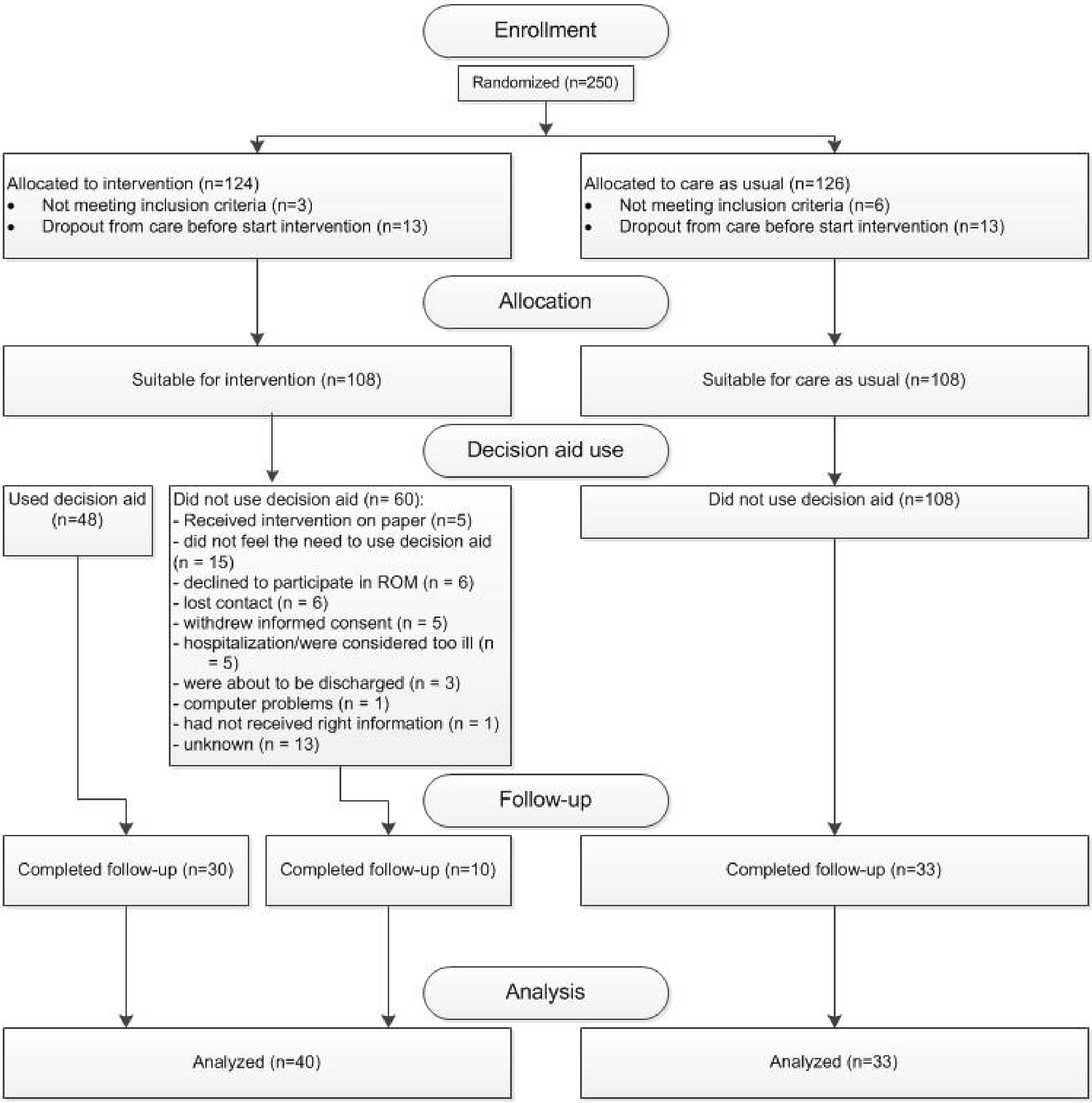
A total of 250 patients (n=124 intervention vs n=126 control) were included in the trial of whom 73 completed the follow-up measurement and were included in the final analysis (response rate 29.2%). Of these 73 patients, 40 were in the intervention and 33 in the control condition. Of the 40 patients in the intervention condition who completed the follow-up measurement, 30 used the decision aid. A detailed overview of the flow of participants is presented in Figure 4.

Table 1. Percentage of clinicians (completely) agreeing with statements about shared decision making and decision aids (n=19).

Item	Agree or completely agree, n (%)
A decision aid will cause patients to ask more questions than they would otherwise have asked	16 (84)
A decision aid will cause patients to be more involved in decision making about treatment[a]	15 (83)
All eligible patients should be invited to use the decision aid	15 (79)
Knowing risks and benefits, most patients want to decide how acceptable treatment is to them	13 (68)
Patients using a decision aid will be much better informed	13 (68)
Patients should see a decision aid before a treatment decision is made	12 (63)
Patients usually want to be an equal partner with physicians in making important treatment decisions	10 (53)
With a decision aid, I will be able to reduce time spent educating patients about treatment[a]	7 (39)
Most patients prefer the clinicians to take responsibility for their medical problems	4 (21)
Using a decision aid will reduce the risk of malpractice	4 (21)
A decision aid will eliminate the need for third-party utilization as second opinion	3 (16)
A decision aid may cause some patients to make the wrong choice	3 (16)
The majority of patients do not wish to be involved in decision making about their treatment	1 (5)

[a] n=18.

Figure 4. Participant flow diagram.



Demographic Variables and Baseline Data

Demographic variables and baseline data of patients included in the analysis are presented in Table 2. Patients in the 2 conditions did not differ in age, Global Assessment of Functioning (GAF), MANSA, HoNOS, PANSS, API, level of education, whether they had a job or were studying, and whether or not they used antipsychotics. However, in the intervention group were fewer females (*P*=.01) and fewer patients with a partner (*P*=.01).

The patients who dropped out of the study and did not complete the follow-up measurement were slightly younger (t_{213}=–2.129, *P*=.03) and were more often men (χ^2_1=5.6, *P*=.02) than the patients who did complete the outcome measurement. They did not differ on any of the other baseline characteristics. Patients in the intervention condition who received the allocated intervention versus those who did not receive the intervention did not differ on all baseline characteristics.

Table 2. Demographic variables and baseline data of study participants.

Variable	Intervention (n=40)	Control (n=33)	*P* [a]
Age (years), mean (SD)	37 (12.35)	40 (13.47)	.35
Sex (female), n (%)	13 (33)	21 (64)	.01
Education (≥ 12 years), n	10 (n=12)	10 (n=12)	.99
Job or study, n (%)	13 (33; n=39)	16 (48)	.23
Partner, n (%)	9 (23; n=39)	18 (55)	.01
Use of antipsychotics, n (%)	29 (73)	22 (67)	.60
Test, mean (SD) [b]			
GAF	61.8 (9.08)	57.4 (10.91)	.06
MANSA	60.7 (9.50)	62.3 (13.26)	.58
HoNOS	7.7 (4.75)	8.4 (4.32)	.53
PANSS total score	13.3 (5.24)	15.4 (5.51)	.13
API	55.7 (12.72)	52.7 (12.96)	.38
Number of patients from the first episode of psychosis team within condition, n (%)	16 (40)	13 (39)	.99

[a] Using Fisher exact test or *t* test.

[b] GAF: Global Assessment of Functioning; MANSA: Manchester Short Assessment of Quality of Life; HoNOS: Health of the Nation Outcome Scales; PANSS: Positive and Negative Syndrome Scale; API: Autonomy Preference Index.

Patient Involvement in Treatment Planning and Their Satisfaction With Care

Intention-to-treat analyses showed that patients in the intervention condition did not differ from patients in the control condition in their perceived involvement in medical decision making (COMRADE) after they had used the Web-based decision aid (COMRADE satisfaction with communication: $F_{1,68}$=0.422, *P*=.52; COMRADE confidence in decision: $F_{1,67}$=0.086, *P*=.77; see also Table 3). This was the primary outcome measure. Patients also did not differ in self-reported satisfaction with care (CSQ) ($F_{1,70}$=0.014, *P*=.91).

Per protocol analyses also showed that patients in the intervention condition who received the allocated intervention and completed the follow-up measure (n=30) did not differ regarding their perceived involvement in medical decision making and in satisfaction with care from patients in the control condition (n=33) (COMRADE satisfaction with communication: $F_{1,57}$=0.155, *P*=.70; COMRADE confidence in decision: $F_{1,56}$=0.413, *P*=.52; CSQ: $F_{1,60}$=0.789, *P*=.34).

In an additional analysis, patients in the intervention condition who received the allocated intervention (n=30) were compared to patients in the intervention condition who did not receive the allocated intervention (n=10). No differences were found for patients' perceived involvement in medical decision making (COMRADE satisfaction with communication: $F_{1,36}$=0.642, *P*=.43; COMRADE confidence in decision: $F_{1,36}$=2.310, *P*=.14). Patients did, however, differ on the secondary outcome self-reported satisfaction with care ($F_{1,37}$=6.306, *P*=.02). Patients who received the allocated intervention were less satisfied than patients who did not.

Table 3. Primary outcome data of patients' perceived involvement in medical decision making at the end of the study using the Combined Outcome Measure for Risk Communication and Treatment Decision-making Effectiveness (COMRADE) test.

COMRADE subscale[a]	Intervention, mean (SD)	Control, mean (SD)	*F* (df)	*P*
Satisfaction with communication (n=73)	38.25 (1.06	37.19 (1.165)	0.422 (1,68)	.52
Confidence in decision (n=70)	38.78 (1.17)	38.72 (1.307)	0.086 (1,67)	.77

[a] Group differences were analyzed using a general linear model with age and partner status as covariates.

Use of and Satisfaction With the Web-Based Decision Aid

Of the 48 patients who used the Web-based decision aid, 12 used their own computer, 12 used the computer at the clinic, and 6 used a computer elsewhere. Furthermore, 13 used the decision aid independently, 16 received assistance from a professional (often their case manager), and 1 received assistance from someone else. First-episode patients used their own computer and used the decision aid without assistance more often than chronic patients did. Of the 48 patients who used the website, 34 (71%) used full functionality of the Web-based decision aid, meaning that patients completed the care needs assessment (first webpage of the website) and looked through the digital catalog with descriptions of treatment modules (second webpage of the website). More than half of them were long-term care patients (27/48, 56%).

In the intervention condition, 29 of 48 patients who used the decision aid (60%) completed questions about their satisfaction with the decision aid. They agreed or completely agreed with the following statements: “I have been well informed about the treatment options offered by Friesland Mental Health Care Service by the decision aid” (22/29, 76%), “The advice presented by the decision aid has helped me to reflect on what I want” (22/29, 76%), “The decision aid was easy to use” (20/28, 71%), “I would recommend the decision aid to others” (20/27, 74%) and “The decision aid helped me to get a clearer view on what my problem areas or points of interest are” (17/28, 61%). Patients were divided on whether the decision aid helped them to better prepare the evaluation meeting with their clinicians, 44% (12/27) said it did help; 56% (15/27) were neutral or said it did not help. Means and standard deviations can be found in Table 4.

Table 4. Secondary outcome data of patients’ satisfaction with the Web-based decision aid.

Question	Mean (SD)[a]
I have been well informed about the treatment options offered by the GGZ Friesland by the decision aid (n=29)	3.93 (0.84)
The advice presented by the decision aid has helped me to reflect on what I want (n=29)	3.86 (0.79)
As a consequence of using the decision aid, I was better prepared for the evaluation meeting with my clinician (n=27)	3.33 (0.78)
The decision aid helped me to get a clearer view on what my problem areas or points of interest are (n=28)	3.61 (0.92)
The decision aid was easy to use (n=28)	3.79 (1.07)
I would recommend the decision aid to others (n=27)	3.89 (0.75)

[a] Scores ranged from 0 (completely disagree) to 5 (completely agree).

Discussion

Principal Findings

In this study, we report on a clinical trial and process evaluation of a Web-based intervention to facilitate shared decision making for people with psychotic disorders.

To be able to explore potential implementation issues and contextual problems within the trial, we conducted a process evaluation. This evaluation showed that no significant problems could be observed in the attitude and beliefs of clinicians. Participating clinicians had an overall positive attitude toward shared decision making. They reported that their patients were generally interested in and capable of participating in medical decision making, they considered patient decision aids be to potentially helpful, and they judged relatively few factors to be hampering in a shared decision-making process. However, problems were observed in the implementation of the intervention. Not all patients in the intervention group were actually offered the possibility to use the decision aid and, more importantly, ROM and treatment evaluation meetings in which the treatment plan was to be discussed in a process of shared decision making did not always take place. Moreover, interviews indicate that the Web-based intervention might have been framed differently to different patients, which may have shaped their expectations and affected their evaluation. An interesting finding in the process evaluation was that patients who perceived their involvement in medical decision making as low were judged by clinicians to be less capable of participating in decision making. This could imply that patients participate less because they are less capable. Nevertheless, we cannot rule out that patients participate less because clinicians consider them less capable and, therefore, provide less opportunities for patients to participate in decision making.

The findings of our trial show that more than one-third of the patients who were provided access to the Web-based decision aid chose to use it and most used full functionality of the decision aid whether they were first-episode patients or long-term patients. Users and nonusers did not differ in demographic variables. At least one-quarter of the patients used their own computer and a similar proportion used the decision aid without assistance. Most of these were first-episode patients. On average, users of the decision aid reported to be rather satisfied with the system. Nevertheless, primary outcome results could not support the assumption that the use of electronic decision aids increases patient involvement in medical decision making, neither in intention-to-treat analyses nor in per protocol analyses. In addition, we did not find a difference in self-reported satisfaction with care between patients who had the opportunity to use the decision aid versus those who did not.

Our outcomes are in-line with the study by Woltmann et al [12] who found no difference in patient satisfaction between intervention and control group. However, they contradict the findings by Hamann et al [3] and Steinwachs et al [13] who found a positive effect of decision aids on patients’ involvement in consultations with their clinicians. This discrepancy can be explained by several reasons. First, the decision aids used in these trials differed in format (Hamann et al [3] used a printed decision aid) and content. Some decision aids primarily concentrated on pharmacological information, whereas others had a broader focus. Second, settings were different. In our study, patients could use the decision aid either in the clinic or at home, with or without assistance, whereas in the trial by Hamann et al [3], patients used the decision aid in a psychiatric ward with assistance of trained nurses. The setting in the study by Steinwachs et al [13] was not described. Third, our response rate was very low. This is partly because of the naturalistic setting of our study. However, response rates are highly dependent on selection criteria used in studies. For example, if

Steinwachs et al [13] included all eligible patients (eg, not excluding patients who were considered unsuitable by their clinician), their response rate would have been comparable. Fourth, the outcome measures used in our study might have been too unspecific, indirect, or insensitive to detect differences in a small sample. The COMRADE measures patients' perceived involvement in medical decision making with a self-report questionnaire that is completed retrospectively. What actually happens during the conversation between patient and clinician remains a black box. Furthermore, research has shown that ratings on patient satisfaction questionnaires tend to be more optimistic than patients' actual evaluations [28,29], implying that there may be less differentiation in the response behavior. Finally, discrepancies could, but are not likely to, be explained by lack of need for shared decision making in our patient sample. Patients' mean score on the API, which indicates their preference for participation in medical decision making, was comparable to or even slightly higher than previous studies in people with schizophrenia [2,3,30].

Strengths and Limitations

Given the problems observed in the process evaluation, the intervention designed for our study appeared not to fit in optimally with the routine practice of the participating clinical care teams. Therefore, the lack of significant effects on our outcome measures cannot be solely attributed to failures intrinsic to the intervention. Future studies might benefit from a stronger integration of shared decision-making interventions in clinical practice by training clinical teams in using (output) from decision aids. A comprehensive overview of the working flow of patients and clinicians is crucial to realize this integration. Given the low response rate and moderate participation rate in this study, it may also be desirable to investigate efficacy of decision aids in a less naturalistic setting in which participating patients are selected more strictly and required to use the decision aid before performing a naturalistic study. In addition, special attention should be paid to the selection of outcome measures used to assess the shared decision-making process. Instruments focusing on satisfaction might suffer from ceiling effects, and instruments such as the COMRADE may be too broad and indirect to detect changes in the decision-making process. A better alternative is to record conversations between clinicians and patients and observe what is actually happening within that conversation. A promising instrument for this may be the recently developed Mappin'SDM [31], which combines patient, clinician, and observer perspectives. It is also important to note that using Web-based decision aids or support systems does not need to be a desirable target for all patients. Although some may benefit from new tools, others might not. It would be most helpful to know what works for whom.

The main limitation of this study is the weak implementation of the study protocol; as a result, it is difficult to draw firm conclusions about the study's outcomes. We tried to prevent this by preparing the participating teams before the start of the trial and keeping closely in touch during the trial (eg, being present at clinical meetings, functioning as helpdesk, sending individual emails to participating clinicians as reminders of specific actions). Another important limitation is the large numbers of dropouts before the follow-up measurement, even though patients were offered a small gift for returning their completed questionnaire.

Our study also has strengths. Most importantly, it affirms previous findings that many people with a severe mental illness can work with electronic decision aids, either with or without assistance, at the clinic or at home. Furthermore, our study provides insight in variation among the population concerning interest in and use of electronic decision aids. Our results suggest that part of the population is not able or does not feel the need to work with these decision aids. Based on our results, the ratio of users versus nonusers could be 50-50. Another strength is that we collected detailed information about allocation and reception of the intervention with varying illness durations, and we included a process evaluation that allowed us to perform a critical analysis on the trial results.

Conclusion

The development of electronic decision aids to facilitate shared medical decision making is encouraged and many people with a psychotic disorder can work with them. This holds for both first-episode patients and long-term care patients, although the latter group might need more assistance. However, effects of decision aids on patient participation in medical decision making have not been consistently demonstrated.

Acknowledgments

This study was supported by the Netherlands Organisation for Health Research and Development (ZonMw); Fonds Psychische Gezondheid; ICT regie; and the Dutch Ministry of Health, Welfare and Sport under the name WEGWEIS (grant number 300020011).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video of the Web-based decision aid.

[AVI File, 125MB - jmir_v15i10e216_app1.avi]

Multimedia Appendix 2

CONSORT-EHEALTH Checklist V1.6.2 [32].

[PDF File (Adobe PDF File), 993KB - jmir_v15i10e216_app2.pdf]

References

1. Drake RE, Deegan PE. Shared decision making is an ethical imperative. Psychiatr Serv 2009 Aug;60(8):1007. [doi: 10.1176/appi.ps.60.8.1007] [Medline: 19648184]
2. Hamann J, Cohen R, Leucht S, Busch R, Kissling W. Do patients with schizophrenia wish to be involved in decisions about their medical treatment? Am J Psychiatry 2005 Dec;162(12):2382-2384. [doi: 10.1176/appi.ajp.162.12.2382] [Medline: 16330606]
3. Hamann J, Langer B, Winkler V, Busch R, Cohen R, Leucht S, et al. Shared decision making for in-patients with schizophrenia. Acta Psychiatr Scand 2006 Oct;114(4):265-273. [doi: 10.1111/j.1600-0447.2006.00798.x] [Medline: 16968364]
4. Adams JR, Drake RE, Wolford GL. Shared decision-making preferences of people with severe mental illness. Psychiatr Serv 2007 Sep;58(9):1219-1221. [doi: 10.1176/appi.ps.58.9.1219] [Medline: 17766569]
5. Hamann J, Mischo C, Langer B, Leucht S, Kissling W. Physicians' and patients' involvement in relapse prevention with antipsychotics in schizophrenia. Psychiatr Serv 2005 Nov;56(11):1448-1450. [doi: 10.1176/appi.ps.56.11.1448] [Medline: 16282267]
6. Kal EF. Paternalism or lack of time? Psychiatr Serv 2009 Oct;60(10):1403. [doi: 10.1176/appi.ps.60.10.1403] [Medline: 19797392]
7. Torrey WC, Drake RE. Practicing shared decision making in the outpatient psychiatric care of adults with severe mental illnesses: redesigning care for the future. Community Ment Health J 2010 Oct;46(5):433-440. [doi: 10.1007/s10597-009-9265-9] [Medline: 19898985]
8. Kalk NJ, Pothier DD. Patient information on schizophrenia on the internet. Psychiatric Bulletin 2008 Nov 01;32(11):409-411. [doi: 10.1192/pb.bp.107.017004]
9. Corrigan PW, Larson JE. Shared decision making: whose decision? Psychiatr Serv 2009 Nov;60(11):1555; author reply 1555-1555; author reply 1556. [doi: 10.1176/appi.ps.60.11.1555] [Medline: 19880481]
10. Drake RE, Deegan PE, Woltmann E, Haslett W, Drake T, Rapp CA. Comprehensive electronic decision support systems. Psychiatr Serv 2010 Jul;61(7):714-717. [doi: 10.1176/appi.ps.61.7.714] [Medline: 20592007]
11. Deegan PE, Rapp C, Holter M, Riefer M. Best practices: a program to support shared decision making in an outpatient psychiatric medication clinic. Psychiatr Serv 2008 Jun;59(6):603-605. [doi: 10.1176/appi.ps.59.6.603] [Medline: 18511580]
12. Woltmann EM, Wilkniss SM, Teachout A, McHugo GJ, Drake RE. Trial of an electronic decision support system to facilitate shared decision making in community mental health. Psychiatr Serv 2011 Jan;62(1):54-60. [doi: 10.1176/appi.ps.62.1.54] [Medline: 21209300]
13. Steinwachs DM, Roter DL, Skinner EA, Lehman AF, Fahey M, Cullen B, et al. A web-based program to empower patients who have schizophrenia to discuss quality of care with mental health providers. Psychiatr Serv 2011 Nov;62(11):1296-1302 [FREE Full text] [doi: 10.1176/appi.ps.62.11.1296] [Medline: 22211208]
14. Research Randomizer. URL: http://www.randomizer.org/
15. van der Krieke L, Emerencia AC, Aiello M, Sytema S. Usability evaluation of a web-based support system for people with a schizophrenia diagnosis. J Med Internet Res 2012;14(1):e24 [FREE Full text] [doi: 10.2196/jmir.1921] [Medline: 22311883]
16. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short Assessment of Quality of Life (MANSA). Int J Soc Psychiatry 1999;45(1):7-12. [Medline: 10443245]
17. Wing JK, Beevor AS, Curtis RH, Park SB, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS). Research and development. Br J Psychiatry 1998 Jan;172:11-18. [Medline: 9534825]
18. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276 [FREE Full text] [Medline: 3616518]
19. Ende J, Kazis L, Ash A, Moskowitz MA. Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. J Gen Intern Med 1989;4(1):23-30. [Medline: 2644407]
20. Edwards A, Elwyn G, Hood K, Robling M, Atwell C, Holmes-Rovner M, et al. The development of COMRADE--a patient-based outcome measure to evaluate the effectiveness of risk communication and treatment decision making in consultations. Patient Educ Couns 2003 Jul;50(3):311-322. [Medline: 12900105]
21. 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]
22. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. BMJ 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]

23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
24. Hennink M, Hutter I, Bailey A. In-depth interviews. In: Hennink M, Hutter I, Bailey A, editors. Qualitative Research Methods. London: Sage; 2011:108-134.
25. Punter A. The physiatrist’s attitude towards shared decision-making: A further analysis among physiatrists on the role of shared decision-making in rehabilitation healthcare [bachelor's thesis]. Enschede: University of Twente; 2008. URL: http://essay.utwente.nl/58854/1/scriptie_A_Punter.pdf [accessed 2013-09-23] [WebCite Cache ID 6JqfNF4wB]
26. Holmes-Rovner M, Valade D, Orlowski C, Draus C, Nabozny-Valerio B, Keiser S. Implementing shared decision-making in routine practice: barriers and opportunities. Health Expect 2000 Sep;3(3):182-191. [Medline: 11281928]
27. Charles C, Gafni A, Whelan T. Self-reported use of shared decision-making among breast cancer specialists and perceived barriers and facilitators to implementing this approach. Health Expect 2004 Dec;7(4):338-348. [doi: 10.1111/j.1369-7625.2004.00299.x] [Medline: 15544686]
28. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med 1998 Nov;47(9):1351-1359. [Medline: 9783878]
29. Collins K, O'Cathain A. The continuum of patient satisfaction--from satisfied to very satisfied. Soc Sci Med 2003 Dec;57(12):2465-2470. [Medline: 14572851]
30. Hamann J, Mendel R, Meier A, Asani F, Pausch E, Leucht S, et al. "How to speak to your psychiatrist": shared decision-making training for inpatients with schizophrenia. Psychiatr Serv 2011 Oct;62(10):1218-1221. [doi: 10.1176/appi.ps.62.10.1218] [Medline: 21969650]
31. Kasper J, Hoffmann F, Heesen C, Köpke S, Geiger F. MAPPIN'SDM--the multifocal approach to sharing in shared decision making. PLoS One 2012;7(4):e34849 [FREE Full text] [doi: 10.1371/journal.pone.0034849] [Medline: 22514677]
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

API: Autonomy Preference Index
CANSAS-P: Camberwell Assessment of Need Short Appraisal Schedule–Self-Report Version
COMRADE: Combined Outcome Measure for Risk Communication and Treatment Decision-making Effectiveness
CSQ: Client Satisfaction Questionnaire
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision)
GAF: Global Assessment of Functioning
HoNOS: Health of the Nation Outcome Scales
MANSA: Manchester Short Assessment of Quality of Life
METiGG: Medisch-ethische Toetsingscommissie instellingen Geestelijke Gezondheidszorg (Dutch medical ethical committee for mental health care)
NTR: NederlandsTrial Register (Dutch trial register)
PANSS: Positive and Negative Syndrome Scale
ROM: Routine Outcome Monitoring

Edited by G Eysenbach; submitted 26.07.13; peer-reviewed by V Shaffer, T Thyvalikakath; comments to author 16.08.13; revised version received 29.08.13; accepted 15.09.13; published 07.10.13.

Please cite as:
van der Krieke L, Emerencia AC, Boonstra N, Wunderink L, de Jonge P, Sytema S
A Web-Based Tool to Support Shared Decision Making for People With a Psychotic Disorder: Randomized Controlled Trial and Process Evaluation
J Med Internet Res 2013;15(10):e216
URL: http://www.jmir.org/2013/10/e216/
doi:10.2196/jmir.2851
PMID:24100091

Original Paper

Internet Addiction Test (IAT): Which is the Best Factorial Solution?

Palmira Faraci[1], PhD; Giuseppe Craparo[1], PhD; Roberta Messina[1]; Sergio Severino[1], PhD

University of Enna "Kore", Faculty of Human and Social Sciences, Enna, Italy

Corresponding Author:
Giuseppe Craparo, PhD
University of Enna "Kore"
Faculty of Human and Social Sciences
Cittadella Universitaria
Enna, 94100
Italy
Phone: 39 0935536536
Fax: 39 0935536943
Email: giuseppe.craparo@unikore.it

Abstract

Background: The Internet Addiction Test (IAT) by Kimberly Young is one of the most utilized diagnostic instruments for Internet addiction. Although many studies have documented psychometric properties of the IAT, consensus on the optimal overall structure of the instrument has yet to emerge since previous analyses yielded markedly different factor analytic results.

Objective: The objective of this study was to evaluate the psychometric properties of the Italian version of the IAT, specifically testing the factor structure stability across cultures.

Methods: In order to determine the dimensional structure underlying the questionnaire, both exploratory and confirmatory factor analyses were performed. The reliability of the questionnaire was computed by the Cronbach alpha coefficient.

Results: Data analyses were conducted on a sample of 485 college students (32.3%, 157/485 males and 67.7%, 328/485 females) with a mean age of 24.05 years (SD 7.3, range 17-47). Results showed 176/485 (36.3%) participants with IAT score from 40 to 69, revealing excessive Internet use, and 11/485 (1.9%) participants with IAT score from 70 to 100, suggesting significant problems because of Internet use. The IAT Italian version showed good psychometric properties, in terms of internal consistency and factorial validity. Alpha values were satisfactory for both the one-factor solution (Cronbach alpha=.91), and the two-factor solution (Cronbach alpha=.88 and Cronbach alpha=.79). The one-factor solution comprised 20 items, explaining 36.18% of the variance. The two-factor solution, accounting for 42.15% of the variance, showed 11 items loading on Factor 1 (Emotional and Cognitive Preoccupation with the Internet) and 7 items on Factor 2 (Loss of Control and Interference with Daily Life). Goodness-of-fit indexes (NNFI: Non-Normed Fit Index; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual) from confirmatory factor analyses conducted on a random half subsample of participants (n=243) were satisfactory in both factorial solutions: two-factor model (χ^2_{132}= 354.17, P<.001, χ^2/df=2.68, NNFI=.99, CFI=.99, RMSEA=.02 [90% CI 0.000-0.038], and SRMR=.07), and one-factor model (χ^2_{169}=483.79, P<.001, χ^2/df=2.86, NNFI=.98, CFI=.99, RMSEA=.02 [90% CI 0.000-0.039], and SRMR=.07).

Conclusions: Our study was aimed at determining the most parsimonious and veridical representation of the structure of Internet addiction as measured by the IAT. Based on our findings, support was provided for both single and two-factor models, with slightly strong support for the bidimensionality of the instrument. Given the inconsistency of the factor analytic literature of the IAT, researchers should exercise caution when using the instrument, dividing the scale into factors or subscales. Additional research examining the cross-cultural stability of factor solutions is still needed.

(J Med Internet Res 2013;15(10):e225) doi:10.2196/jmir.2935

KEYWORDS

IAT; Internet; addiction; factorial structure; psychometric properties; structural validity

Introduction

Overview

The current overview of global Internet usage provides a striking picture of the extent of the phenomenon. Because of a steady strengthening between computer technology and traditional communication processes [1-3], Internet users' growth reached 566.4% from 2000 to 2012. Because the majority of online users have become members of chats, forums, and social networks, the rise and popularity of the Internet is strongly linked to its use in communication and socialization processes. For this, the medium has become an ever-increasing part of many people's day-to-day lives [4], changing the way to communicate. According to several researchers, aseptic characteristics of Computer Mediated Communication (CMC) made virtual relationships "shallow and impersonal" [3] and online anonymity produces a psychological state characterized by the abandonment of social pressures [5].

Internet addiction [4,6-8] is generally categorized under the label of "technological addiction" and is defined by Kandell as a "psychological dependence on the Internet regardless of the type of activity once logged on" [9]. Enough agreement exists on the association between Internet addiction and material and psychological consequences [10], such as the neglect of academic, work, and domestic responsibilities, disruption of relationships, social isolation, and financial problems [11]. Furthermore, literature [12-21] has amply demonstrated that pre-existing familial and social problems, as well as psychological and psychiatric disturbances are more prevalent among dependent Internet users.

Internet Addiction Test (IAT) by Young

One of the most common diagnostic instruments for Internet addiction was proposed by Young in 1996. The author pioneered the study on Internet addiction, developing a structured Internet Addiction Test (IAT) on the basis of the DSM-IV criteria (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition) for pathological gambling [22]. In its first version, IAT comprised eight questions and was administered to a group of subjects recruited through various announcements in newspapers, online forums, and websites. If respondents answered "yes" to five or more of the criteria, they were classified as Dependents. A total of 396 subjects fell into the Internet-Addicted user category, while 100 respondents were labelled as Non-Addicted. Most relevant results revealed that Internet-Addicted users spent approximately eight times the number of hours per week as that of Non-Dependents in using the Internet. Moreover, different from Non-Dependents (who used the medium essentially to manage email, look for information, or download software), Dependent users spent most of the time in synchronous communication environments, chat rooms, and MUDs (multi-user domains). This caused severe impairment in academic, relationship, financial, and occupational life areas.

Later, Young extended the previous version of IAT [12]. The new scale exhibits the following characteristics:

- It comprises 20 items rated in a five-point Likert scale (from 1 - not at all, to 5 - always).
- As with the first diagnostic questionnaire, this measurement is derived from the DSM–IV criteria for pathological gambling and alcoholism and it measures the extent of individual's problems due to the Internet use in daily routine, social life, productivity, sleeping patterns, and feelings.
- On the basis of the total score obtained on the test, the individual is placed into one of three categories: average online user (from 20 to 39) who has a full control of his or her usage; experiences frequent problems because of excessive Internet use (from 40 to 69); or has significant problems because of Internet use (from 70 to 100).

Though the IAT is one of the most common instruments to assess Internet addiction, its use remains problematic. Indeed, empirical researches on Internet addiction provided conflicting results on its psychometric properties; moreover, the instrument has not been subjected to rigorous and systematic psychometric investigations [23].

Widyanto and McMurran administered the IAT on 86 subjects recruited online. The factor analysis of the IAT items revealed six factors (salience, excessive use, neglect work, anticipation, lack of control, neglect social life), with good internal consistency and concurrent validity [11]. In a more recent study, conducted on 236 Internet chatters, Ferraro, Caci, D'Amico, and Di Blasi found a six-factor solution, with an explained variance of 55.6%. The six factors were named as follows: compromised social quality of life, compromised individual quality of life, compensatory usage of the Internet, compromised academic/working careers, compromised time control, and excitatory usage of the Internet [24]. Although both surveys converge toward a six-factor solution, these factors did not correspond to the same items in the two studies [25]. Furthermore, Barke, Nyenhuis, and Kröner-Herwig administered the German version of the IAT in a large sample of students [26]. Factor analysis revealed a stable two-factor solution: Factor 1, "Emotional and Cognitive Preoccupation with the Internet", which explained 21.03% of the variance for the offline sample and 26.73% for the online sample, and Factor 2, "Loss of Control and Interference with Daily Life", which explained 20.97% of the variance for the offline sample and 19.99% of the variance for the online sample. The first factor encompasses items on the emotional and cognitive elements related to use of Internet. The second factor is composed of items on "(unsuccessful) attempts at curbing online time and detrimental consequences of the Internet use for daily functioning" (p. 541 [26]). The two-factor solution fit well with data also in a study conducted by Watters, Keefer, Kloosterman, Summerfeldt, and Parker in a large sample of Canadian high-school students [27]. Finally, exploratory and confirmatory factor analysis applied on the Arabic [28] and French [25] versions of the IAT revealed that a one-factor model fits the data very well.

The heterogeneity of these results could be attributed to several causes, such as the fact that many studies have used this scale in various settings [29], focusing on subjects of different ages and nationalities.

The aim of the present study is to provide a contribution in assessing the psychometric properties of the IAT in a sample of Italian college students, specifically testing its factor structure stability across cultures.

Methods

Participants and Procedure

Of the 521 Italian adults screened, 36 had one or more items with missing values and were not included in data analyses. Thus, participants totalled 485 (32.3%, 157/485 males and 67.7%, 328/485 females) with a mean age of 24.05 years (SD 7.3, range 17-47). The group of participants were recruited on a voluntary basis.

Confirmatory factor analyses were performed on a random subsample (sample 2) of 243 participants (35.8%, 87/243 male and 64.2%, 156/243 female), ranging in age from 18 to 50 years (mean 22.12, SD 5.9).

Data Analyses

In order to determine the dimensional structure underlying the questionnaire, data from the 485 participants were subjected to exploratory factor analysis. With the 20-item questionnaire, we were able to satisfy the minimum 10 participants-per-item ratio that is usually recommended; a number of 24.25 subjects per item largely ensured that reliable factors would emerge.

Prior to exploratory factor analysis, data were inspected to ensure items were significantly correlated, using Bartlett's Test of Sphericity. Also, in order to evaluate whether items share sufficient variance to justify factor extraction, KMO's Test of Sampling Adequacy was used. Sampling adequacy values greater than .80 and .90 are considered excellent, values between .50 and .60 marginally acceptable, and values less than .50 unacceptable [30].

Principal axis factoring with oblique rotation (promax criterion) was selected as the method of factor extraction. To determine the number of factors, both Kaiser's [31] criterion (items with eigenvalues greater than 1) and the Scree test [32] were used. Random data parallel analysis [33] was also performed. The eigenvalues derived from the actual data were compared to the eigenvalues derived from the random data. Factors were retained as long as the *i*th eigenvalue from the actual data was greater than the *i*th eigenvalue from the random data [34].

The reliability of the questionnaire, in terms of internal consistency, was computed by the Cronbach alpha coefficient. Corrected item-scale correlations were examined assuring they exceeded .30, recommended as the standard for supporting internal consistency [35].

The IAT factor structure that emerged from exploratory factor analysis was verified using the structural equation modelling technique. In particular, a confirmatory factor analysis was conducted on the data from the random subsample of participants (sample 2). Least Square, which is applicable when data do not meet the assumption of multivariate normality, was selected as the procedure for estimation.

The closeness of the hypothetical model to the empirical data was statistically evaluated through multiple goodness-of-fit indexes. Chi-square is sensitive to sample size and may be significant when the actual differences between the observed and implied model covariances are slight [36]. Therefore, we did not use this statistic as an evaluation of absolute fit, but referred to the ratio of chi-square to degrees of freedom (χ^2/df [37]), the Non-Normed Fit Index (NNFI [38]), the Comparative Fit Index (CFI [39]), and the Standardized Root Mean Square Residual (SRMR [39]) to evaluate adequacy of fit of each model. We also reported the Root Mean Square Error of Approximation (RMSEA [40]) to provide an indication of the global fit of the model. Model testing was accomplished using the EQS (version 6.1) structural equations modeling software package [41]. Higher values for the CFI and NNFI are considered good (>.90, acceptable and >.95, desirable [42]). The RMSEA is an index of misfit per degree of freedom; lower values are preferred (<.08, acceptable, <.05, desirable [42]). The SRMR is the average standardized deviation in the model-based reproduced covariances in contrast to those observed in the data; lower values are optimal (<.10, acceptable, <.05, desirable [42]).

Results

Participants

A series of analyses was conducted to examine the psychometric properties of the questionnaire, including reliability and both exploratory and confirmatory factor analyses. Results showed 176/485 (36.3%) participants with IAT score from 40 to 69, revealing excessive Internet use, and 11/485 (1.9%) participants with IAT score from 70 to 100, suggesting significant problems because of Internet use.

Exploratory Factor Analysis

The KMO's Test of Sampling Adequacy was .94 and Bartlett's Test of Sphericity (χ^2_{190}=4014.0) was significant (P<.001), indicating that the IAT items were appropriate for a factor analysis.

We employed Horn's [33] parallel analysis (PA) for determining the number of factors to retain because it has been shown empirically to give accurate results [43]. This criterion involves comparison of eigenvalues for data under study with those extracted from and averaged over a large number of random data sets (we used 1000) based on the same number of variables and subjects. If eigenvalue I for data under study exceeds the average over a large number of random data-based eigenvalues I, that factor is retained. One then proceeds to factor II and so on, retaining only the number of factors for which real data-based eigenvalues exceed averages derived from random data. Parallel analysis determined five factors to be extracted. The resulting number of factors is evidently over-defined, with two factors comprised by only two indicators, one item failed to load .30 or greater in any factor, and 11 items loaded simultaneously on two factors without a difference of at least .30 between loading on the primary factor and loading on other factors.

As a consequence of these poor findings, we followed the eingenvalues-greater-than-one criterion, extracting three factors but rotation (both orthogonal and oblique) failed to converge. Examination of the scree plot suggested two factors to be extracted. Inspection of factor loadings revealed 18 items to have been appropriate, having pattern coefficients of .35 or greater, which is generally regarded as the standard for pattern coefficient cutoff criteria [44]. Item 17 (“Do you try to cut down the amount of time you spend online and fail?”) and item 8 (“Does your job performance or productivity suffer because of the Internet?”) presented double loadings and were eliminated. The two-factor solution, accounting for 42.15% of the variance, showed 11 items loading on Factor 1 (Emotional and Cognitive Preoccupation with the Internet), and 7 items on Factor 2 (Loss of Control and Interference with Daily Life); we utilized the same wording proposed by Barke, Nyenhuis, and Kröner-Herwig [26]. Table 1 depicts the pattern coefficients for the two-factor solution. Factors intercorrelation according to the results of exploratory factor analysis was .65. Correlations between the two-factor mean scores (ie, sum of the items/number of items) was .64 (*P*<.01). To be thorough, an exploratory factor analysis requesting one factor was also performed. The eigenvalue and variance accounted for the factor were 7.24 and 36.18%, respectively. Table 2 reports factor loadings of the IAT item for the one-factor solution.

Reliability

The reliability of the IAT was assessed for both one- and two-factor structure models. Internal consistency was assessed with coefficient alpha for the entire sample of 485 participants. Satisfactory results were evident for both one-factor solution (Cronbach alpha=.91, see Table 2) and two-factor solution (Factor 1 Cronbach alpha=.88 and Factor 2 Cronbach alpha=.79; see Table 3).

Table 1. Factor loadings of the IAT items for the two-factor solution.

Items[a]	Factor 1[b]	Factor 2[c]
20. Do you feel depressed, moody, or nervous when you are offline, which goes away once you are back online?	.940	
15. Do you feel preoccupied with the Internet when offline or fantasize about being online?	.694	
3. Do you prefer the excitement of the Internet to intimacy with your partner?	.678	
19. Do you choose to spend more time online over going out with others?	.649	
18. Do you try to hide how long you’ve been online?	.628	
11. Do you find yourself anticipating when you go online again?	.623	
12. Do you feel that life without the Internet would be boring, empty, and joyless?	.622	
13. Do you snap, yell, or act annoyed if someone bothers you while you are online?	.518	
10. Do you block disturbing thoughts about your life with soothing thoughts of the Internet?	.473	
4. Do you form new relationships with fellow online users?	.443	
14. Do you lose sleep due to late night log-ins?	.414	
2. Do you neglect household chores to spend more time online?		.803
1. Do you feel that you stay online longer than you intend?		.761
16. Do you find yourself saying “just a few more minutes” when online?		.595
6. Does your work suffer because of the amount of time you spend online?		.549
5. Do others in your life complain to you about the amount of time you spend online?		.542
9. Do you become defensive or secretive when someone asks what you do online?		.403
7. Do you check your email before something else that you need to do?		.372
% explained variance	36.08	6.07

[a] Items are ordered by factor loading rather than item number.

[b] Factor 1: Emotional and Cognitive Preoccupation with the Internet

[c] Factor 2: Loss of Control and Interference with Daily Life

Table 2. Factor loadings of the IAT items and corrected item-total correlations for the one-factor solution.

Items[a]	Loadings	Item-total correlations
11. Do you find yourself anticipating when you go online again?	.705	.670
15. Do you feel preoccupied with the Internet when offline or fantasize about being online?	.699	.647
5. Do others in your life complain to you about the amount of time you spend online?	.687	.666
6. Does your work suffer because of the amount of time you spend online?	.680	.656
13. Do you snap, yell, or act annoyed if someone bothers you while you are online?	.674	.640
18. Do you try to hide how long you’ve been online?	.664	.621
20. Do you feel depressed, moody, or nervous when you are offline, which goes away once you are back online?	.662	.606
8. Does your job performance or productivity suffer because of the Internet?	.656	.622
19. Do you choose to spend more time online over going out with others?	.646	.603
10. Do you block disturbing thoughts about your life with soothing thoughts of the Internet?	.636	.606
14. Do you lose sleep due to late night log-ins?	.611	.573
17. Do you try to cut down the amount of time you spend online and fail?	.610	.581
12. Do you feel that life without the Internet would be boring, empty, and joyless?	.597	.558
16. Do you find yourself saying “just a few more minutes” when online?	.589	.577
2. Do you neglect household chores to spend more time online?	.550	.548
9. Do you become defensive or secretive when someone asks what you do online?	.529	.517
4. Do you form new relationships with fellow online users?	.486	.461
3. Do you prefer the excitement of the Internet to intimacy with your partner?	.450	.401
1. Do you feel that you stay online longer than you intend?	.417	.424
7. Do you check your email before something else that you need to do?	.300	.295
% explained variance	36.18	
Cronbach alpha		.91

[a] Items are ordered by factor loading rather than item number.

Table 3. Corrected item-total correlations.

Item[a]	Factor 1[b]	Factor 2[c]
Item 20	.708	
Item 15	.668	
Item 3	.491	
Item 19	.631	
Item 18	.616	
Item 11	.692	
Item 12	.595	
Item 13	.627	
Item 10	.588	
Item 4	.467	
Item 14	.535	
Item 2		.603
Item 1		.520
Item 16		.550
Item 6		.603
Item 5		.619
Item 9		.472
Item 7		.325
Cronbach alpha	.88	.79

[a] Items are ordered by factor rather than item number.

[b] Factor 1: Emotional and Cognitive Preoccupation with the Internet

[c] Factor 2: Loss of Control and Interference with Daily Life

Confirmatory Factor Analysis

The confirmatory factor analyses (CFA) conducted on sample 2 (n=243) showed the acceptable goodness-of-fit indexes for the two-factor model (χ^2_{132}=354.17; $P<.001$, χ^2/df=2.68, NNFI=.99, CFI=.99, RMSEA=.02 [90% CI 0.000-0.038], and SRMR=.07). All manifest variables loaded significantly ($P<.05$) on their hypothesized latent factors. Figure 1 shows the standardized parameter estimates.

According to the results of the CFA, the latent factors are highly correlated to each other. Specifically, they share 70.22% of common variance indicating poor discriminant validity between extracted factors and maybe a more parsimonious solution could be obtained.

Consequently, confirmatory analysis was performed on all IAT items to test for unidimensionality. The completely standardized factor loadings are reported in Table 4. Table 5 contains results for both two-factor and one-factor models specified and tested.

The comparative fit of the models was assessed with the Akaike Information Criterion (AIC [45,46]), which is used for model comparison, with the smallest value being indicative of the best fitting model. AIC for the one-factor model was 145.79, AIC for the two-factor model was 90.17, providing greater support for the bidimensionality of the instrument.

Table 4. Standardized factor loadings of the IAT items for the one-factor solution.

Items	Loadings	Residuals
1. Do you feel that you stay online longer than you intend?	.406	.914
2. Do you neglect household chores to spend more time online?	.484	.875
3. Do you prefer the excitement of the Internet to intimacy with your partner?	.475	.880
4. Do you form new relationships with fellow online users?	.377	.926
5. Do others in your life complain to you about the amount of time you spend online?	.675	.738
6. Does your work suffer because of the amount of time you spend online?	.668	.745
7. Do you check your email before something else that you need to do?	.347	.938
8. Does your job performance or productivity suffer because of the Internet?	.670	.742
9. Do you become defensive or secretive when someone asks what you do online?	.507	.862
10. Do you block disturbing thoughts about your life with soothing thoughts of the Internet?	.618	.786
11. Do you find yourself anticipating when you go online again?	.610	.793
12. Do you feel that life without the Internet would be boring, empty, and joyless?	.546	.838
13. Do you snap, yell, or act annoyed if someone bothers you while you are online?	.633	.774
14. Do you lose sleep due to late night log-ins?	.584	.812
15. Do you feel preoccupied with the Internet when offline or fantasize about being online?	.650	.760
16. Do you find yourself saying “just a few more minutes” when online?	.563	.827
17. Do you try to cut down the amount of time you spend online and fail?	.582	.813
18. Do you try to hide how long you’ve been online?	.586	.810
19. Do you choose to spend more time online over going out with others?	.586	.810
20. Do you feel depressed, moody, or nervous when you are offline, which goes away once you are back online?	.594	.804

Table 5. Fit indices for the one-factor and two-factor models.

Model	χ^2	df	*P* value	NFI[a]	NNFI[b]	CFI[c]	SRMR[d]	RMSEA[e]	90% CI
One-factor model	483.79	169	<.001	.895	.984	.986	.070	.024	0.000-0.039
Two-factor model	354.17	132	<.001	.906	.989	.991	.067	.020	0.000-0.038

[a] NFI: Normed Fit Index

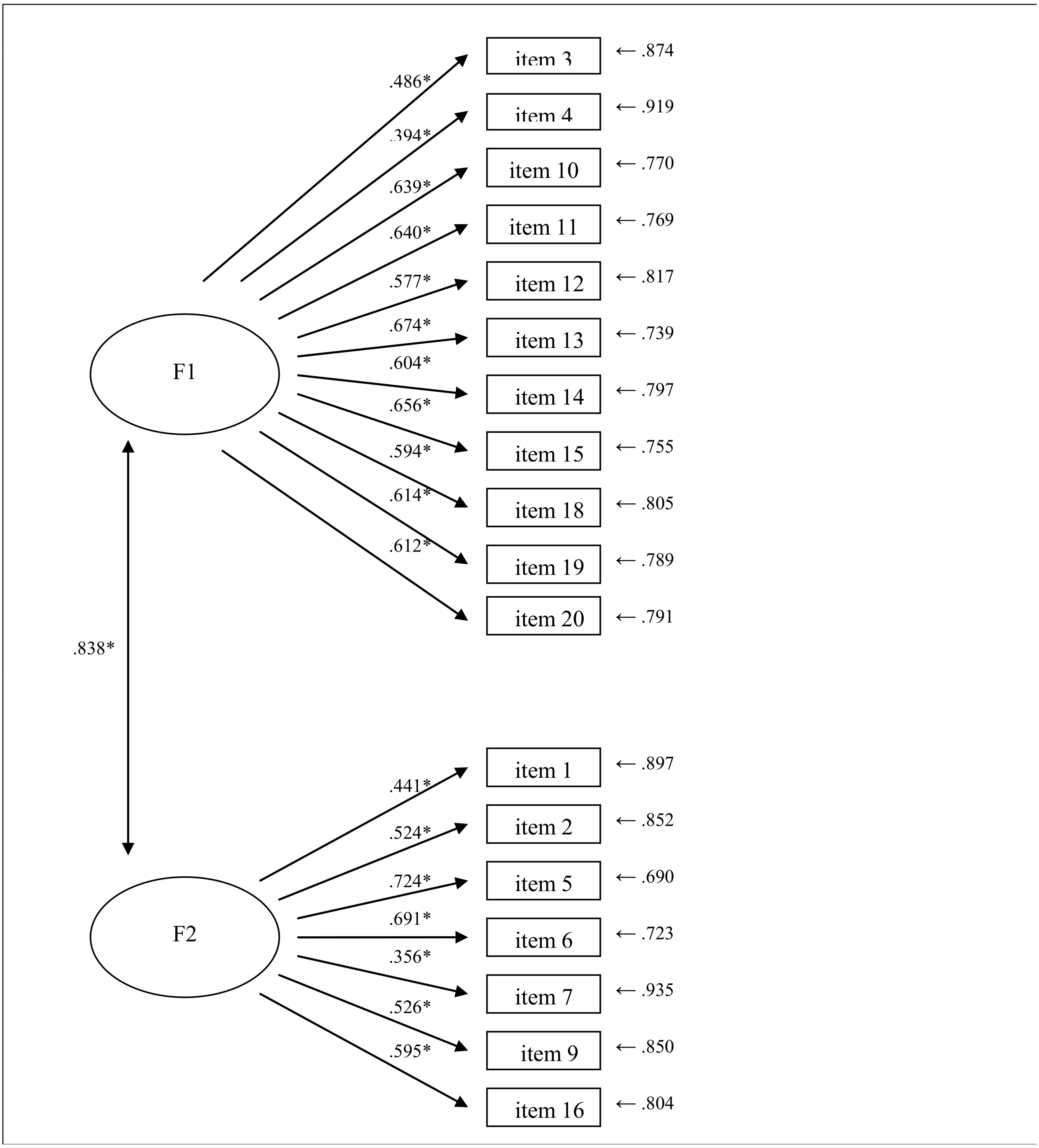
[b] NNFI: Non-Normed Fit Index

[c] CFI: Comparative Fit Index

[d] SRMR: Standardized Root Mean Square Residual

[e] RMSEA: Root Mean Square Error of Approximation

Figure 1. IAT empirical model (standardized solution). Note: F1 = Emotional and Cognitive Preoccupation with the Internet; F2 = Loss of Control and Interference with Daily Life. * P<.05.



Discussion

Principal Findings

The present study examined the model of Internet addiction as assessed by a widely used self-report measure, the IAT. In line with many previous studies suggesting the need to test the factor structure stability across cultures and samples of commonly used instruments in several fields of psychological research [47-52], we sought to document the factor structure of the scale, with the final aim to enhance our understanding of the Internet addiction construct.

Knowledge of the structure of the IAT and its consistency over cultures and languages can serve a number of useful purposes: advance theory regarding the place of the disorder within the nosology of psychiatric conditions, hence contributing to the development of accurate and valid assessment tools.

Extant research on the factor structure of IAT has done much to highlight key issues in the dimensionality of the construct, yet several concerns warrant further empirical attention. Indeed, although it remains one of the most broadly used measures of Internet addiction worldwide, its factor structure remains questionable. Thus, factor analytic research on the IAT is important for the psychometric evaluation of the instrument and for clarifying the nature of the Internet addiction construct itself.

Many studies have documented psychometric properties of the IAT, with markedly different factor analytic results. Consensus on the optimal overall structure has yet to emerge since previous analyses have found between one- and six-factor solutions for the IAT.

Our study was aimed at determining the most parsimonious and veridical representation of the structure of Internet addiction as measured by the IAT. Based on our findings, support was provided for both single- and two-factor models (Factor 1: Emotional and Cognitive Preoccupation with the Internet; Factor 2: Loss of Control and Interference with Daily Life) with slightly strong support for the bidimensionality of the instrument. Nevertheless, the two-factor solution presents some limitations due to the resulting high association between emerged factors. Indeed, different dimensions are generally expected not to be highly correlated, indicating that the subscales measure several aspects of the investigated construct. However, the revealed high associations between factors is understandable because of the unavoidable conceptual connection of the questionnaires' subscales, also found in previous studies [26]. Otherwise, the more parsimonious solution, though usable, would be less effective for a detailed assessment of Internet addiction with consequential loss of salient information.

Limitations

Overall, our findings should be interpreted with some caution because the sample contained only college students. This condition is tempered by the fact that they are an at-risk population in which intense Internet use is common and potentially consequential [9,53]. Clearly, more research needs to be conducted with larger and more varied samples of participants to further establish the structural validity of the instrument.

Conclusions

In summary and in closing, on the basis of the present results combined with inconsistency of the factor analytic literature of the IAT, it seems apparent that researchers should be aware of these psychometric issues and exercise caution when using the IAT, dividing the scale into factors or subscales. Preliminary evidence of scale validity is encouraging; however, additional research examining the cross-cultural stability of factor solutions is still needed.

Conflicts of Interest

None declared.

References

1. Caretti V. In: La Barbera D, editor. Addiction. Milan: Raffaello Cortina; 2010.
2. Severino S. La comunicazione del sapere a distanza. Enna: La Moderna; 2008.
3. Parks MR, Floyd K. Making friends in cyberspace. J Commun 1996;46:80-97.
4. Griffiths MD. Internet Addiction - Time to be taken seriously? Addict Res 2000;8(5):413-418.
5. Severino S, Messina R. Group Dynamics in Online and Face-to-face Interactions: an Experimental Study on Learning Methods. Sociology Mind 2011;1(2):65-73.
6. Holden C. 'Behavioral' addictions: do they exist? Science 2001 2;294(5544):980-982. [doi: 10.1126/science.294.5544.980] [Medline: 11691967]
7. Schimmenti A, Guglielmucci F, Barbasio C, Granieri A. Attachment disorganization and dissociation in virtual worlds: a study on problematic internet use among players of online role playing games. Clinical Neuropsychiatry 2012;9(5):195-202.
8. Griffiths MD. Does Internet and Computer "Addiction" Exist? Some Case Study Evidence. CyberPsychology & Behavior 2000 Apr;3(2):211-218. [doi: 10.1089/109493100316067]
9. Kandell JJ. Internet Addiction on Campus: The Vulnerability of College Students. CyberPsychology & Behavior 1998 Jan;1(1):11-17. [doi: 10.1089/cpb.1998.1.11]
10. Griffiths MD. Internet addiction: Does it really exist? In Gackenbach, J. (ed.), Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications. New York: Academic Press; 2007.
11. Widyanto L, McMurran M. The psychometric properties of the internet addiction test. Cyberpsychol Behav 2004;7(4):443-450. [doi: 10.1089/cpb.2004.7.443] [Medline: 15331031]
12. Young KS. Internet Addiction: The Emergence of a New Clinical Disorder. CyberPsychology & Behavior 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
13. Kim K, Ryu E, Chon MY, Yeun EJ, Choi SY, Seo JS, et al. Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: a questionnaire survey. Int J Nurs Stud 2006;43(2):185-192. [doi: 10.1016/j.ijnurstu.2005.02.005] [Medline: 16427966]
14. De Berardis D, D'Albenzio A, Gambi F, Sepede G, Valchera A, Conti CM, et al. Alexithymia and its relationships with dissociative experiences and Internet addiction in a nonclinical sample. Cyberpsychol Behav 2009 Feb;12(1):67-69. [doi: 10.1089/cpb.2008.0108] [Medline: 19132914]
15. Tao R, Huang X, Wang J, Zhang H, Zhang Y, Li M. Proposed diagnostic criteria for internet addiction. Addiction 2010;105(3):556-564. [doi: 10.1111/j.1360-0443.2009.02828.x] [Medline: 20403001]

16. Ko CH, Yen JY, Chen CC, Chen SH, Yen CF. Proposed diagnostic criteria of Internet addiction for adolescents. J Nerv Ment Dis 2005;193(11):728-733. [Medline: 16260926]
17. Xiuqin H, Huimin Z, Mengchen L, Jinan W, Ying Z, Ran T. Mental health, personality, and parental rearing styles of adolescents with Internet addiction disorder. Cyberpsychol Behav 2010;13(4):401-406. [doi: 10.1089/cyber.2009.0222] [Medline: 20712498]
18. Craparo G. Internet addiction, dissociation, and alexithymia. Procedia - Social and Behavioral Sciences 2011 Jan;30:1051-1056. [doi: 10.1016/j.sbspro.2011.10.205]
19. Sanders CE, Field TM, Diego M, Kaplan M. The relationship of Internet use to depression and social isolation among adolescents. Adolescence 2000;35(138):237-242. [Medline: 11019768]
20. Oktan V. The predictive relation between emotion management skills and Internet Addiction. Soc Behav Personal 2011;39(10):1425-1430.
21. Lei L, Wu Y. Adolescents' paternal attachment and Internet use. Cyberpsychol Behav 2007;10(5):633-639. [doi: 10.1089/cpb.2007.9976] [Medline: 17927530]
22. Young K. Addictive use of the Internet: A case that breaks the stereotype. Psychol Rep 1996;79:899-902. [Medline: 8969098]
23. Demetrovics Z, Szeredi B, Rózsa S. The three-factor model of Internet addiction: the development of the Problematic Internet Use Questionnaire. Behav Res Methods 2008 May;40(2):563-574. [Medline: 18522068]
24. Ferraro G, Caci B, D'Amico A, Di Blasi M. Internet addiction disorder: an Italian study. Cyberpsychol Behav 2007;10(2):170-175. [doi: 10.1089/cpb.2006.9972] [Medline: 17474832]
25. Khazaal Y, Billieux J, Thorens G, Khan R, Louati Y, Scarlatti E, et al. French validation of the internet addiction test. Cyberpsychol Behav 2008;11(6):703-706. [doi: 10.1089/cpb.2007.0249] [Medline: 18954279]
26. Barke A, Nyenhuis N, Kröner-Herwig B. The German version of the internet addiction test: a validation study. Cyberpsychol Behav Soc Netw 2012;15(10):534-542. [doi: 10.1089/cyber.2011.0616] [Medline: 23002984]
27. Watters CA, Keefer CV, Kloosterman PH, Summerfeldt LJ, Parker JDA. Examining the structure of the Internet Addiction Test in adolescents: A bifactor approach. Computers in Human Behavior 2013 Nov;29(6):2294-2302. [doi: 10.1016/j.chb.2013.05.020]
28. Hawi NS. Arabic validation of the Internet addiction test. Cyberpsychol Behav Soc Netw 2013;16(3):200-204. [doi: 10.1089/cyber.2012.0426] [Medline: 23441685]
29. Frangos CC, Frangos CC, Sotiropoulos I. A Meta-analysis of the Reliability of Young's Internet Addiction Test. In: WCE 2012. 2012 Presented at: World Congress on Engineering; July 4 - 6, 2012; London, UK.
30. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis (4th ed.). New Jersey: Prentice-Hall; 1995.
31. Kaiser HF. A note on Guttmans's Lower Bound for the number of common factors. Br J Math Stat 1961;14(1):1.
32. Cattell RB. The scree test for the number of factors. Multivariate Behavioral Research 1966 Apr;1(2):245-276. [doi: 10.1207/s15327906mbr0102_10]
33. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30:179-185. [Medline: 14306381]
34. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. Behav Res Methods Instrum Comput 2000 Aug;32(3):396-402. [Medline: 11029811]
35. De Vellis R. Scale development: Theory and applications (2nd ed.). Thousand Oaks, CA: Sage Publications; 2003.
36. Kline RB. Principles and practice of structure equation modelling. New York: Guilford Press; 1998.
37. Jöreskog KG, Sörbom D. Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software; 1993.
38. Bentler PM, Bonett. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88(3):588-606. [doi: 10.1037//0033-2909.88.3.588]
39. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-246. [Medline: 2320703]
40. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. Multivariate Behavioral Research 1990 Apr;25(2):173-180. [doi: 10.1207/s15327906mbr2502_4]
41. Bentler PM. EQS 6 structural equations program manual. Encino: Multivariate Software; 2006.
42. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
43. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. Psychol Bull 1986;99:432-442.
44. Comrey AL, Lee HB. A first course in factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
45. Akaike H. A new look at the statistical model identification. IEEE Trans. Automat. Contr 1974 Dec;19(6):716-723. [doi: 10.1109/TAC.1974.1100705]
46. Akaike H. Factor analysis and AIC. Psychometrika 1987 Sep;52(3):317-332. [doi: 10.1007/BF02294359]
47. Craparo G, Faraci P, Fasciano S, Carrubba S, Gori A. A factor analytic study of the Boredom Proneness Scale (BPS). Clinical Neuropsychiatry 2013;10(3-4):164-170.
48. Craparo G, Faraci P, Rotondo G, Gori A. The Impact of Event Scale -Revised: Psychometric properties of the Italian version in a sample of flood victims. Neuropsychiatr Dis Treat 2013;9:1427-1432. [doi: 10.2147/NDT.S51793]

49. Gori A, Giannini M, Socci S, et al. Assessing social anxiety disorder: psychometric properties of the Italian Social Phobia Inventory (I-SPIN). Clinical Neuropsychiatry 2013;10(1):37-42.
50. Manna G, Faraci P, Como MR. Factorial structure and psychometric properties of the Sensation Seeking Scale - Form V (SSS-V) in a sample of Italian adolescents. Europe's Journal of Psychology 2013;9(2):276-288. [doi: 10.5964/ejop.v9i2.500]
51. Triscari MT, Faraci P, D'Angelo V, Urso V. Paura di volare: struttura fattoriale di due questionari. Psicoterapia Cognitiva e Comportamentale 2011;17(2):211-226.
52. Faraci P, Triscari MT, D'Angelo V, Urso V. Fear of flying assessment: A contribution to the Italian validation of two self-report measures. Rev Psych 2011;18(2):91-100.
53. Jelenchick LA, Becker T, Moreno MA. Assessing the psychometric properties of the Internet Addiction Test (IAT) in US college students. Psychiatry Res 2012 Apr 30;196(2-3):296-301 [FREE Full text] [Medline: 22386568]

Abbreviations

AIC: Akaike Information Criterion
CFA: Confirmatory Factor Analysis
CFI: Comparative Fit Index
CMC: Computer Mediated Communication
IAT: Internet Addiction Test
NFI: Normed Fit Index
NNFI: Non-Normed Fit Index
RMSEA: Root Mean Square Error of Approximation
SRMR: Standardized Root Mean Square Residual

Edited by G Eysenbach; submitted 31.08.13; peer-reviewed by F Liga, A Bruno, I Formica; comments to author 19.09.13; accepted 27.09.13; published 09.10.13.

Please cite as:
Faraci P, Craparo G, Messina R, Severino S
Internet Addiction Test (IAT): Which is the Best Factorial Solution?
J Med Internet Res 2013;15(10):e225
URL: http://www.jmir.org/2013/10/e225/
doi:10.2196/jmir.2935
PMID:24184961

Original Paper

For Whom Does It Work? Moderators of Outcome on the Effect of a Transdiagnostic Internet-Based Maintenance Treatment After Inpatient Psychotherapy: Randomized Controlled Trial

David Daniel Ebert[1,2], PhD; Mario Gollwitzer[3], PhD; Heleen Riper[1,4], PhD; Pim Cuijpers[1,4], PhD; Harald Baumeister[5,6], PhD; Matthias Berking[1,2], PhD

[1]Innovation Incubator, Division Health-Training.Online, Leuphana University Lueneburg, Lüneburg, Germany
[2]Division of Clinical Psychology and Psychotherapy, Department of Psychology, Philipps University of Marburg, Marburg, Germany
[3]Division of Methodology and Social Psychology, Department of Psychology, Philipps University of Marburg, Marburg, Germany
[4]Department of Clinical Psychology & EMGO Institute, VU Amsterdam, Amsterdam, Netherlands
[5]Department of Medical Psychology and Medical Sociology, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[6]Department of Rehabilitation Psychology and Psychotherapy, Institute of Psychology, University of Freiburg, Freiburg, Germany

Corresponding Author:
David Daniel Ebert, PhD
Innovation Incubator
Division Health-Training.Online
Leuphana University Lueneburg
Main Building, Ground Floor
Rotenbleicherweg 67
Lüneburg, 21335
Germany
Phone: 49 17616781057
Fax: 49 64212824053
Email: Ebert@leuphana.de

Abstract

Background: Recent studies provide evidence for the effectiveness of Internet-based maintenance treatments for mental disorders. However, it is still unclear which participants might or might not profit from this particular kind of treatment delivery.

Objective: The study aimed to identify moderators of treatment outcome in a transdiagnostic Internet-based maintenance treatment (TIMT) offered to patients after inpatient psychotherapy for mental disorders in routine care.

Methods: Using data from a randomized controlled trial (N=400) designed to test the effectiveness of TIMT, we performed secondary analyses to identify factors moderating the effects of TIMT (intervention) when compared with those of a treatment-as-usual control condition. TIMT involved an online self-management module, asynchronous patient–therapist communication, a peer support group, and online-based progress monitoring. Participants in the control condition had unstructured access to outpatient psychotherapy, standardized outpatient face-to-face continuation treatment, and psychotropic management. Self-reports of psychopathological symptoms and potential moderators were assessed at the start of inpatient treatment (T1), at discharge from inpatient treatment/start of TIMT (T2), and at 3-month (T3) and 12-month follow-up (T4).

Results: Education level, positive outcome expectations, and diagnoses significantly moderated intervention versus control differences regarding changes in outcomes between T2 and T3. Only education level moderated change differences between T2 and T4. The effectiveness of the intervention (vs control) was more pronounced among participants with a low (vs high) education level (T2-T3: B=–0.32, SE 0.16, *P*=.049; T2-T4: B=–0.42, SE 0.21, *P*=.049), participants with high (vs low) positive outcome expectations (T2-T3: B=–0.12, SE 0.05, *P*=.02) and participants with anxiety disorder (vs mood disorder) (T2-T3: B=–0.43, SE 0.21, *P*=.04). Simple slope analyses revealed that despite some subgroups benefiting less from the intervention than others, all subgroups still benefited significantly.

Conclusions: This transdiagnostic Internet-based maintenance treatment might be suitable for a wide range of participants differing in various clinical, motivational, and demographic characteristics. The treatment is especially effective for participants with low education levels. These findings may generalize to other Internet-based maintenance treatments.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 28632626; http://www.controlled-trials.com/isrctn/pf/28632626 (Archived by WebCite at http://www.webcitation.org/6IqZjTLrx).

(J Med Internet Res 2013;15(10):e191) doi:10.2196/jmir.2511

KEYWORDS

maintenance treatment; continuation treatment; Internet-based intervention; transdiagnostic treatment; mental disorders/inpatient psychotherapy; guided self-help; randomized controlled trial; relapse prevention; predictors; moderators

Introduction

Despite strong evidence for the efficacy of psychotherapy for common mental health disorders [1,2], long-term outcome of psychotherapeutic interventions are still a major concern [3-6]. Psychological treatments following acute phase psychotherapy that aim to maintain achieved changes (ie, maintenance phase treatments) have been shown to enhance outcome sustainability (eg, major depressive disorder [7,8], obsessive compulsive disorder [9], and personality disorders [10,11]). However, such interventions are difficult to disseminate owing to high intervention costs and limited clinician availability.

The use of the Internet to provide guided self-help maintenance phase treatments may help to overcome this unmet maintenance need. Internet-based guided self-help strategies for the maintenance phase of psychotherapies have several advantages over face-to-face maintenance approaches. These include (1) greater potential for the integration of acquired skills in daily life because of an emphasis on the patient’s active role in (guided) self-help treatment [12], (2) elimination of waiting periods between acute and maintenance treatment, (3) elimination of travel time and costs for both patients and clinicians, (4) access to the programs on a 24/7 basis, and (5) lower costs.

Several studies have shown promising results with delivering maintenance phase treatments over the Internet [13-20]. For example, our group developed a form of Internet-based continuation phase psychotherapy, a transdiagnostic Internet-based maintenance treatment (TIMT) following inpatient psychotherapy [15,20]. TIMT was designed to increase long-term outcomes of inpatients treated in a routine care setting for common mental health disorders, such as major depressive, anxiety, posttraumatic stress, obsessive compulsive, eating, or somatoform disorders. Recently, TIMT was evaluated in a pragmatic randomized controlled trial (RCT), comparing TIMT in addition to treatment as usual (TAU) to TAU only (N=400). In this study, participants in the TIMT plus TAU condition showed a better maintenance of inpatient treatment effects (ie, differences in change of psychopathological symptom severity) from inpatient discharge to 3-month follow-up (between-group effect size: d=0.38, P<.001) and 12-month follow-up (between-group effect size: d=0.55, P<.001) than TAU-only controls [15].

Although there is evidence for the general effectiveness of Internet-based maintenance phase treatments, little is known about which patients might or might not benefit from this particular kind of treatment delivery. Investigating the moderating effects of patient characteristics on Internet-based maintenance phase treatment effectiveness is crucial for identifying appropriate populations and for customizing interventions to the specific needs of patient subgroups. More knowledge regarding who is likely or unlikely to profit from these interventions should also help in identifying relevant mechanisms of change as well as allocating health care resources on an evidence-based level [21].

Only a few studies to date have investigated moderators of Internet-based intervention outcomes for mental health problems. In 1 of these studies, Warmerdam and colleagues [22] explored moderators of Internet-based cognitive behavioral therapy (CBT) and Internet-based problem-solving therapy for depressive symptoms. None of the variables investigated in this study (demographic variables, illness severity, dysfunctional attitudes, and problem-solving skills) moderated the differential effectiveness of the 2 treatments. In a study comparing the effects of Internet-based CBT to group-based face-to-face CBT, Spek et al [23] found that participants high in altruism performed better in group CBT than in Internet-based CBT (no significant findings for age, gender, education, neuroticism, extraversion, agreeableness, openness and conscientiousness, pretreatment severity, previous episodes of depression, and marital status). When comparing responses to online CBT for depression compared to a waitlist control group Button et al [24] found that higher pretreatment severity of depressive symptoms were associated with a greater benefit of treatment. In another study, de Graaf and colleagues [25] explored pretreatment and short-term improvement variables as moderators of unsupported Internet-based CBT outcomes, usual primary care (TAU), and CBT combined with TAU for depression. They found that patients with higher levels of extreme positive responding to questionnaires had a better outcome in Internet-based CBT compared to TAU, whereas those with parental psychiatric history or with a major depressive disorder diagnosis had a better outcome in Internet-based CBT plus TAU compared to TAU.

The aim of the present study was to identify moderating factors on the effects of TIMT after inpatient psychotherapy. Using data from a pragmatic RCT on the effectiveness of TIMT (ISRCTN:28632626) [15], we conducted secondary analyses to identify demographic, clinical, and motivational variables that moderate the effects of TIMT on change in psychopathological symptom severity.

Given the current lack of data on moderators of Internet-based continuation phase treatment effects, we used an exploratory approach including a wide range of potential pretreatment moderators [21]. Our choice of moderators was based on (1) results of previous studies investigating moderators of face-to-face continuation treatment outcomes [26,27], (2) results of previous predictors/moderators in Internet-based intervention outcome studies [22,23,25,28,29], (3) predictors of relapse/long-term outcome studies [5,27,30-36], and (4) theoretical assumptions attributed to intervention characteristics. The final list of potential moderators investigated in the present study included (1) demographics, such as age, gender, education level, and computer/Internet literacy; (2) clinical characteristics, such as diagnoses, remission status, age of first onset, comorbid personality disorder, and reliable change during inpatient treatment, and (c) motivational variables, such as self-efficacy and positive outcome expectations.

The primary research questions of this study were:

1. Do any of the pretreatment factors included in this study moderate the effectiveness of TIMT compared with TAU?
2. If moderating effects are found, do participants characterized by disadvantageous scores on identified moderators still benefit from TIMT?

Methods

Study Design

We performed secondary analyses using data from a pragmatic RCT comparing TIMT in addition to TAU following inpatient psychotherapy to TAU only (N=400) [15]. The RCT was conducted in a German clinic providing routine mental health care. Study outcomes were assessed by using self-report measures that were completed at inpatient admission (T1), end of inpatient treatment/beginning of TIMT (T2), 3 months after discharge/end of TIMT (T3), and 12 months after inpatient treatment completion (T4). The study was powered to find a small to moderate effect size in the main effect analyses, which was considered to be the smallest relevant difference to health care decision makers in this context. All procedures were approved by the university and the hospital institutional review boards. Design and results of the effectiveness trial are described in detail in a previously published study [15].

Participants and Procedures

We recruited potential participants from 2189 patients treated for a variety of mental disorders between July 2008 and October 2009 in the study hospital. Patients were eligible for the study if they (1) were age 18 years or older, (2) met criteria for a mental disorder according to the *International Classification of Diseases, Tenth Revision* (*ICD-10*) [37], (3) spoke German sufficiently, (4) had basic reading and writing skills, and (5) had access to a computer with an Internet connection. Exclusion criteria were (1) a psychotic diagnosis, (2) acute alcohol or substance dependence, and (3) a significant risk for suicide.

Participants who gave full written informed consent were randomly assigned to receive TAU only (control) or TAU plus TIMT (intervention). In total, 58 of 400 (14.5%) participants did not complete the T3 assessment and 113 (28.5%) did not complete the T4 assessment. Participants who did not provide data at 1 of the follow-ups did not differ from participants without missing data on baseline psychopathological symptom severity scores or any other clinical characteristics (all P values >.10), except for age (noncompleters on average 2.21 years younger than completers, P=.02). No significant interactions were found between missing pattern and outcome using pattern mixture analyses [38]. Thus, missing data appear not to bias the results. Figure 1 summarizes participant enrollment and flow throughout the study [15].

Interventions

Inpatient Treatment

Inpatient treatment was based on CBT [39]. Participants received 1 session of individual therapy (50 minutes) and an average of 6 sessions of group therapy (90 minutes) per week. Interventions were supplemented with sports therapy and physiotherapy, as well as medical treatment (including pharmacotherapy) when necessary. Treatment was delivered by 6 experienced therapists and 14 therapists in training. Duration of treatment ranged between 22 and 98 days (mean 46.30, SD 8.17).

Treatment as Usual Condition

Following inpatient treatment, all participants had unstructured access to outpatient psychotherapy and standardized outpatient group-based, face-to-face, maintenance treatment [40] as typically provided by the referring agencies. In addition, there was no restriction on the use of medication during the study period.

Treatment as Usual Plus Transdiagnostic Internet-Based Maintenance Treatment Condition

In addition to TAU, the intervention group had TIMT for 12 weeks. The main focus of TIMT is to support patients in the sustained utilization of skills acquired during treatment. For this purpose, TIMT works to help participants identify activities that they have found helpful and systematically integrate these into their daily life routines. Because TIMT aims to enhance whatever strategy patients experienced as helpful, it can be used to maintain treatment outcome regardless of which psychopathology the patient is suffering from and regardless of the kind of treatment the patient received before. TIMT consists of 5 core components. The first component is the generation of a personal development plan. This process is conducted during the last 10 days of inpatient treatment in which TIMT participants complete 3 sessions of blended (face-to-face and online) standardized goal-setting and action planning instead of inpatient TAU. Participants develop a detailed plan including (1) highly relevant personal goals they want to achieve during the intervention phase, and (2) implementation intentions [41], including details on how and when they will achieve these goals. The second and central component of TIMT is the completion of a structured Web diary in which participants evaluate the realization of their personal goals weekly and set specific goals for the next week. The third component of TIMT is an online peer support group. Subgroups consisting of 3 to 6 participants are asked to give asynchronous online feedback to one another on their Web diaries. The fourth component of TIMT is coach support, involving weekly asynchronous written online feedback from a therapist regarding a participants' Web diary. Coaches differed in their level of formal training, ranging from master's level psychology students (n=1) and psychotherapists-in-training (n=1) to experienced CBT-trained psychotherapists with more than 10 years of professional experience (n=3). Coaches were supervised once a week by a licensed senior therapist, as is usual in the study hospital. Coaches were advised not to spend more than 30 minutes per week on support per patient. Total duration of support rendered was 231 minutes on average per patient (range: 10-490, SD 128). Finally, TIMT included weekly online monitoring of psychopathological symptoms.

Treatment Received

The intervention and control group did not differ in types of treatment received except for frequency of sedatives taken. Participants in the intervention group were less likely to take sedatives than controls (*P*<.001) [15].

Measures

Moderators

In total, we included 11 pretreatment participant characteristics: age, sex, education, main diagnosis, comorbid personality disorder, remission status at the end of inpatient treatment, reliable change in the primary outcome during inpatient treatment, years since first disorder onset, Internet/computer literacy, positive outcome expectations, and health-related self-efficacy.

Information on sex, age, and education were extracted from the inpatient clinic patient files. All self-report data were assessed using an online-based assessment tool. Diagnoses and year of first disorder onsets were assessed during the intake interview. All interviewers were experienced psychotherapists who were either psychologists or physicians with a master's degree or higher, trained extensively in administering the structured clinical interviews of the German version of the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) (*DSM-IV*) [42]. Participants were classified as being remitted at inpatient discharge (yes/no) when individual scores in psychopathological symptom severity (primary outcome measure as described subsequently) exceeded a raw score value of 0.685 [43]). Reliable change in symptom severity (yes/no) was determined according to the widely used reliable change index of Jacobson and Truax [44]. Individual reliable change scores less than -1.96 were considered to reflect reliable (positive) change; scores equal to or greater than -1.96 reflected no reliable change. Participants were classified as Internet/computer illiterate if they checked the not at all response category for the item "I am used to sending and receiving emails" (1=not at all; 4=completely true). All other participants were classified as Internet/computer literate. Participants were coded as having a low education level if they reported 9 years of school education, as a medium education level if they reported 10 years of school education and a corresponding degree, or as a high education level if they reported a minimum of 13 years of school/college education and a corresponding degree.

Positive outcome expectations were assessed by using the respective subscale of the Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV) [45]. This scale consists of 4 items measuring participants' expectations regarding the effectiveness of their inpatient treatment (eg, "I think that finally my problems will be solved"). Response scales ranged from 0 (=do not agree) to 4 (=agree completely). Higher scores reflect higher positive outcome expectations. Construct validity of the scale was demonstrated in several studies [45]. In the present study, internal consistency (Cronbach alpha) was .77.

Health-related self-efficacy was assessed by using the self-efficacy subscale of the 49-item short form of the Hamburg Modules for the Assessment of Psychosocial Health (HEALTH-49) questionnaire [43].This scale includes 5 items measuring expected persistence and success in several domains (eg, "Despite my discomfort, I achieve the personal goals that I set for myself" score inverted for scale calculation; 0=not true, 4=very true). Higher scores denote lower self-efficacy. Internal consistency (Cronbach alpha) was .86 in the present study.

Dependent Variable

The primary outcome from the effectiveness trial was change in general psychopathological symptom severity (symptom severity) from discharge (T2) to 3- and 12-month follow-ups (T3, T4). Symptom severity was assessed by using the HEALTH-49, a widely used measure of symptom severity in Germany [43]. The HEALTH-49 general psychopathological symptom severity scale consists of 18 items related to somatoform complaints (7 items), depressiveness (6 items), and phobic anxiety (5 items). Participants were asked to rate the severity to which they had suffered from the presented symptoms in the previous 2 weeks (0=not at all; 4=very much). Reliability and construct validity have been established in several studies based on large clinical and nonclinical samples (1548 psychotherapy inpatients, 5630 primary care patients, see [43]). In the present study, internal consistency (Cronbach alpha coefficients) at baseline was .87 for the overall general psychopathological symptom severity score, 0.90 for depressive symptoms, 0.86 for somatoform complaints, and 0.86 for phobic anxiety.

Statistical Analyses

Group differences regarding baseline characteristics were compared via chi-square tests for categorical variables and *t* tests for continuous variables. Interactions between pretreatment participant characteristics (moderators) and interindividual differences in intraindividual changes across measurement occasions were modeled and tested via multilevel mixed-effect models. Change in symptom severity over time was dummy coded and treated as a fixed level-1 (ie, within-subjects) effect (dummy 1: T1-T2, dummy 2: T2-T3, dummy 3: T2-T4). Treatment conditions (0=control condition, 1=intervention condition) was treated as a fixed level-2 (ie, between-subjects) effect. More important for the present purpose, interactions between moderator and treatment condition, all cross-level interaction effects (condition × T1-T2, T2-T3, T2-T4; moderator × T1-T2, T2-T3, T2-T4), and 3-way interaction effects (moderator × condition × T1-T2, T2-T3, T2-T4) were also included in the models. A 3-way interaction effect of moderator × condition × T2-T3 or moderator × condition × T2-T4 would indicate that the magnitude of the intervention effect varies as a function of the moderator. The model imposed no restrictions on the covariance matrix for measurement occasions. Thus, no model assumptions were tested. We standardized continuous predictors so that regression coefficients were estimated for participants with average scores on the putative moderator.

To increase interpretability and allow for testing nonlinear effects, categorical variables with more than 2 categories (ie, diagnosis, years since first disorder onset, education) were recoded into a maximum of 3 meaningful categories. Because of low prevalence rates, we excluded diagnoses other than depression, anxiety disorders, and adjustment disorders. All continuous moderators (ie, age, self-efficacy, positive outcome expectations) were standardized so that regression coefficients refer to participants with average scores on each moderator.

Aiming at an intention-to-treat (ITT) design, we included all participants randomly assigned to conditions. We employed a full information maximum likelihood (FIML) estimation, which allows for all available data to be included without replacement or imputation of missing values. The FIML estimation for mixed models is especially robust with respect to missing data [46].

Additionally, we conducted follow-up simple slope analyses for each significant 3-way interaction effect [47] to probe the relevant lower-order effects. In this method, the slope and the significance of the intervention main effect is evaluated for conditional values of the moderator. For significant 3-way interactions of continuous moderators, simple slopes were calculated for the mean and one standard deviation above and below the mean [48].

Effect sizes for each significant moderator were calculated based on comparing the effect of control versus intervention groups on symptom severity scores, with participants grouped by the significant moderator variable. Cohen’s *d* scores [49] were calculated by standardizing the differences between baseline and follow-up by the pooled standard deviation of baseline scores.

To verify whether the results of the ITT analyses would be sustained among the intervention completers sample only, we subsequently repeated all mixed-effects models with participants who stayed within key treatment parameters (completed at least 6 of 12 Web diary entries or more than 25 posts, n=177).

To clarify the generalizability of our findings, we assessed all potential moderators also from patients who were treated during the recruitment period in the study center, but did not participate in the trial (not invited, declined to participate, not fulfilling inclusion criteria) but gave informed consent to use their data for research purposes (n=1789). Study participants and nonparticipants were compared using chi-square tests for categorical variables and *t* tests for continuous variables.

Finally, if a significant moderator effect contradicted our a priori expectations, we conducted post hoc simple slope analyses for the control and intervention groups separately to identify the reasons for the effect. All analyses were performed with SPSS 19 (IBM Corp, Armonk, NY, USA).

Descriptive Data

Table 1 shows descriptive statistics for the dependent variable general psychopathological symptom severity. Table 2 shows descriptive data for all moderator variables. Table 1 and parts of Table 2 have been reported in previous studies [15]. Consistent with random assignment, no differences were found between intervention and control group on any of the pretreatment variables.

Figure 1. Participant flow and study dropouts at each stage of the study.

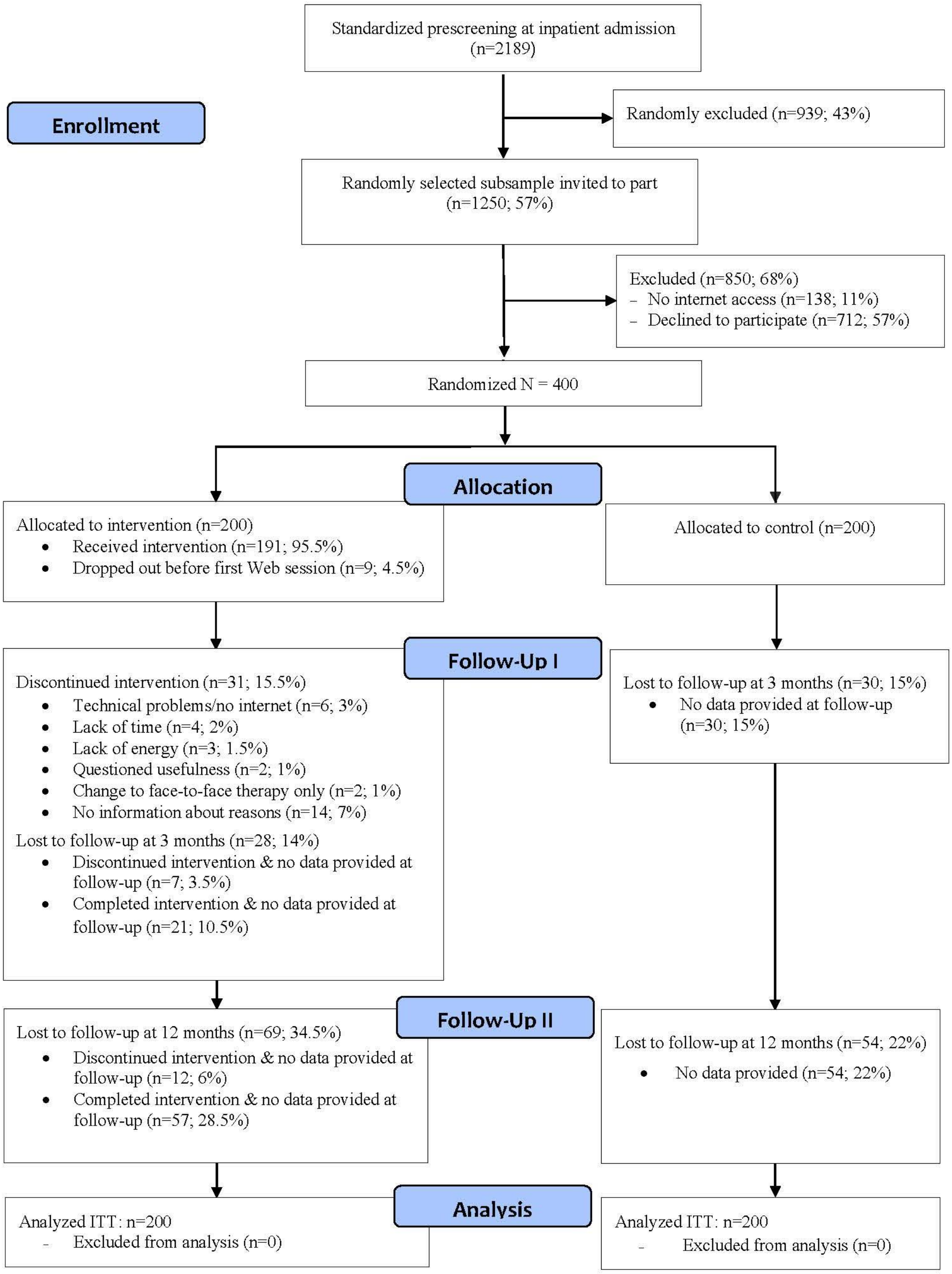


Table 1. Descriptives for primary trial main outcome, psychopathological symptom severity as measured by the general psychopathological symptom severity subscale of the HEALTH-49 questionnaire.

Assessment points	Time of Assessment	Intervention (n=200)		Control (n=200)	
		Mean	SD	Mean	SD
T1	Inpatient admission	1.50	0.69	1.49	0.71
T2	Inpatient discharge	0.83	0.64	0.83	0.66
T3	3-month follow-up	0.71	0.61	0.96	0.69
T4	12-month follow-up	0.78	0.69	1.12	0.84

Table 2. Descriptives for pretreatment moderator variables.

Variables	Intervention (n=200)	Control (n=200)	Nonparticipants[a] (n=1789)
Age, mean (SD)	45.09 (8.88)	45.45 (9.80)	47.12 (9.45)
Sex (female), n (%)	147 (73.5)	151 (75.5)	1360 (76.0)
Education, n (%)			
High	80 (40.0)	78 (39.0)	498 (27.8)
Medium	93 (46.5)	91 (45.5)	779 (43.5)
Low	26 (13.0)	31 (15.5)	509 (28.5)
Existing Internet literacy (%)	178 (89.0)	167 (83.5)	1132 (67.5)[b]
Disorder, n (%)			
Mood disorder	108 (54.0)	113 (56.5)	918 (51.3)
Anxiety	19 (9.5)	18 (9.0)	206 (11.5)
Adjustment	53 (26.5)	38 (19.0)	405 (22.6)
Other	20 (10.0)	31 (15.5)	260 (14.5)
Comorbid personality disorder, n (%)	20 (10.0)	22 (11.0)	175 (9.8)
Years since first disorder onset (years) n (%)			
< 1	44 (22.0)	47 (23.5)	430 (24.2)
1-5	55 (27.5)	44 (22.0)	444 (24.9)
> 5	96 (48.0)	105 (52.5)	906 (50.9)
Reliable change during inpatient treatment, n (%)	100 (50.0)	90 (45.0)	1052 (58.8)
Remission at discharge, n (%)	94 (47.0)	93 (46.5)	787 (44.0)
Self-efficacy, mean (SD)	1.47 (0.83)	1.49 (0.87)	1.58 (0.90)
Positive outcome expectations, mean (SD)	3.86 (0.74)	3.92 (0.66)	3.72 (0.78)

[a] All differences between conditions were nonsignificant. If percentages do not reach 100, it is due to missing data.

[b] n=1676.

Results

Moderators of Treatment Outcome

Overview

The subsequent tables show the mixed-effect model results based on ITT for the interactions between pretreatment participant characteristics (moderators), intervention condition, and changes in symptom severity. Intercepts represent the estimated level of symptom severity at baseline (discharge, T2). The regression coefficient of the moderator represents differences in symptom severity between participants differing in 1 unit of the hypothesized moderator at baseline. The regression coefficient of T1-T2 represents the average difference in symptom severity between inpatient admission (T1) and inpatient discharge (T2) in the control group, the regression coefficient of T2-T3 represents the average difference in symptom severity between discharge (T2) and 3-month follow-up (T3) in the control group, and the regression coefficient of T2-T4 represents the average difference in symptom severity between discharge (T2) and 1-year follow-up (T4) in the control group. The regression coefficient of the condition represents differences in symptom severity between the intervention and the control condition at discharge (T2). The cross-level interactions condition × T1-T2, T2-T3, T2-T4 represent intervention versus control group differences in changes over time.

As expected, we found (1) a significant decrease in symptom severity between T1 and T2 in both conditions (T1-T2), (2) no interaction between T1-T2 and the intervention condition, (3) a significant condition × T2-T3 interaction effect showing that symptom severity remained low in the intervention group between T2 and T3 but increased in the control group, and (4) a significant T2-T4 × condition interaction effect showing that symptom severity remained low in the intervention group between T2 and T4 but increased in the control group (Table 4). The regression coefficients of the moderator × T1-T2, T2-T3, T2-T4 interaction effects represent moderator effects on changes in symptom severity across measurement occasions. Finally, the regression coefficient of condition × moderator × T1-T2, T2-T3, T2-T4 interaction effects represent moderator effects on intervention versus control condition differences on change scores over time.

Dichotomous Moderator Variables

Table 3 shows results for dichotomous moderator variables. The results revealed no moderator effects on intervention versus control group differences on changes in symptom severity over time (see Table 3, condition × moderator × T1-T2, T2-T3, T2-T4). Thus, none of the dichotomous moderators reliably altered the effectiveness of intervention versus control on symptom severity over time. The intervention was superior to control with regard to outcome sustainability, irrespective of sex, Internet literacy, reliable changes during inpatient treatment, comorbid personality disorder, or remission status at T2.

Table 3. Multilevel results of the interactions between pretreatment participant characteristics (dichotomic moderator variables), intervention condition, and change in psychopathological symptom severity (dummy coded) for the intention-to-treat sample (N=400) using full maximum likelihood estimation.

Interaction terms	Sex[a]			Internet literacy[b]			Reliable change[c]			Comorbid PD[d]			Remission status[e]		
	B	SE	*P*	B	SE	*P*	B	SE	*P*	B	SE	*P*	B	SE	*P*
Intercept[f]	0.94	0.09	<.001	0.91	0.11	<.001	0.98	0.06	<.001	0.78	0.05	<.001	0.30	0.04	<.001
Moderator	–0.16	0.11	.13	–0.11	0.12	.39	–0.34	0.09	<.001	0.42	0.15	.004	0.98	0.06	<.001
T1-T2 (dummy 1)[g]	0.60	0.08	<.001	0.69	0.10	<.001	0.25	0.03	<.001	0.66	0.04	<.001	0.84	0.06	<.001
T2-T3 (dummy 2)[h]	0.20	0.08	.008	0.19	0.09	.04	0.00	0.05	.99	0.16	0.04	<.001	0.36	0.05	<.001
T2-T4 (dummy 3)[i]	0.48	0.10	<.001	0.40	0.12	.001	0.12	0.06	.05	0.30	0.05	<.001	0.44	0.07	<.001
Condition[j]	–0.35	0.13	.006	–0.13	0.18	.48	0.05	0.09	.58	0.02	0.07	.78	0.01	0.06	.90
Condition×T1-T2	0.04	0.12	.750	0.07	0.16	.66	–0.06	0.05	.25	0.04	0.06	.53	–0.04	0.08	.61
Condition×T2-T3	–0.24	0.11	.03	–0.30	0.14	.04	–0.22	0.07	.002	–0.24	0.06	<.001	–0.30	0.07	<.001
Condition×T2-T4	–0.45	0.14	.001	–0.65	0.18	<.001	–0.41	0.10	<.001	–0.34	0.07	<.001	–0.30	0.10	.002
Moderator×T1-T2	0.09	0.10	.35	–0.03	0.11	.80	0.92	0.05	<.001	0.02	0.13	.87	–0.31	0.08	<.001
Moderator×T2-T3	–0.04	0.09	.61	–0.02	0.10	.83	0.36	0.07	<.001	0.04	0.12	.75	–0.37	0.07	<.001
Moderator×T2-T4	–0.22	0.11	.05	–0.11	0.13	.41	0.41	0.09	<.001	0.08	0.16	.61	–0.25	0.10	.01
Condition×moderator	0.46	0.15	.002	0.14	0.19	.47	–0.07	0.13	.57	–0.23	0.21	.28	–0.02	0.09	.83
Cond×mod×T1-T2[k]	–0.03	0.14	.84	–0.06	0.17	.72	0.05	0.07	.47	–0.25	0.19	.20	0.11	0.12	.36
Cond×mod×T2-T3[k]	–0.02	0.12	.89	0.06	0.15	.71	–0.08	0.10	.46	–0.12	0.18	.49	0.12	0.10	.22
Cond×mod×T2-T4[k]	0.12	0.16	.47	0.34	0.20	.09	0.06	0.14	.67	–0.22	0.23	.34	–0.11	0.14	.41

[a] Sex (0=female; 1=male).

[b] Existing Internet literacy (0=no; 1=yes).

[c] Reliable change: reliable change during inpatient treatment (0=no; 1=yes).

[d] Comorbid PD: comorbid personality disorder (0=no; 1=yes).

[e] Remission status: remission status at baseline (T2) (0=in remission; 1=not in remission).

[f] Intercept: general psychopathological symptom severity in control at baseline (T2).

[g] T1-T2: dummy-coded change in general psychopathological symptom severity from T1 to T2.

[h] T2-T3: dummy-coded change in general psychopathological symptom severity from T2 to T3.

[i] T2-T4: dummy-coded change in general psychopathological symptom severity from T2 to T4.

[j] Condition (0=control; 1=intervention).

[k] Cond × mod: condition × moderator.

Trichotomous Moderator Variables

Table 4 shows mixed-effects model results for the 3 trichotomous moderator variables education level, diagnoses, and years since first disorder onset. Three significant 3-way interaction effects were found. Education dummy 2 (low vs high education) interacted with condition × T2-T3, and with condition × T2-T4. These interactions indicate that a greater intervention effect was found among participants with low compared to high education level (see Figure 2). Participants low in education showed a larger intervention vs control condition difference on changes in symptom severity between discharge and 3-month follow-up and between discharge and 1-year follow-up. Post hoc analyses demonstrated that although simple slopes for the intervention main effects (condition × T2-T3, T2-T4) were lower among high-educated participants compared to low-educated participants, the intervention main effect was still significant (simple slope high-educated participants T2-T3: B=–0.17, SE 0.08, *P*=.04; T2-T4: B=–0.25, SE 0.11, *P*=.03; simple slope low-educated participants T2-T3: B=–0.49, SE 0.14, *P*<.001; T2-T4: B=–0.66, SE 0.18, *P*<.001). Moreover, diagnoses dummy 1 (mood disorders vs anxiety disorders) interacted with condition × T2-T3. Participants diagnosed with an anxiety disorder showed a larger intervention versus control group difference on changes in symptom severity between discharge and 3-month follow-up than participants diagnosed with a mood disorder (see Figure 3). Post hoc analyses demonstrate that although simple slopes for the intervention main effect (condition × T2-T3) were lower among participants with a mood disorder compared to participants with an anxiety disorder, the intervention main effect was significant in both groups (simple slope mood disorder T2-T3: B=–0.21, SE 0.07, *P*=.004; simple slope anxiety disorder T2-T3: B=–0.64, SE 0.02, *P*<.001). Diagnoses dummy 1 did not moderate the association between treatment and change in symptom severity from discharge to 1-year follow-up.

Years since disorder onset did not moderate the effect of treatment on any intervention versus control group differences on change scores. Thus, transdiagnostic Internet-based maintenance treatment is effective irrespective of years since first disorder onset.

Figure 2. Estimated course of symptoms based on simple slope mixed-effect model analysis for significant moderators effect of education (0=high education, n=159; 1=low education, n=57) at inpatient admission (T1), inpatient discharge/begin transdiagnostic Internet-based maintenance treatment (T2), 3-month follow-up/end transdiagnostic Internet-based maintenance treatment (T3), and 12-month follow-up (T4).

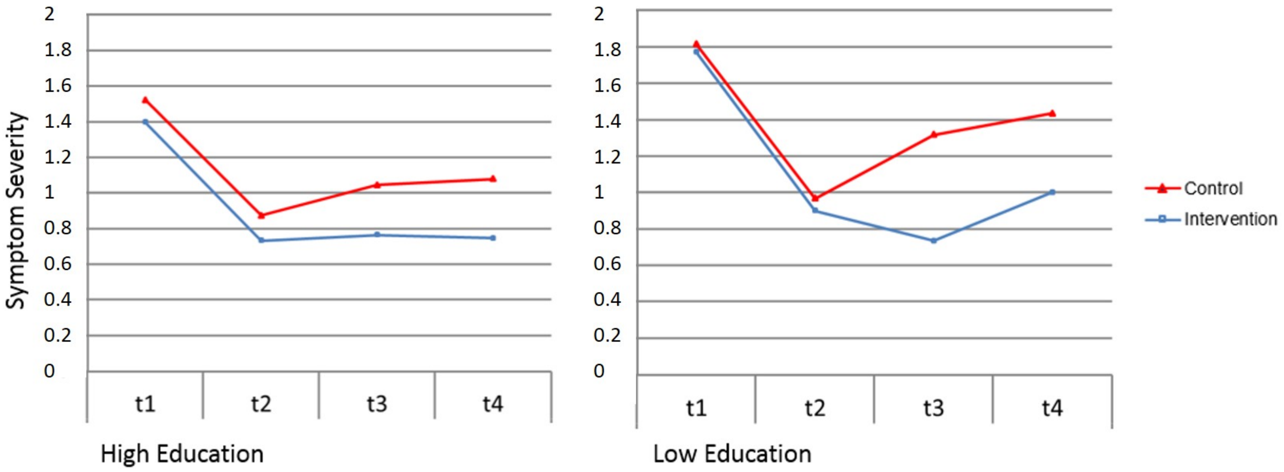


Figure 3. Estimated course of symptoms based on simple slope mixed-effect model analyses for significant moderator effect of diagnoses (0=mood disorder, n=221; 1=anxiety disorder, n=37) at inpatient admission (T1), inpatient discharge/begin transdiagnostic Internet-based maintenance treatment (T2), 3-month follow-up/end transdiagnostic Internet-based maintenance treatment (T3), and 12-month follow-up (T4).

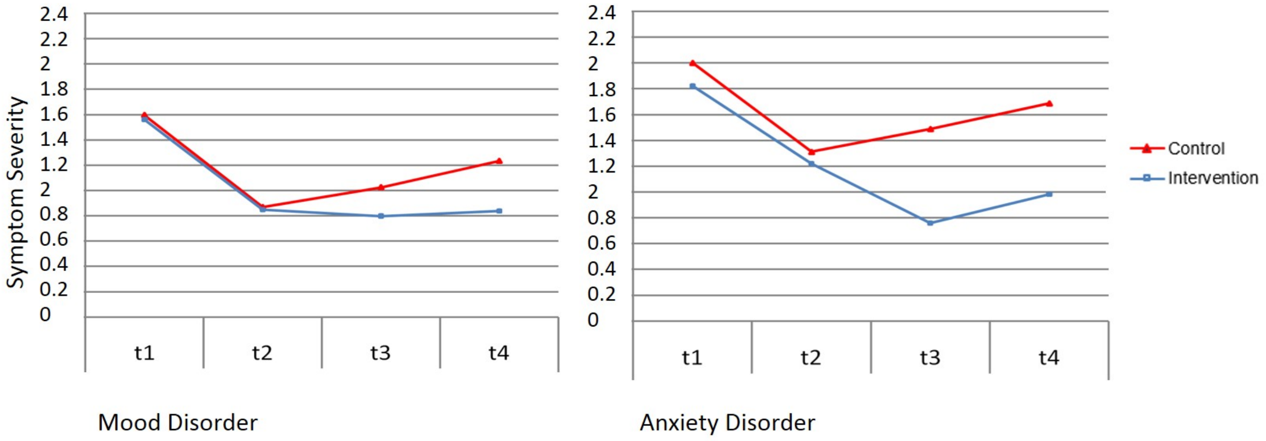


Table 4. Multilevel results for interactions between pretreatment participant characteristics (trichotomous moderator variables), intervention condition, and change in psychopathological symptom severity (dummy coded) for the intention-to-treat sample (N=400) using full maximum likelihood estimation.

Interaction terms	Education level[a]			Diagnoses[b]			Years since onset[c]		
	B	SE	*P*	B	SE	*P*	B	SE	*P*
Intercept[d]	0.88	0.07	<.001	0.87	0.06	<.001	0.88	0.09	<.001
Moderator dummy 1	–0.16	0.10	.11	0.44	0.16	.005	–0.04	0.11	.687
Moderator dummy 2	0.08	0.14	.55	–0.34	0.12	.004	–0.13	0.13	.313
T1-T2[e]	0.63	0.07	<.001	0.73	0.06	<.001	0.64	0.09	<.001
T2-T3[f]	0.17	0.06	.005	0.16	0.05	.002	0.10	0.08	.18
T2-T4[g]	0.22	0.08	.004	0.36	0.06	<.001	0.31	0.10	.002
Condition[h]	–0.14	0.10	.17	–0.02	0.08	.82	–0.28	0.13	.02
Condition×T1-T2	0.03	0.09	.76	–0.02	0.08	.85	0.09	0.12	.43
Condition×T2-T3	–0.17	0.08	.04	–0.21	0.07	.004	–0.20	0.10	.06
Condition×T2-T4	–0.25	0.11	.03	–0.38	0.09	<.001	–0.33	0.14	.02
Moderator dummy 1×T1-T2	0.03	0.09	.72	–0.04	0.15	.80	–0.04	0.11	.70
Moderator dummy 1×T2-T3	–0.04	0.08	.64	0.02	0.15	.89	0.04	0.09	.68
Moderator dummy 1×T2-T4	0.07	0.11	.48	0.01	0.19	.96	–0.03	0.12	.80
Moderator dummy 2×T1-T2	0.17	0.13	.17	–0.09	0.11	.40	0.20	0.12	.10
Moderator dummy 2×T2-T3	0.10	0.11	.35	0.03	0.10	.75	0.14	0.11	.19
Moderator dummy 2×T2-T4	0.35	0.14	.02	–0.19	0.13	.14	0.02	0.14	.88
Cond×mod×dummy 1[i]	0.22	0.14	.11	–0.07	0.22	.74	0.36	0.15	.02
Cond×mod×dummy 2[i]	0.19	0.20	.34	0.08	0.16	.63	0.27	0.18	.13
Cond×mod×dummy 1×T1-T2[i]	–0.03	0.13	.82	–0.07	0.22	.75	–0.07	0.15	.63
Cond×mod×dummy 1×T2-T3[i]	–0.08	0.11	.50	–0.43	0.21	.04	–0.03	0.13	.82
Cond×mod×dummy 1×T2-T4[i]	–0.12	0.15	.42	–0.24	0.26	.37	0.01	0.17	.93
Cond×mod×dummy 2×T1-T2[i]	0.05	0.18	.77	0.10	0.15	.52	–0.14	0.17	.43
Cond×mod×dummy 2×T2-T3[i]	–0.32	0.16	.049	–0.03	0.14	.83	–0.09	0.15	.56
Cond×mod×dummy 2×T2-T4[i]	–0.42	0.21	.049	0.15	0.18	.41	0.03	0.19	.89

[a] Education level dummy 1 (0=high education level; 1=medium education level), education level dummy 2 (0=high education level; 1=low education level).

[b] Diagnoses dummy 1 (0=mood disorder; 1=anxiety disorder), diagnoses dummy 2 (0=mood disorder; 1=adjustment disorder).

[c] Years since onset: years since disorder onset dummy 1 (0=1-5 years; 1=>5 years), years since disorder onset dummy 2 (0=1-5 years; 1=<1 year).

[d] Intercept: general psychopathological symptom severity in control at baseline (T2).

[e] T1-T2: dummy-coded change in general psychopathological symptom severity from T1 to T2.

[f] T2-T3: dummy-coded change in general psychopathological symptom severity from T2 to T3.

[g] T2-T4: dummy-coded change in general psychopathological symptom severity from T2 to T4.

[h] Condition (0=control; 1=intervention).

[i] Cond × mod × dummy: condition × moderator × dummy.

Continuous Moderator Variables

Table 5 shows mixed-effect model results for the continuous moderator variables age, self-efficacy, and positive outcome expectations. One significant 3-way interaction was found. Positive outcome expectations interacted with condition × T2-T3. This interaction indicates that more positive outcome expectations were associated with stronger intervention effects between discharge and 3-month follow-up (see Figure 4). Follow-up analyses revealed that although simple slopes for the intervention main effect (condition × T2-T3) were lower among participants with a moderate (mean) positive outcome expectation than for participants with a high (mean + 1 SD) positive outcome expectations, the intervention effect was still significant (simple slope moderate positive outcome expectations T2-T3: B=–0.25, SE 0.05, *P*<.001; simple slope high positive outcome expectations T2-T3: B=–0.36, SE 0.07, *P*<.001). For participants with a low positive outcome expectations (mean – 1 SD), the simple slope for the intervention main effect (condition × T2-T3) was lower and no longer significant (simple slope low positive outcome expectations T2-T3: B=–0.13, SE 0.07, *P*=.09). Only 14.4% of participants (57/400) expressed low positive outcome expectations. Therefore, the drop to nonsignificance was likely because of low power. Moreover, simple slope analyses for this participant group showed that the intervention main effect on change in symptom severity from discharge to 1-year follow-up was significant (simple slope low positive outcome expectations T2-T4: B=–0.38, SE 0.10, *P*<.001). Although short-term effects were not significant, participants with low positive outcome expectations benefited in the long term from the intervention. There was no interaction between positive outcome expectations and change in symptom severity from discharge to 1-year follow-up and no interaction effect including the other continuous variables age and self-efficacy. Thus, TIMT seems to be effective irrespective of age and self-efficacy.

Table 5. Multilevel results for interactions between pretreatment participant characteristics (continuous moderator variables), intervention condition, and change in psychopathological symptom severity (dummy coded) for intention-to-treat sample (N=400) using full maximum likelihood estimation.

Interaction terms	Age[a]			Self efficacy[a]			Positive outcome expectations[a]		
	B	SE	*P*	B	SE	*P*	B	SE	*P*
Intercept[b]	0.83	0.05	.<001	0.82	0.04	<.001	0.84	0.04	<.001
Moderator[c]	–0.12	0.04	.004	0.44	0.03	<.001	–0.23	0.05	<.001
T1-T2[d]	0.67	0.04	<.001	0.67	0.04	<.001	0.66	0.04	<.001
T2-T3[e]	0.17	0.04	<.001	0.17	0.04	<.001	0.16	0.04	<.001
T2-T4[f]	0.31	0.05	<.001	0.31	0.05	<.001	0.30	0.05	<.001
Condition[g]	–0.01	0.06	.85	0.01	0.05	.89	–0.01	0.06	.84
Condition×T1-T2	0.02	0.06	.79	0.01	0.06	.88	0.02	0.06	.75
Condition×T2-T3	–0.24	0.05	<.001	–0.25	0.05	<.001	–0.25	0.05	<.001
Condition×T2-T4	–0.35	0.07	<.001	–0.36	0.07	<.001	–0.35	0.07	<.001
Moderator×T1-T2	0.01	0.04	.79	–0.18	0.04	<.001	0.08	0.04	.08
Moderator×T2-T3	0.07	0.04	.07	–0.16	0.04	.002	0.07	0.04	.08
Moderator×T2-T4	0.05	0.05	.26	–0.05	0.05	.26	0.03	0.05	.62
Condition×moderator	0.03	0.06	.60	–0.04	0.05	.42	0.12	0.06	.06
Condition×moderator×T1-T2	0.03	0.06	.60	0.05	0.06	.39	–0.10	0.06	.09
Condition×moderator×T2-T3	0.04	0.05	.50	0.07	0.05	.22	–0.12	0.05	.02
Condition×moderator×T2-T4	0.00	0.07	.98	–0.10	0.07	.15	0.03	0.07	.65

[a] All continuous variables standardized.

[b] Intercept: general psychopathological symptom severity in control at baseline (T2).

[c] Moderators (0=mean; 1=mean + 1 SD).

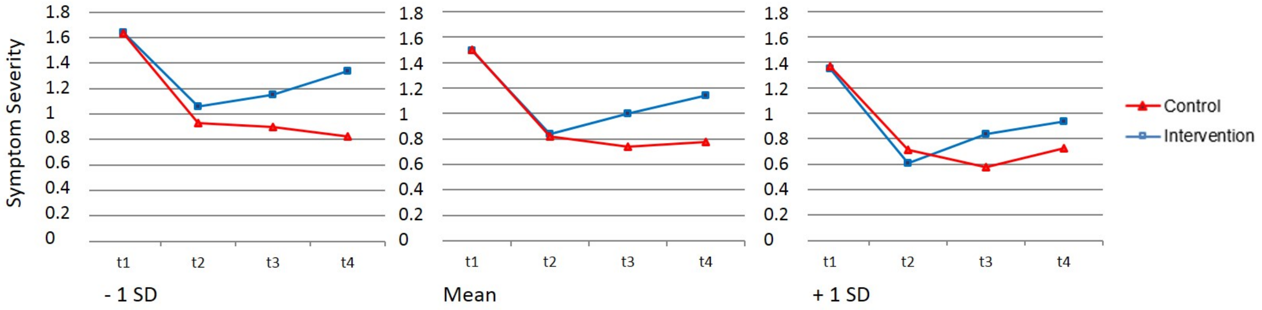
[d] T1-T2: dummy-coded change in general psychopathological symptom severity from T1 to T2.

[e] T2-T3: dummy-coded change in general psychopathological symptom severity from T2 to T3.

[f] T2-T4: dummy-coded change in general psychopathological symptom severity from T2 to T4.

[g] Condition (0=control; 1=intervention).

Figure 4. Estimated course of symptoms based on simple slope mixed-effect model analyses for significant moderator positive outcome expectations (mean vs mean – 1 SD vs mean + 1 SD) at inpatient admission (T1), inpatient discharge/begin transdiagnostic Internet-based maintenance treatment (T2), 3-month follow-up/end transdiagnostic Internet-based maintenance treatment (T3), and 12-month follow-up (T4).



Effect Sizes

Effect sizes (Cohen's *d*) for each significant moderator were calculated based on comparing the effect of control versus intervention condition on symptom severity, with participants grouped by parameter values on each significant moderator variable. A mean effect size of *d*=0.22 was found for participants with high education and *d*=0.80 for participants with low education for control versus intervention group differences in change of psychopathological symptom severity from discharge to 3-month follow-up. For change from discharge to 1-year follow-up, a mean effect size of *d*=0.30 for high-educated participants and a mean effect size of *d*=0.57 for low-educated participants was found. With diagnoses as the moderator, control versus intervention group differences in change from discharge to 3-month follow-up were *d*=0.33 for participants with a mood disorder and *d*=1.02 for participants with an anxiety disorder. With positive outcome expectations as moderator, control versus intervention group differences in change from discharge to 3-month follow-up were *d*=0.58 for participants with high positive outcome expectations, *d*=0.39 for participants with mean positive outcome expectations, and *d*=0.20 for participants with low positive outcome expectations.

Intervention Completers Sample

The results of the following intervention completers analyses closely paralleled those of the ITT analyses. Most of the significant 3-way interactions were also significant in the completers sample (B=–0.45 to –0.12, SE 0.05-0.21, *P*=.03-.046). Only the interaction of education dummy 2 with condition × T2-T4 was no longer significant at follow-up (B=–0.43, SE 0.22, *P*=0.05). None of the nonsignificant interactions in the ITT analyses was significant in the completers sample (B=–0.11 to 0.02, SE 0.07-0.20, *P*=.08-.97).

Generalizability

As shown in Table 1 and partly reported in previous studies [15], study participants did not differ from nonparticipants (n=1789) regarding sex (χ^2_1=0.4, *P*=.52), years since first disorder onset (χ^2_2=0.1, *P*=.93), existing comorbid personality disorder (χ^2_1=0.5, *P*=.46), or remission status at the end of inpatient treatment (χ^2_1=1.0, *P*=.32) or initial psychopathological symptom severity at inpatient admission (study participant symptom severity T1: mean 1.49, SD 0.70; nonparticipant symptom severity T1: mean 1.52, SD 0.84, $t_{679.03}$=–0.69, *P*=.54). Study participants were significantly younger than nonparticipants (with an average difference of 1.7 years, t_{2135}=–3.54, *P*<.001), had higher self-efficacy (t_{2182}=–2.11, *P*=.04, *d*=0.15) had a slightly higher education level (χ^2_2=40.81, *P*<.001, Kendall's tau coefficient=0.11), had higher positive outcome expectations ($t_{625.6}$=4.07, *P*<.001, *d*=0.27). Compared to nonparticipants, a greater percentage of participants had access to the Internet (χ^2_1=47.3, *P*<.001, phi coefficient=0.15) were Internet literate (χ^2_1=62.7, *P*<.001, phi coefficient=0.17), and relatively fewer showed reliable change during inpatient treatment (χ^2_1=5.3, *P*=.02, phi coefficient=0.05).

Post Hoc Analyses

The moderator effect of education contradicted our a priori expectation of higher educated participants benefiting to a greater extent from the Internet-based intervention than lower educated participants. Thus, we conducted further post hoc simple slope analyses for the control group and the intervention group separately to identify possible explanations for this effect. For participants in the control group, we found no significant interaction between education and changes in symptom severity from discharge to 3-month follow-up (education dummy 2 × T2-T3 interaction, B=0.10, SE 0.11, *P*=.35), but we found a significant interaction between education and changes from discharge to 12-month follow-up (education dummy 2 × T2-T4, B=0.35, SE 0.14, *P*=.02). Less-educated participants had a greater risk for deterioration from discharge to 1-year follow-up than more-educated participants did. In contrast, we found no significant interaction of low compared to high education level in the intervention group, neither for changes in symptom severity from discharge to 3-month follow-up (B=0.21, SE 0.17, *P*=.07) nor for changes from discharge to 12-month follow-up (B=0.06, SE 0.15, *P*=.68). In contrast to the control group, less-educated intervention participants did not show a greater risk for deterioration in symptom severity than more-educated participants, indicating that participating in TIMT can effectively reduce this risk factor.

Discussion

Principal Results and Comparison With Prior Work

In the present study, we aimed to identify moderators of treatment outcome for TIMT following inpatient psychotherapy. Education level, positive outcome expectations, and mental health diagnoses were identified as significant moderators of TIMT's effects on psychopathological symptom severity. Findings indicate that the effects of TIMT on general psychopathological symptom severity were more pronounced among participants with a low (vs high) education level. Participants with high positive outcome expectations profited in the short term (until 3-month follow-up) more than participants with low positive outcome expectations. However, this effect was not significant at 1-year follow-up. Moreover, participants with a mood disorder benefited less from the intervention than did participants with an anxiety disorder; however, this effect was also not significant at 1-year follow-up. Simple slope analyses revealed that even when some groups profited less from participating, treatment effects in these subgroups were still significant, except for the subgroup of participants with low positive outcome expectation at 3-month follow-up.

Other pretreatment variables did not interact with TIMT's effects indicating that TIMT might be superior to TAU only with regard to outcome sustainability irrespective of age, gender, comorbid personality disorder, years since disorder onset, self-efficacy, remission status at the end of inpatient treatment, reliable change in psychopathological symptom severity during inpatient treatment, and Internet literacy. However, given that these analyses were exploratory and the study was not powered to find small interaction effects, these null findings should be interpreted with caution.

The finding that participants with low education benefited more from using TIMT than participants with high education contrasts with findings from a study investigating moderators in face-to-face continuation phase psychotherapy in which education did not interact with treatment outcome [27]. Moreover, the finding is also in contrast to 3 other studies, that found that high education was associated with a better treatment outcome in Internet-based intervention studies [23,29,50]. There are several possible explanations for the contrast between the current findings and findings from previous studies: First, these differences can be explained with variances in treatment type (acute vs maintenance phase; disorder-specific vs transdiagnostic), different type of acute phase treatment (outpatient vs inpatient), study population, and design. Second, it could also be hypothesized that inpatients with low education might display a higher risk for deterioration after inpatient discharge than those with high education because of their more pronounced difficulties with transferring the acquired skills into their daily life. Therefore, they might profit to a greater extent from a maintenance intervention than participants with high education. This assumption is in-line with a risk-reduction model of continuation phase treatments [8], assuming that such concepts may effectively reduce an increased risk for relapse or recurrence because of a nonchangeable vulnerability (eg, education, genetic predisposition, developmental conditions) by helping participants to reduce the consequences of such risk factors. Post hoc simple slope analyses revealed that in this study the control group of participants with low education were more likely to deteriorate compared to highly educated participants, whereas in the intervention group no such interaction could be found, indicating that participating in TIMT can effectively reduce this risk factor. Moreover, the inpatient treatments present the rather unique opportunity to introduce patients to the online-based intervention face-to-face and to teach them the necessary skills for using the intervention successfully. Therefore, 1 possible mechanism responsible for the findings in studies in which participants with low education profited less from Internet-based treatments (ie, low Internet skills) no longer has any effect. However, as this study is 1 of the first studies investigating moderators of outcomes in maintenance phase treatments following inpatient psychotherapy, future studies are clearly needed to further clarify the moderating role of education for treatment outcome.

On the basis of our data, we can only speculate on possible explanations as to why participants with anxiety disorder profited to a greater extent (in the short term) than participants with depression. These results are consistent with findings showing that effect sizes are typically larger for Internet interventions targeting anxiety than interventions targeting depression. In a review of 26 RCTs, Griffiths and colleagues [51] found that effect size differences ranged from 0.42 to 0.65 for interventions involving participants with clinically significant symptoms of depression, and 0.29 to 1.74 for interventions involving participants with a diagnosed anxiety disorder. Unlike guidelines for the treatment of depression [52,53], current guidelines for the treatment of anxiety disorders [1,54] do not recommend continuation phase psychological treatments following acute phase psychotherapy. Our findings, however, suggest that participants with anxiety disorder can benefit from an Internet-based maintenance treatment following inpatient psychotherapy. With regard to the subgroup effects for depression, future studies should try to examine treatment strategies to improve TIMT's outcome, especially for this high-risk group [3]

The significant finding for positive outcome expectancies regarding change differences from discharge to 3-month follow-up is consistent with the idea that high expectancies for change are associated with better treatment outcome [45,55]. However, these change differences turned insignificant at 1-year follow-up. Therefore, its use as a predictive indicator for treatment allocation seems limited.

Limitations

To validly interpret the results of this study, several limitations should be considered. First, as in most moderator studies, the analyses in this study were exploratory with participants not being randomized based on potential moderators of interest. Despite the limitations of this procedure, a growing recognition among methodologists has developed about its importance for fostering empirically founded hypotheses to be tested in future studies before clinical application [21]. Second, additional unmeasured variables (eg, participants' genetic markers, developmental histories, self-regulation skills, coping strategies, attribution style, personality traits) may also moderate TIMT's effects, which should be considered in subsequent studies. Third, as in most longitudinal studies, missing values had to be considered a relevant threat to the validity of the analyses. However, the adjustment used to address missing data (FIML) is especially robust with regard to missing data in mixed models [46]. Fourth, TIMT was a multicomponent intervention (ie, personal development plan, Web diary, peer support group, coach support, monitoring). Thus, the extent that the effects of specific components were moderated by studied variables is still unclear. Fifth, the sample size did not provide sufficient power to detect significant findings for potential moderator variables with subgroups of small sizes. Because of this limitation, other diagnoses in addition to mood disorder, adjustment, and anxiety disorders as potential moderating variables could not be included in the conducted analyses. Therefore, no generalization can be made for participants with other diagnoses. Likewise, the sample size did not provide sufficient power to examine moderators separately for different diagnostic status or gender. Therefore, it remains unclear whether moderators of TIMT's effects vary across subpopulation (eg, different moderators for male and female or for different primary diagnoses). Compared with the nonparticipant cohort, individuals with low education were underrepresented in the study sample. Thus, the finding that TIMT was especially effective for participants with a low education level may only be generalizable to low-educated participants who are interested in participating in such an intervention. Finally, the sample included in this study was recruited in only 1 inpatient hospital, which clearly limits the generalizability of the findings to other patient populations.

Strengths

Strengths of the study include (1) its large sample size compared to other studies, (2) a TAU control condition, which allowed us to specify which participants might and might not benefit from TIMT compared to treatment provided by routine health care services, (3) inclusion and exclusion criteria were kept to a minimum to maximize the ecological validity, and (4) generalizability of findings was assessed by comparing the moderator sample with a large sample of participants representing basically all patients treated in the study site.

Conclusions

Transdiagnostic Internet-based guided self-help interventions may represent a cost-effective, far-reaching method for implementing maintenance phase treatments. Findings from the current study suggest that TIMT following inpatient psychotherapy helps patients differing in various characteristics to maintain treatment outcome. It is especially effective for participants with low education levels. Although some subgroups were identified as having profited less from the intervention than others, all subgroups benefited significantly. Future studies should replicate our results before clinical application.

Acknowledgments

David Ebert and Matthias Berking contributed to the design of the study. David Ebert and Mario Gollwitzer contributed to the analysis of data. All authors contributed to the interpretation of data, to the final draft of the paper, and approved the final version for publication. This study was funded by the Dr Ebel Fachkliniken–Vogelsbergklinik, Grebenhain, Germany and the European Union (EFRE: CCI 2007DE161PR001).

Conflicts of Interest

The first author (DE) developed the intervention under study.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [56].

[PDF File (Adobe PDF File), 1MB - jmir_v15i10e191_app1.pdf]

References

1. American Psychiatric Association (APA). Practice Guideline for the Treatment of Patients with Panic Disorder, 2nd ed. Washington, DC: American Psychiatric Association (APA); 2009:1-90.
2. 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]
3. Vittengl JR, Clark LA, Dunn TW, Jarrett RB. Reducing relapse and recurrence in unipolar depression: a comparative meta-analysis of cognitive-behavioral therapy's effects. J Consult Clin Psychol 2007 Jun;75(3):475-488 [FREE Full text] [doi: 10.1037/0022-006X.75.3.475] [Medline: 17563164]
4. Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, et al. Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland. Health Technol Assess 2005 Nov;9(42):1-174 [FREE Full text] [Medline: 16266559]

5. Halmi KA, Agras WS, Mitchell J, Wilson GT, Crow S, Bryson SW, et al. Relapse predictors of patients with bulimia nervosa who achieved abstinence through cognitive behavioral therapy. Arch Gen Psychiatry 2002 Dec;59(12):1105-1109. [Medline: 12470126]
6. Chung T, Maisto SA. Relapse to alcohol and other drug use in treated adolescents: review and reconsideration of relapse as a change point in clinical course. Clin Psychol Rev 2006 Mar;26(2):149-161. [doi: 10.1016/j.cpr.2005.11.004] [Medline: 16364524]
7. Bockting CL, Schene AH, Spinhoven P, Koeter MW, Wouters LF, Huyser J, et al. Preventing relapse/recurrence in recurrent depression with cognitive therapy: a randomized controlled trial. J Consult Clin Psychol 2005 Aug;73(4):647-657. [doi: 10.1037/0022-006X.73.4.647] [Medline: 16173852]
8. Jarrett RB, Vittengl JR, Clark LA. Preventing recurrent depression. In: Whisman MA, editor. Adapting Cognitive Therapy for Depression: Managing Complexity and Comorbidity. New York: Guilford Press; 2008:132-156.
9. McKay D, Todaro JF, Neziroglu F, Yaryura-Tobias J. Evaluation of a naturalistic maintenance program in the treatment of obsessive-compulsive disorder: A preliminary investigation. Journal of Anxiety Disorders 1996 May;10(3):211-217. [doi: 10.1016/0887-6185(96)00006-0]
10. Wilberg T, Karterud S, Pedersen G, Urnes O, Irion T, Brabrand J, et al. Outpatient group psychotherapy following day treatment for patients with personality disorders. J Pers Disord 2003 Dec;17(6):510-521. [Medline: 14744077]
11. Leirvåg H, Pedersen G, Karterud S. Long-term continuation treatment after short-term day treatment of female patients with severe personality disorders: Body awareness group therapy versus psychodynamic group therapy. Nord J Psychiatry 2010 Apr;64(2):115-122. [doi: 10.3109/08039480903487525] [Medline: 20392134]
12. Klaw E, Humphreys K. Facilitating client involvement in self-help groups. In: Koocher GP, Norcross JC, Hill SS, editors. Psychologists' Desk Reference, 2nd ed. New York: Oxford University Press; 2005:502-505.
13. Barnes C, Harvey R, Mitchell P, Smith M, Wilhelm K. Evaluation of an online relapse prevention program for bipolar disorder. Disease Management & Health Outcomes 2007;15(4):215-224. [doi: 10.2165/00115677-200715040-00003]
14. Bauer S, Wolf M, Haug S, Kordy H. The effectiveness of Internet chat groups in relapse prevention after inpatient psychotherapy. Psychother Res 2011 Mar;21(2):219-226. [doi: 10.1080/10503307.2010.547530] [Medline: 21347978]
15. Ebert D, Tarnowski T, Gollwitzer M, Sieland B, Berking M. A transdiagnostic Internet-based maintenance treatment enhances the stability of outcome after inpatient cognitive behavioral therapy: a randomized controlled trial. Psychother Psychosom 2013;82(4):246-256. [doi: 10.1159/000345967] [Medline: 23736751]
16. Golkaramnay V, Bauer S, Haug S, Wolf M, Kordy H. The exploration of the effectiveness of group therapy through an Internet chat as aftercare: a controlled naturalistic study. Psychother Psychosom 2007;76(4):219-225. [doi: 10.1159/000101500] [Medline: 17570960]
17. Gulec H, Moessner M, Mezei A, Kohls E, Túry F, Bauer S. Internet-based maintenance treatment for patients with eating disorders. Professional Psychology: Research and Practice 2011;42(6):479-486. [doi: 10.1037/a0025806]
18. Holländare F, Johnsson S, Randestad M, Tillfors M, Carlbring P, Andersson G, et al. Randomized trial of Internet-based relapse prevention for partially remitted depression. Acta Psychiatr Scand 2011 Oct;124(4):285-294. [doi: 10.1111/j.1600-0447.2011.01698.x] [Medline: 21401534]
19. Kordy H, Theis F, Wolf M. [Modern information and communication technology in medical rehabilitation. Enhanced sustainability through Internet-delivered aftercare]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2011 Apr;54(4):458-464. [doi: 10.1007/s00103-011-1248-7] [Medline: 21465402]
20. Ebert DD, Hannig W, Tarnowski T, Sieland B, Götzky B, Berking M. Web-based rehabilitation aftercare following inpatient psychosomatic treatment. Rehabilitation (Stuttg) 2013 Jun;52(3):164-172. [doi: 10.1055/s-0033-1345191] [Medline: 23761205]
21. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [Medline: 12365874]
22. Warmerdam L, van Straten A, Jongsma J, Twisk J, Cuijpers P. Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. J Behav Ther Exp Psychiatry 2010 Mar;41(1):64-70. [doi: 10.1016/j.jbtep.2009.10.003] [Medline: 19913781]
23. 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]
24. Button KS, Wiles NJ, Lewis G, Peters TJ, Kessler D. Factors associated with differential response to online cognitive behavioural therapy. Soc Psychiatry Psychiatr Epidemiol 2012 May;47(5):827-833. [doi: 10.1007/s00127-011-0389-1] [Medline: 21541696]
25. de Graaf LE, Hollon SD, Huibers MJ. Predicting outcome in computerized cognitive behavioral therapy for depression in primary care: A randomized trial. J Consult Clin Psychol 2010 Apr;78(2):184-189. [doi: 10.1037/a0018324] [Medline: 20350029]
26. Bockting CL, Spinhoven P, Koeter MW, Wouters LF, Visser I, Schene AH, DELTA study group. Differential predictors of response to preventive cognitive therapy in recurrent depression: a 2-year prospective study. Psychother Psychosom 2006;75(4):229-236. [doi: 10.1159/000092893] [Medline: 16785772]

27. Vittengl JR, Clark LA, Jarrett RB. Moderators of continuation phase cognitive therapy's effects on relapse, recurrence, remission, and recovery from depression. Behav Res Ther 2010 Jun;48(6):449-458 [FREE Full text] [doi: 10.1016/j.brat.2010.01.006] [Medline: 20163785]
28. Andersson G, Bergström J, Holländare F, Ekselius L, Carlbring P. Delivering cognitive behavioural therapy for mild to moderate depression via the Internet: predicting outcome at 6-month follow-up. Verhaltenstherapie 2004;14(3):185-189. [doi: 10.1159/000080914]
29. Riper H, Kramer J, Keuken M, Smit F, Schippers G, Cuijpers P. Predicting successful treatment outcome of web-based self-help for problem drinkers: secondary analysis from a randomized controlled trial. J Med Internet Res 2008;10(4):e46 [FREE Full text] [doi: 10.2196/jmir.1102] [Medline: 19033150]
30. Brown TA, Barlow DH. Long-term outcome in cognitive-behavioral treatment of panic disorder: clinical predictors and alternative strategies for assessment. J Consult Clin Psychol 1995 Oct;63(5):754-765. [Medline: 7593868]
31. Katschnig H, Amering M. The long-term course of panic disorder and its predictors. J Clin Psychopharmacol 1998 Dec;18(6 Suppl 2):6S-11S. [Medline: 9872707]
32. Jarrett RB, Kraft D, Doyle J, Foster BM, Eaves GG, Silver PC. Preventing recurrent depression using cognitive therapy with and without a continuation phase: a randomized clinical trial. Arch Gen Psychiatry 2001 Apr;58(4):381-388 [FREE Full text] [Medline: 11296099]
33. Judd LL, Paulus MP, Zeller P. The role of residual subthreshold depressive symptoms in early episode relapse in unipolar major depressive disorder. Arch Gen Psychiatry 1999 Aug;56(8):764-765. [Medline: 10435613]
34. Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, et al. Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. J Affect Disord 1998 Sep;50(2-3):97-108. [Medline: 9858069]
35. ten Doesschate MC, Bockting CL, Koeter MW, Schene AH, DELTA Study Group. Prediction of recurrence in recurrent depression: a 5.5-year prospective study. J Clin Psychiatry 2010 Aug;71(8):984-991. [doi: 10.4088/JCP.08m04858blu] [Medline: 20797379]
36. Bockting CL, Spinhoven P, Koeter MW, Wouters LF, Schene AH, Depression Evaluation Longitudinal Therapy Assessment Study Group. Prediction of recurrence in recurrent depression and the influence of consecutive episodes on vulnerability for depression: a 2-year prospective study. J Clin Psychiatry 2006 May;67(5):747-755. [Medline: 16841624]
37. Dilling H, Mombour W. In: Schmidt MH, editor. Internationale Klassifikation psychischer Störungen, ICD-10. Deutsche Ausgabe. [International classification of mental diseases, ICD-10. German edition]. Bern: Huber; 1991.
38. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods 1997;2(1):64-78. [doi: 10.1037/1082-989X.2.1.64]
39. Barlow DH. Clinical Handbook of Psychological Disorders: A Step-by-Step Treatment Manual, 4th edition. New York: Guilford Press; 2008.
40. Kobelt A, Grosch E, Lamprecht F, Hentschel HJ. In: Schmid-Ott G, editor. Ambulante psychosomatische Nachsorge: Integratives Trainingsprogramm nach stationärer Rehabilitation. [Outpatient psychosomatic maintenance treatment: integrative training program after inpatient rehabilitation]. Stuttgart: Schattauer; 2002.
41. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement. A meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-119. [doi: 10.1016/S0065-2601(06)38002-1]
42. Wittchen HU, Zaudig M, Fydrich T. SKID-I/II Strukturiertes Klinisches Interview für DSM-IV. Deutsche Version. [SKID-I/II Structured Clinical Interview for DSM-IV. German version]. Weinheim: Beltz; 1997.
43. Rabung S, Harfst T, Kawski S, Koch U, Wittchen HU, Schulz H. [Psychometric analysis of a short form of the "Hamburg Modules for the Assessment of Psychosocial Health" (HEALTH-49)]. Z Psychosom Med Psychother 2009;55(2):162-179. [Medline: 19402020]
44. 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]
45. Schulte D. Messung der Therapieerwartung und Therapieevaluationvon Patienten (PATHEV). Zeitschrift für Klinische Psychologie und Psychotherapie 2005 Jul 1;34(3):176-187. [doi: 10.1026/1616-3443.34.3.176]
46. Little RJA. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. Journal of the American Statistical Association 1995 Sep;90(431):1112. [doi: 10.2307/2291350]
47. Preacher KJ, Curran PJ, Bauer DJ. Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis. Journal of Educational and Behavioral Statistics 2006 Jan 01;31(4):437-448. [doi: 10.3102/10769986031004437]
48. Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Mahwah, NJ: L Erlbaum Associates; 2003.
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988:567.
50. Warmerdam EH. Online treatment of adults with depression: Clinical effects, economic evaluation, working mechanisms and predictors. 2010. URL: http://dare.ubvu.vu.nl//handle/1871/16313 [accessed 2012-05-08] [WebCite Cache ID 67VTDho6i]

51. 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]
52. American Psychiatric Association (APA). Practice Guideline for the Treatment of Patients with Major Depressive Disorder, 3rd ed. Arlington, VA: American Psychiatric Association (APA); 2010:1-128.
53. National Collaborating Centre for Mental Health. Depression: The Treatment and Management of Depression in Adults. Clinical Guideline no 90. London, UK: National Institute for Health and Clinical Excellence (NICE); 2009:64.
54. National Collaborating Centre for Mental Health, National Collaborating Centre for Primary Care. Generalised Anxiety Disorder and Panic Disorder (with or Without Agoraphobia) in Adults. Clinical Guideline no 113. London, UK: National Institute for Health and Clinical Excellence (NICE); 2011:56.
55. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [Medline: 11132119]
56. 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

CBT: cognitive behavior therapy
FIML: full information maximum likelihood
ITT: intention-to-treat
RCT: randomized controlled trial
TAU: treatment as usual
TIMT: transdiagnostic Internet-based maintenance treatment

Edited by G Eysenbach; submitted 23.12.12; peer-reviewed by M Wolf, Z Zhuohong; comments to author 18.01.13; revised version received 15.07.13; accepted 29.07.13; published 10.10.13.

Please cite as:
Ebert DD, Gollwitzer M, Riper H, Cuijpers P, Baumeister H, Berking M
For Whom Does It Work? Moderators of Outcome on the Effect of a Transdiagnostic Internet-Based Maintenance Treatment After Inpatient Psychotherapy: Randomized Controlled Trial
J Med Internet Res 2013;15(10):e191
URL: http://www.jmir.org/2013/10/e191/
doi:10.2196/jmir.2511
PMID:24113764

Original Paper

Internet-Based, Culturally Sensitive, Problem-Solving Therapy for Turkish Migrants With Depression: Randomized Controlled Trial

Burçin Ünlü Ince[1,2*], MSc; Pim Cuijpers[1,2,3*], PhD; Edith van 't Hof[1,2,4], PhD; Wouter van Ballegooijen[1,2], PhD; Helen Christensen[5], PhD; Heleen Riper[1,2,3], PhD

[1]Department of Clinical Psychology, Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, Netherlands
[2]EMGO Institute for Health and Care Research (EMGO+), VU University Medical Center, Amsterdam, Netherlands
[3]Division of Online Health Training, Innovation Incubator, Leuphana University, Lueneburg, Germany
[4]GGZ inGeest, Regional Mental Health Service Centre, VU University Medical Centre, Amsterdam, Netherlands
[5]Black Dog Institute, University of New South Wales, Sydney, Australia
[*]these authors contributed equally

Corresponding Author:
Burçin Ünlü Ince, MSc
Department of Clinical Psychology
Faculty of Psychology and Education
VU University Amsterdam
Van der Boechorststraat 1
Amsterdam, 1081 BT
Netherlands
Phone: 31 205983688
Fax: 31 205988758
Email: b.unlu@vu.nl

Abstract

Background: Turkish migrants living in the Netherlands have a high prevalence of depressive disorders, but experience considerable obstacles to accessing professional help. Providing easily accessible Internet treatments may help to overcome these barriers.

Objective: The aim of this study was to evaluate the effectiveness of a culturally sensitive, guided, self-help, problem-solving intervention through the Internet for reducing depressive symptoms in Turkish migrants.

Methods: A two-armed randomized controlled trial was conducted. The primary outcome measure was the severity of depressive symptoms; secondary outcome measures were somatic symptoms, anxiety, quality of life, and satisfaction with the treatment. Participants were assessed online at baseline, posttest (6 weeks after baseline), and 4 months after baseline. Posttest results were analyzed on the intention-to-treat sample. Missing values were estimated by means of multiple imputation. Differences in clinical outcome between groups were analyzed with a *t* test. Cohen's *d* was used to determine the between-groups effect size at posttreatment and follow-up.

Results: Turkish adults (N=96) with depressive symptoms were randomized to the experimental group (n=49) or to a waitlist control group (n=47). High attrition rates were found among the 96 participants of which 42% (40/96) did not complete the posttest (6 weeks) and 62% (59/96) participants did not complete the follow-up assessment at 4 months. No significant difference between the experimental group and the control group was found for depression at posttest. Recovery occurred significantly more often in the experimental group (33%, 16/49) than in the control group (9%, 4/47) at posttest (P=.02). Because of the high attrition rate, a completers-only analysis was conducted at follow-up. The experimental group showed significant improvement in depression compared to the control group both at posttest (P=.01) and follow-up (P=.01).

Conclusions: The results of this study did not show a significant effect on the reduction of depressive symptoms. However, the effect size at posttest was high, which might be an indicator of the possible effectiveness of the intervention when assessed in a larger sample and robust trial. Future research should replicate our study with adequately powered samples.

Trial Registration: Dutch Trial Register: NTR2303. http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2303 (Archived by WebCite at http://www.webcitation.org/6IOxNgoDu).

(J Med Internet Res 2013;15(10):e227) doi:10.2196/jmir.2853

KEYWORDS

depression; randomized controlled trial; ethnic groups; Internet; psychotherapy

Introduction

Depressive disorders are highly prevalent [1,2] and are significantly associated with an impaired quality of life [3,4]. It is estimated that the prevalence of depression varies in different ethnic populations. For example, a European study showed that the prevalence of depressive symptoms among adult ethnic minorities was significantly higher than among native people [5]. Lower socioeconomic conditions and discrimination against ethnic minorities have been found to be important predictors for these differences. Research shows that Turkish people in the Netherlands, one of the largest ethnic minority groups in the country, have the highest 1-month prevalence of depressive and/or anxiety disorders (18.7%) in comparison with Dutch (6.6%) and Moroccan (9.8%) people [6]. Furthermore, it has been found that young women of Turkish and South-Asian descent in the Netherlands are at increased risk for committing suicide. Social oppression is perceived as one of the risk factors contributing to this higher suicidal risk [7].

Despite the fact that ethnic minorities encounter a higher risk for depression compared to the ethnic majority, they seem to receive less professional help from mental health care services than native people in Western countries [8,9]. Several reasons have been found for this lower uptake. For example, people from ethnic minorities seek mental health care at a later and more advanced stage of their mental health problems. They also have a higher chance of dropping out from therapy prematurely [10]. To lower the access threshold, it is important to apply effective recruitment strategies and to provide culturally sensitive interventions for ethnic minorities.

Psychotherapy, such as cognitive behavior therapy [11,12] and problem-solving therapy [13], has found to be effective in the treatment of adult depression, but little is known about whether this effectiveness also holds for ethnic minorities. Data are mostly obtained from studies among white, middle-income populations, leaving ethnic minorities underrepresented in clinical research [14].

However, a recent meta-analysis taking ethnic minorities into consideration showed a first indication that psychotherapy may be equally as effective in ethnic minorities as in native populations [15]. Therefore, this finding would justify strategies for lowering the access threshold to psychotherapy for ethnic minorities with depression. One such way could be the delivery of depression interventions by Internet. Because the Internet can overcome several barriers to treatment uptake, it could help in reaching out to ethnic minorities with unmet needs for treatment. It can lower the access threshold and provide anonymity and considerable flexibility in terms of time and place.

Internet interventions have proved to be effective in the treatment of depressive symptoms and the prevention of depression, as shown in a number of studies [16-18]. However, it is unclear whether this evidence for Internet-based interventions can be generalized to ethnic minority groups. So far, few studies have focused on ethnic minorities in online trials. For example, a recent Australian study showed promising results for an Internet-based cognitive behavior therapy in the treatment of depression in Chinese migrants [19]. The participants in the experimental group evaluated the Internet-based treatment as acceptable and reported significantly reduced depressive symptoms (Cohen's d=0.93) up to 3 months after treatment compared to a control group.

In the Netherlands, one such successful Internet-based, guided, self-help intervention based on problem-solving therapy is Alles Onder Controle (AOC; Everything under Control). AOC has been shown to be clinically effective in the reduction of depressive symptoms with a moderate effect size (Cohen's d=0.50) [20,21]. This intervention appears also to be cost-effective as shown by Warmerdam and colleagues [22]. For the purpose of this study, AOC was adapted to the specific needs of Turkish people living in the Netherlands (AOC-TR) [23]. We investigated the effectiveness of AOC-TR by means of a randomized controlled trial with a similar design as the trial of Warmerdam and colleagues [21]. We hypothesized that Turkish adult migrants in the experimental group would show a significant reduction in depressive complaints compared to those in a waitlist control group. To the best of our knowledge, this is the first study evaluating a culturally sensitive, Internet-based, self-help intervention for Turkish migrants with depressive complaints.

Methods

Trial Design

A two-armed randomized controlled trial was conducted to examine the effectiveness of AOC-TR, comparing the experimental group with a waiting list control group. The experimental group obtained direct access to the intervention and the waiting list control group received access after 4 months. Ethical approval was granted by an independent medical ethics committee (METc VUmc registration no: 2010/82). The trial is registered in the Dutch Trial Register (NTR2303). A detailed description of the trial design has been published earlier [23].

Participants and Inclusion/Exclusion Criteria

Participants aged 18 years or older with depressive symptoms as measured by a Center for Epidemiologic Studies Depression Scale (CES-D) score ≥16 [24] and a Turkish background (participant or at least 1 parent was born in Turkey) were included in the trial. To be included, participants also needed to have access to a computer with Internet, have an email address, and have returned a signed informed consent form.

Exclusion took place if the participant was suicidal, which was assessed in 2 steps as part of the online screening. First, the suicide item on the Beck Depression Inventory II (BDI-II) was presented [25,26]. The BDI-II is validated among Dutch [27,28] and Turkish populations [29,30]. Second, if the response was affirmative, the suicide risk was measured with the suicidality section of the Mini-international Neuropsychiatric Interview (MINI) [31,32] in Dutch [33] or Turkish [34]. Participants with a relatively high risk were advised to contact their general practitioner or were referred to the online portal for suicide prevention [35].

Recruitment

Recruitment took place from June 16, 2010 to March 15, 2012. Participants were recruited among the adult Turkish migrant population via several recruitment strategies. The following strategies were applied: advertisements in Dutch and Turkish national newspapers, magazines, and community websites; banners on health-related websites for migrants; and through social media. Information brochures were distributed at Turkish associations in the Netherlands, mental health care organizations, and sociocultural organizations. The recruitment took place in 2 languages, Dutch and Turkish. Facebook was the most effective recruitment strategy during the trial.

Recruitment on Facebook took place between January 1, 2011 and March 15, 2012. A personal profile and a fan page about the trial were created on Facebook, where pictures from the research website, information about the project, and status updates were shared. Facebook groups related to Turkish migrants and Turkish groups focusing on (general) health and psychology were joined. Next, random people from these groups were invited to join the fan page and friend requests were sent. A total of 584 friend requests were accepted by these invited people. Afterwards, friends of our friends list and people from the Facebook groups began to add our Facebook profile, which resulted in 3308 friends on the research profile by the end of the trial. Friends from our friends list and from the joined Facebook groups sent us messages or, if the researcher was online, chat conversations took place with them. We received or had chat conversations (about diverse topics, including application to the trial) with 348 people.

The advertisements contained a link to our research website with detailed information about the trial [36,37]. Interested parties could apply by sending an email to the researcher, who then returned a digital information brochure about the study, the informed consent form, and a unique Web link for an online screening questionnaire.

Intervention

The original version of the self-guided, problem-solving intervention (AOC [20]) was adapted to a culturally sensitive intervention (AOC-TR) in collaboration with the Trimbos Institute (Netherlands Institute of Mental Health and Addiction). First, the intervention was translated from Dutch into Turkish, and then both versions were culturally adapted. Although there are multiple descriptions of cultural adaptation of psychotherapy to specific populations, it has been defined as the modification of intervention protocols according to the clients' values, contexts, and worldviews [38]. Culture-specific adaptations in our intervention included several components: (1) the participants' preferred language, (2) describing psychological problems in terms of idioms of distress (eg, using symptoms of depression instead of the term depression), (3) explicitly discussing migration and culture by using culture-specific cases and problems that are recognizable for the target group concerned, and (4) including recognizable examples of persons with similar problems (eg, a young woman who migrated 2 years ago and can't find her way in the Netherlands). After adapting the intervention from the original Dutch version, 2 native Turkish persons evaluated the interventions both for language- and cultural-specific items in close collaboration with the first author who is a Turkish person herself. Finally, recommendations from these reviewers in terms of culture and language were incorporated in the interventions. Screenshots of the interventions are shown in Multimedia Appendices 1 and 2.

The AOC-TR consists of 5 sessions over 5 weeks. During the intervention, participants indicate what they think is important in their lives, they make a list of their problems and worries, and they categorize their problems into 3 groups: (1) unimportant problems, which are not related to what they think is important in their lives, (2) important and solvable problems, which are approached by a systematic problem-solving approach consisting of 6 steps, and (3) important but unsolvable problems, such as having lost someone through death or having a chronic general medical disease and making a plan for how to live with it. The core of the intervention is the 6-step problem-solving procedure, which teaches to use this technique during the course for several of their important and solvable problems. The idea is that by mastering this technique people will regain mastery of their problems and ultimately their lives.

The participants received feedback on their homework assignments in brief weekly emails in either Turkish or Dutch from the researcher (BÜI).

Control Condition

The control condition was a waiting list comparator; participants in this condition did not receive access to the intervention after randomization. However, they were provided with access to the intervention 4 months after the baseline measures.

Outcome Measures

Overview

Assessments took place before randomization (T0), after completing the treatment (8 weeks, T1), and 4 months after baseline (T2). All assessments were offered in the preferred language of the participant, either Dutch or Turkish.

Primary Outcome Measure: Depressive Symptoms

Depression severity was measured with the CES-D [24] including 20 self-rated items, each scored from 0 to 3. The Dutch [39], Turkish [40], and online [41] versions of the CES-D have been proven to have good psychometric properties in terms of validity and reliability. In the current study, the internal consistency was good (Cronbach alpha =.87 at baseline).

Secondary Outcome Measures

Anxiety

The anxiety scale of the Hospital Anxiety and Depression Scale (HADS) was used to measure symptoms of anxiety [42]. The HADS consists of an anxiety scale with a total of 7 items. Each item is scored on a 4-point Likert scale within a range of 0 to 3 (low to high). The HADS has proven to be a valid and reliable instrument in various normal and clinical Dutch [43] and Turkish samples [44]. The Cronbach alpha coefficient was .78 at baseline in the current study.

Somatic Symptoms

To measure somatic symptoms, the somatization subscale on the Symptom Checklist-90-Revised (SCL-90-R) was used [45]. This is a 5-point rating scale containing 12 items. Dutch [46] and Turkish translations [47] were used for this study, both having good reliability and validity. In the current study, the Cronbach alpha coefficient was .86 at baseline.

Quality of Life

Quality of life was measured using the EuroQol Questionnaire (EQ-5D) [48,49] in the official Dutch and Turkish translations, both of which have been validated [50,51]. The last item on the EQ-5D, the EQ visual analogue scale (EQ-VAS), was used in which the health state of the participant is measured by a thermometer-like scale from 0 (worst) to 100 (best health state).

Satisfaction With the Treatment

Participants were asked to define their satisfaction with each lesson by asking, “Was this lesson useful to you?” in Dutch and Turkish. The answers could be rated on a 5-point Likert scale. The score per item ranged from 1 (not at all) to 5 (very much). The Cronbach alpha coefficient was .90 at T1.

Additional Measures

Sociodemographic information (sex, age, country of birth of participant and participant’s parents, educational level, employment, and long-term relationship or partner status) and additional information were collected about how the participants were referred to the trial, why they chose an Internet-based intervention, and whether they use the Internet for health-related topics.

Sample Size

The sample size was calculated on an expected difference of d=0.45 between the experimental and control groups. This expected difference was based on effect sizes derived from previous effect studies on Internet-guided problem-solving therapy for depression [21]. To achieve a power of 0.80 and an alpha of .05, we needed 78 participants at baseline in each condition (N=156). In keeping with our hypothesis, the primary and secondary outcomes were analyzed with a 1-tailed t test as in the study of Warmerdam and colleagues [21].

Randomization

Participants were randomly assigned to the experimental or the control group after baseline assessment. The allocation schedule was generated by an independent researcher using a computerized system.

Analyses

Overview

The study was carried out in accordance with the CONSORT guidelines [52]. Differences in demographic characteristics were computed with a chi-square test. For small samples, the likelihood ratio test was performed. Clinical outcomes, differences in baseline, posttest, and follow-up mean scores (at T0, T1, and T2) were analyzed with a t test.

Missing Values

Only posttreatment data were analyzed according to the intention-to-treat principle. Missing values were handled using the multiple imputation technique in SPSS Statistics version 20.0 (IBM Corp, Armonk, NY, USA). All variables (except nominal variables) were included as predictors and generated 100 imputations. Analyses were performed using pooled data.

Effect Sizes

For comparison of the 2 means, Cohen’s d was used to determine the between-group effect size at posttreatment and follow-up [53]. Cohen’s d was calculated as the difference between the posttest mean scores of the intervention and the control group divided by the pooled standard deviation. Effect sizes of 0.8 are assumed to be large, effect sizes of 0.5 are moderate, and effect sizes of 0.2 are assumed to be small [53].

Clinically Significant Change

Analyses of clinically significant change on the CES-D were conducted according to the Jacobson and Truax formula [54]. This method evaluates 2 criteria for each participant. The first is whether each participant’s CES-D score improved such that it is unlikely to be due to chance (reliable change index, RCI). The RCI is a function of a participant’s pretest and posttest scores, the standard deviation of the population before treatment, and the test-retest reliability of the measure [54,55]. A participant is considered to have experienced reliable change if his or her RCI is greater than 1.96 [56]. The second criterion evaluated for participants shown to have reliable change is whether their posttreatment symptom level places them at a score of 16 or lower on the CES-D. Clinically significant change was determined if the participant had recovered and shown reliable improvement over time.

Per-Protocol Analysis

Per-protocol analyses were performed for participants who completed all the measurements and all 5 lessons of the course (if randomized to the experimental condition).

Results

Participants

Figure 1 shows the flow of participants through the trial. A total of 287 individuals applied for participation. However, 66 of them did not complete the screening. The screening questionnaire was filled in by 221 individuals, of whom 125 were excluded primarily because of suicidal ideations (64/125, 51.2%). A total of 96 individuals met all inclusion criteria and were randomized to 1 of the 2 conditions.

Table 1 provides the baseline characteristics of study participants. The mean age of the participants was 35.2 years (SD 9.3) and 62% (59/96) were women. Most participants were born in Turkey (91%, 87/96) and preferred the Turkish language for study participation (89%, 85/96). More than three-quarters of the participants (78%, 75/96) were recruited through the Internet. The most important reason for choosing an Internet intervention was flexibility of use (62%, 59/96), followed by privacy and anonymity (23%, 22/96).

The mean score at baseline for all the participants on the CES-D was 29.9 (SD 9.6, range 11-52). There were no statistically significant differences between the experimental and control group at baseline on any of the demographic and secondary outcomes.

Attrition

Of the 96 original participants, a total of 40 (42%) participants did not complete the posttest (6 weeks), and 59 participants (62%) did not complete the follow-up assessment at 4 months. Reasons for the high attrition rates are not known. There were no significant differences in attrition rates between the experimental (47%, 23/49) and the control group (36%, 17/47) at posttest (*P*=.29). However, at follow-up, the experimental group (74%, 36/49) had a higher attrition rate than the control group (49%, 23/47; *P*=.01).

Effects of the Intervention at Posttest

Intention-to-Treat Analysis

Table 2 shows the outcomes for the primary (CES-D) and secondary (HADS, SCL-90, and EQ-5D) measures at posttreatment. The results show no difference between the experimental and the control group at posttest for the primary outcome assessed with the CES-D (*P*=.07; Cohen’s *d*=0.37, 95% CI –0.03 to 0.78). We did not find any significant differences between the 2 groups on the secondary outcomes.

Clinically Significant Change

Data on clinically significant change are shown in Table 3. In the intention-to-treat sample, the experimental group (32.9%) had significantly higher recovery rates on the CES-D than the control group (9.4%, *P*=.02) at posttest. However, no differences between the experimental and control group were found for improvement or clinically significant change.

Per-Protocol Analysis

The outcomes for participants who fulfilled the protocol for intervention and outcome assessments are shown in Table 4. Several significant outcomes at posttest assessments can be observed. At posttest, the experimental group showed a significantly greater improvement in depressive symptoms compared to the control group (*P*<.001) with a large effect size of *d*=1.68 (95% CI 0.69-2.67). Differences were also found in favor of the experimental group for reduction of anxiety symptoms (*P*<.001), with a large effect size of *d*=1.48 (95% CI 0.51-2.45) and also in somatization symptoms (*P*<.001), with a large effect size of *d*=1.37 (95% CI 0.41-2.33) compared to the control group.

Completers-Only Analysis

Table 5 shows the outcomes for responding participants at posttest assessments in comparison with the control condition. Results show a significantly greater improvement in depressive symptoms in the experimental group than the control group at posttest (*P*<.001), with a large effect size of *d*=0.72 (95% CI 0.18-1.26). We did not find any significant differences between the 2 groups on the secondary outcomes at posttest.

Sessions Attended and Satisfaction With Treatment

A total of 18 of 49 (37%) participants who were assigned to the experimental group did not start the treatment. Of those who started, 12 of 49 (26%) participants completed 1 to 2 lessons, 9 of 49 (18%) participants completed 3 or 4 lessons, and 10 of 49 (20%) participants completed all 5 lessons. Participants who completed the treatment expressed moderate satisfaction (total score mean 2.75, SD 0.96) with the intervention.

Effects of the Intervention at Follow-Up: Based on Completers-Only Sample

Because of high attrition, we conducted completers-only analysis for follow-up results at 4 months after the start of the intervention. Results are shown in Tables 3-5. As is shown, the experimental group did significantly better on the primary and secondary outcomes analyses (clinically significant change, per-protocol analysis, and completers-only analysis).

Table 1. Demographic characteristics and baseline test scores at T0 (N=96).

Demographic characteristics and baseline tests	Total (N=96)	Experimental group (n=49)	Control group (n=47)	*P* value[a]
Age (years), mean (SD)	35.2 (9.3)	34.9 (8.9)	35.6 (9.8)	.72
Gender (female), % (n)	62 (59)	65 (32)	57 (27)	.43
Born in Turkey, % (n)	91 (87)	92 (45)	89 (42)	.68
Long-term relationship, % (n)	64 (61)	71 (35)	55 (26)	.10
Educational level, % (n) [b]				
Low	27 (26)	35 (17)	19 (9)	
Middle	41 (39)	31 (15)	51 (24)	
High	32 (31)	35 (17)	30 (14)	.09
Preference for Turkish language, % (n)	89 (85)	88 (43)	89 (42)	.81
Recruitment channel, % (n)				
Internet	78 (75)	78 (38)	79 (37)	
Internet through Facebook	99 (74)	100 (38)	97 (36)	
Newspaper	1 (1)	0 (0)	2 (1)	
Magazine	1 (1)	2 (1)	0 (0)	
Friends or family	6 (6)	10 (5)	2 (1)	
Other	14 (13)	10 (5)	17 (8)	.89
Employed, % (n)	52 (49)	45 (22)	57 (27)	.22
Reason for choosing Internet intervention, % (n)				
Privacy/anonymity	23 (22)	22 (11)	23 (11)	
Flexibility	62 (59)	51 (25)	72 (34)	
Other	16 (15)	27 (13)	4 (2)	.01
Use of the Internet for health information, % (n)				
Physical complaints	7 (7)	8 (4)	6 (3)	
Psychological complaints	10 (10)	8 (4)	13 (6)	
Physical and psychological complaints	67 (64)	71 (35)	62 (29)	
None	16 (15)	12 (6)	19 (9)	.65
Test outcomes, mean (SD)				
Depression (CES-D)	29.9 (9.6)	29.6 (9.2)	30.1 (10.1)	.79
Anxiety (HADS)	13.0 (4.1)	13.3 (4.3)	12.7 (3.9)	.52
Somatization (SCL-90)	30.0 (8.6)	31.0 (9.2)	29.0 (8.0)	.25
Quality of life (EQ-VAS)[b]	60.4 (21.2)	57.7 (21.5)	63.14 (20.8)	.27

[a] CES-D was analyzed with a 1-tailed *t* test. The other tests were analyzed with a 2-tailed *t* test.

[b] Total: N=74; experimental group: n=38; control group: n=36. This item was the last of the assessment, which was not filled in by every participant.

Figure 1. CONSORT flow diagram of the participation progress through the trial.

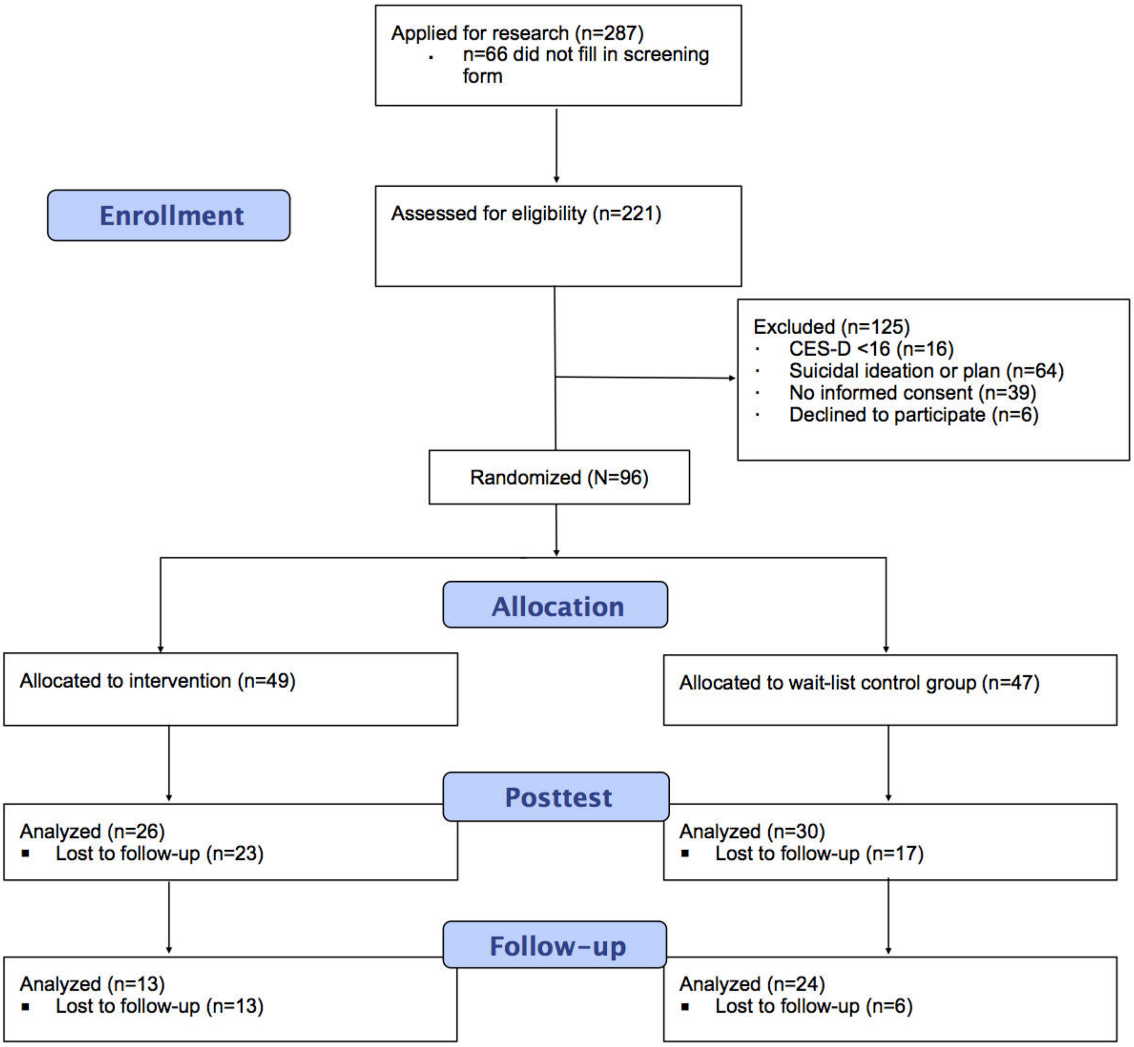


Table 2. Study outcomes at posttest including posttest between-group effect size (Cohen’s *d*): intention-to-treat analysis.

Outcome and group	Pretest		Posttest		*P* value	Mean difference	Cohen’s *d* (95% CI)
	n	Mean	n	Mean			
Primary outcome							
Depression (CES-D)							
Experimental group	49	29.6	49	23.0			
Control group	47	30.1	47	27.2	.07	–4.25	0.37 (–0.03, 0.78)
Secondary outcomes							
Anxiety (HADS)							
Experimental group	49	13.3	49	11.0			
Control group	47	12.7	47	11.7	.23	–0.76	0.25 (–0.16, 0.65)
Somatization (SCL-90)							
Experimental group	49	31.0	49	28.1			
Control group	47	29.0	47	28.0	.48	0.13	0.15 (–0.26, 0.55)
Quality of life (EQ-VAS)							
Experimental group	38	57.7	38	65.4			
Control group	36	63.1	36	65.7	.48	–0.28	0.15 (–0.26, 0.55)

Table 3. Clinically significant change analyses of depression tested by the CES-D.

Clinically significant change analyses	Posttest, n (%)		*P* value	Follow-up, n (%)		*P* value
	Experimental group	Control group		Experimental group	Control group	
Intention-to-treat analysis						
Recovery	15.6 (32.9)	4.3 (9.4)	.02			
Improvement	18.5 (37.8)	9.4 (20.0)	.07[b]			
Clinically significant change	11.8 (24.9)	2.6 (5.7)	.05			
Completers only						
Recovery	10 (38.5)	1 (3.3)	<.001	6 (46.2)	3 (12.5)	.01
Improvement	12 (46.2)	4 (13.3)	.01	7 (53.8)	4 (16.7)	.01
Clinically significant change	6 (23.1)	1 (3.3)	.01	5 (38.5)	2 (8.3)	.01

[a] Recovery was defined as having a CES-D score below 16. Improvement was defined as having a reliable change if the individual RCI is greater than 1.96. Clinically significant change was determined if both recovery and improvement took place.

[b] For this analysis, the *P* value of the chi-square analysis is provided.

Table 4. Study outcomes of participants at posttest and follow-up including between-group effect size (Cohen’s *d*): per protocol (n=30).

Per protocol	Posttest				4-month follow-up			
	n	Mean (SD)	*P*	Cohen’s *d* (95% CI)	n	Mean (SD)	*P*	Cohen’s *d* (95% CI)
Primary outcome								
Depression (CES-D)								
Experimental group	6	15.3 (9.9)			6	19.0 (13.9)		
Control group	24	29.5 (8.8)	<.001	1.68 (0.69, 2.67)	24	30.1 (11.3)	.02	1.13 (0.19, 2.07)
Secondary outcomes								
Anxiety (HADS)								
Experimental group	6	7.3 (2.9)			6	7.8 (4.9)		
Control group	24	12.4 (3.5)	<.001	1.48 (0.51, 2.45)	24	12.1 (3.7)	.01	1.26 (0.31, 2.21)
Somatization (SCL-90)								
Experimental group	6	18.8 (6.2)			6	19.2 (6.9)		
Control group	24	28.5 (8.1)	.001	0.37 (0.41, 2.33)	24	28.5 (8.5)	.01	1.27 (0.32, 2.22)
Quality of life (EQ-VAS)								
Experimental group	5	79.4 (28.9)			5	82.3 (23.0)		
Control group	20	63.6 (21.7)	.07	0.95 (–0.06, 1.97)	20	66.4 (23.2)	.11	0.83 (–0.18, 1.84)

Table 5. Study outcomes of participants at posttest and follow-up including follow-up between-group effect size (Cohen’s *d*): completers only (n=56).

Completers only	Posttest				4-month follow-up			
	n	Mean (SD)	*P*	Cohen’s *d* (95% CI)	n	Mean (SD)	*P*	Cohen’s *d* (95% CI)
Primary outcome								
Depression (CES-D)								
Experimental group	26	21.38 (10.5)			13	21.23 (10.79)		
Control group	30	28.27 (8.71)	.01	0.72 (0.17, 1.26)	24	30.08 (11.27)	.01	0.94 (0.23, 1.65)
Secondary outcomes								
Anxiety (HADS)								
Experimental group	26	10.54 (4.00)			13	9.69 (4.92)		
Control group	30	11.87 (3.76)	.10	0.45 (–0.08, 0.98)	24	12.08 (3.67)	.05	0.69 (0.00, 1.39)
Somatization (SCL-90)								
Experimental group	26	26.00 (10.02)			13	25.31 (9.78)		
Control group	30	27.70 (7.76)	.24	0.32 (–0.21, 0.85)	24	28.54 (8.52)	.15	0.51 (–0.18, 1.19)
Quality of life (EQ-VAS)								
Experimental group	19	70.95 (19.52)			10	78.80 (22.50)		
Control group	23	64.17 (21.54)	.15	0.46 (–0.16, 1.07)	19	66.42 (23.19)	.13	0.61 (–0.17, 1.39)

Discussion

Principal Results

Our results show no significant difference in improvement of depressive complaints in the experimental group compared to the control group on intention-to-treat analysis at posttest. This may be because our study was underpowered [57] (see also the limitations section). However, we found an effect size for the primary outcome (depression) of *d*=1.68 at posttest and *d*=1.13 at follow-up, which is an indication that the intervention could be effective with a sample size indicative of sufficient power. We did not find any differences on secondary outcomes. Recovery occurred significantly more often in the experimental group than in the control group at posttest (*P*=.01) but clinically significant change was not (*P*=.09). Completers-only analyses showed that the results of the analysis for the primary outcome differed from the imputation data, suggesting that the analysis was influenced by data imputation.

Comparison With Previous Work

Our results did not support the effectiveness of the Internet-based, guided, self-help intervention, in contrast to the original version (AOC) [20]. As mentioned before, AOC was previously shown to be clinically effective in the general Dutch population in the reduction of depressive symptoms with a moderate effect size (Cohen's d=0.50). However, the effect size at posttest in the current study was high, which might be an indicator of the possible effectiveness of the intervention for Turkish migrants when assessed in a larger sample.

The inclusion of ethnic minorities in clinical research has been a challenge for many years. Ethnic minorities are generally underrepresented in scientific and clinical analyses and are known to be a hard-to-reach population for research purposes [8,58]. Although previous research shows that ethnic minorities may have participated in Internet interventions (eg, [59]), randomized controlled trials on the effectiveness of psychotherapy for common mental disorders, such as depression, are still sparse. Ethnic minorities are underrepresented in mental health research and literature about ethnic differences in this field is very small [8]. In our recent meta-analysis, we found only 56 randomized controlled trials on the psychological treatment of depression in adults reporting the proportion of participants from ethnic minorities, of which none of them made distinct comparisons between ethnic populations [15].

It is generally believed that ethnic minorities are less willing to participate in clinical research; however, very small differences between ethnic minorities and majorities are found in the willingness to participate in health research [60]. Other factors, such as higher costs associated with the recruitment, the exclusion criterion of insufficient ability to speak the second (native) language, the shortage of ethnic minority coordinators in trials, and stereotypes and myths, are considered to be important barriers for their participation [61].

Our study shows that recruitment of ethnic minorities is possible when appropriate recruitment strategies are applied. For example, almost 80% of our participants were recruited through the Internet (primarily on Facebook). Traditional media, such as advertisements in newspapers or banners on websites, appear not to be successful recruitment strategies for this target group, although many studies have applied this strategy successfully for recruiting participants for randomized controlled trials and studies in routine practice among the general population [21,62]. Flexibility and privacy of the Internet were the main reasons for respondents to agree to participate. The use of social media in research is a relatively new development, and may potentially prove more effective for recruiting ethnic minorities in research trials. The contact through social media and the visibility of the researcher seemed to lower the threshold for participation in research and for help seeking. Although we did not find significant results from the intervention, the current trial shows that the Internet (1) is an effective way to reach hard-to-reach populations, (2) lowers the threshold to get in contact with a professional, (3) can be an effective recruitment strategy for clinical trials, and (4) is potentially an effective way to deliver cognitive behavioral therapy for ethnic minorities.

Furthermore, our participants consisted primarily of first-generation migrants who had a preference for the Turkish (native) language. Offering the intervention and assessment measures in 2 languages may have been another successful strategy to lower the threshold for study participation. Generally, participants are only included in intervention studies when they can read and speak the language of the country they live in [61].

Another argument for the low-access threshold of our intervention might be found in the large number of applicants with suicidal ideation (30%, 64/221). We had to exclude these applicants (51.2% of excluded group) because they are a high-risk group not suited to our guided self-help intervention. In keeping with the protocol, we referred these individuals to their primary care physician or to the online portal for suicide prevention [35]. These applicants were primarily women (59.4%) with a mean age of 33.5 years (range 18-53).This is a rather high number when compared with those excluded because of suicidal ideation among primarily native population studies [63,64]. The number also appears high when compared to prevalence rates at the population level. For example, in the Netherlands, 8.3% of the Dutch population have ever had suicidal ideation in their lifetime and 2.2% have attempted suicide [65]. As mentioned in the introduction, young Turkish women are at increased risk of committing suicide [7]. It seems that our study has reached a large number of this group. Future Internet research could focus more specifically on this high-risk group. These individuals experience a high burden of disease and unmet needs, but appear to be reachable by Internet. Thus, the delivery of psychotherapy through the Internet appears to be a promising way to target hard-to-reach ethnic minority groups.

Limitations

This study has several limitations. First, the attrition ratio was high at posttest. We compensated for this high attrition by means of multiple imputation. Attrition was even higher at follow-up; therefore, we decided not to apply multiple imputation for this time point. Instead, we conducted completers-only analyses. High attrition remains a common problem in Internet interventions, with rates of up to 50% [66,67]. Analyses showed no differences in study attrition rates at posttest between the experimental group and control group. However, at follow-up, the experimental group had a higher attrition rate than the control group. Reminders in the form of emails were sent (maximum 5 times per assessment), but this did not result in a low attrition rate for study dropout. Reasons for this high attrition rate are not known; as a result, we can only guess why this happened. It is possible that participants in the experimental group stopped with the trial after finishing the intervention. Filling in the posttreatment questionnaires might not have been regarded as an obligatory part of the trial. Furthermore, the control group was waiting to receive access to the intervention, they could have perceived filling in the questionnaires as an obligation to partake in the intervention.

Second, although we reached a relatively large number of Turkish migrants, recruitment and inclusion were challenging and complex. One of the challenges was to find an appropriate recruitment strategy. In the end, we did find one (ie, Facebook), but we had only limited time left for recruitment because of the overall period available for this study. This could be one of the main reasons for not having been able to obtain the required sample size. Another problem was that we had to exclude most of the eligible participants for reasons such as high suicidality risk. Thus, we did not achieve the target sample size (N=200) during the study period, which may have resulted in an underpowered study. In turn, this may have been the reason that significant effects were not detected and it limited the generalizability of our results.

Third, our target population focused on the online population because participants were required to have access to the Internet and an email address to be included in our trial. Moreover, it is known that almost 80% of the Turkish population in the Netherlands has Internet access [68]. However, our population may have differed from most Turkish people in another way. When we look at the demographic characteristics of our sample, we notice that younger women (mean 35.2 years) with a middle to higher educational level (70%) took part. This is a higher proportion than the Turkish population in the Netherlands, of which 30.1% had at least a middle educational level [69]. Our sample conforms to the sample characteristics of nonmigrant populations in similar trials, in which women (aged 35-55 years) with higher educational levels have taken part [20,21].

Furthermore, although participants were required to have access to a computer with Internet and have an email address, we did not assess the reading and Internet comfort level of the participants. Given that most of our respondents were recruited through Facebook, we assume that respondents at least were able to understand our intervention and questionnaires. Next, respondents could choose the language in which they wanted to follow the intervention and answer the questionnaires (Turkish or Dutch).

Finally, we used only self-report assessments to measure the severity of depressive symptoms in participants. We used self-report on purpose because we wanted to keep the access barrier for study participation as low as possible. Diagnostic interviews are an extra burden for participants and it is not yet possible to conduct them through the Internet. Therefore, our study lacks a diagnosis of depression in the study participants. Research has shown that online self-report questionnaires have good validity (eg, [41]) and yield scores equivalent to paper-and-pencil questionnaires (eg, [70]).

Future Research and Implications

The results of this study have promising implications for the clinical field. Our study is one of the first to assess the clinical effectiveness of guided self-help interventions by Internet for Turkish migrants with depressive complaints. In addition, the guided self-help intervention for Turkish migrants in the Netherlands could also be suitable for Turkish populations in other EU countries or in Turkey itself, where guided self-help is not yet common practice. It may be a welcome intervention both for clinicians and for minorities because there is a lack of evidence-based culturally sensitive psychotherapy for ethnic minorities and there is a high threshold to these services.

Future research should replicate our findings with adequately powered samples for posttest and follow-up measurement to assess the clinical effectiveness in a robust manner. Future research should focus on monitoring participants who drop out prematurely from the study at follow-ups to evaluate the reasons for withdrawal. It is also important to evaluate the impact of culturally sensitive components in Internet interventions for ethnic minority populations with depression.

Conclusions

The results of this study did not show a significant effect on the reduction of depressive symptoms. However, the effect size at posttest was high, which might be an indicator of the possible effectiveness of the intervention when assessed in a larger sample and robust trial. Future research should replicate our study with adequately powered samples.

Acknowledgments

This study is supported by the VU University and the Trimbos-institute, the Netherlands Institute of Mental Health and Addiction. BÜI, PC, WvB, and HR contributed to the design of this study. BÜI adapted the intervention for the Turkish population and drafted the manuscript. All authors contributed to the further writing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the Turkish version of AOC-TR.

[JPG File, 122KB - jmir_v15i10e227_app1.JPG]

Multimedia Appendix 2

Screenshot of the Dutch version of AOC-TR.

[JPG File, 164KB - jmir_v15i10e227_app2.JPG]

Multimedia Appendix 3

CONSORT-EHEALTH checklist V1.6.2 [71].

[PDF File (Adobe PDF File), 980KB - jmir_v15i10e227_app3.pdf]

References

1. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
2. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004(420):38-46. [doi: 10.1111/j.1600-0047.2004.00329.x] [Medline: 15128386]
3. Ustün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJ. Global burden of depressive disorders in the year 2000. Br J Psychiatry 2004 May;184:386-392 [FREE Full text] [Medline: 15123501]
4. Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, et al. Impact of psychiatric disorders on health-related quality of life: general population survey. Br J Psychiatry 2007 Apr;190:326-332 [FREE Full text] [doi: 10.1192/bjp.bp.106.025106] [Medline: 17401039]
5. Missinne S, Bracke P. Depressive symptoms among immigrants and ethnic minorities: a population based study in 23 European countries. Soc Psychiatry Psychiatr Epidemiol 2012 Jan;47(1):97-109. [doi: 10.1007/s00127-010-0321-0] [Medline: 21110000]
6. de Wit MA, Tuinebreijer WC, Dekker J, Beekman AJ, Gorissen WH, Schrier AC, et al. Depressive and anxiety disorders in different ethnic groups: a population based study among native Dutch, and Turkish, Moroccan and Surinamese migrants in Amsterdam. Soc Psychiatry Psychiatr Epidemiol 2008 Nov;43(11):905-912. [doi: 10.1007/s00127-008-0382-5] [Medline: 18587679]
7. van Bergen D, Smit JH, van Balkom AJ, Saharso S. Suicidal behaviour of young immigrant women in the Netherlands. Can we use Durkheim's concept of'fatalistic suicide' to explain their high incidence of attempted suicide? Ethnic and Racial Studies 2009 Feb;32(2):302-322. [doi: 10.1080/01419870802315043]
8. Office of the Surgeon General (US), Center for Mental Health Services (US), National Institute of Mental Health (US). Mental health: Culture, race, and ethnicity: A supplement to mental health: A report of the Surgeon General. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2001. URL: http://www.ncbi.nlm.nih.gov/books/NBK44243/ [accessed 2013-10-07] [WebCite Cache ID 6KC2IJZP9]
9. Leong FT, Kalibatseva Z. Cross-cultural barriers to mental health services in the United States. Cerebrum 2011 Mar;2011:5 [FREE Full text] [Medline: 23447774]
10. Trujillo M. Multicultural aspects of mental health. Primary Psychiatry 2008;15(4):65-84.
11. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
12. Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, et al. A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. Health Technol Assess 2001;5(35):1-173 [FREE Full text] [Medline: 12387733]
13. Cuijpers P, van Straten A, Warmerdam L. Problem solving therapies for depression: a meta-analysis. Eur Psychiatry 2007 Jan;22(1):9-15. [doi: 10.1016/j.eurpsy.2006.11.001] [Medline: 17194572]
14. Miranda J, Nakamura R, Bernal G. Including ethnic minorities in mental health intervention research: a practical approach to a long-standing problem. Cult Med Psychiatry 2003 Dec;27(4):467-486. [Medline: 14727681]
15. Ünlü Ince B, Riper H, Van ‘t Hof E, Cuijpers P. The effects of psychotherapy for depression in ethnic minorities: a metaregression analysis. Psychiatric Services 2013:- (forthcoming).
16. van't Hof E, Cuijpers P, Waheed W, Stein DJ. Psychological treatments for depression and anxiety disorders in low- and middle- income countries: a meta-analysis. Afr J Psychiatry (Johannesbg) 2011 Jul;14(3):200-207 [FREE Full text] [doi: 10.4314/ajpsy.v14i3.2] [Medline: 21863204]
17. 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]
18. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
19. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]

20. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
21. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008;10(4):e44 [FREE Full text] [doi: 10.2196/jmir.1094] [Medline: 19033149]
22. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
23. Unlü B, Riper H, van Straten A, Cuijpers P. Guided self-help on the Internet for Turkish migrants with depression: the design of a randomized controlled trial. Trials 2010;11:101 [FREE Full text] [doi: 10.1186/1745-6215-11-101] [Medline: 21047442]
24. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
25. Beck AT, Steer RA. Manual for the Beck Depression Inventory. San Antonio, TX: Psychological Corporation New York; 1987.
26. Beck AT, Steer RA, Brown GN. Beck Depression Inventory–II (BDI–II). San Antonio, TX: Psychological Corp; 1996.
27. Zitman FG, Griez EJL. Standaardisering depressievragenlijsten. Tijdschrift voor Psychiatrie 1989;31:114-125.
28. Van der Does AJW. BDI-II-NL. Handleiding. De Nederlandse versie van de Beck Depression Tweede Editie. Lisse: Harcourt Test Publishers; 2002.
29. Hisli N. Beck depresyon envanterinin geçerliliği üzerine bir çalışma. Psikoloji Dergisi 1988;22:118-126.
30. Uslu RI, Kapci EG, Oncu B, Ugurlu M, Turkcapar H. Psychometric properties and cut-off scores of the Beck Depression Inventory-II in Turkish adolescents. J Clin Psychol Med Settings 2008 Sep;15(3):225-233. [doi: 10.1007/s10880-008-9122-y] [Medline: 19104967]
31. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
32. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 1997;12(5):224-231. [doi: 10.1016/S0924-9338(97)83296-8]
33. van Vliet IM, de Beurs E. [The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders]. Tijdschr Psychiatr 2007;49(6):393-397 [FREE Full text] [Medline: 17614093]
34. Engeler A. MINI Araçlari Türkçe Uyarlama. Turkey: GlaxoSmithKline; 2004.
35. Stichting 113Online: zelfmoordpreventie. URL: http://www.113online.nl/ [accessed 2013-07-26] [WebCite Cache ID 6IOYV4c2C]
36. Alles Onder Controle TR. URL: http://www.allesondercontroletr.net [accessed 2013-07-26] [WebCite Cache ID 6IOaDsz6j]
37. Her Şey Kontrol Altında. 2011 URL: http://www.herseykontrolaltinda.net [accessed 2013-07-26] [WebCite Cache ID 6IOaWw88g]
38. Bernal G, Jiménez-Chafey MI, Domenech Rodríguez MM. Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. Professional Psychology: Research and Practice 2009;40(4):361-368. [doi: 10.1037/a0016401]
39. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. Psychol Med 1997 Jan;27(1):231-235. [Medline: 9122304]
40. Spijker J, van der Wurff FB, Poort EC, Smits CH, Verhoeff AP, Beekman AT. Depression in first generation labour migrants in Western Europe: the utility of the Center for Epidemiologic Studies Depression Scale (CES-D). Int J Geriatr Psychiatry 2004 Jun;19(6):538-544. [doi: 10.1002/gps.1122] [Medline: 15211532]
41. Donker T, van Straten A, Marks I, Cuijpers P. A brief Web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res 2009;11(3):e19 [FREE Full text] [doi: 10.2196/jmir.1134] [Medline: 19632977]
42. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
43. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med 1997 Mar;27(2):363-370. [Medline: 9089829]
44. Aydemir Ö, Güvenir T, Kuey L, Kültür S. Validity and reliability of Turkish version of Hospital Anxiety and Depression Scale. Türk Psikiyatri Dergisi 1997;8:280-287.

45. Derogatis LR. SCL-90-R Administration, scoring and procedures Manual-ll (2nd ed). Baltimore: Clinical Psychometric Research; 1983.
46. Arrindell WA, Ettema JHM. Symptom Checklist. Handleiding bij een multidimensionele psychopathologie-indicator. Lisse: Swets & Zeitlinger; 2003.
47. Dağ I. Belirti Tarama Listesi (Scl-90-R)’nin Üniversite Öğrencileri için güvenirliği ve geçerliği [Reliability and validity of the Symptom Check List (SCL-90-R) for university students]. Türk Psikiyatri Dergisi 1991;2:5-12.
48. Brooks R. EuroQol: the current state of play. Health Policy 1996 Jul;37(1):53-72. [Medline: 10158943]
49. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]
50. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd 2005 Jul 9;149(28):1574-1578. [Medline: 16038162]
51. Eser E, Dinç G, Cambaz S. EURO-QoL (EQ-5D) indeksinin toplum standartları ve psikometrik özellikleri: Manisa kent toplumu örneklemi. 2007 Presented at: 2 Sağlıkta Yaşam Kalitesi Kongresi; April 5-7, 2007; Izmir.
52. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
53. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
54. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [Medline: 2002127]
55. Bouma J, Ranchor AV, Sanderman R, van Sonderen E. Het meten van symptomen van depressie met de CES-D: een handleiding [Measuring Depression with the CES-D: A Manual]. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen; 1995.
56. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. J Consult Clin Psychol 1999 Jun;67(3):300-307. [Medline: 10369050]
57. Hackshaw A, Kirkwood A. Interpreting and reporting clinical trials with results of borderline significance. BMJ 2011;343:d3340. [Medline: 21727163]
58. Griner D, Smith TB. Culturally adapted mental health intervention: A meta-analytic review. Psychotherapy (Chic) 2006;43(4):531-548. [doi: 10.1037/0033-3204.43.4.531] [Medline: 22122142]
59. Muñoz RF, Lenert LL, Delucchi K, Stoddard J, Perez JE, Penilla C, et al. Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. Nicotine Tob Res 2006 Feb;8(1):77-87. [doi: 10.1080/14622200500431940] [Medline: 16497602]
60. Wendler D, Kington R, Madans J, Van Wye G, Christ-Schmidt H, Pratt LA, et al. Are racial and ethnic minorities less willing to participate in health research? PLoS Med 2006 Feb;3(2):e19 [FREE Full text] [doi: 10.1371/journal.pmed.0030019] [Medline: 16318411]
61. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. Health Soc Care Community 2004 Sep;12(5):382-388. [doi: 10.1111/j.1365-2524.2004.00507.x] [Medline: 15373816]
62. Riper H, Kramer J, Conijn B, Smit F, Schippers G, Cuijpers P. Translating effective web-based self-help for problem drinking into the real world. Alcohol Clin Exp Res 2009 Aug;33(8):1401-1408. [doi: 10.1111/j.1530-0277.2009.00970.x] [Medline: 19413646]
63. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461 [FREE Full text] [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
64. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. PLoS One 2012;7(5):e38021 [FREE Full text] [doi: 10.1371/journal.pone.0038021] [Medline: 22741027]
65. Ten Have M, van Dorsselaer S, de Graaf R. Prevalence and risk factors for first onset of suicidal behaviors in the Netherlands Mental Health Survey and Incidence Study-2. J Affect Disord 2013 May;147(1-3):205-211. [doi: 10.1016/j.jad.2012.11.005] [Medline: 23200296]
66. 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]
67. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
68. Foquz Media 2008. URL: http://www.foquz.nl/etnomarketing/persberichten/internet10-03-2008.html [accessed 2013-08-30] [WebCite Cache ID 6JGLDysWx]
69. Centraal Bureau voor de Statistiek. 2013 Jul 13. Statline 2013 URL: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71822NED&D1=0&D2=a&D3=0&D4=1,3-9&D5=0-2,4,8-10&D6=0&D7=l&HD=081216-1451&HDR=T,G6,G2,G5,G3&STB=G1,G4 [WebCite Cache ID 6JGLPIkOD]

70. Carlbring P, Brunt S, Bohman S, Austin D, Richards J, Öst LG, et al. Internet vs paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Computers in Human Behavior 2007 May;23(3):1421-1434. [doi: 10.1016/j.chb.2005.05.002]
71. 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

AOC: Alles Onder Controle (Everything under Control)
AOC-TR: Alles Onder Controle TR (Turkish adapted version)
BDI-II: Beck Depression Inventory II
CES-D: Center for Epidemiologic Studies Depression Scale
EQ-5D: EuroQol Questionnaire 5D
EQ-VAS: EuroQol visual analogue scale
MINI: Mini-international Neuropsychiatric Interview
NTR: Dutch Trial Register
RCI: reliable change index
SCL-90-R: Symptom Checklist-90-Revised.

Edited by G Eysenbach; submitted 26.07.13; peer-reviewed by D Linares, A Aguilera; comments to author 16.08.13; revised version received 30.08.13; accepted 09.09.13; published 11.10.13.

Please cite as:
Ünlü Ince B, Cuijpers P, van 't Hof E, van Ballegooijen W, Christensen H, Riper H
Internet-Based, Culturally Sensitive, Problem-Solving Therapy for Turkish Migrants With Depression: Randomized Controlled Trial
J Med Internet Res 2013;15(10):e227
URL: http://www.jmir.org/2013/10/e227/
doi:10.2196/jmir.2853
PMID:24121307

Original Paper

The Complex Relationship of Realspace Events and Messages in Cyberspace: Case Study of Influenza and Pertussis Using Tweets

Anna C Nagel[1], MPH; Ming-Hsiang Tsou[2], PhD; Brian H Spitzberg[3], PhD; Li An[2], PhD; J Mark Gawron[4], PhD; Dipak K Gupta[5], PhD; Jiue-An Yang[2], MA; Su Han[2], MS; K Michael Peddecord[1], PhD; Suzanne Lindsay[1], PhD; Mark H Sawyer[6,7], MD

[1]Graduate School of Public Health, San Diego State University, San Diego, CA, United States
[2]Department of Geography, San Diego State University, San Diego, CA, United States
[3]School of Communication, San Diego State University, San Diego, CA, United States
[4]Department of Linguistics, San Diego State University, San Diego, CA, United States
[5]Department of Political Science, San Diego State University, San Diego, CA, United States
[6]Division of Pediatrics Infectious Diseases, University of California San Diego School of Medicine, La Jolla, CA, United States
[7]Epidemiology and Immunization Services Branch, City of San Diego Health and Human Services Agency, San Diego, CA, United States

Corresponding Author:
Ming-Hsiang Tsou, PhD
Department of Geography
San Diego State University
Storm Hall #326
5500 Campanile Dr
San Diego, CA, 92182
United States
Phone: 1 619 594 0205
Fax: 1 619 594 4938
Email: mtsou@mail.sdsu.edu

Abstract

Background: Surveillance plays a vital role in disease detection, but traditional methods of collecting patient data, reporting to health officials, and compiling reports are costly and time consuming. In recent years, syndromic surveillance tools have expanded and researchers are able to exploit the vast amount of data available in real time on the Internet at minimal cost. Many data sources for infoveillance exist, but this study focuses on status updates (tweets) from the Twitter microblogging website.

Objective: The aim of this study was to explore the interaction between cyberspace message activity, measured by keyword-specific tweets, and real world occurrences of influenza and pertussis. Tweets were aggregated by week and compared to weekly influenza-like illness (ILI) and weekly pertussis incidence. The potential effect of tweet type was analyzed by categorizing tweets into 4 categories: nonretweets, retweets, tweets with a URL Web address, and tweets without a URL Web address.

Methods: Tweets were collected within a 17-mile radius of 11 US cities chosen on the basis of population size and the availability of disease data. Influenza analysis involved all 11 cities. Pertussis analysis was based on the 2 cities nearest to the Washington State pertussis outbreak (Seattle, WA and Portland, OR). Tweet collection resulted in 161,821 flu, 6174 influenza, 160 pertussis, and 1167 whooping cough tweets. The correlation coefficients between tweets or subgroups of tweets and disease occurrence were calculated and trends were presented graphically.

Results: Correlations between weekly aggregated tweets and disease occurrence varied greatly, but were relatively strong in some areas. In general, correlation coefficients were stronger in the flu analysis compared to the pertussis analysis. Within each analysis, flu tweets were more strongly correlated with ILI rates than influenza tweets, and whooping cough tweets correlated more strongly with pertussis incidence than pertussis tweets. Nonretweets correlated more with disease occurrence than retweets, and tweets without a URL Web address correlated better with actual incidence than those with a URL Web address primarily for the flu tweets.

Conclusions: This study demonstrates that not only does keyword choice play an important role in how well tweets correlate with disease occurrence, but that the subgroup of tweets used for analysis is also important. This exploratory work shows potential in the use of tweets for infoveillance, but continued efforts are needed to further refine research methods in this field.

(J Med Internet Res 2013;15(10):e237) doi:10.2196/jmir.2705

KEYWORDS

Twitter; infoveillance; infodemiology; cyberspace; syndromic surveillance; influenza; pertussis; whooping cough

Introduction

Background

Use of the Internet has shifted from being solely a one-way transfer of information to an interactive multidimensional channel. Cyberspace resides as a source of information accessible to the user who is able to contribute to cyberspace through social media and online communities [1]. Infodemiology is the study of the distribution and causal factors of information in cyberspace and its ability to improve public health [2]. The Internet provides many resources for infodemiology, including search engine queries (ie, Google Flu Trends [3]), publications, marketing campaigns, and user-generated content, such as blogs and social media status updates [2]. Researchers are pioneering a variety of methods and applications using these resources for disease detection (see [4] for overview). This study focuses on the infodemiology of pertussis-related (also called whooping cough) and influenza-related status updates on Twitter (tweets).

Every year millions of Americans become infected with the flu, resulting in illness, missed work and school days, and death. Deaths from seasonal influenza occur mostly in young children and the elderly, largely because of flu complications and the exacerbation of existing conditions, such as congestive heart failure [5]. Influenza causes a substantial economic burden associated with loss in productivity because of missed work and health care costs [6]. Pertussis infects a much smaller population, but can result in severe complications, especially among those who are young and unvaccinated. Approximately 57% of infants under 1 year of age are hospitalized for pertussis, and the risk is greater the younger the child [7]. The most common complication among hospitalized infants is apnea (67%), or pauses in breathing that may result in cyanosis, followed by pneumonia (23%). Death and violent convulsions occur in approximately 1.6% of hospitalized infants, and brain disease (encephalopathy) occurs in approximately 0.4% [7]. As of December 29, 2012, Washington State had experienced 4744 pertussis cases, 5.9 times more than the prevalence for the same time period in 2011 (807) [8]. The early notification of disease outbreaks greatly increases the ability of affected communities to control and treat an epidemic. Traditional surveillance methods are a vital factor in the control of diseases, but there is often a time lag between the reporting of individual cases and the accumulation of these data into a report [9].

Related Work

The Internet has become a fundamental tool for geographic information system (GIS) technology. Devices enabled with Global Positioning System (GPS) receivers and the Internet allow for precise geographic information of events for a variety of uses, including those focused on public health. For example, Love Clean Streets is used to alert authorities of problems in the community, such as graffiti and potholes [10]. HealthMap maps disease occurrences based on a variety of sources, including user reports [11]. Noise pollution can be analyzed based on pedestrian audio recordings from their GPS-enabled devices [12]. Researchers have used information contained in tweets to detect earthquakes in Japan [13]. Each Twitter user was labeled as a sensor; the sensor was either positive (the user tweeted earthquake-related information) or negative (they did not tweet information). Through these methods, the authors were able to detect an earthquake with 96% probability and notify authorities before the Japan Meteorological Agency [13]. Another study aimed to analyze Twitter activity during a fire outside of Marseille, France, in 2009. The researchers found tweets were accurate and timely, but not for all phases of the event. They concluded more work needed to be done in this field to fully utilize its potential [14].

Recently, other innovative syndromic surveillance methods using the Internet have been developed [15-25]. Syndromic surveillance plays a crucial role in early disease detection. In its simplest form, syndromic surveillance aims to detect a signal indicating a possible disease outbreak before the traditional surveillance methods of diagnosing and reporting diseases. The signal is usually either a symptom or symptom surrogate [26], such as pharmaceutical prescriptions [17]. Researchers in Japan found a high correlation between prescription drug purchases from over 5000 pharmacies and influenza activity reported by official sentinel surveillance [17]. Infoveillance, a component of infodemiology, is the monitoring of online texts. Online information acts as a signal of disease occurrence or public interest related to a disease [2]. These methods can be extended to investigate the public's understanding of health topics, such as misunderstandings of dosing instructions and the resulting misuse of antibiotics [25].

The availability of public health-related Internet data has inspired many innovative studies. One study assessed the usefulness of social media for the surveillance of intentional and unintentional foodborne-illness outbreaks [9]. The authors concluded social media can play an important role in identifying clusters of foodborne illness faster than traditional methods. Many people with foodborne illness do not seek medical attention; however, they may be more likely to report symptoms online because of its ease and convenience [9]. Although limitations exist, the authors concluded that a system to exploit the large amount of data available on social media platforms in real time would be useful for detecting foodborne-illness outbreaks [9].

Twitter is the leading service of choice for disease tracking in social media. One notable study used tweets to track public concern and flu activity in the United States during the 2009 influenza A (H1N1) pandemic [22]. The researchers used tweets containing disease transmission, disease countermeasure, pork consumption, and vaccine-related keywords to track public concern. In several cases, the percentage of tweets with these keywords changed in response to news events and official disease reports [22]. A second subset of keywords was used to train a prediction model. Estimates from this model were compared with regional influenza-like illness (ILI) cases reported by the Centers for Disease Control and Prevention (CDC) and showed a close correlation. The real-time estimates from this study can be determined 1 to 2 weeks before traditional surveillance methods [22]. Researchers used Twitter to assess public concern during the 2009 H1N1 outbreak [19]. One component of the study used these more than 2 million tweets to investigate the adoption of the World Health Organization's (WHO) terminology of H1N1 compared to swine flu, the initial term used. Over the study period, the percentage of tweets using H1N1 increased from 8.8% to 40.5% [19]. The authors concluded Twitter is a valuable tool for infodemiology, which can help health professionals to realize and address the public's concerns [19].

Objectives

In this study, we aimed to explore the realspace health events that influence ideas and messages in cyberspace and, in turn, determine to what extent these cyberspace messages affect the real world. More specifically, we investigated how the 2012-2013 influenza season (estimated by ILI reports) and the 2012 pertussis outbreak in Washington State are reflected in cyberspace, measured by the production of keyword-specific tweets. We also examined the extent to which these tweets act as a signal of disease occurrence or public interest, and investigated how keyword choice and specific subgroups of tweets correlate with disease occurrence, and how the scale at which disease incidence data are collected (ie, city vs state level) affects correlations with tweets collected at the city level.

Methods

Data Collection

This paper extends previous explorations of the innovative Visualizing Information Space in Ontological Networks (VISION) framework using the 2012-2013 flu season and the 2012 pertussis outbreak as case studies. The VISION framework was developed by our research team to better understand the connection between space, time, and messages [27]. Two information-mining tools were created: one for collecting webpage information and the other for collecting tweets. The focus of this paper was devoted exclusively to tweets under the assumption that tweet activity would be more dynamically indicative of disease diffusion compared to webpage content. Twitter provides a large source of publicly available data. Twitter has more than 140 million active users producing millions of tweets (messages of 140 characters or less) on a daily basis [28]. In this study, we tap into this resource with geo-search-enabled Twitter Tools created to operate with the Twitter Search and Streaming application programming interfaces (APIs) [29]. Our tools, in combination with the Twitter APIs, return a Microsoft Excel spreadsheet of tweets associated with a keyword (in the tweet text, the user name, or the title of a linking Web page) and within a specified geographical range. Additional information is provided with each tweet, such as the user name, the time it was created, and the location. The location is based on either the user's self-proclaimed hometown or latitude and longitude coordinates if the user was tweeting from a GPS-enabled device.

In this study, tweets were collected for the keywords flu, influenza, pertussis, and whooping cough. Although the word "flu" is nested within the word influenza, our search tool treated these as separate search terms. Tweets were obtained from within a 17-mile radius of the city center of 11 US cities (Boston, MA; Chicago, IL; Cleveland, OH; Denver, CO; Fort Worth, TX; Jacksonville, FL; Nashville-Davidson, TN; New York, NY; Portland, OR; San Diego, CA; and Seattle, WA). Cities were chosen based on their population and the availability of sufficient ILI data at the city or county level. A radius of 17 miles was specified to cover a large urban area while avoiding overlapping with nearby cities. Figure 1 shows the geographical location of each of the cities of interest. For the flu and influenza keywords, tweet collection began on August 31, 2012, and continued through March 4, 2013, resulting in 161,821 flu tweets and 6174 influenza tweets. During analysis, focus was on tweets from the CDC's Morbidity Mortality Weekly Report (MMWR) weeks 37 to 45 depending on when ILI data became available for each city (weeks starting September 1 to November 4, 2012) to MMWR week 9 (ending March 2, 2013).

The resulting tweets were compared to weekly ILI rates at the city or corresponding county level. These reports are the percentage of patients seen for influenza-like symptoms (fever greater ≥100°F and a cough and/or sore throat in the absence of a known cause other than influenza) compared to all patient visits for the week [30]. It is worth noting that ILI reporting is optional. The CDC does not report ILI data below the state level; therefore, ILI cases were obtained from individual city or county health department Web sites and, in the case of San Diego, from a contact at the County of San Diego Health and Human Services Agency [31]. In a few cases, an ILI report was missing for a particular week. In such cases, the previous and following week's ILI percentages were averaged. In a case in which 2 weeks in a row were missing, the ILI rate for the previous week was used for the first missing week and the second missing week was derived from the following week's ILI rate.

Tweet collection began on June 3, 2012 for the pertussis and whooping cough keywords and ended December 1, 2012, resulting in 160 pertussis tweets and 1167 whooping cough tweets. Tweets were compared to pertussis cases in Washington State. The prevalence and incidence of probable and confirmed pertussis cases in Washington State were reported on a weekly basis on the Washington State Department of Health website [8]. Tweet collection focused on 2 of the 11 cities nearest to the Washington State outbreak: Seattle and Portland.

Figure 1. The 11 cities of interest (using a 17-mile radius) for which tweets including the keywords flu and influenza (all 11 cities) and pertussis and whooping cough (primarily Seattle and Portland) were used in the study.

Analysis

The tweeting rate (number of tweets per 100,000 individuals in each city) was determined. For consistency, the city population was estimated within a 17-mile radius of the city center. The tweeting rate and disease data were then represented as bar graphs so fluctuations in each could easily be visually compared. For visualization purposes, the maximum pertussis and whooping cough tweet rates (in both Portland and Seattle) and the maximum disease incidences were rescaled and set equal to one another. For flu and influenza, the maximum tweet rate and ILI were rescaled and set equal to one another in each city.

The association between weekly aggregated tweets and disease incidence was determined using Pearson correlation coefficients in R version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria) for each of the 11 cities for influenza and Portland and Seattle for pertussis. In addition, tweets were further subdivided to determine the type of tweets that best correlated with disease cases. As mentioned previously, 4 keywords were used for tweet collection: flu, influenza, pertussis, and whooping cough. Each of these 4 datasets was segregated into nonretweets, retweets, tweets without a URL Web address, and tweets with a URL Web address. Fisher z-transformation was used to assess significant differences in correlation coefficients among various groups of tweets or keywords. Correlation coefficients were compared between tweets from the 2 keywords used for each disease, from different cities, and for the 4 tweet subgroups listed previously.

Results

Flu and Influenza Tweets

Correlation coefficients among flu and influenza tweet rates per 100,000 population and estimated flu incidence based on ILI reports are displayed in Table 1. Significance testing between 2 correlations was only performed when both correlations were significant individually. When comparing flu and influenza tweets from the all tweets group (column 1), correlations for both keywords were significant in 6 cities: Denver, Fort Worth, Jacksonville, Nashville-Davidson, San Diego, and Seattle. Significant differences between correlations were found in 4 of these cities (Denver, Jacksonville, San Diego, and Seattle). Among all tweets, correlations in all 4 cities were significantly higher for flu compared to influenza tweets.

Correlations differed between subgroups after subdividing the tweets for each keyword into nonretweets, retweets, tweets with a URL Web address, and those without a URL Web address. The *P* values from the Fisher z-transformations for the nonretweet versus retweet comparison and the comparison between tweets with a URL versus those without are presented in Table 1. For the flu keyword, 6 cities (Denver, Fort Worth, Jacksonville, Nashville-Davidson, San Diego, and Seattle) had significant correlation coefficients for both the nonretweet and the retweet groups. Significantly higher correlations were seen among the nonretweet group for all 6 cities (P<.001 for each comparison). Differences between significant nonretweet and retweet correlations for the influenza keyword were not significant. For the flu keyword, significantly larger correlations (P<.05 for each comparison) were found among tweets without a URL Web address compared to those with a URL Web address in 6 of the 8 cities (Boston, Cleveland, Denver, Fort Worth, Nashville-Davidson, and Seattle) in which both correlations being compared were significant. For influenza, 5 cities (Denver, Fort Worth, Nashville-Davidson, New York, and Seattle) had significant correlations for both tweets with a URL Web address and those without, but none of these comparisons showed significant differences between correlations.

The 11 cities used for tweet and ILI comparison were distributed across the continental United States, allowing for the investigation of geographical variations. Figures 2 and 3 show the weekly tweeting rate and ILI report percentages from MMWR week 39 (starting September 23, 2012) to MMWR week 9 (ending March 2, 2013) as barcharts for the flu and influenza keywords, respectively. The barcharts are organized in the table according to the corresponding city's geographical region. The first column is more generally the western states, the second northeastern, and the third column southern states. Weekly changes in tweeting rate and ILI report percentages can be seen from MMWR week 51 (starting December 16, 2012) to MMWR week 2 (starting January 6, 2013) in Figure 4.

In Figures 2 and 3, the total (unsubdivided) tweets are shown in which the corresponding correlation coefficients were pulled from all tweets. The black bars indicate missing tweets during MMWR week 52. The black bar shows the tweets actually collected, but there were likely more. The maximum ILI and tweet rates for each city were rescaled and set equal to one another for better visualization, but this limits the viewer's ability to compare frequency values between cities. Instead, we suggest focusing on the general trends and correlations between ILI rates and tweets within each city.

Figure 2. Barcharts indicating trends in all tweets containing the keyword flu (pink) and influenza-like illness (ILI) rates (blue) beginning MMWR weeks 37-45 (starting September 1 to November 4, 2012 depending on when ILI data became available for a particular city) and ending MMWR week 9 (March 2, 2013). The black bar indicates a week in which tweets were missing. Significant correlations are bolded.

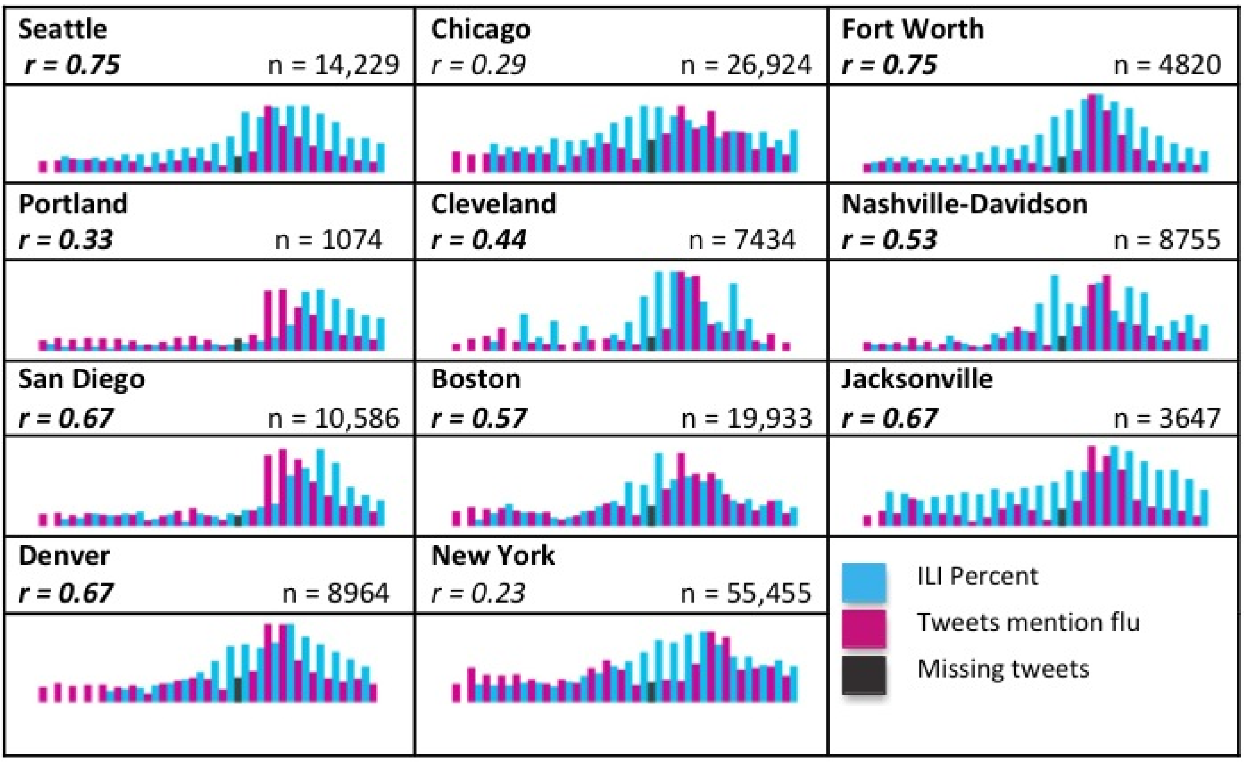


Figure 3. Barcharts indicating trends in all tweets containing the keyword influenza (pink) and influenza-like illness (ILI) rates beginning MMWR weeks 37-45 (starting September 1 to November 4, 2012 depending on when ILI data became available for a particular city) and ending MMWR week 9 (March 2, 2013). The black bar indicates a week in which tweets were missing. Significant correlations are shown in bold.

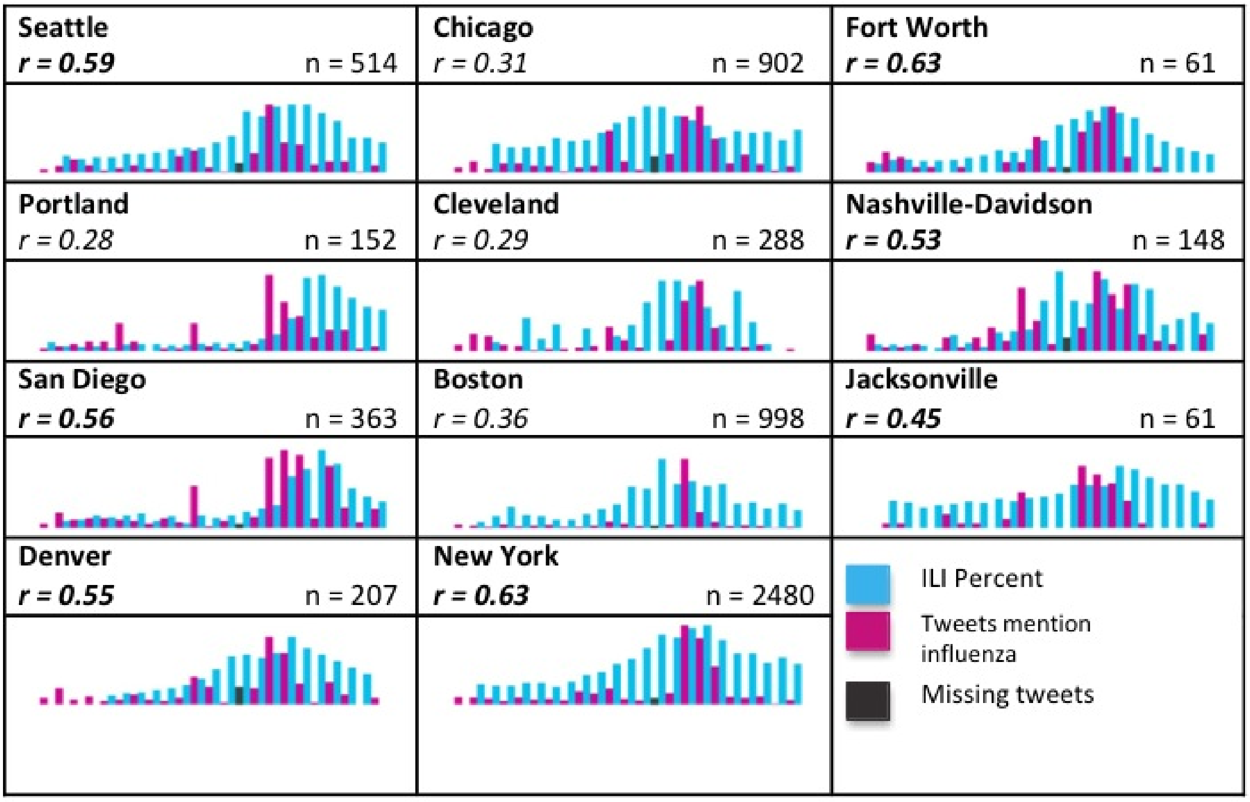


Figure 4. Weekly changes in influenza-like illness (ILI) rates and the rate of tweets including the keyword flu per 100,000 people starting from MMWR week 51 (December 16 to December 22, 2012) through MMWR week 2 (January 6 to January 12, 2013) mapped across the 11 cities from which tweets were collected. Larger circles represent higher rates.

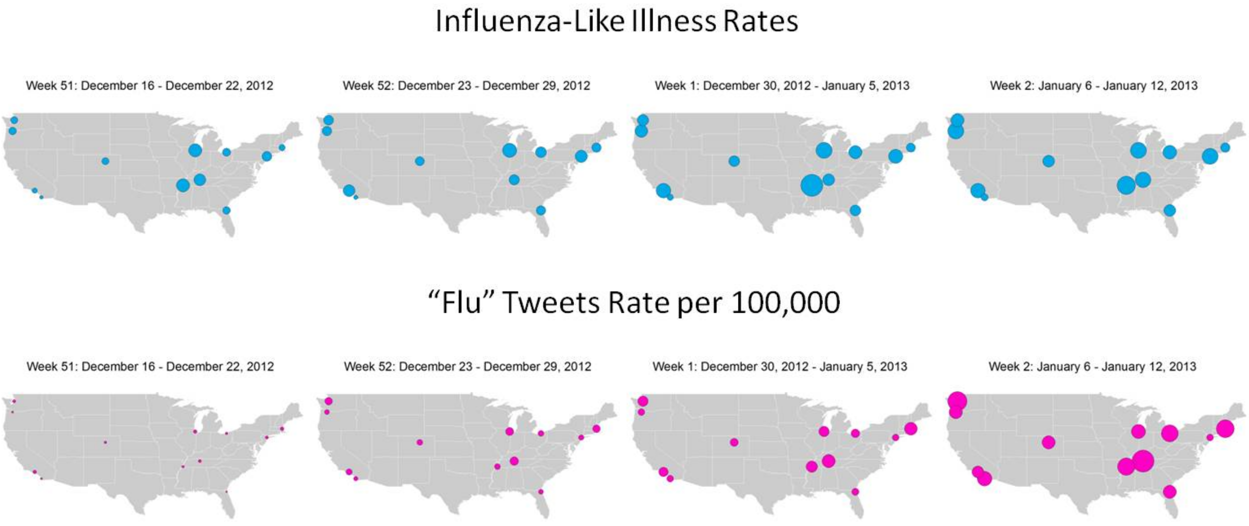


Table 1. Correlation coefficients between tweets (and tweet subgroups) and influenza-like illness (ILI) reports for each city for the flu and influenza keywords.

City	All tweets *r*	Nonretweet *r*	Retweets *r*	*P* value[a]	Tweets with URL *r*	Tweets without URL *r*	*P* value[b]	Total tweets n
Flu								
Boston	.57	.57	.48	<.001	.49	.60	<.001	19,933
Chicago	.29	.31	.19	<.001	.14	.40	<.001	26,924
Cleveland	.44	.49	.30	<.001	.40	.46	.004	7434
Denver	.67[c]	.69	.53	<.001	.62	.69	<.001	8964
Fort Worth	.75	.75	.67	<.001	.65	.77	<.001	4820
Jacksonville	.67[c]	.71	.32	<.001	.63	.67	.06	3647
Nashville-Davidson	.53	.61	.35	<.001	.37	.66	<.001	8755
New York	.23	.23	.17	<.001	.29	.17	<.001	55,455
Portland	.33	.49	.33	<.001	.37	.52	<.001	1074
San Diego	.67[c]	.70	.55	<.001	.66	.68	.07	10,586
Seattle	.75[c]	.77	.67	<.001	.73	.75	.01	14,229
Influenza								
Boston	.36	.41	.32	.10	.34	.46	.07	998
Chicago	.31	.30	.27	.66	.26	.41	.02	902
Cleveland	.29	.31	.25	.59	.35	-.06	.001	288
Denver	.55[c]	.60	.42	.17	.49	.60	.27	207
Fort Worth	.63	.65	.08	.24	.52	.48	.85	61
Jacksonville	.45[c]	.45	.27	.63	.53	.28	.27	61
Nashville-Davidson	.53	.53	.29	.35	.48	.49	.94	148
New York	.63	.65	.61	.11	.63	.58	.07	2480
Portland	.28	.31	.08	.35	.09	.59	.001	152
San Diego	.56[c]	.58	.48	.30	.58	.31	.01	363
Seattle	.59[c]	.67	.42	<.001	.55	.63	.19	514

[a] From Fisher z-transformation comparing nonretweet and retweet correlation coefficients.

[b] From Fisher z-transformation to determine significant differences among correlation coefficients of tweets with a URL compared to those without a URL Web address.

[c] Significant differences between the flu and influenza correlation coefficients for all tweets when both correlations being compared were significant.

Pertussis and Whooping Cough Tweets

The weekly pertussis and whooping cough tweets also resulted in varying levels of correlation within cities, and by keywords and tweet subgroups (listed in Table 2). Significant correlations between tweets and pertussis incidence in Washington State were only found among tweets collected using the whooping cough keyword. This may be driven by the relatively small number of tweets for the pertussis keyword. The whooping cough keyword appeared to be more highly correlated with pertussis incidence than pertussis keyword tweets, probably reflecting the colloquial nature of tweet language. Further interpretation focuses solely on the whooping cough tweet results. Similar to the flu analysis, Fisher z-transformation was used when testing for significant differences between 2 individually significant correlations. Among the all tweets group, tweets from Portland were significantly more highly correlated with disease incidence than tweets from Seattle ($P<.001$).

Tweets were divided into nonretweets, retweets, tweets with a URL Web address, and tweets without a URL Web address, but trends between these groups were not as obvious as with the flu analysis. Correlations for nonretweets and retweets were both significant in Portland. Although nonretweets appeared to be more highly correlated with disease incidence, the difference was not significant ($P=.39$). On the other hand, correlations were significant for both tweets with a URL Web address and those without a URL Web address among tweets from Seattle. In this case, tweets without a URL Web address were significantly more highly correlated with pertussis incidence in Washington State than tweets with a URL Web address (*P*=.01).

Figure 5 gives a visual representation of all pertussis and whooping cough tweets compared to pertussis incidence from MMWR weeks 23 to 48 (June 3 and ending December 1, 2012). Significant correlations among the all tweets group and disease incidence are shown in bold. The highest tweeting rate and the largest weekly incidence were rescaled and set equal to one another for better visualization. This figure illustrates the difference in tweeting rates between the 2 cities. There were relatively few pertussis tweets (top row) compared to whooping cough tweets (bottom row). Referring back to Table 2, it is apparent that tweets from Portland were more highly correlated with pertussis incidence in Washington State than those from Seattle for the whooping cough keyword.

Table 2. Correlation coefficients between tweets (and tweet subgroups) and pertussis incidence in Washington State in Seattle and Portland for the pertussis and whooping cough keywords.

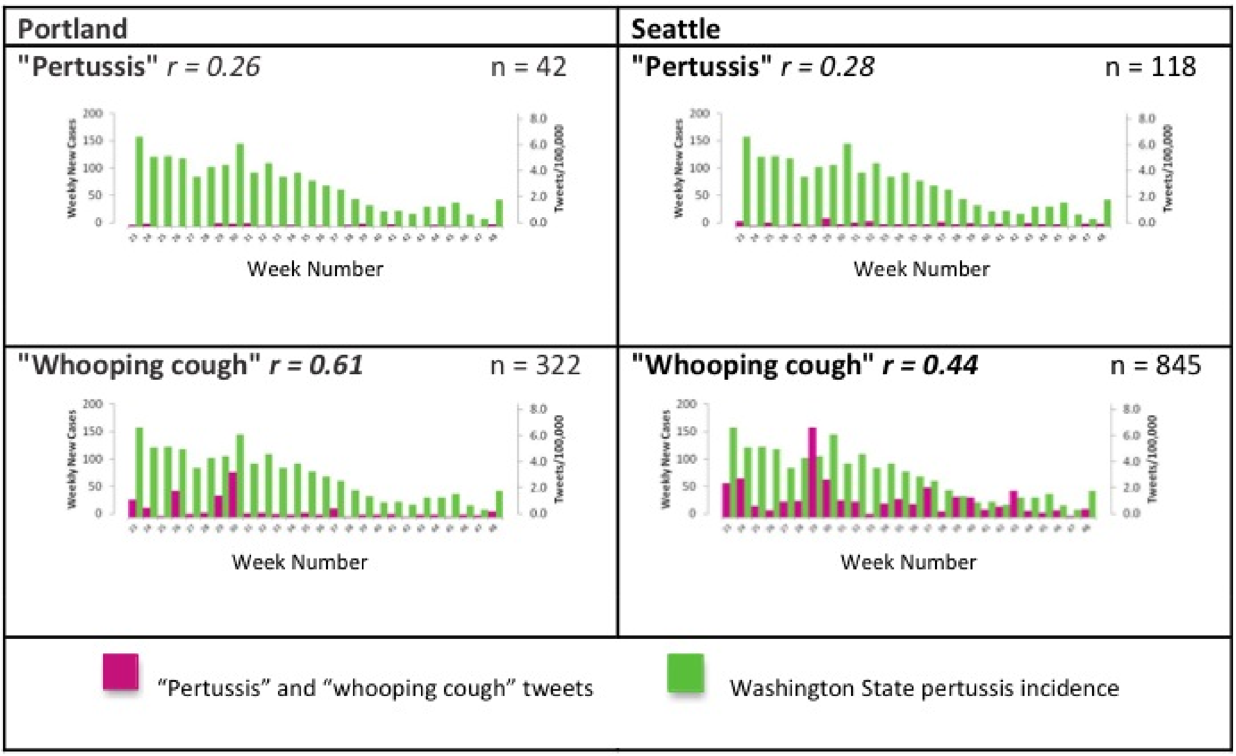
City	All tweets *r*	Nonretweets *r*	Retweets *r*	*P* value[a]	Tweets with URL *r*	Tweets without URL *r*	*P* value[b]	Total tweets n
Pertussis								
Portland	.26	.21	.26	.90	.27	.13	.69	42
Seattle	.28	.21	.31	.65	.25	.15	.58	118
Whooping cough								
Portland	.61[c]	.56	.49	.39	.53	.38	.28	322
Seattle	.44[c]	.47	.37	.11	.41	.44	.01	845

[a] From Fisher z-transformation comparing nonretweet and retweet correlation coefficients.

[b] From Fisher z-transformation to determine significant differences among correlation coefficients of tweets with a URL compared to those without a URL Web address.

[c] Significant differences between the Seattle and Portland correlation coefficients for all tweets.

Figure 5. Barcharts indicating trends in all tweets containing the keywords pertussis and whooping cough (pink) and pertussis incidence in Washington State (green) for Portland and Seattle beginning MMWR weeks 23-48 (June 3, 2013 to December 1, 2013). Significant correlations are bolded.



Discussion

Principal Findings

This study demonstrates that tweets can function as a signal of disease activity and public interest. In this paper, we outline the differences in the ability of varying groups of tweets to temporally track influenza and pertussis incidence measured in the community. Correlations between tweets and local disease activity were variable, but were relatively strong in some areas and for specific subgroups of tweets, such as nonretweets and those without a URL Web address. Another key finding is the benefit of using the vernacular term for diseases, such as whooping cough rather than pertussis.

The 2012-2013 influenza season was unique in its temporal and spatial spread. The 2012-2013 flu season peaked earlier than it had in almost a decade [32] and cases were initially seen on the east coast. Referring to Table 1 and the charts in Figures 2 and 3, stronger correlations occurred in western cities compared to northeastern cities. Because of the higher number of tweets and the more significant correlations, our interpretation will focus on Figure 2. For the most part, both tweeting rates and ILI rates are low for the first two-thirds of the time period, then peak and then decrease in the last third. Among the 9 cities that had significant correlations between tweets and ILI rates, tweeting rates peaked before ILI rates in 5 cities. This shows tweets have the potential to act as a signal of a flu outbreak before traditional ILI reporting methods. However, this was only evident in 5 cities; in the other 4, ILI peaked before or at the same time as the tweets. Further investigation is needed to determine what combination of keywords or tweet subgroups can indicate a looming outbreak before traditional methods.

Nonretweets and those tweets without a URL Web address were more highly correlated with ILI activity compared to retweets and tweets with a URL Web address, respectively. Retweets and tweets with a URL Web address are arguably less likely to be about the individual tweeting them, and are likely used for sharing information created by others. Nonretweets and tweets without a URL Web address may have correlated better with flu activity than their respective counterparts because users are tweeting about themselves and possibly indicating that they have the flu. When Twitter users indicate they are infected with the flu, it is not possible to know if this is after a health provider's diagnosis or, more likely, the users' own interpretations of their symptoms.

On the other hand, the pertussis outbreak in Washington State provided a unique opportunity for analyzing the interaction between cyberspace and realspace after the disease incidence had peaked. In general, correlation coefficients between tweets and pertussis incidence were lower than those for tweets and ILI rates, but this may have been because of fewer tweets and the nature of the diseases. Figure 5 shows just how low the pertussis tweet frequencies were compared to whooping cough. We assumed tweets from Seattle would correlate better with pertussis incidence in Washington given that Seattle is in the state, but this was not the case. In fact, correlations were higher outside of Washington State (in Portland) for 4 of 5 tweet categories (all except tweets without a URL Web address).

Because of the larger tweet number and significant results, our interpretations will focus on the whooping cough keyword. Like the influenza analysis, tweets without a URL Web address were more strongly correlated with pertussis incidence than tweets with a URL Web address, but this was only significant in Seattle. Correlations among nonretweets remained higher compared to retweets, although this was not significant in either city.

Referring back to Figure 5, it can be seen that the pertussis incidence starts high and then slowly decreases over the 25 weeks. Overall, whooping cough tweets increased until week 29 and then decreased for the rest of the time period. The peak in tweets during week 29 was likely driven by media reports on a CDC press release about the pertussis epidemic in Washington State [33]. In fact, many of the tweets during that week mentioned the press release and were accompanied by a URL Web address. It seemed in this analysis the media played a larger role in tweet production than in the flu analysis. If nothing else, these findings suggest that health communication campaigns can penetrate social media in measurable ways. Because tweet collection began after the number of new pertussis cases peaked, we are unable to conclude whether tweets could signal a potential pertussis outbreak. However, because of the correlation between tweets and pertussis in Portland, for example, further explorations in this area may prove to be worthwhile.

Differences between trends seen for influenza and pertussis may have been driven by several factors. First, the 2 diseases are measured differently. Pertussis is a mandatory reported disease, meaning the Washington State pertussis incidence rates were based on true diagnosed cases. Influenza cases, on the other hand, were measured by a proxy. The ILI diagnoses can only estimate possible influenza cases, but may also be an indicator of other respiratory diseases. There were fewer pertussis-related tweets than influenza-related tweets. This may reflect the real world prevalence of the diseases. Each year, many people become infected with the flu, whereas relatively few are infected with pertussis. The commonality of flu and the rarity of pertussis make outbreaks a topic of interest in the media.

The age range of infection may also play a role in differing trends between influenza and pertussis correlations with tweets. People of all ages become infected with the flu, including a large adult population. However, pertussis is more common and most severe in infants and young children. This younger population may be less likely or unable to tweet about their illness. Also, the CDC press release resulted in a large increase in tweets, many of which contained a URL Web address linking to an article on this topic.

In addition, we were able to investigate how well tweets collected at the city level correlated with disease at the city or state level. As outlined previously, we have several conjectures as to why the flu analysis showed better correlations with disease occurrence than the pertussis analysis, but it is also important to emphasize that tweets were collected at the city level and compared to city ILI rates or state-level pertussis incidence. We chose to focus on the city level because of the importance of a quick local response in keeping the disease from spreading. Usually, ILI rates are aggregated into larger regions because of limitations in reporting at a smaller scale. However, 1 study during the 2009 H1N1 influenza season reported that ILI rates from their 8 sentinel sites that were part of a university health care system correlated well with state and regional ILI rates and were available sooner [34].

Differences in correlation coefficients may have also been driven by tweets from Seattle and Portland, which may not have been representative of Washington State pertussis activity. As a basis for contrast and control, it would have been beneficial to have either collected tweets from more cities in Washington State or to have obtained pertussis incidence at the city level for Seattle and Portland. This could also have shed light on the reasons why tweets from Portland were more highly correlated with pertussis incidence than those from Seattle. Further exploration of the geography of disease outbreak becomes an important direction for future research.

A few similarities between the influenza and pertussis analyses are also evident. For both, higher correlations were seen among nonretweets versus retweets. A hypothesis for this trend has already been given for influenza, but in the case of pertussis, an explanation is elusive at present. Both the influenza and pertussis explorations indicated the keyword chosen for collecting tweets played a vital role in correlation coefficients. As expected, there were more whooping cough than pertussis tweets, possibly because whooping cough is the colloquial term, whereas pertussis is used primarily by health professionals. A similar trend was seen between the flu and influenza tweets; both were used interchangeably by the general public, but flu may have been the preferred term among Twitter users because of the character limit of each tweet.

Limitations

Limitations in this study were experienced for both tweet collection and disease reporting. Server issues interrupted the VISION information-mining tool and resulted in missing flu and influenza tweets during MMWR week 52 (the week is indicated by black bars in Figures 2 and 3). We suspect that if some tweets were not missing, correlations between tweets and ILI rates would have been slightly different and possibly higher. On the other end, ILI reporting is optional and the health care providers who supplied ILI rates varied between cities. For example, in some cities ILI was reported by emergency departments, whereas ILI was reported by primary care physicians in others. These 2 sources may have different ILI rates, but the general trend over time is likely similar in both. For the pertussis and whooping cough keywords, tweet collection did not begin until the number of new pertussis cases in Washington State had already peaked. Obviously, this makes it impossible to evaluate whether tweets in this case could detect an outbreak before traditional methods. However, in Portland, for example, moderate to strong correlations were observed after the outbreak, indicating an association between whooping cough tweets and pertussis. Further exploration is needed to determine whether this trend would have preceded the peak in new pertussis cases.

The number of keywords used in this study was fairly restricted because of the exploratory nature of this work. Additional keywords may greatly influence the correlations observed between tweets and disease occurrence. These additional keywords may include those in other languages, especially in cities with a large population of non-English or multilingual speakers. Another limitation is the possible misclassification of tweets by location. A previous study indicated only approximately 2.02% of the 23.8 million tweets collected worldwide during 2 separate weeks in October and November 2011 were accompanied by a GPS location [35]. For those tweets without a GPS location, we relied on the user's self-proclaimed hometown; however, the meaning of hometown can vary. The same study investigated the accuracy of user-supplied data in the United States by comparing tweets that contained both a GPS location and a user-supplied location. The state was determined for both and found to match approximately 88% of the time [35]. Although this study focused on tweets at the city level, previous work indicates self-proclaimed hometowns may be reliable. Furthermore, tweets were collected during the flu season and during the pertussis epidemic; however, it may be beneficial to collect tweets throughout the year to better determine how well tweets can detect the initial outbreak.

Conclusions

Because of the ever-changing nature of cyberspace, and specifically social media, the use of Internet data for infodemiology and infoveillance research provides many challenges. The meanings of messages change over time and within spatial variations leaving a complex system for researchers to navigate. However, exploratory results from this study indicate a strong association between tweets in cyberspace and the real world events of disease occurrence.

In future work, we aim to further investigate the actual tweet content and its association with disease incidence at the city, state, and country level. In addition, attention needs to be given to the impact the media have on the population's tweeting rate; for example, Twitter users may be inspired to tweet in reaction to a particular news story. Further investigation may indicate what type of tweets or specific words within these tweets best correlate with disease activity and should be used to detect outbreaks of disease in real time. Research has shown that although infoveillance methods are still relatively new, their impact in detecting outbreaks is becoming more demonstrable. Well-developed infoveillance methods may detect disease diffusion weeks before traditional methods and at much lower cost, allowing health services to better prepare for and prevent disease. Continued efforts in this field are needed to reach the potential of infodemiology to improve the public's health and, specifically, its application in syndromic surveillance.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No 1028177, project titled CDI-Type II: Mapping Cyberspace to Realspace: Visualizing and Understanding the Spatiotemporal Dynamics of Global Diffusion of Ideas and the Semantic Web. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

None declared.

References

1. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
2. 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]
3. 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]
4. 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]
5. Centers for Disease Control and Prevention. Seasonal Influenza (Flu). 2009 Jul 01. Clinical signs and symptoms of influenza URL: http://www.cdc.gov/flu/professionals/acip/clinical.htm [accessed 2012-08-22] [WebCite Cache ID 6AARGU6oS]
6. Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 2007 Jun 28;25(27):5086-5096. [doi: 10.1016/j.vaccine.2007.03.046] [Medline: 17544181]
7. Centers for Disease Control and Prevention. Pertussis (Whooping cough). 2012 Jun 20. Infants and Children URL: http://www.cdc.gov/pertussis/about/complications.html [accessed 2013-01-01] [WebCite Cache ID 6DNS7KfjK]

8. Washington State Department of Health. Weekly pertussis update for Washington State URL: http://www.doh.wa.gov/Portals/1/Documents/Pubs/384-254-PertussisUpdate.pdf [accessed 2013-01-02] [WebCite Cache ID 6DNm2NEOp]
9. Henning K. Center for Disease Control and Prevention. 2004 Sep 24. What is syndromic surveillance? URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a3.htm [accessed 2009-03-28] [WebCite Cache ID 5fde6JBKx]
10. Love Clean Streets. URL: http://www.lovecleanstreets.org/Reports/Home [accessed 2012-12-12] [WebCite Cache ID 6CqU9SNz0]
11. HealthMap. URL: http://healthmap.org/en/ [accessed 2012-12-12] [WebCite Cache ID 6CqUFFIuc]
12. Kanjo E. NoiseSPY: a real-time mobile phone platform for urban noise monitoring and mapping. Mobile Netw Appl 2009 Nov 20;15(4):562-574. [doi: 10.1007/s11036-009-0217-y]
13. Sakaki T, Okazaki M, Mastsuo Y. Earthquake shakes Twitter users: Real-time event detection by social sensors. 2010 Presented at: International World Wide Web Conference Committee; Apr 26-30, 2010; Raleigh, NC p. 978-988.
14. De Longueville B, Smith R, Luraschi G. "OMG, from here, I can see the flames!": a use case of mining Location Based Social Networks to acquire spatio-temporal data on forest fires. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. 2009 Presented at: International Workshop on Location Based Social Networks; Nov 4-6, 2009; Seattle, WA p. 73-80.
15. Arranz Izquierdo J, Leiva Rus A, Carandell Jäger E, Pujol Buades A, Méndez Castell MC, Salvà Fiol A, et al. [Syndromic surveillance of Influenza-like illness in primary care: a complement to the sentinel surveillance network for periods of increased incidence of Influenza]. Aten Primaria 2012 May;44(5):258-264. [doi: 10.1016/j.aprim.2011.03.008] [Medline: 21924796]
16. Culotta A. Lightweight methods to estimate influenza rates and alcohol sales volume from Twitter messages. Lang Resources & Evaluation 2012 May 13;47(1):217-238. [doi: 10.1007/s10579-012-9185-0]
17. Ohkusa Y, Sugawara T, Taniguchi K, Okabe N. Real-time estimation and prediction for pandemic A/H1N1(2009) in Japan. J Infect Chemother 2011 Aug;17(4):468-472. [doi: 10.1007/s10156-010-0200-3] [Medline: 21387184]
18. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B. Predicting flu trends using Twitter data. 2011 Presented at: The First International Workshop on Cyber Networking Systems; Apr 10-15, 2011; Shanghai.
19. 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]
20. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012 Jan;86(1):39-45 [FREE Full text] [doi: 10.4269/ajtmh.2012.11-0597] [Medline: 22232449]
21. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
22. 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]
23. Sofean M, Smith M. A real-time disease surveillance architecture using social networks. Stud Health Technol Inform 2012;180:823-827. [Medline: 22874307]
24. Newkirk RW, Bender JB, Hedberg CW. The potential capability of social media as a component of food safety and food terrorism surveillance systems. Foodborne Pathog Dis 2012 Feb;9(2):120-124. [doi: 10.1089/fpd.2011.0990] [Medline: 22217109]
25. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
26. Henning K. Centers for Disease Control and Prevention. September 24, 2004 URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a3.htm [accessed 2009-03-28] [WebCite Cache ID 5fde6JBKx]
27. Tsou MH, Yang JA, Lusher D, Han S, Spitzberg B, Gawron JM, et al. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 U.S. presidential election. 2012 Presented at: AutoCarto International Symposium on Automated Cartography; Sep 16-18, 2012; Columbus, OH p. 16-18.
28. Twitter blog. Twitter. 2012 Mar 21. Twitter turns six URL: http://blog.twitter.com/2012/03/twitter-turns-six.html [accessed 2012-08-01] [WebCite Cache ID 69bD02ViQ]
29. Twitter Developers. 2012. Using the Twitter Search API URL: https://dev.twitter.com/docs/using-search [accessed 2012-12-12] [WebCite Cache ID 6CqV8ZCWd]
30. Centers for Disease Control and Prevention. Influenza (flu): Overview of influenza surveillance in the United States. 2012 Oct 05. URL: http://www.cdc.gov/flu/pdf/weekly/overview.pdf [accessed 2012-12-12] [WebCite Cache ID 6CqRLtUmB]
31. SDSU Mapping Ideas. 2012. 2012-2013 Weekly ILI URL: http://mappingideas.sdsu.edu/2012-2013-weekly-ILI/ [accessed 2013-05-07] [WebCite Cache ID 6GRd7NXLG]
32. Centers for Disease Control and Prevention. Press Briefing Transcript. 2012 Dec 03. Telebriefing on US influenza activity and vaccination rates for current season URL: http://www.cdc.gov/media/releases/2012/t1203_influenza_activity.html [accessed 2013-03-26] [WebCite Cache ID 6FOjF5E9A]

33. Centers for Disease Control and Prevention. Press Briefing Transcript. 2012 Jul 19. Pertussis epidemic in Washington State - 2012 URL: http://www.cdc.gov/media/releases/2012/t0719_pertussis_epidemic.html [accessed 2013-04-15] [WebCite Cache ID 6FtCSk6K9]
34. Baker AW, Enfield K, Mehring B, Turner JC, Sifri CD. Local influenza-like illness surveillance at a university health system during the 2009 H1N1 influenza pandemic. Am J Infect Control 2012 Sep;40(7):606-610. [doi: 10.1016/j.ajic.2011.12.009] [Medline: 22418609]
35. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. "Right time, right place" health communication on Twitter: value and accuracy of location information. J Med Internet Res 2012;14(6):e156 [FREE Full text] [doi: 10.2196/jmir.2121] [Medline: 23154246]

Abbreviations

API: application programming interface
CDC: Centers for Diseases Control and Prevention
GIS: geographic information system
GPS: Global Positioning System
ILI: influenza-like illness
MMWR: Morbidity Mortality Weekly Report
VISION: Visualizing Information Space in Ontological Networks
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.05.13; peer-reviewed by F Chen, O Leal Neto, T Hernández; comments to author 11.06.13; revised version received 06.08.13; accepted 16.09.13; published 26.10.13.

Please cite as:
Nagel AC, Tsou MH, Spitzberg BH, An L, Gawron JM, Gupta DK, Yang JA, Han S, Peddecord KM, Lindsay S, Sawyer MH
The Complex Relationship of Realspace Events and Messages in Cyberspace: Case Study of Influenza and Pertussis Using Tweets
J Med Internet Res 2013;15(10):e237
URL: http://www.jmir.org/2013/10/e237/
doi:10.2196/jmir.2705
PMID:24158773

Original Paper

Integrating Patient Reported Outcomes With Clinical Cancer Registry Data: A Feasibility Study of the Electronic Patient-Reported Outcomes From Cancer Survivors (ePOCS) System

Laura Ashley[1], PhD; Helen Jones[2], RN; James Thomas[3], BSc; Alex Newsham[2], MA; Amy Downing[2], PhD; Eva Morris[2], PhD; Julia Brown[4], MSc; Galina Velikova[2], BMBS, PhD; David Forman[5], PhD; Penny Wright[2], PhD

[1]School of Social, Psychological and Communication Sciences, Faculty of Health and Social Sciences, Leeds Metropolitan University, Leeds, United Kingdom

[2]Leeds Institute of Cancer and Pathology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

[3]National Cancer Intelligence Network, London, United Kingdom

[4]Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, United Kingdom

[5]Section of Cancer Information, International Agency for Research on Cancer, World Health Organization, Lyon, France

Corresponding Author:
Laura Ashley, PhD
School of Social, Psychological and Communication Sciences
Faculty of Health and Social Sciences
Leeds Metropolitan University
Calverley Building, City Campus
Leeds, LS1 3HE
United Kingdom
Phone: 44 (0)113 8124970
Fax: 44 (0)113 206 8512
Email: l.j.ashley@leedsmet.ac.uk

Abstract

Background: Routine measurement of Patient Reported Outcomes (PROs) linked with clinical data across the patient pathway is increasingly important for informing future care planning. The innovative electronic Patient-reported Outcomes from Cancer Survivors (ePOCS) system was developed to integrate PROs, collected online at specified post-diagnostic time-points, with clinical and treatment data in cancer registries.

Objective: This study tested the technical and clinical feasibility of ePOCS by running the system with a sample of potentially curable breast, colorectal, and prostate cancer patients in their first 15 months post diagnosis.

Methods: Patients completed questionnaires comprising multiple Patient Reported Outcome Measures (PROMs) via ePOCS within 6 months (T1), and at 9 (T2) and 15 (T3) months, post diagnosis. Feasibility outcomes included system informatics performance, patient recruitment, retention, representativeness and questionnaire completion (response rate), patient feedback, and administration burden involved in running the system.

Results: ePOCS ran efficiently with few technical problems. Patient participation was 55.21% (636/1152) overall, although varied by approach mode, and was considerably higher among patients approached face-to-face (61.4%, 490/798) than by telephone (48.8%, 21/43) or letter (41.0%, 125/305). Older and less affluent patients were less likely to join (both *P*<.001). Most non-consenters (71.1%, 234/329) cited information technology reasons (ie, difficulty using a computer). Questionnaires were fully or partially completed by 85.1% (541/636) of invited participants at T1 (80 questions total), 70.0% (442/631) at T2 (102-108 questions), and 66.3% (414/624) at T3 (148-154 questions), and fully completed at all three time-points by 57.6% (344/597) of participants. Reminders (mainly via email) effectively prompted responses. The PROs were successfully linked with cancer registry data for 100% of patients (N=636). Participant feedback was encouraging and positive, with most patients reporting that they found ePOCS easy to use and that, if asked, they would continue using the system long-term (86.2%, 361/419). ePOCS was not administratively burdensome to run day-to-day, and patient-initiated inquiries averaged just 11 inquiries per month.

Conclusions: The informatics underlying the ePOCS system demonstrated successful proof-of-concept – the system successfully linked PROs with registry data for 100% of the patients. The majority of patients were keen to engage. Participation rates are likely to improve as the Internet becomes more universally adopted. ePOCS can help overcome the challenges of routinely collecting PROs and linking with clinical data, which is integral for treatment and supportive care planning and for targeting service provision.

(J Med Internet Res 2013;15(10):e230) doi:10.2196/jmir.2764

KEYWORDS

cancer; oncology; patient reported outcomes; patient reported outcome measures; health-related quality of life; survivorship; cancer registry; electronic data capture; health information technology; Internet

Introduction

In recent decades, the number of people living with and beyond cancer has increased substantially [1]. Although there is increasing understanding of survivorship outcomes, these will not remain static as complex new treatments with unknown long-term effects are introduced, and as the proportion of older survivors with comorbid health and social care problems increases [2]. In addition, new models for follow-up are being encouraged that will include fewer patients being reviewed in hospital and more patients self-managing their care [3]. The escalating costs of cancer care in times of fiscal tightness augur challenging decisions for service planners [4]. These decisions must be determined by up-to-date real-world evidence, and this is increasingly likely to include patient reported outcomes (PROs) [5-7]. PROs may be collected to evaluate survivors' reintegration in society, long-term needs, support requirements, and quality of life and have application in multiple arenas: macro (population surveillance), meso (cancer service delivery), and micro (individual patient care) [8]. Health and social care providers need to find sustainable, cost-efficient methods for collecting PROs regularly, routinely, and at scale from across the whole patient pathway in order to inform the evaluation of future treatments and service planning. It is vital that providers also find a means to efficiently and reliably link PROs to patients' clinical and treatment data, to help identify clinical predictors of survivorship difficulties, and thus facilitate risk stratification and targeted service provision.

Cancer registries and increasingly electronic health records (EHR) [9,10] provide clinical, treatment, and some sociodemographic data but do not routinely include PROs. A number of large-scale mailed surveys have reported cancer survivors' functional and psychosocial well-being, lifestyle behaviors, and supportive care needs with some of the surveys using cancer registries for identification of survivors [11]. Traditionally, cancer registries have been used to record incidence, prevalence, and survival using data collected prospectively for all cancer patients. The role of registries is evolving as registration data are now being linked with other large electronic datasets, providing a rich source of population-based data to inform service planning [12]. A recent review has demonstrated increasing use of cancer registries in quality of life studies of cancer survivors worldwide [13].

Online surveys are an obvious way forward (inexpensive and widely used), but the challenges to their use in health care include ensuring the involvement of all patients, obtaining meaningful patient consent, combining PROs with medical information, and maintaining data security. At the Eindhoven cancer registry in the Netherlands, an online system has been successfully established, complementing a mailed paper questionnaire alternative, to collect and link PROs data to registry data from patients identified via the registry post registration: the Patient Reported Outcomes Following Initial treatment and Long-term Evaluation of Survivorship (PROFILES) system [14-16]. In order to collect PROs across the whole patient pathway, patients need to be recruited soon after diagnosis and thus pre-registration. One strategy for identifying and recruiting patients close to diagnosis is via hospitals. Use of electronic PROs early in the patient pathway in cancer outpatient consultations has been shown to be feasible, acceptable, and beneficial for patients [17-21]. Online systems have also been used, with some success, for remote monitoring of patients on follow-up [22]. Web-based PROs systems linked to EHRs/cancer registries that patients consent to join close to diagnosis in the clinical setting, potentially offer a sustainable and scalable way forward for routinely collecting and linking PROs with medical data over time. Ongoing increases in Internet usage should help enable this approach. In Great Britain, 73% of adults now use the Internet every day, although only 37% of those over 65 years old use a computer daily [23]. Therefore, although UK-based online PROs data collection systems for home use are technologically achievable, their feasibility would need to be carefully evaluated prior to implementation in regard to patient response rates, as well as reliability and validity of data collected, data security, and administrative burden [24].

We have designed and built a potentially UK-scalable system for administering Patient Reported Outcome Measures (PROMs) online at specified post-diagnostic time-points to patients identified and consented in the clinical setting, for linking and storing the collected PROs data with patients' clinical data in the regional registry, and for semiautomating the associated patient monitoring and correspondence. This is the first such system developed in the United Kingdom and is known as electronic Patient-reported Outcomes from Cancer Survivors (ePOCS). A comprehensive description of the design and development of ePOCS has been published open-access and includes a graphical representation of the system components and data flows as well as details concerning data linkage [25]. The two key components of the ePOCS system are QTool, a custom-designed Web-based password-protected questionnaire administration and management system, and the Tracker, a custom-designed database for monitoring patients' QTool activity and generating study correspondence (eg, invitations to complete questionnaires, reminders), which is housed on a secure registry server. In brief, patients complete PROMs using QTool, which is accessed via a public-facing website (Figures 1-4). The PROMs data are subsequently linked to patients' clinical data transferred from the EHR to the registry and stored in the National Cancer Data Repository. Monitoring of and communications with patients (primarily by email) are semiautomated via the Tracker (Figure 5).

This study aimed to test the technical and clinical feasibility of the novel ePOCS system by running it in two UK National Health Service (NHS) settings over 2 years. Feasibility outcomes included system informatics performance, patient recruitment, retention, representativeness and questionnaire completion (response rate), patient feedback, and administration burden involved in running the system.

Figure 1. Screenshot of the website homepage of the ePOCS system.

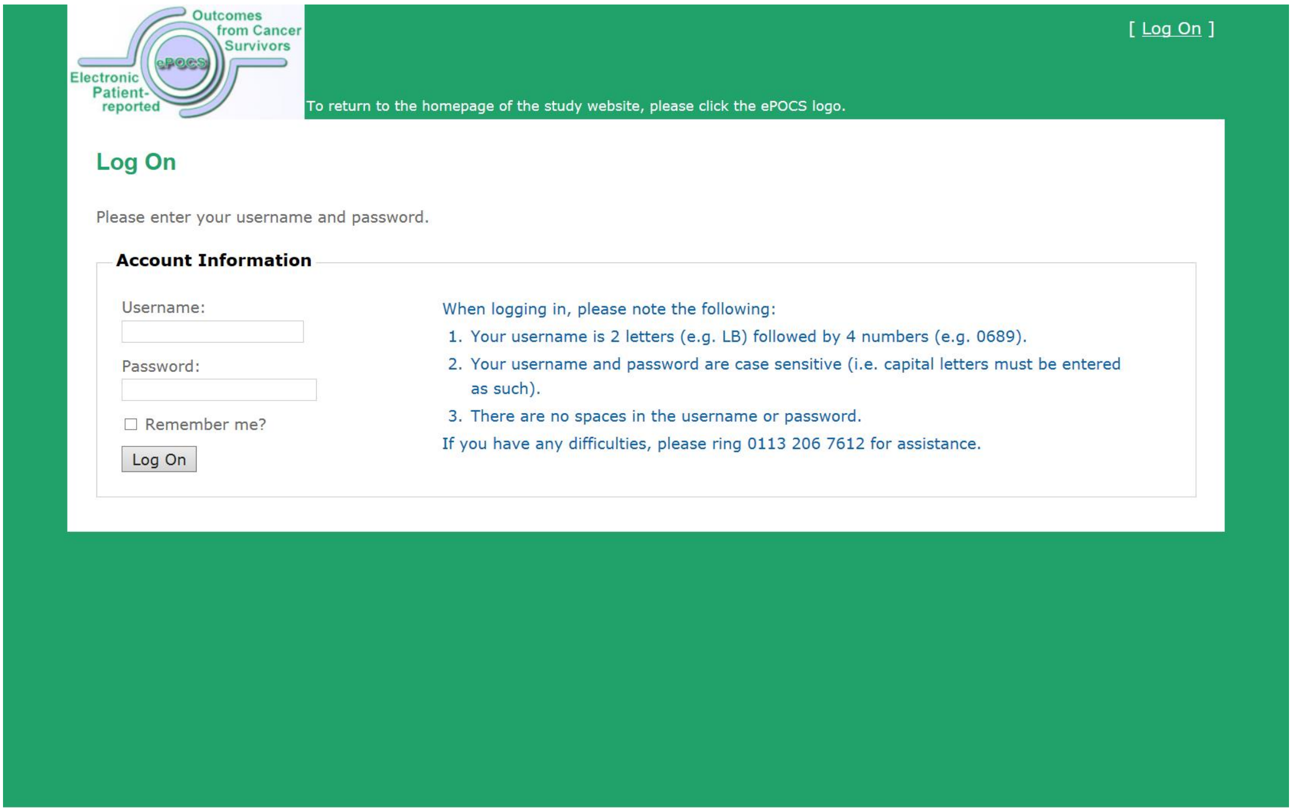
A study to measure the experiences of people living with and beyond cancer

Welcome to the ePOCS Study

If you are taking part in the study and want to access your questionnaires CLICK HERE

636 people have now joined the study

Figure 2. Screenshot of the login page of the ePOCS system.



To return to the homepage of the study website, please click the ePOCS logo.

Log On

Please enter your username and password.

Account Information

Username:

Password:

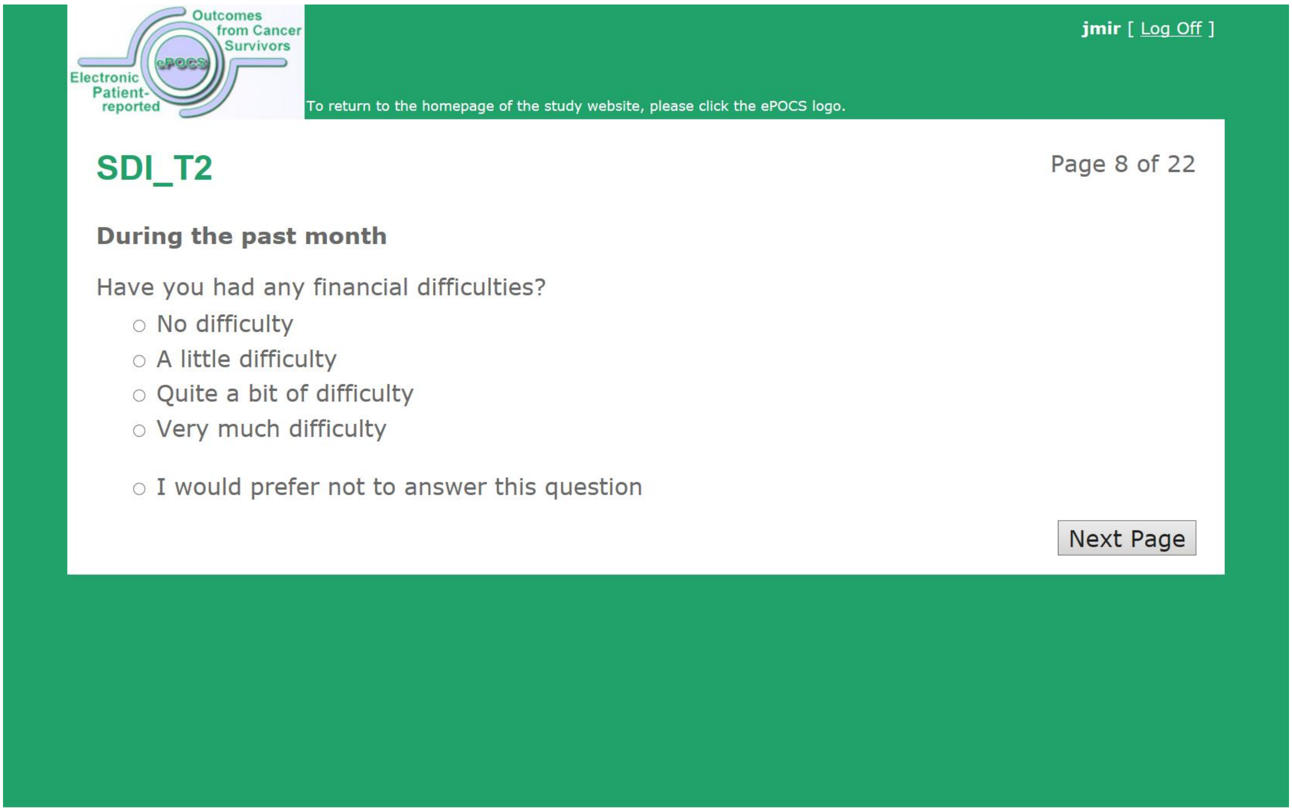
☐ Remember me?

When logging in, please note the following:

1. Your username is 2 letters (e.g. LB) followed by 4 numbers (e.g. 0689).
2. Your username and password are case sensitive (i.e. capital letters must be entered as such).
3. There are no spaces in the username or password.

If you have any difficulties, please ring 0113 206 7612 for assistance.

Figure 3. Screenshot of an ePOCS system questionnaire item (item 7 from the 21-item Social Difficulties Inventory).



To return to the homepage of the study website, please click the ePOCS logo.

SDI_T2

Page 8 of 22

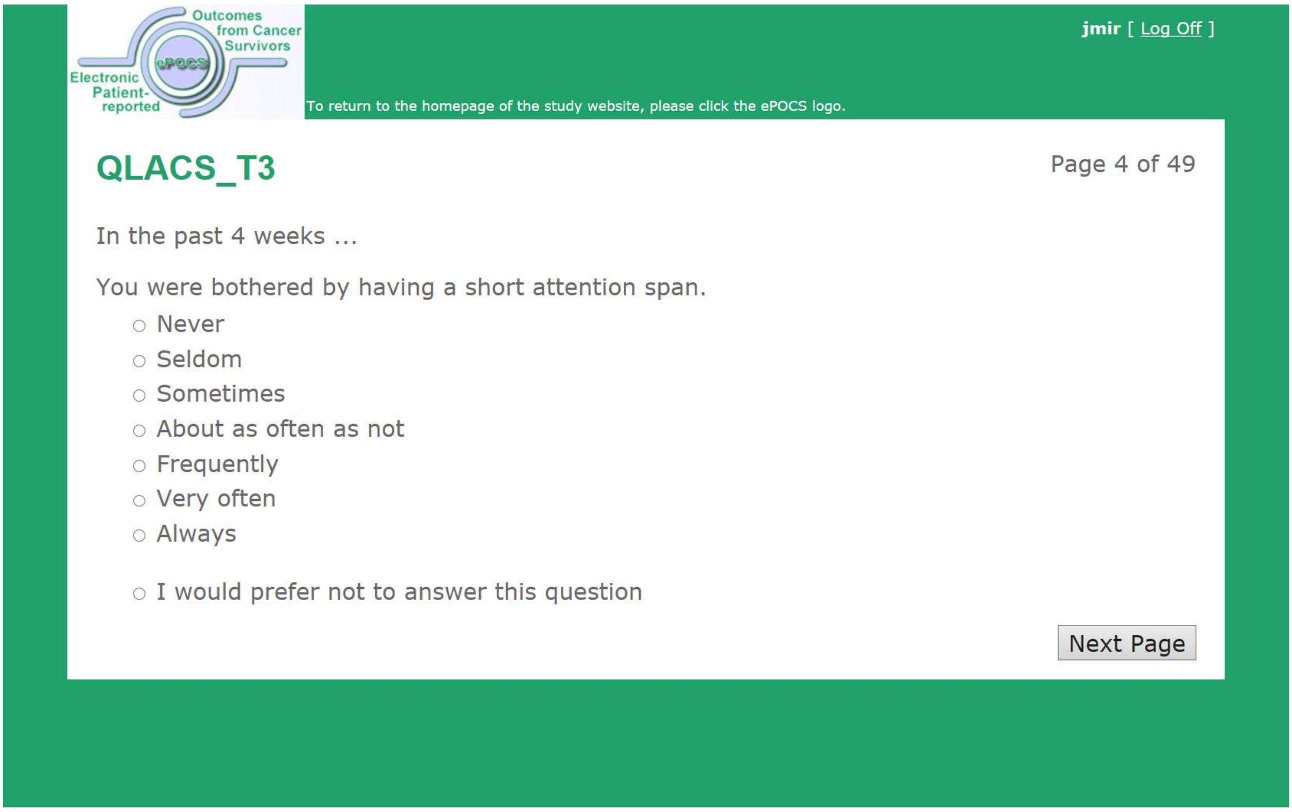
During the past month

Have you had any financial difficulties?

○ No difficulty
○ A little difficulty
○ Quite a bit of difficulty
○ Very much difficulty

○ I would prefer not to answer this question

Figure 4. Screenshot of an ePOCS system questionnaire item (item 3 from the 47-item Quality of Life in Adult Cancer Survivors scale).



To return to the homepage of the study website, please click the ePOCS logo.

QLACS_T3

Page 4 of 49

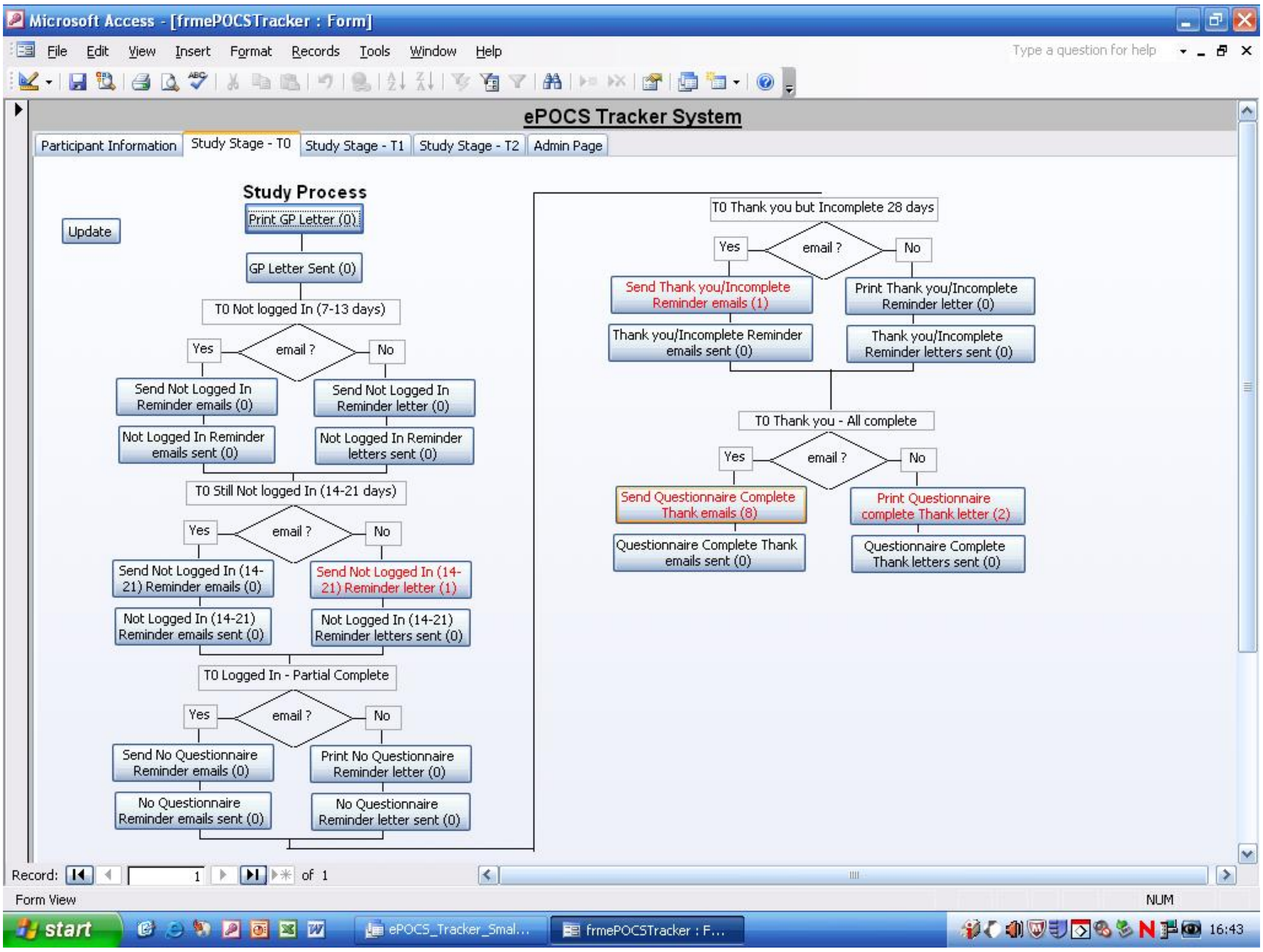
In the past 4 weeks ...

You were bothered by having a short attention span.

- ○ Never
- ○ Seldom
- ○ Sometimes
- ○ About as often as not
- ○ Frequently
- ○ Very often
- ○ Always

○ I would prefer not to answer this question

Figure 5. Screenshot of part of the ePOCS system Tracker, used to generate and send required daily patient correspondence (due invitations, reminders, etc, appear in red).



Methods

Overview

Following ethical approval from the NHS Leeds (East) Research Ethics Committee (ref. 10/H1306/65), a prospective, repeated-measures feasibility study was run in the Yorkshire Cancer Network (YCN) Cancer Centre and one YCN Cancer Unit, in the United Kingdom. The comprehensive protocol for the feasibility study has been published open-access [26].

Patients

Adult patients were eligible if diagnosed with potentially curable breast, colorectal, or prostate cancer within the last 6 months and were English literate. The target was to approach all eligible patients during the recruitment period (November 2010 to September 2011).

Recruitment Procedures

Eligible patients were identified during discussions in routine multidisciplinary meetings and/or through consultation of medical notes by NHS research nurses and/or oncology clinicians, who then initially approached patients about study participation. Wherever feasible, patients were approached and informed about the study in-person, typically during a routine hospital appointment. Where this was not possible (eg, patient missed their appointment), patients were sent a letter about the study signed by their consultant, or were sometimes telephoned if the patient knew the recruiting research nurse/clinician. Recruiting research nurses and clinicians completed a paper form for each approached patient, on which they recorded, among other things, the mode and location of approach. Participants provided written informed consent, and their consent status was recorded in the EHR [27] by the recruiting research nurse/clinician who also provided participants with their ePOCS username and password. Patients who chose not to join the study were not required to provide a reason why, but where patients volunteered a reason, this was recorded by the recruiting research nurse/clinician on the paper form (as was “reason not given”). After consent, participants were followed up by the ePOCS research team.

Follow-Up Procedures

When joining the study, patients were asked to provide an email address that was used for all study correspondence (eg, invitations to complete questionnaires, reminders). For patients who did not provide an email address, all study correspondence was mailed. Participants were asked to complete questionnaires comprising multiple PROMs within 6 months of diagnosis (T1), and at 9 months (T2) and 15 months (T3) post diagnosis. At T2 and T3, participants had up to 6 weeks to complete the questionnaire. The PROMs chosen for each questionnaire were typical of those likely to be used in future applications of the system and covered various psychosocial and quality-of-life issues ([28-35]; see Measures section below). The total number of questionnaire items that participants were asked to complete ranged between 80 and 154, dependent on time-point and cancer site. At each time-point, a maximum of three email/letter reminders were sent, and patients received a communication thanking them for their participation (for those with any outstanding PROMs, this included notification of the questionnaire closing date). Prior to contacting participants at T2 and T3, the ePOCS research team verified patient health status.

A pen-and-paper feedback questionnaire devised by the researchers and a prepaid addressed envelope were mailed to all retained participants post T3. The feedback questionnaire contained a mix of 28 closed and open questions chiefly covering ease of use of various aspects of the system at the different time-points, perceived positive aspects of the system, and suggestions for system improvement. The full feedback questionnaire is reproduced in Multimedia Appendix 1. The questionnaire was mailed in order to keep feedback on the system distinct from the system itself and to encourage as wide and representative a response as possible, by facilitating inclusion of patients who, although in the study, did not engage or were no longer engaging with the online ePOCS system.

Throughout the study, the ePOCS research team diligently maintained a “patient contact” log of all patient-initiated inquiries to the team. For each inquiry (ie, each instance of contact), among other things, the date, mode of communication (eg, email, telephone, letter), and a detailed reason for the contact were recorded on a database. The ePOCS system Tracker automatically generates all the administrative actions due each day (ie, all the patient invitations, reminders, thank yous, etc, that need sending on that date). In order to test that the Tracker was correctly generating all the necessary correspondence, the research team also manually worked out all the administrative actions due each day for 6 months at the study start and when the first participants reached T2 and T3.

Measures

Illness Perception Questionnaire-Revised

The Illness Perception Questionnaire-Revised (IPQ-R) [28] assesses patients’ personal beliefs and expectations about their illness (eg, about its controllability and consequences) and comprises nine subscales, eight of which were used in this study. The IPQ-R (minus the omitted “causes” subscale) comprises 38 statements (eg, “my cancer is a serious condition”, “my cancer will improve in time”) rated on a scale of 1 (strongly disagree) to 5 (strongly agree), and 14 symptoms (eg, “breathlessness”, “headaches”) rated on a yes/no scale, with respect to patients’ views at the present moment.

EuroQol-5D, Version 2

The EuroQol-5D, version 2 (EQ-5Dv2) [29] is a 6-item generic measure of health status that assesses mobility, self-care, usual activities, pain/discomfort, and anxiety/depression using a 3-option response format according to the severity of problems experienced that day (no problems, some problems, severe problems). The EQ-5Dv2 also includes a visual analogue scale on which health state today is rated from 0 (worst imaginable health state) to 100 (best imaginable health state).

Medical Outcomes Study 36-Item Short-Form Health Survey, Version 2

The Medical Outcomes Study 36-Item Short-Form Health Survey, version 2 (SF-36v2) [30] is a generic measure of health status and functioning that assesses eight domains including physical functioning, pain, and mental health. The measure comprises 36 items (eg, “have you been happy”, “did you feel worn out”) rated on a variety of Likert-type response scales (eg, excellent to poor, all of the time to none of the time), primarily with respect to the past 4 weeks.

Social Difficulties Inventory

The Social Difficulties Inventory (SDI-21) [31] assesses various everyday difficulties commonly experienced by cancer patients, including relationship difficulties, domestic problems, and financial worries; 21 questions (eg, “have you felt isolated”, “have you had any financial difficulties”) are answered on a 0 (no difficulty) to 3 (very much) scale with respect to the past month.

European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire

The European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ) (EORTC-QLQ) [32-34] is a cancer-specific measure assessing health-related functioning and symptoms that includes a generic core questionnaire and numerous diagnosis specific modules. This study used the breast (EORTC-QLQ-BR23), colorectal (EORTC-QLQ-CR29), and prostate (EORTC-QLQ-PR25) modules, each of which contain between 23 and 29 questions. All EORTC-QLQ items (eg, “did you have a dry mouth”, “has weight gain been a problem for you”) are rated on a scale of 1 (not at all) to 4 (very much) with respect to the past week or month.

Quality of Life in Adult Cancer Survivors Scale

The Quality of Life in Adult Cancer Survivors (QLACS) scale [35] measures health-related quality of life in seven generic and five cancer-specific domains, including cognitive problems, social avoidance, and appearance. QLACS comprises 47 items (eg, “you felt tired a lot”, “you had difficulty doing activities that require concentrating”) rated on a scale of 1 (never) to 7 (always) with respect to the past 4 weeks.

Psychometric information about the measures and further associated references are provided in the published study protocol [26], as well as information about other study questions that are not part of a standard validated PROM (eg, questions about employment status, use of health, and social services). Patients were asked to provide information about their ethnicity, relationship status, and level of education in the T1 questionnaire, and other sociodemographic details (eg, gender, age, postcode) and clinical information (eg, date and type of cancer diagnosis, treatment regimens) were obtained from participants’ medical records (following their explicit permission, recorded on the consent form).

Study Outcomes

Informatics Performance: How Successful Are the ePOCS System Informatics?

Technical success and reliability were evaluated by (1) calculating the proportion of patients with successful linkage of ePOCS PROs data and registry data, (2) comparing, and subsequently exploring any discrepancy between, the manually worked out administrative actions due each day (eg, the invitations, reminders required) with those generated automatically by the system Tracker, and (3) examining the number and type of information technology (IT)-related patient inquiries recorded in the “patient contact” log.

Recruitment and Representativeness: Do All Patients Join Up to Use the ePOCS System?

Recruitment (ie, consent rate [CR]) was assessed by calculating the proportion of eligible patients recruited relative to all eligible patients approached. Potential differences in CR by mode of approach (face-to-face, letter, telephone), and location (Cancer Centre, Cancer Unit) were also explored. The representativeness of recruited patients was assessed by examining differences in sociodemographic and clinical characteristics between eligible consenting patients and eligible approached patients who did not join the study. The types and frequency of reasons for nonparticipation, recorded by the consenting research nurses and clinicians, were also analyzed.

Retention, Representativeness, and Questionnaire Completion: Do All Patients Complete ePOCS Questionnaires Fully and Repeatedly Over Time?

Retention was assessed by calculating the proportion of consented patients still in study relative to all consented patients, and the representativeness of retained patients was assessed by examining differences in sociodemographic and clinical characteristics between patients who remained in study and patients who withdrew from the study.

Questionnaire completion, or the response rate (RR), was assessed at all 3 time-points in two ways: RR1 is the number of fully and partially completed questionnaires / all eligible patients approached minus those who have died, and RR2 is the number of fully and partially completed questionnaires / all eligible consented patients minus those who have died.

We defined a fully completed questionnaire as one in which all the items have been answered (ie, responded to, as patients could choose to answer that they “prefer not to answer”), and a partially completed questionnaire as one in which less than all of the items have been answered (ie, one or more of the items had no response).

Associations between patient characteristics and questionnaire completion were explored.

For each PROM at each time-point, the proportion of missing data, median completion time, and psychometric reliability were also assessed. Missing data were calculated as the number of “prefer not to answer” item responses within the total PROM dataset (number of items in PROM multiplied by the number of patients who *fully* completed the PROM).

Patient Feedback: What Do Patients Think About Providing Data via ePOCS?

Participant opinion regarding ePOCS was evaluated from the post-T3 feedback questionnaire. Closed questions were analyzed using proportions. Free-text comments were read by the ePOCS research team (HJ, LA, PW) and following discussion key themes were agreed. The text was imported into QSR NVivo 9 with the main coding undertaken by HJ. Coding consistency, coding saturation, and consensus discussions were undertaken by PW and HJ. Quotes were grouped and examples chosen to best represent the majority opinion for each theme.

Administration Burden: Is It Administratively Onerous to Run the ePOCS System?

Administrative burden was assessed by examining (1) the successful functioning of the Tracker system in automatically generating the administrative actions due each day (eg, the invitations, reminders required) (see feasibility outcome 1), as this minimizes workload, (2) proportion of patients providing an email address for study correspondence, as this too reduces workload compared to printing and mailing study correspondence, and (3) the dates, types, and frequency of patient-initiated inquiries recorded in the "patient contact" log.

Quantitative data were analyzed using IBM-SPSS Statistics-19. Group differences were examined using chi-square tests, *t* tests, and binary logistic regression (alpha=0.05). Socioeconomic status was determined using Index of Multiple Deprivation (IMD) scores and quintiles calculated from patients' postcodes obtained from their medical records (February 2012 release) [36]. PROMs internal reliability was assessed using Cronbach alpha (≥0.70 acceptable).

Results

Informatics Performance: How Successful Are the ePOCS System Informatics?

The ePOCS system successfully linked PROs data with clinical registry data for 100% of patients (N=636). Two key problems were identified in the day-to-day running of ePOCS from the comparison of the manually worked out administrative actions due each day (eg, the number of invitations, reminders required and for which particular patients) with those generated automatically by the system Tracker. Some programming updates to the EHR that were not notified in advance to the ePOCS team affected registry data transfers, resulting in void or inaccurate actions generated in the Tracker database. These were resolved quickly. The second problem concerned date of definitive diagnosis. The time-points for questionnaire completion were determined from patients' date of diagnosis at the time of consent, which was entered into QTool (the questionnaire administration component of the ePOCS system) to guide the timing of questionnaire administration for each patient. However, within the hospital EHR, patient diagnoses can change following diagnostic tests, and when this happened, a new diagnosis date was transferred to the Tracker (the patient monitoring and correspondence component of the ePOCS system), which was different from the original diagnosis date entered into QTool, thus causing QTool-Tracker asynchrony. This resulted in 8 (0.7%) missed invitations of 1227 due. Additional system programming prevented the "original" diagnosis date taken at the time of consent and used in QTool from being overwritten in the Tracker, thus resolving the problem. The majority of IT-related inquiries from participants using the system (n=86) concerned issues with logging on, and notably, confusion between similar-looking (eg, zero/letter o) and case-sensitive letters/numbers in usernames/passwords.

Recruitment and Representativeness: Do All Patients Join Up to Use the ePOCS System?

Of 1152 eligible patients approached, 636 consented to participate (55.21%). Patient recruitment is detailed in Figure 6. The most effective recruitment strategy was face-to-face in clinic (61.4%, 490/798) compared with letter (41.0%, 125/305), and telephone (48.8%, 21/43). For 6 patients the mode of approach was not recorded. Recruitment was higher at the Cancer Centre (61.1%, 510/835) than at the Unit (39.7%, 126/317), and there was a significant association between recruitment strategy and location, with letters employed more frequently at the Unit (38.3%, 121/316) than Centre (22.2%, 184/830) (all three χ^2, $P<.001$).

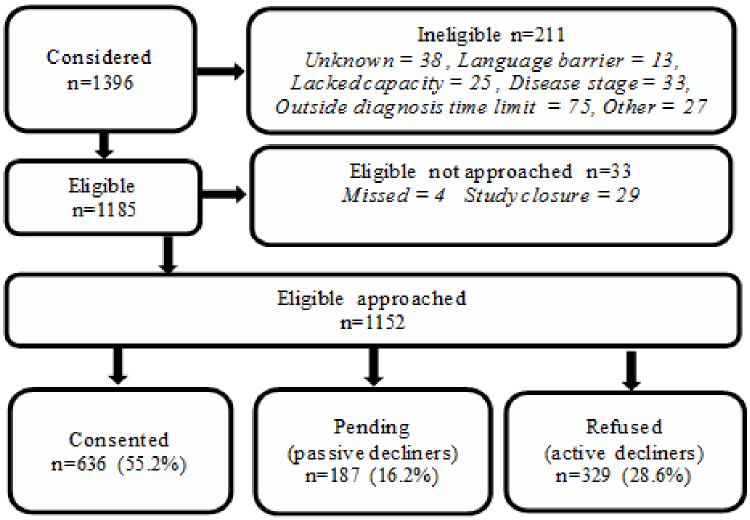
Participants (mean 61.3, SD 11.09 years) were significantly younger than declining patients (mean 66.0, SD 12.05 years; $t_{1150}=-6.903$, $P<.001$), and significantly more affluent (missing value=1; $\chi^2_4=22.106$, $P<.001$, n=1151). No differences were found by gender ($P=.88$), diagnosis ($P=.21$) or time post diagnosis ($P=.21$). Only active decliners had the opportunity to provide a reason for declining participation. Of these, 61/329 (18.5%) provided no reason for nonparticipation. The majority (71.1%, 234/329) gave IT reasons for nonparticipation (eg, no computer/Internet access, do not like computers). Participant characteristics are shown in Table 1.

Table 1. Clinical and sociodemographic characteristics of participants.

Characteristic	Cancer Centre n=510 n (%)	Cancer Unit n=126 n (%)	Total N=636 n (%)
Collected at time of consent (N=636)			
Cancer diagnosis			
Breast	228 (44.7)	69 (54.8)	297 (46.7)
Colorectal	170 (33.3)	22 (17.5)	192 (30.2)
Prostate	112 (22.0)	35 (27.8)	147 (23.1)
Gender and age			
Men, median age 66 years (range 23-92)	223 (43.7)	51 (40.5)	274 (43.1)
Women, median age 58 years (range 24-88)	287 (56.3)	75 (59.5)	362 (56.9)
Index of Multiple Deprivation Quintile (1) [a]			
20% most deprived	98 (19.2)	21 (16.8)	119 (18.7)
20-40% most deprived	98 (19.2)	25 (20.0)	123 (19.4)
20% middle deprived	79 (15.5)	17 (13.6)	96 (15.1)
20-40% least deprived	131 (25.7)	38 (30.4)	169 (26.6)
20% least deprived	104 (20.4)	24 (19.2)	128 (20.2)
Email address			
Yes	408 (80.0)	120 (95.2)	528 (83.0)
No	102 (20.0)	6 (4.8)	108 (17.0)
Collected from T1 participant self-report (n=540)			
Ethnicity (1) [a]			
White British	409 (97.6)	114 (95.0)	523 (97.0)
White other	5 (1.2)	5 (4.2)	10 (1.9)
British minority ethnic group	5 (1.2)	1 (0.8)	6 (1.1)
Relationship status			
Single	22 (5.2)	7 (5.8)	29 (5.4)
Married/Co-habiting/Civil partnership	316 (75.2)	95 (79.2)	411 (76.1)
Widowed	43 (10.2)	9 (7.5)	52 (9.6)
Separated/Divorced	26 (6.2)	7 (5.8)	33 (6.1)
Other	13 (3.1)	2 (1.7)	15 (2.8)
Highest educational qualification (15) [a]			
No formal qualifications	109 (26.7)	15 (12.9)	124 (23.6)
School qualifications	103 (25.2)	34 (29.3)	137 (26.1)
University degree/s	82 (20.0)	26 (22.4)	108 (20.6)
Vocational qualification/s	52 (12.7)	13 (11.2)	65 (12.4)
Other	63 (15.4)	28 (24.1)	91 (17.3)
Employment status prior to cancer diagnosis			
Full-time employment	141 (33.6)	35 (29.2)	176 (32.6)
Part-time employment	60 (14.3)	19 (15.8)	79 (14.6)
Homemaker	15 (3.6)	3 (2.5)	18 (3.3)
Retired	187 (44.5)	54 (45.0)	241 (44.6)
Other	17 (4.0)	9 (7.5)	26 (4.8)

[a] Value in parentheses is the number of missing values.

Figure 6. Flow chart of study recruitment.



Retention, Representativeness, and Questionnaire Completion: Do All Patients Complete ePOCS Questionnaires Fully and Repeatedly Over Time?

Almost all participants (95.1%, 605/636) were still enrolled in the study at T3, with 12 deaths and 19 withdrawals accounting for attrition. Participants who withdrew were older (mean 69.7, SD 9.34 years) than those who stayed in study (mean 60.8, SD 10.97 years; t_{622}=3.51, *P*<.001). Reasons for withdrawal included IT-related issues (n=5) and lack of relevance (n=2) (other, n=5; no reason given, n=7).

At T1, 85.1% of invited participants fully or partially completed the questionnaire, and at T2 and T3, this value was 70.0% and 66.3% respectively (see Table 2). Of the 636 consented participants, 597 were invited to complete the questionnaire at all three time-points. The 39 participants not invited to complete the questionnaire at all time-points included those who had actively withdrawn from the study (n=19), who had died (n=12), or for whom there was a technical error and a time-point invitation was not generated (n=8). Of the 597 invited at all time-points, 57.6% (344/597) fully completed the questionnaire at all three time-points. Sixty-four (10.7%) completed no questionnaire items at any of the time-points, and the remaining 31.7% (189/597) completed some proportion of the total number of questionnaire items across the three time-points (ie, were "partial" completers).

Age (*P*=.57), recruitment strategy (*P*=.10), and recruitment location (*P*=.06) were not associated with full questionnaire completion. Patients were more likely to fully complete the questionnaires at all three time-points if they were male (*P*=.02), more affluent (*P*<. 01), were diagnosed with prostate cancer (*P*=.01), or provided an email address (*P*=.02). Of patients who fully completed the questionnaire at T1, 86.7% (431/497) provided an email address, and at T2 and T3, this value was 86.0% (355/413) and 86.2% (338/392) respectively. Entering these variables (missing value=1) into a binary logistic regression analysis (full-completers versus partial and noncompleters) resulted in the IMD quintile being the only significant predictor in the model (χ^2_8=25.41, *P*=.001, n=596), with the three more socioeconomically deprived groups being less likely to fully complete the questionnaires compared with the most affluent group. The percentage of variance explained by the model was minimal (Cox and Snell R square=0.042; 4.2% variance explained).

Of the 344 participants who fully completed the questionnaire at all three time-points, 82 required no reminders (3 missing values) (24.0%, 82/341). Ninety-two required one or more reminders at all three time-points (27.0%, 92/341), and the remaining 167 (49.0%, 167/341) received one or more reminders at one or two time-points. Among participants who fully completed all three questionnaires, age (*P*=.66), gender (*P*=1.0), and IMD quintile (*P*=.11) were not associated with reminder status (ie, received a reminder at any time-point versus needed no reminder at any time-point). Having provided an email address or not was related to reminder status, with those not providing an email address being more likely to require a reminder (missing values=3; χ^2_1=4.750, *P*=.03, n=341).

The rates of missing data (ie, in fully completed PROMs, so patients choosing "I would prefer not to answer this question"), completion times, and internal reliability for each PROM at each time-point are shown in Table 3. Missing data were minimal, ranging from just 0.29% of EQ-5Dv2 items (T1) to a still modest 3.15% of SDI-21 items (T2) and were largely attributable to patients opting not to answer questions about sexual matters. Time taken to complete individual PROMs ranged from a median of 1:24 minutes:seconds (IQR=00:50) for the 6-item EQ-5Dv2 (T3) to 12:46 minutes:seconds (IQR=7:30) for the 66-item IPQ-R (T1). Overall, the PROMs demonstrated acceptable internal reliability.

Patient Feedback: What Do Patients Think About Providing Data via ePOCS?

Feedback questionnaires were sent to 599 of the 605 T3 participants (2 died and 4 withdrew during the T3 window). Feedback was returned by 71.6% of participants (429/599) with most returns from those who had completed all questionnaires (69.9%, 300/429). Most participants reported that they found it *very easy* or *easy* to get to the ePOCS website (item 6) (T1=94.9%, 373/393; T2=98.1%, 352/359; T3=97.9%, 328/335), to log on with their username and password (item 7) (T1=94.1%, 367/390; T2=96.4%, 347/360; T3=97.6%, 323/331), and to get to the questionnaires (item 8) (T1=98.2%, 386/393; T2=91.6%, 329/359; T3=97.9%, 328/335) at all three time-points. Participants who had required help with the system had mainly received this from partners and family. Most participants favored the electronic system over paper questionnaires (item 18) (79.7%, 337/423), and most participants stated that they would *very likely* or *definitely* continue using ePOCS to complete questionnaires for the next 10-15 years if asked (item 17) (86.2%, 361/419).

Most participants responded positively when asked what they liked about the electronic system (item 13) (69.0%, 296/429): “Easy and relaxed, able to complete at your own time, in your own environment” [colorectal cancer patient, male, 62 years old], “It was convenient and easy to use with the option of reviewing answers (given) when required. I liked the option to be able to leave the system but come back to complete later” [prostate cancer patient, male, 65 years old], and “It was easy to use (once I had logged on with help from my husband). I am not very computer literate but could easily use the system” [breast cancer patient, female, 47 years old].

In many cases participants indicated that they preferred ePOCS to a paper system: “It is interactive. I liked receiving an email telling me it was time to complete the questionnaire. I appreciated receiving an email reminding me to complete the questionnaire when I had not done so. I liked the ‘paperless’ system” [colorectal cancer patient, female, 46 years old], and “Very easy to access. Less trouble than using pen and paper and having to post the result” [colorectal cancer patient, male, 73 years old].

However, 27.0% (116/429) of participants did not provide a response to item 13, and a small number of participants made indifferent or negative comments (4.0%:17/429): “I did not have any likes/dislikes about the system, it was like any other questionnaire” [prostate cancer patient, male, 79 years old] and “Nothing at all. I hate computers and prefer a written system like this” [breast cancer patient, female, 56 years old].

For participants who indicated that they would have preferred a paper system (item 18), the reasons mainly concerned lack of computer knowledge, not having to rely on others for help, and finding it easier to get an overview of a whole questionnaire: “No experience of computers and related points” [colorectal cancer patient, female, 88 years old], “I would prefer paper because my daughter has a busy life and can’t always help me and I couldn’t do it myself” [breast cancer patient, female, 51 years old], and “Easier to preview questions and review answers” [colorectal cancer patient, male, 41 years old].

About a third of participants (33.8%, 145/429) commented on how ePOCS might be changed (item 14), with most improvement comments (75.9%, 110/145) concerning the number, type, repetition, and layout of the questions: “Rather too many questions and some feeling of overlap” [breast cancer patient, female, 70 years old].

Although not asked to, some participants volunteered reasons for their participation in the study, and altruism and a sense of belonging to a community were commonly cited: “If it helps in any way to achieve better treatment and after care, I’m all for it” [colorectal cancer patient, male, 76 years old], and “I liked answering the questions because I felt it gave me more of an understanding of my condition and I didn’t feel like it was just me with these symptoms” [breast cancer patient, female, 40 years old].

Administration Burden: Is It Administratively Onerous to Run the ePOCS System?

The Tracker database was easy and quick to use with all required daily correspondence (invitations, reminders, thank-yous) automatically generated and populated with the appropriate patient’s details, ready for sending (after health status was verified). As most participants provided an email address (83.0%, 528/636), sending the required numerous reminders (Table 2) was not onerous. There were proportionately few occasions when participants contacted the ePOCS research team (n=281), averaging 11 contacts per month. The reasons for the patient inquiries are given in Figure 7.

Table 2. Questionnaire completion, reminders sent, and response rates at all time-points (number of items in questionnaire [number varies dependent upon diagnostic group] T1=80, T2=102-108, T3=148-154).

	Time 1 (T1)[a]	Time 2 (T2)[b]	Time 3 (T3)[c]
Invitation to complete the questionnaire not given			
Died	0	5	12
Actively withdrew	0	9	19
Technical error	0	5	3
Invitation to complete the questionnaire given	636	617	602
Questionnaire – fully completed	520	417	394
Reminders sent [d]			
0	238 (45.8%)	209 (50.1%)	208 (52.8%)
1	168 (32.3%)	119 (28.5%)	95 (24.1%)
2	80 (15.4%)	55 (13.2%)	61 (15.5%)
3	31 (6.0%)	32 (7.7%)	28 (7.1%)
Missing	3 (0.6%)	2 (0.5%)	2 (0.5%)
Questionnaire – partially completed	21	25	20
Reminders sent [d]			
0	0 (0%)	0 (0%)	0 (0%)
1	1 (4.8%)	7 (28.0%)	10 (50.0%)
2	6 (28.6%)	8 (32.0%)	9 (45.0%)
3	14 (66.7%)	9 (36.0%)	1 (5.0%)
Missing	0 (0%)	1 (4.0%)	0 (0%)
Questionnaire – no items completed	95	175	188
Reminders sent [d]			
0	0 (0%)	0 (0%)	0 (0%)
1	3 (3.2%)	10 (5.7%)	5 (2.7%)
2	5 (5.3%)	5 (2.9%)	4 (2.1%)
3	87 (91.6%)	160 (91.4%)	177 (94.1%)
Missing	0 (0%)	0 (0%)	2 (1.1%)
Response rate (RR)			
RR1[e]	541/1152 (47.0%)	442/1147 (38.5%)	414/1140 (36.3%)
RR2[f]	541/636 (85.1%)	442/631 (70.0%)	414/624 (66.3%)

[a] T1 window – between date of consent and 6 months post diagnosis.

[b] T2 window – a 6-week window for completion with the midpoint at 9 months post diagnosis.

[c] T3 window – a 6-week window for completion with the midpoint at 15 months post diagnosis.

[d] Reminders were not sent to those who contacted the ePOCS team and actively withdrew after the invitation/reminder was sent.

[e] RR1=number of fully and partially completed questionnaires/all eligible patients approached minus those who died.

[f] RR2=number of fully and partially completed questionnaires/all eligible consented patients minus those who died.

Table 3. Time to complete, missing data, and psychometric reliability for standard validated ePOCS PROMs (in addition to the standard validated PROMs shown here, participants also completed other questions, eg, about sociodemographic information, employment, and the financial costs of cancer).

PROM (n items)		Fully completed, n	Completion time[a], min:sec		Missing data[b]	Internal reliability, Cronbach α[c]	
			Median	Range	Total %	N scales α≥.70	α range
IPQ-R (66)	T1	531	12:46	04:44–410:32	0.80	7/7 (100%)	.78–.90
EQ-5Dv2 (6)	T1	526	01:54	00:39–29:05	0.29	n/a	n/a
	T2	426	01:30	00:31–54:20	0.31	n/a	n/a
	T3	402	01:24	00:29–63:42	0.46	n/a	n/a
SF-36v2 (36)	T2	432	08:31	02:59–39:14	0.39	8/8 (100%)	.83–.95
	T3	400	07:44	02:35–262:37	0.35	8/8 (100%)	.85–.95
SDI-21 (21)	T2	423	03:47	01:17–29:13	3.15	4/4 (100%)	.72–.89
EORTC-QLQ-BR23 (23)	T2	196	03:56	01:24–32:57	2.13	4/5 (80%)	.69–.92
	T3	183	03:36	01:14–40:04	2.68	4/5 (80%)	.64–.92
EORTC-QLQ-CR29 (29)	T2	117	06:35	03:07–712:33	1.38	2/5 (40%)	.45–.90
	T3	104	05:45	02:36–26:12	1.39	3/5 (60%)	.69–.83
EORTC-QLQ-PR25 (25)	T2	117	04:27	02:10–28:41	0.68	3/5 (60%)	.41–.82
	T3	111	04:13	02:01–44:41	1.23	2/5 (40%)	.43–.80
QLACS (47)	T3	407	09:56	03:36–288:52	2.25	12/12 (100%)	.75–.94

[a] Completion time descriptive statistics are based on participants who started and completed a PROM on the same calendar day.

[b] Missing data (ie, patients' choosing to respond "I would prefer not to answer this question") per PROM is based on the number of patients who fully completed that PROM.

[c] Spearman-Brown reliability coefficient for 2-item subscales.

Figure 7. Reasons for participant inquiries to the ePOCS team over the study period.

Discussion

Principal Findings

This study has tested and demonstrated the technical and clinical feasibility of an innovative electronic system for collecting PROs online and linking them with clinical registry data. In general, the results showed that the system informatics performed successfully, demonstrated encouraging rates of patient recruitment, retention, and questionnaire completion, revealed predominately positive feedback from patients, and showed a low administration burden involved in running the system. However, patients who joined and stayed in the study were not wholly representative of all invited/recruited patients.

The informatics underlying the ePOCS system demonstrated successful proof-of-concept. The system successfully linked PROs with registry data for 100% of patients. The set-up work involved in establishing the linkage capacity was undertaken over several weeks during system building by a member of the registry IT team (but took approximately 2 working days compressed). When the system is running, as in the current study, an output of linked data can be instantaneously "pulled off" the system. Importantly, the labor involved in the initial linkage work is not impacted by the number of patients in the system and would therefore remain modest even if use of the system were scaled up considerably. The system also worked efficiently with relatively few day-to-day running problems. Speedy resolution of problems was possible due to the close working relationship developed between all parties: the ePOCS research team, the design teams of the QTool and Tracker components of the ePOCS system, and the EHR and registry IT teams. The modest number of IT-related inquiries from patients mainly concerned confusion with some letters/numbers in usernames and passwords and was easily resolved; importantly, this can be avoided when running the system in the future by more careful design of the composition of usernames and passwords.

Over half the patients invited joined up to use the ePOCS system. The participation rate of 55.21% (636/1152) is encouraging when compared with response rates for mailed cancer patient surveys (31%-64%) [37], although in a recent prospective longitudinal paper-based cancer patient survey using a similar recruitment strategy to ePOCS, participation was as high as 77% [38]. Among participants approached face-to-face in clinic, rather than via letter or telephone, the recruitment rate was considerably higher (61.4%) than the overall rate, indicating, perhaps not surprisingly, that in-person invitations to participate yield the best rates of patient consent. Although it would thus be ideal to employ face-to-face clinic-based recruitment, this is likely to be logistically and financially unfeasible if the system is being used to collect data from very large numbers of patients (eg, nationwide) and/or from numerous geographically spread locations. In these cases, mailed invitations, or a proportion of mailed invitations, will likely have to be used.

Patients who declined to join the feasibility study were older and less affluent (and patients who withdrew from the study were also older) and most commonly cited IT reasons for nonparticipation. Older age and socioeconomic deprivation are characteristics commonly associated with study nonparticipation [39], and these participation biases may have been exacerbated in ePOCS as older age and deprivation are also associated with lower computer/Internet use [23,40]. Encouragingly, there were no differences between consenting and declining patients on the basis of gender, diagnosis, or time post diagnosis. In the long term, adoption of the Internet will be almost universal. In the short term, to avoid bias and discrimination, online systems can be used alongside other methods [24,41], such as pen-and-paper, as done in PROFILES [14], where mailed questionnaires complement the electronic system and participation rates have been around 70% [15,16], or the telephone, as tested for use in individual patient management [42].

Almost all invited participants fully or partially completed all the questionnaires (89.3%), despite the relatively large number of items at each time-point (ie, between 80 and 154). Deprivation was negatively associated with questionnaire completion, and IT issues may also have influenced this, although the low number of feedback questionnaires returned from non/partial-responders makes this impossible to determine. Receiving reminders led to improved completion of questionnaires, although the impact diminished somewhat at T2 and T3. Among participants who fully completed all three questionnaires, patients who did not provide an email address were more likely to need a reminder. Participants who receive invitations in the mail have to make a special effort to go to a computer, boot up, and find the ePOCS website prior to logging on to complete questionnaires, whereas participants who receive an email invitation are already online and can simply use the weblink in the email to log on and complete questionnaires. The time taken to complete individual PROMs varied considerably (eg, 6-item EQ-5Dv2 at T1 ranged from 00:39 to 29:05 minutes:seconds), suggesting that some participants broke off partway through to complete other activities. As has been observed elsewhere (eg, [43]), the median completion time for PROMs given at multiple time-points decreased with repeated administrations (eg, EQ-5Dv2, SF-36v2). This may be due to a learning effect, although as the time period between questionnaires is considerable, it may also be a result of less Internet-confident participants opting not to respond at later time-points, thus reducing the overall median completion times.

Participant feedback was generally positive and endorsed the ePOCS approach. Over 90% of participants found it *very easy* or *easy* to access the ePOCS website, to log on, and to access the questionnaires, at all three time-points. Encouragingly, most respondents preferred the online ePOCS system to pen-and-paper questionnaires, and the most common suggestions for changing the system were concerned with the questionnaires/PROMS administered (eg, number and repetition of items) rather than the system itself. Impressively, 86.2% of feedback respondents indicated that, if asked, they would likely continue using the ePOCS system to complete questionnaires in the long term. It must be kept in mind, however, that feedback questionnaires were administered only to patients who joined the study, and almost 70% of the returned feedback was from patients who had engaged with the system and completed all questionnaires. The positive feedback does not therefore provide a full and balanced picture of what all patients would think about such a system, and the results must be extrapolated cautiously.

The ePOCS system was not administratively onerous to run. The Tracker successfully automatically generated all the administrative actions due each day (eg, the invitations, reminders required). Checking the Tracker daily and sending all the due correspondence, which was mostly via email, took a member of the research team between just 15 and 30 minutes. In the ePOCS study, 2 members of the research team shared this task (to allow for sickness/holiday absence). The number of participant inquiries received over the course of the study averaged a modest 11 per month, even with 636 patients in study, indicating that the time required to provide support to patients using the system is not burdensome. This is similar to the experience of the PROFILES system in which only around 2% of patients contact the PROFILES helpdesk [14]. Like any such e-system, ePOCS could never be entirely automated, but administration is relatively undemanding and could be run day-to-day by trained administrative assistants. Importantly, unlike a paper-based counterpart, ePOCS would remain administratively undemanding and affordable if scaled up. The low cost of online questionnaires is clearly demonstrated in a study investigating the cost of survey response by mode of administration; sizeable differences were found in the administrative costs of paper-based, Web-based, and mixed-mode surveys with estimates of costs per RR2 (as defined in ePOCS) of US $4.78, $0.64, and $3.61 respectively [44].

ePOCS provides an infrastructure to routinely collect and link PROs to clinical and cancer registration data. Preparatory ePOCS work indicated patients' disinclination to consent when critical treatment decisions are being considered (not tested in this study) or close to the time of diagnosis [45]. Reluctance to consent near diagnosis was not confirmed in this study. Therefore, patients may be asked very early in their cancer pathway (ideally face-to-face by their clinical team) to provide PROs, as long as particular consideration is given at times of critical decision making. If ePOCS or a similar system were to be introduced from diagnosis onwards, the PROs data would provide a real-time feed of the patient/survivor experience to supplement data from other existing sources [46]. Additional programming could enable linkage to other EHRs, registries, or to the new English National Cancer Online Registration Environment (EnCORE), which pulls patient-level data from several local and national feeds.

IT is playing an increasing role in the delivery of high-quality cancer care [47]. Feasibility of online PROs assessment for use in individual patient management in clinical practice has been demonstrated in two recent studies. In a study of online toxicity reporting from home during routine chemotherapy [48], the participation rate was 75%, and on average monthly compliance was 83% and weekly compliance 62%. Although the consent rate of 75% was considerably higher than the ePOCS consent rate of 55.21% (636/1152), this may be accounted for by the exclusion of non-Internet users in the study eligibility criteria. In a study using an online system for collecting PROs in between clinic visits [20], the participation rate was 68% and patients completed a median of 71% of assigned questionnaires. Although this consent rate is also higher than for ePOCS, patients had the option to complete the questionnaires in clinic using a laptop provided. The ePOCS system was designed and built to link PROs with clinical and cancer registration data. The approach could be used for data collected for clinical purposes with a transfer of PROs to registries along with other clinical data, subject to governance approvals. For example, a system such as ePOCS could be introduced in routine practice for individual patient management, close to diagnosis and during treatment, as used in many centres [18]; be used in personalized medicine, providing real-time data informing treatment and symptom management strategies [19]; or be used in follow-up where remote monitoring to identify patients' late treatment effects and supportive care needs would form part of risk-stratified pathways of care [3,49]. Use of the ePOCS system in clinical practice would, however, require substantial additional software programming to enable live linkage with the EHR, and the development of a training program for staff as well as protocols/algorithms for the use and interpretation of PROs by clinicians. In addition, PROMs used in patient care tend to be different in scope and psychometrics to those needed for epidemiological research, and some may not be interchangeable [7,50,51]. The PROMs used in this study were chosen for their applicability in survivorship research and performed well. If ePOCS were rolled out for individual survivorship management in addition to epidemiological data collection, a mixed combination of carefully reviewed PROMs will have to be agreed upon.

Strengths and Limitations

To our knowledge, this is the first study to test and report on the feasibility of a system like ePOCS. Strengths of the study include multisite recruitment, the large number of patients invited to participate (over 1000), and the examination of patients' use of the system on multiple, longitudinal occasions. We also consider an important strength to be the focus on technical and administrative feasibility which, alongside patient consent and response rates, are of key importance for those seeking to use and run a system like ePOCS.

The principal limitation of the study is that we were unable to test the system in a more authentic context. Unavoidably, in order to obtain patients' informed consent, the current study was presented to patients as being about the ePOCS system. However, future studies that use the ePOCS system will be presented to patients with emphasis on the PROs to be collected and analyzed, and the ePOCS system will be mentioned only secondarily as the data collection tool. It will be important to examine patient recruitment, retention, and response rates in future PROs studies that simply use the ePOCS system rather than aim to test it. The second notable limitation concerns the minimal feedback obtained from patients who declined participation and who consented but did not complete questionnaires. In health information technology (HIT) research generally there is a lack of studies exploring the perspectives and experiences of patients who choose not to engage with HIT and/or who withdraw participation. We are currently planning a study with such patients, aimed at understanding and overcoming modifiable barriers to patients' acceptance and use of HIT.

Conclusions

Routine collection of PROs is integral for planning patient-centred, compassionate, and personalized health care. This study has shown that the ePOCS system performs well, is accepted by the majority of patients, and is an efficient means to collect and collate PROs data at scale. Although IT usage is not currently universal, every year more patients will become Internet users. Until then, and for those who choose not to engage with e-systems, conventional alternatives will also have to be offered. This should not hold back plans for introducing systems such as ePOCS, as the majority of patients are keen to engage and provide information they believe will help future cancer patients.

Acknowledgments

This study was funded by Macmillan Cancer Support, with research nurses supported by the West Yorkshire Comprehensive Local Research Network, and sponsored by the University of Leeds. We thank the patients who participated in this study, the ePOCS research nurses, the study steering group, Johnathan Joffe for clinical support, and the Psychosocial Oncology and Clinical Practice Research Group.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Copy of the feedback questionnaire used to obtain patient opinion on the ePOCS system.

[PDF File (Adobe PDF File), 374KB - jmir_v15i10e230_app1.pdf]

References

1. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, ICBP Module 1 Working Group. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet 2011 Jan 8;377(9760):127-138 [FREE Full text] [doi: 10.1016/S0140-6736(10)62231-3] [Medline: 21183212]
2. Parry C, Kent EE, Mariotto AB, Alfano CM, Rowland JH. Cancer survivors: a booming population. Cancer Epidemiol Biomarkers Prev 2011 Oct;20(10):1996-2005 [FREE Full text] [doi: 10.1158/1055-9965.EPI-11-0729] [Medline: 21980007]
3. Jefford M, Rowland J, Grunfeld E, Richards M, Maher J, Glaser A. Implementing improved post-treatment care for cancer survivors in England, with reflections from Australia, Canada and the USA. Br J Cancer 2013 Jan 15;108(1):14-20 [FREE Full text] [doi: 10.1038/bjc.2012.554] [Medline: 23257892]
4. Sullivan R, Peppercorn J, Sikora K, Zalcberg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. Lancet Oncol 2011 Sep;12(10):933-980. [doi: 10.1016/S1470-2045(11)70141-3] [Medline: 21958503]
5. Department of Health, Macmillan Cancer Support, NHS Improvement. The National Cancer Survivorship Initiative Vision. London; 2010. URL: http://www.ncsi.org.uk/wp-content/uploads/NCSI-Vision-Document.pdf [accessed 2013-08-13] [WebCite Cache ID 6Iq4G6LOU]
6. Basch E, Abernethy AP, Mullins CD, Reeve BB, Smith ML, Coons SJ, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol 2012 Dec 1;30(34):4249-4255. [doi: 10.1200/JCO.2012.42.5967] [Medline: 23071244]
7. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013 Jan 4:-. [doi: 10.1007/s11136-012-0344-y] [Medline: 23288613]
8. Lipscomb J, Donaldson MS, Hiatt RA. Cancer outcomes research and the arenas of application. J Natl Cancer Inst Monogr 2004(33):1-7. [doi: 10.1093/jncimonographs/lgh038] [Medline: 15504917]
9. Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. Clin Epidemiol 2011;3:259-272 [FREE Full text] [doi: 10.2147/CLEP.S23690] [Medline: 22135501]
10. Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. J Med Internet Res 2005;7(1):e3 [FREE Full text] [doi: 10.2196/jmir.7.1.e3] [Medline: 15829475]
11. Lerro CC, Stein KD, Smith T, Virgo KS. A systematic review of large-scale surveys of cancer survivors conducted in North America, 2000-2011. J Cancer Surviv 2012 Jun;6(2):115-145. [doi: 10.1007/s11764-012-0214-1] [Medline: 22389171]
12. Parkin DM. The evolution of the population-based cancer registry. Nat Rev Cancer 2006 Aug;6(8):603-612. [doi: 10.1038/nrc1948] [Medline: 16862191]
13. Thong MS, Mols F, Stein KD, Smith T, Coebergh JW, van de Poll-Franse LV. Population-based cancer registries for quality-of-life research: a work-in-progress resource for survivorship studies? Cancer 2013 Jun 1;119 Suppl 11:2109-2123. [doi: 10.1002/cncr.28056] [Medline: 23695923]
14. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, Profiles Registry Group. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. Eur J Cancer 2011 Sep;47(14):2188-2194. [doi: 10.1016/j.ejca.2011.04.034] [Medline: 21621408]
15. Mols F, Oerlemans S, Vos AH, Koster A, Verelst S, Sonneveld P, et al. Health-related quality of life and disease-specific complaints among multiple myeloma patients up to 10 yr after diagnosis: results from a population-based study using the PROFILES registry. Eur J Haematol 2012 Oct;89(4):311-319. [doi: 10.1111/j.1600-0609.2012.01831.x] [Medline: 22762785]
16. Mols F, Thong MS, Vissers P, Nijsten T, van de Poll-Franse LV. Socio-economic implications of cancer survivorship: results from the PROFILES registry. Eur J Cancer 2012 Sep;48(13):2037-2042. [doi: 10.1016/j.ejca.2011.11.030] [Medline: 22196035]

17. Wright EP, Selby PJ, Crawford M, Gillibrand A, Johnston C, Perren TJ, et al. Feasibility and compliance of automated measurement of quality of life in oncology practice. J Clin Oncol 2003 Jan 15;21(2):374-382. [Medline: 12525532]
18. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012 Oct;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
19. Hirsch B, Abernethy A. Structured Decision-Making: Using Personalized Medicine to Improve the Value of Cancer Care. JPM 2012 Dec 27;3(1):1-13. [doi: 10.3390/jpm3010001]
20. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901. [doi: 10.1002/pon.3087] [Medline: 22544513]
21. 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]
22. Sebrow D, Lavery HJ, Brajtbord JS, Hobbs A, Levinson AW, Samadi DB. Monitoring validated quality of life outcomes after prostatectomy: initial description of novel online questionnaire. Urology 2012 Feb;79(2):314-319. [doi: 10.1016/j.urology.2011.08.075] [Medline: 22137540]
23. Office for National Statistics. Internet Access - Households and Individuals, Statistical Bulletin. 2013. URL: http://www.ons.gov.uk/ons/dcp171778_322713.pdf [accessed 2013-08-13] [WebCite Cache ID 6Iq4M7f5r]
24. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]
25. Ashley L, Jones H, Thomas J, Forman D, Newsham A, Morris E, et al. Integrating cancer survivors' experiences into UK cancer registries: design and development of the ePOCS system (electronic Patient-reported Outcomes from Cancer Survivors). Br J Cancer 2011 Nov 8;105 Suppl 1:S74-S81 [FREE Full text] [doi: 10.1038/bjc.2011.424] [Medline: 22048035]
26. Ashley L, Jones H, Forman D, Newsham A, Brown J, Downing A, et al. Feasibility test of a UK-scalable electronic system for regular collection of patient-reported outcome measures and linkage with clinical cancer registry data: the electronic Patient-reported Outcomes from Cancer Survivors (ePOCS) system. BMC Med Inform Decis Mak 2011;11:66 [FREE Full text] [doi: 10.1186/1472-6947-11-66] [Medline: 22029686]
27. Newsham AC, Johnston C, Hall G, Leahy MG, Smith AB, Vikram A, et al. Development of an advanced database for clinical trials integrated with an electronic patient record system. Comput Biol Med 2011 Aug;41(8):575-586. [doi: 10.1016/j.compbiomed.2011.04.014] [Medline: 21703607]
28. Moss-Morris R, Weinman J, Petrie K, Horne R, Cameron L, Buick D. The Revised Illness Perception Questionnaire (IPQ-R). Psychology & Health 2002 Jan;17(1):1-16. [doi: 10.1080/08870440290001494]
29. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
30. Ware JE, Kosinski M, Dewey JE. SF-36 health survey manual & interpretation guide. Lincoln, RI: QualityMetric Inc; 2000.
31. Wright P, Smith A, Roberts K, Selby P, Velikova G. Screening for social difficulties in cancer patients: clinical utility of the Social Difficulties Inventory. Br J Cancer 2007 Oct 22;97(8):1063-1070 [FREE Full text] [doi: 10.1038/sj.bjc.6604006] [Medline: 17895887]
32. Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J Clin Oncol 1996 Oct;14(10):2756-2768. [Medline: 8874337]
33. Sprangers MA, te Velde A, Aaronson NK. The construction and testing of the EORTC colorectal cancer-specific quality of life questionnaire module (QLQ-CR38). European Organization for Research and Treatment of Cancer Study Group on Quality of Life. Eur J Cancer 1999 Feb;35(2):238-247. [Medline: 10448266]
34. van Andel G, Bottomley A, Fosså SD, Efficace F, Coens C, Guerif S, et al. An international field study of the EORTC QLQ-PR25: a questionnaire for assessing the health-related quality of life of patients with prostate cancer. Eur J Cancer 2008 Nov;44(16):2418-2424. [doi: 10.1016/j.ejca.2008.07.030] [Medline: 18774706]
35. Avis NE, Smith KW, McGraw S, Smith RG, Petronis VM, Carver CS. Assessing quality of life in adult cancer survivors (QLACS). Qual Life Res 2005 May;14(4):1007-1023. [Medline: 16041897]
36. Department for Communities and Local Government. English Indices of Deprivation: Guidance Document. London; 2010. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6222/1871538.pdf [accessed 2013-08-13] [WebCite Cache ID 6Iq4QODTf]
37. Vangeest JB, Johnson TP. Using incentives in surveys of cancer patients: do "best practices" apply? Cancer Causes Control 2012 Dec;23(12):2047-2052. [doi: 10.1007/s10552-012-0082-z] [Medline: 23076587]
38. Armes J, Crowe M, Colbourne L, Morgan H, Murrells T, Oakley C, et al. Patients' supportive care needs beyond the end of cancer treatment: a prospective, longitudinal survey. J Clin Oncol 2009 Dec 20;27(36):6172-6179. [doi: 10.1200/JCO.2009.22.5151] [Medline: 19884548]
39. Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242 [FREE Full text] [doi: 10.1002/cncr.23157] [Medline: 18008363]

40. Paul CL, Carey ML, Hall AE, Lynagh MC, Sanson-Fisher RW, Henskens FA. Improving access to information and support for patients with less common cancers: hematologic cancer patients' views about Web-based approaches. J Med Internet Res 2011;13(4):e112 [FREE Full text] [doi: 10.2196/jmir.1894] [Medline: 22189354]
41. Bhinder S, Chowdhury N, Granton J, Krahn M, Tullis DE, Waddell TK, et al. Feasibility of internet-based health-related quality of life data collection in a large patient cohort. J Med Internet Res 2010;12(3):e35 [FREE Full text] [doi: 10.2196/jmir.1214] [Medline: 20719740]
42. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, et al. Effect of telecare management on pain and depression in patients with cancer: a randomized trial. JAMA 2010 Jul 14;304(2):163-171 [FREE Full text] [doi: 10.1001/jama.2010.944] [Medline: 20628129]
43. Velikova G, Wright EP, Smith AB, Cull A, Gould A, Forman D, et al. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. J Clin Oncol 1999 Mar;17(3):998-1007. [Medline: 10071295]
44. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. Eval Rev 2009 Oct;33(5):464-480. [doi: 10.1177/0193841X09340214] [Medline: 19605623]
45. Ashley L, Jones H, Velikova G, Wright P. Cancer patients' and clinicians' opinions on the best time in secondary care to approach patients for recruitment to longitudinal questionnaire-based research. Support Care Cancer 2012 Dec;20(12):3365-3372. [doi: 10.1007/s00520-012-1518-4] [Medline: 22722886]
46. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer 2012 Oct 9;107(8):1220-1226. [doi: 10.1038/bjc.2012.408] [Medline: 22996611]
47. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
48. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol 2013 Jul 10;31(20):2580-2585. [doi: 10.1200/JCO.2012.47.6804] [Medline: 23733753]
49. Maher EJ. Managing the consequences of cancer treatment and the English National Cancer Survivorship Initiative. Acta Oncol 2013 Feb;52(2):225-232. [doi: 10.3109/0284186X.2012.746467] [Medline: 23237053]
50. Ahmed S, Berzon RA, Revicki DA, Lenderking WR, Moinpour CM, Basch E, International Society for Quality of Life Research. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Med Care 2012 Dec;50(12):1060-1070. [doi: 10.1097/MLR.0b013e318268aaff] [Medline: 22922434]
51. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314. [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]

Abbreviations

CR: consent rate
EHR: Electronic Health Record
ePOCS: electronic Patient-reported Outcomes from Cancer Survivors (system)
HIT: health information technology
IMD: Index of Multiple Deprivation
IT: information technology
NHS: National Health Service
PROFILES: Patient Reported Outcomes Following Initial treatment and Long-term Evaluation of Survivorship (system)
PROs: patient reported outcomes
PROMs: patient reported outcome measures
RR: response rate
YCN: Yorkshire Cancer Network

Edited by G Eysenbach; submitted 11.06.13; peer-reviewed by T Smith, C Valle, C Yuan; comments to author 30.06.13; revised version received 13.08.13; accepted 07.09.13; published 25.10.13.

Please cite as:
Ashley L, Jones H, Thomas J, Newsham A, Downing A, Morris E, Brown J, Velikova G, Forman D, Wright P
Integrating Patient Reported Outcomes With Clinical Cancer Registry Data: A Feasibility Study of the Electronic Patient-Reported Outcomes From Cancer Survivors (ePOCS) System
J Med Internet Res 2013;15(10):e230
URL: http://www.jmir.org/2013/10/e230/
doi:10.2196/jmir.2764
PMID:24161667

Original Paper

Telephone Versus Internet Administration of Self-Report Measures of Social Anxiety, Depressive Symptoms, and Insomnia: Psychometric Evaluation of a Method to Reduce the Impact of Missing Data

Erik Hedman[1,2], PhD; Brjánn Ljótsson[2], PhD; Kerstin Blom[3], MSc; Samir El Alaoui[3], MSc; Martin Kraepelien[3], MSc; Christian Rück[3], PhD; Gerhard Andersson[3,4], PhD; Cecilia Svanborg[3], PhD; Nils Lindefors[3], PhD; Viktor Kaldo[3], PhD

[1]Osher Center for Integrative Medicine, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[2]Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[3]Division of Psychiatry, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[4]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

Corresponding Author:
Erik Hedman, PhD
Osher Center for Integrative Medicine
Department of Clinical Neuroscience
Karolinska Institutet
Retzius väg 8
Stockholm, 171 77
Sweden
Phone: 46 08 524 800 00
Fax: 46 8 779 54 16
Email: kire.hedman@ki.se

Abstract

Background: Internet-administered self-report measures of social anxiety, depressive symptoms, and sleep difficulties are widely used in clinical trials and in clinical routine care, but data loss is a common problem that could render skewed estimates of symptom levels and treatment effects. One way of reducing the negative impact of missing data could be to use telephone administration of self-report measures as a means to complete the data missing from the online data collection.

Objective: The aim of the study was to compare the convergence of telephone and Internet administration of self-report measures of social anxiety, depressive symptoms, and sleep difficulties.

Methods: The Liebowitz Social Anxiety Scale-Self-Report (LSAS-SR), Montgomery-Åsberg Depression Rating Scale-Self-Rated (MADRS-S), and the Insomnia Severity Index (ISI) were administered over the telephone and via the Internet to a clinical sample (N=82) of psychiatric patients at a clinic specializing in Internet-delivered treatment. Shortened versions of the LSAS-SR and the ISI were used when administered via telephone.

Results: As predicted, the results showed that the estimates produced by the two administration formats were highly correlated (r=.82-.91; P<.001) and internal consistencies were high in both administration formats (telephone: Cronbach alpha=.76-.86 and Internet: Cronbach alpha=.79-.93). The correlation coefficients were similar across questionnaires and the shorter versions of the questionnaires used in the telephone administration of the LSAS-SR and ISI performed in general equally well compared to when the full scale was used, as was the case with the MADRS-S.

Conclusions: Telephone administration of self-report questionnaires is a valid method that can be used to reduce data loss in routine psychiatric practice as well as in clinical trials, thereby contributing to more accurate symptom estimates.

(J Med Internet Res 2013;15(10):e229) doi:10.2196/jmir.2818

KEYWORDS

Internet; telephone; self-report measures; missing data; method validation

Introduction

Self-report measures are widely used in both routine psychiatric care and in clinical trials as they have several advantages including psychometric properties similar to clinician-administered instruments [1], low cost, and the potential to administer the instruments over the Internet [2,3]. However, a common problem in these settings is data loss. As pointed out by Claassen et al [4], even in randomized controlled trials attrition rates can be 30-40% and in effectiveness studies on regular care patients this number is likely to be even higher. In an effectiveness study on Internet-based cognitive behavior therapy (ICBT) for panic disorder, we found that as few as 32% of patients completed self-report assessments at six-month follow-up, despite several text message reminders [5]. Data loss lowers the statistical power and as attrition could be non-randomly distributed (eg, persons with more severe symptoms may be more likely not to fill out self-assessments), this could render skewed estimates of symptom levels and treatment effects [6]. A common way of handling this problem is through statistical procedures such as multiple imputation or the use of full-information maximum likelihood estimation models [7]. These missing data strategies do however have some important disadvantages, including non-testable assumptions of the randomness of the missing data patterns, difficulty of dealing with non-normally distributed covariates, and the complexity of the computational process [8].

One way of reducing the negative impact of missing data without the disadvantages of advanced imputational methods could be to increase completion rates of self-reports through the use of telephone assessments, that is, telephone administration of self-report measures as a means to complete the data missing from the online data collection. Several studies have demonstrated that diagnostic assessment interviews can be conducted over the telephone with high convergent validity with face-to-face interviews [9,10]. However, a diagnostic interview or a clinician's expert rating of the patient's level of symptoms is not equivalent to a telephone-administered self-report, where the clinician's impact on the ratings is put to a minimum by using standardized questions and answers that are read verbatim to the patient. The body of evidence is limited when it comes to how accurate this kind of telephone administration of self-report instruments is in comparison to the standard way of self-assessment. We have found only two studies investigating this. In these studies, it was shown that the Penn-State Worry Questionnaire (PSWQ), the Beck Depression Inventory (BDI), and the General Health Questionnaire (GHQ) could be completed over the telephone providing similar estimates as when administered as self-assessment using paper and pencil [11,12]. Another problem when providing self-assessments over the telephone with patients who have failed to complete standard self-assessments is that the patient's motivation to devote a substantial amount of time for a telephone interview might be limited. This problem could be even more pronounced in long-term follow-ups. Against this background, it is important to use instruments with few items when conducting telephone-administered assessments with self-report measures.

To our knowledge, no prior study has investigated whether telephone and Internet administration of self-report measures produce equivalent results in the assessment of social anxiety, depressive symptoms, and sleep difficulties. More knowledge in this regard could lead to more effective strategies for handling data loss in clinical routine psychiatric care as well as in clinical trials. Also, investigation of whether it is possible to use shortened, and thus more efficient, versions of the full-length scales over the telephone has to our knowledge not been done.

The main aim of this study was to compare the convergence of telephone and Internet administration of self-report measures of social anxiety, depressive symptoms, and sleep difficulties. The Liebowitz Social Anxiety Scale - Self-Report (LSAS-SR) [1], Montgomery Åsberg Depression Rating Scale - Self-Rated (MADRS-S) [13], and the Insomnia Severity Index (ISI) [14] were used. We hypothesized that the estimates produced by the two administration formats would be highly correlated.

As a secondary aim, we wanted to explore three different strategies for developing an interview version of a self-report measure. The first is the most straightforward, as the same questions and response options are used in the interview as in the self-report measures. This was used when comparing Internet-administered MADRS-S to a telephone interview version of the same measure. The second strategy emerged from the need to keep the telephone interviews short. We explored this by reducing the number of items for the LSAS-SR and ISI when the measures were telephone-administered. Thus, we compared the Internet-administered full-scale self-report versions against shortened or full-scale telephone-administered versions of the same measure. The third strategy was to use a different measure within the same symptom domain in the telephone interview compared to when administering the measure via the Internet. The main reason to use this strategy is when the nature of the questions and answers in the self-report measure are deemed somewhat difficult to administer verbally over the telephone. This is the case with MADRS-S, which has long questions and answers are given on a 7-point scale with four anchor labels that are also quite long and unique for each question. Specifically, we investigated whether a shortened telephone-administered version of the Hospital Anxiety and Depression Scale (HADS) [15], deemed to be easier to administer via the telephone, could be as highly correlated with Internet-administered MADRS-S as the telephone-administered MADRS-S.

Methods

Design

This study employed a repeated measurement design where participants provided data in both administration formats, that is, telephone and Internet. Participants completed the Internet-administered self-report questionnaire first, followed by a telephone-administered assessment with the same questionnaire, shortened or full-scale. The average time between assessments was 3.1 days (SD 2.2) and the range was 0 to 7 days. The sample (N=82) comprised three cohorts: (1) participants seeking treatment for social anxiety disorder (SAD), denoted SAD sample (n=14), (2) participants seeking treatment for depression (DEP), denoted DEP sample (n=35), and (3) participants diagnosed with insomnia (Insomnia sample, n=33). Type of self-report measure used and whether the full version of the measure was telephone-administered were as follows: the SAD sample completed the full version of the LSAS-SR via the Internet and a short version of LSAS-SR via telephone; the DEP sample completed the full version of the MADRS-S via the Internet and the full version via telephone; and the Insomnia sample completed the full versions of the ISI and MADRS-S via Internet and short versions of the ISI and HADS via telephone.

Recruitment and Participants

Participants were recruited from a series of patients seeking treatment at the Internet-based Cognitive Behavior Therapy Clinic (ICBT clinic) located at the Karolinska University Hospital Huddinge (Psychiatry Southwest) in Stockholm, Sweden. The ICBT clinic provides Internet-based CBT, which is a treatment that essentially can be described as guided online CBT-bibliotherapy with therapist contact through an Internet-based messaging system resembling email [16]. The ICBT clinic treatment context has been described previously in greater detail [5]. Participants were self-referred and could apply through the official website of the ICBT clinic. Only participants who completed the two assessments within one week (on the Internet and via telephone) were included in the present study. Table 1 presents a demographic description of the participants.

Table 1. Description of the participants.

		SAD[a]sample (n=14)	DEP[b]sample (n=35)	Insomnia sample (n=33)
Age, mean (SD)		31.7 (12.5)	36.5 (10.1)	47.2 (13.6)
Gender				
	Women (%)	9 (64.3)	21 (60.0)	24 (72.7)
	Men (%)	5 (35.7)	14 (40.0)	9 (27.3)
Marital status				
	Married or de facto (%)	7 (50.0)	21 (60.0)	21 (63.3)
	Not married (%)	7 (50.0)	14 (40.0)	12 (36.7)
Parental status				
	Parent	4 (28.6)	16 (45.7)	23 (69.7)
	Not parent	10 (71.4)	19 (54.3)	10 (30.3)
Education				
	Did not finish high school (%)	4 (28.6)	7 (20.0)	1 (3.0)
	Finished high school (%)	8 (57.1)	7 (20.0)	8 (24.2)
	University (%)	2 (14.3)	21 (60.0)	24 (72.7)

[a] SAD: social anxiety disorder

[b] DEP: depression

Measures

Social Anxiety

The LSAS-SR was used to assess social anxiety. The LSAS-SR measures fear in and avoidance of 24 social situations (13 performance and 11 interaction situations) that are assumed to be difficult for people suffering from social anxiety disorder. The LSAS-SR is highly correlated with the clinician-administered Liebowitz Social Anxiety Scale (r=.85) [1]. LSAS-SR has high internal consistency (Cronbach alpha=.95), as well as high test-retest reliability over 12 weeks (r=.83) [17]. The convergent and discriminant validity of LSAS-SR has been shown to be strong and the scale is sensitive to change and is therefore often used in treatment research [17]. When administered via the telephone, a shortened version of the LSAS-SR was used. This short version was derived through factor analysis based on previously collected clinical data from patients with SAD at the ICBT clinic (N=684). Ten situations (rated for both fear and avoidance) were chosen for the short version, based on their correlations with the total scale score while ensuring that items from all factors that emerged in the principal components analysis were represented, in order to avoid making the short version narrower in measurement scope than the full version. The correlations with the full scale were r=.96 (total), r=.95 (fear), and r=.95 (avoidance). The included items, as numbered in the full scale, were 4, 9, 10, 11, 12, 15, 16, 17, 20, and 23.

Depressive Symptoms

We used the MADRS-S and a shortened version of the HADS to assess depressive symptoms. The full version of the MADRS-S was used partly because of its brevity in terms of number of items, partly as each of the items in the scale covers rather different dimensions of depressive symptoms, making item reduction difficult. MADRS-S consists of nine items measuring nine different symptoms and each symptom is rated on a 7-point scale with four predefined anchor labels and three non-defined anchor labels in between. The test-retest reliability of MADRS-S is high with *r* ranging from .80 to .94. In a comparative study, Svanborg and Åsberg [13] showed that MADRS-S correlated highly (*r*=.87) with BDI [18].

The HADS consists of two subscales: one that measures depressive symptoms and one that measures general anxiety. Each subscale has 7 items, each rated 0-3, yielding a total score between 0 and 42. The HADS has good convergent validity as the depressive symptoms subscale is highly correlated with the clinician-administered Montgomery-Åsberg Depression Rating Scale (*r*=.81) [15]. In a review of 71 studies investigating the psychometric properties of HADS, Bjelland et al [19] found that Cronbach alpha exceeded .60 in all of them, indicating stable and adequate internal consistency. In the present study, only a subset of items of the scale assessing depressive symptoms was used and this short version of HADS was solely telephone-administered and compared to the Internet-administered MADRS-S. We chose the four items of the HADS depression scale deemed most suitable for telephone-administration. These items, as numbered from the original scale, were 2, 4, 8, and 10.

Sleep Difficulties

The ISI was used to assess sleep difficulties. The ISI is a 7-item instrument assessing the severity of initial, middle, and late insomnia; sleep satisfaction; interference of insomnia with daytime functioning; noticeability of sleep problems by others; and distress about sleep difficulties. A 5-point scale (0-4) is used to rate each item, yielding a total score of 0 to 28. The ISI has adequate psychometric properties including high internal consistency (Cronbach alpha=.74) and is moderately correlated with other measures of sleep behaviors [14]. The items chosen for the shortened telephone version of ISI were items 1a, 1b, 1c, 2, and 3. These five items were chosen as they correspond to the DSM-IV diagnostic criteria of insomnia.

Procedures

Participants completed Internet-administered assessments through the Internet-based platform of the ICBT clinic. Previous research has shown that the LSAS-SR, MADRS-S, HADS, and ISI can be administered via the Internet with psychometric properties equivalent to the paper-and-pencil versions [2,3,20]. Participants in the SAD sample filled out the LSAS-SR, while those in the DEP sample completed the MADRS-S, and the Insomnia sample completed the MADRS-S and the ISI. After this had been done, participants were contacted by a licensed psychologist or by a student at the master level psychology program who conducted the telephone assessment by reading the questions to the participant and recording the response. As described in the design, this meant that the SAD sample was administered a short version of the LSAS-SR, the DEP sample was administered the full version of the MADRS-S, and the Insomnia sample was administered both a short version of the HAD depression scale and a short version of the ISI. There were eight assessors in total and they followed a structured interview guide after having received education on how to conduct the telephone assessments. The interviewer first informed the respondent on how to give their answers, then read the instructions and questions of the respective instrument exactly as presented in the scale. For items with predefined anchor labels, the clinician read the corresponding text to the participant. Clinicians reading the self-report instrument to the participant were instructed not to make any form of independent assessment of the symptoms or to give any further explanation of how to interpret the question, but to only record the participant’s response.

Statistical Analyses

Analyses were conducted using SPSS, version 20. Cronbach alpha was used to calculate internal consistency. Pearson’s zero-order product-moment correlation was used to analyze intercorrelations across administration formats. Data were standardized prior to correlation analyses by subtracting the mean score from each raw score and dividing by the standard deviation. To provide an estimate of how raw scores from the telephone-administered assessment translated into the full scale as completed online, linear regression analyses were conducted where Internet scores were regressed on telephone scores. *Z* tests were used to investigate differences in correlation coefficients between the measures.

Results

Internal Consistency

The alpha values for each questionnaire and administration format are presented in Table 2. Cronbach alpha ranged between .76 and .86 for telephone administration and .79 and .93 for Internet administration. The differences in internal consistency across administration format were small, with the largest difference being found for the LSAS-SR (telephone: Cronbach alpha=.86 vs Internet: Cronbach alpha=.93).

Correlation Between Administration Formats

The scores from the telephone and Internet administered self-report questionnaires were all highly and significantly (*P*<.001 for all measures) correlated indicating strong positive associations of the two formats. The correlation coefficients were as follows: LSAS-SR (Internet) with short LSAS-SR (telephone), *r*=.82; ISI (Internet) with short ISI (telephone), *r*=.91; MADRS-S (Internet) with MADRS-S (telephone), *r*=.83; MADRS-S (Internet) with short HADS (telephone), *r*=.70. *Z* tests did not indicate any significant differences between correlations across questionnaires (*P*>.50).

Regression Coefficients to Predict Internet Self-Report From Telephone Self-Report

In order to obtain an estimate of how the telephone-administered raw scores best translate into the Internet-administered version, regression coefficients were calculated for each measure where the Internet-administered scores were regressed on the telephone-administered scores. Table 3 presents the beta coefficients (ie, the change in the Internet-administered scales for a one-point increase in the telephone-administered scales), as well as the intercepts (ie, the score on the Internet-administered version when the telephone score equals zero). All beta coefficients were statistically significant indicating that the Internet-administered LSAS-SR can be predicted from the telephone-administered short version of

LSAS-SR ($t_{1,12}$=4.92, P<.001), the Internet-administered MADRS-S can be predicted from the telephone-administered MADRS-S ($t_{1,36}$=8.76, P<.001) as well as from the short HADS depression scale ($t_{1,31}$=5.43, P<.001), and finally, the Internet-administered ISI can be predicted from the telephone-administered ISI ($t_{1,31}$=12.09, P<.001). Thus, missing Internet ratings can be estimated by using the general formula: Internet score = intercept + beta * telephone score.

Table 2. Internal consistencies (Cronbach alpha) for the two administration formats for each questionnaire.

Measure	Telephone administration[f]	Internet administration
LSAS-SR[a]	.86	.93
MADRS-S[b]	.76	.79
HADS[c,d]	.85	-
ISI[e]	.83	.87

[a] LSAS-SR: Liebowitz Social Anxiety Scale – Self-Report

[b] MADRS-S: Montgomery Åsberg Depression Rating Scale – Self-Rating

[c] HADS: Hospital Anxiety and Depression Scale

[d] HADS was only administered via telephone

[e] ISI: Insomnia Severity Index

[f] Shortened versions of the LSAS-SR, ISI, and HADS were used when administered on the telephone

Table 3. Mean, SD, and regression coefficients to predict Internet self-report from telephone self-report.

Measure	Administration format[f]		Regression coefficients, Internet data regressed on telephone data		
	Telephone, mean (SD)	Internet, mean (SD)	Intercept	Beta	P value of regression beta coefficient
LSAS-SR[a]	34.1 (9.6)	70.0 (20.0)	6.25	1.87	<.001
MADRS-S[b]	24.4 (7.2)	26.9 (7.0)	7.36	0.80	<.001
HADS[c,e]	2.6 (2.5)	-	5.77	2.88	<.001
ISI[d]	8.3 (4.7)	12.2 (5.9)	2.81	1.14	<.001

[a] LSAS-SR: Liebowitz Social Anxiety Scale - Self-Report

[b] MADRS-S: Montgomery Åsberg Depression Rating Scale - Self-Rated

[c] HADS: Hospital Anxiety and Depression Scale

[d] ISI: Insomnia Severity Index

[e] HADS was only administered via telephone and predicts MADRS-S in the regression results presented in the table

[f] Shortened versions of the LSAS-SR, ISI, and HADS were used when administered on the telephone

Discussion

Principal Findings

The aim of this study was to compare the convergence of telephone and Internet administration of self-report measures of social anxiety (LSAS-SR), depressive symptoms (MADRS-S), and sleep difficulties (ISI). As predicted, the results showed that the estimates produced by the two administration formats were highly correlated. The correlation coefficients were similar across questionnaires and the shorter versions of the questionnaires used in telephone administration of the LSAS-SR and ISI performed, in general, equally well compared to when the full scale was used, as was the case with the MADRS-S. The analysis also showed that a shortened telephone-administered version of a different scale assessing the same symptom domain could be used to predict Internet-administered self-report scores. In other words, shortened HADS could be used to predict the full MADRS-S with similar effectiveness as when the full telephone-administered MADRS-S was used to predict the Internet-administered MADRS-S. These findings suggest that providing self-report questionnaires over the telephone, in their full or shortened form, is a valid administration format for measures commonly used to assess social anxiety, depressive symptoms, and sleep difficulties.

As outlined in the introduction, prior research in this area is scarce and to our knowledge this is the first study to compare the psychometric properties of self-report measures administered via the telephone and the Internet. However, two prior studies have investigated the correlations between telephone and paper-and-pencil-administered self-report instruments. The present study has similar estimates on measures of association as in the study by Senior and colleagues [11] investigating the worry measure, PSWQ, and the depression inventory, BDI. Evans and colleagues [12] reported a correlation coefficient of .83 when comparing the GHQ administered over the telephone and as paper-and-pencil self-report, which is close to the estimates found in this study. This is further indication that Internet is a valid way of providing self-report questionnaires, which has also been previously demonstrated [2,21].

We regard the findings of the present study as relevant from a clinical as well as from a research perspective as they show that telephone administration can be a valid substitute for conventional use of self-report measures. As missing data is a substantial problem in both routine psychiatric practice and clinical trials, the findings of this study are important as they support the use of telephone interviewing of patients who have failed to provide self-report data. This, in turn, can lead to lower attrition rates and thereby more accurate estimates of symptom levels and treatment effects. As mentioned in the introduction, this type of handling of missing data has some advantages compared to using statistical procedures such as multiple imputation. A direct comparison between these forms of data replacements was beyond the scope of this paper, but should be investigated in future studies. One potential problem when trying to reach patients who have not completed self-report assessments is that their willingness to spend large amounts of time on the telephone being interviewed might be reduced. Therefore, a major implication of this study is that it is also possible to replace the full Internet-administered self-report version of the respective scales with shorter versions (LSAS-SR and ISI) or even with another set of questions within the same symptom category (HADS). This may further increase the possibility of reducing attrition rates, as not more than three or four minutes are required to complete the telephone assessments.

Limitations

There were some limitations to this study. First and most importantly, there was no randomization of the order in which participants completed questionnaires. However, previous research has demonstrated limited effect of order [12]. Second, there was some time lag (a maximum of one week) between assessment points allowing for true natural fluctuations in symptom levels to occur. Considering that no treatment was initiated between the assessment points and that previous studies have found that social anxiety and depressive symptoms tend to be stable for this short period of time if untreated [22], this was nevertheless deemed as acceptable. It also reduced the risk of recall bias. Third, this study used a clinical sample, which may reduce the generalizability of the findings to non-clinical populations. It is however difficult to argue for a plausible mechanism for this potential difference and telephone assessment as data replacement method is probably most useful in clinical settings.

Conclusions

In spite of these limitations, we regard the results of this study as important as they show that telephone administration of self-report measures of social anxiety, depressive symptoms, and sleep difficulties can be a valid method of administration. This procedure can be used to reduce data loss in routine psychiatric practice as well as in clinical trials, thereby contributing to more accurate symptom estimates.

Acknowledgments

We express our warmest gratitude to research nurse, Monica Hellberg, for important clinical assistance.

Conflicts of Interest

None declared.

References

1. Fresco DM, Coles ME, Heimberg RG, Liebowitz MR, Hami S, Stein MB, et al. The Liebowitz Social Anxiety Scale: a comparison of the psychometric properties of self-report and clinician-administered formats. Psychol Med 2001 Aug;31(6):1025-1035. [Medline: 11513370]
2. Hedman E, Ljótsson B, Rück C, Furmark T, Carlbring P, Lindefors N, et al. Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. Computers in Human Behavior 2010 Jul;26(4):736-740. [doi: 10.1016/j.chb.2010.01.010]
3. Thorndike FP, Ritterband LM, Saylor DK, Magee JC, Gonder-Frederick LA, Morin CM. Validation of the insomnia severity index as a web-based measure. Behav Sleep Med 2011;9(4):216-223. [doi: 10.1080/15402002.2011.606766] [Medline: 22003975]
4. Claassen C, Kurian B, Trivedi MH, Grannemann BD, Tuli E, Pipes R, et al. Telephone-based assessments to minimize missing data in longitudinal depression trials: a project IMPACTS study report. Contemp Clin Trials 2009 Jan;30(1):13-19 [FREE Full text] [doi: 10.1016/j.cct.2008.08.001] [Medline: 18761427]
5. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of Internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Feb 14:1-1. [doi: 10.1111/acps.12079] [Medline: 23406572]
6. Allott K, Chanen A, Yuen HP. Attrition bias in longitudinal research involving adolescent psychiatric outpatients. J Nerv Ment Dis 2006 Dec;194(12):958-961. [doi: 10.1097/01.nmd.0000243761.52104.91] [Medline: 17164636]
7. 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]

8. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393 [FREE Full text] [Medline: 19564179]
9. Hajebi A, Motevalian A, Amin-Esmaeili M, Hefazi M, Radgoodarzi R, Rahimi-Movaghar A, et al. Telephone versus face-to-face administration of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, for diagnosis of psychotic disorders. Compr Psychiatry 2012 Jul;53(5):579-583. [doi: 10.1016/j.comppsych.2011.06.001] [Medline: 21820651]
10. Crippa JA, de Lima Osório F, Del-Ben CM, Filho AS, da Silva Freitas MC, Loureiro SR. Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. Perspect Psychiatr Care 2008 Oct;44(4):241-247. [doi: 10.1111/j.1744-6163.2008.00183.x] [Medline: 18826462]
11. Senior AC, Kunik ME, Rhoades HM, Novy DM, Wilson NL, Stanley MA. Utility of telephone assessments in an older adult population. Psychol Aging 2007 Jun;22(2):392-397. [doi: 10.1037/0882-7974.22.2.392] [Medline: 17563195]
12. Evans M, Kessler D, Lewis G, Peters TJ, Sharp D. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? Psychol Med 2004 Jan;34(1):157-162. [Medline: 14971636]
13. Svanborg P, Åsberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [Medline: 8140903]
14. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [Medline: 11438246]
15. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
16. Andersson G. Using the Internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180. [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]
17. Baker SL, Heinrichs N, Kim HJ, Hofmann SG. The liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. Behav Res Ther 2002 Jun;40(6):701-715. [Medline: 12051488]
18. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
19. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [Medline: 11832252]
20. Andersson G, Kaldo-Sandström V, Ström L, Strömgren T. Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. J Psychosom Res 2003 Sep;55(3):259-262. [Medline: 12932800]
21. Carlbring P, Brunt S, Bohman S, Austin D, Richards J, Öst LG, et al. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Computers in Human Behavior 2007 May;23(3):1421-1434. [doi: 10.1016/j.chb.2005.05.002]
22. Reich J, Goldenberg I, Vasile R, Goisman R, Keller M. A prospective follow-along study of the course of social phobia. Psychiatry Res 1994 Dec;54(3):249-258. [Medline: 7792329]

Abbreviations

BDI: Beck Depression Inventory
DEP: depression
GHQ: General Health Questionnaire
HADS: Hospital Anxiety and Depression Scale
ICBT: Internet-based cognitive behavior therapy
ISI: Insomnia Severity Index
LSAS-SR: Liebowitz Social Anxiety Scale - Self-Report
MADRS-S: Montgomery-Åsberg Depression Rating Scale - Self-Rated
PSWQ: Penn-State Worry Questionnaire
SAD: social anxiety disorder

Edited by G Eysenbach; submitted 05.07.13; peer-reviewed by P Cipresso, G Kolostoumpis; comments to author 29.07.13; revised version received 06.08.13; accepted 29.08.13; published 18.10.13.

Please cite as:
Hedman E, Ljótsson B, Blom K, El Alaoui S, Kraepelien M, Rück C, Andersson G, Svanborg C, Lindefors N, Kaldo V
Telephone Versus Internet Administration of Self-Report Measures of Social Anxiety, Depressive Symptoms, and Insomnia: Psychometric Evaluation of a Method to Reduce the Impact of Missing Data
J Med Internet Res 2013;15(10):e229
URL: http://www.jmir.org/2013/10/e229/
doi:10.2196/jmir.2818
PMID:24140566

Original Paper

Can Anonymous Posters on Medical Forums be Reidentified?

Victoria Bobicev[1], PhD; Marina Sokolova[2,3], PhD; Khaled El Emam[2,3], PhD; Yasser Jafer[3], MAS; Brian Dewar[4], MLIS; Elizabeth Jonker[2], BA(Hons); Stan Matwin[3], PhD

[1]Technical University of Moldova, Chisinau, Moldova, Republic Of
[2]CHEO Research Institute, Inc, Ottawa, ON, Canada
[3]University of Ottawa, Ottawa, ON, Canada
[4]Privacy Analytics, Inc, Ottawa, ON, Canada

Corresponding Author:
Marina Sokolova, PhD
CHEO Research Institute, Inc
401 Smyth Road
Ottawa, ON, K1H 8L1
Canada
Phone: 1 613 737 7600 ext 4104
Fax: 1 613 731 1374
Email: sokolova@uottawa.ca

Abstract

Background: Participants in medical forums often reveal personal health information about themselves in their online postings. To feel comfortable revealing sensitive personal health information, some participants may hide their identity by posting anonymously. They can do this by using fake identities, nicknames, or pseudonyms that cannot readily be traced back to them. However, individual writing styles have unique features and it may be possible to determine the true identity of an anonymous user through author attribution analysis. Although there has been previous work on the authorship attribution problem, there has been a dearth of research on automated authorship attribution on medical forums. The focus of the paper is to demonstrate that character-based author attribution works better than word-based methods in medical forums.

Objective: The goal was to build a system that accurately attributes authorship of messages posted on medical forums. The Authorship Attributor system uses text analysis techniques to crawl medical forums and automatically correlate messages written by the same authors. Authorship Attributor processes unstructured texts regardless of the document type, context, and content.

Methods: The messages were labeled by nicknames of the forum participants. We evaluated the system's performance through its accuracy on 6000 messages gathered from 2 medical forums on an in vitro fertilization (IVF) support website.

Results: Given 2 lists of candidate authors (30 and 50 candidates, respectively), we obtained an *F* score accuracy in detecting authors of 75% to 80% on messages containing 100 to 150 words on average, and 97.9% on longer messages containing at least 300 words.

Conclusions: Authorship can be successfully detected in short free-form messages posted on medical forums. This raises a concern about the meaningfulness of anonymous posting on such medical forums. Authorship attribution tools can be used to warn consumers wishing to post anonymously about the likelihood of their identity being determined.

(J Med Internet Res 2013;15(10):e215) doi:10.2196/jmir.2514

KEYWORDS

privacy; personal health information; medical forums; text data mining

Introduction

Consumers have many opportunities to share their or their family's personal health stories online, for example, through social networks or disease-specific forums. Such sharing might include disclosing personally identifiable information (eg, names, addresses, dates) coupled with health information (eg, symptoms, treatments, medical care) [1-3]. In fact, 19% to 28% of all Internet users participate in medical online forums, health-focused groups, and communities, and visit health-dedicated Web sites [4,5]. This shared health information can potentially be seen by a larger audience because 58% of Internet users report searching for health information [6].

To protect their identity when posting sensitive information online, consumers may post anonymously. Anonymity can be achieved by using a fake identity or by using a pseudonym or nickname. However, such methods for ensuring anonymity may not be very effective. There is evidence that online consumers reuse their usernames or handles across multiple sites, which makes it easier to figure out their true identity [7]. Even if a consumer creates a unique identity for posting information on a particular medical forum, text analysis techniques can combine textual data from different forums and correlate the ones that have been written by the same author. If any of those texts has the poster's true identity, then even the anonymous posts can be reidentified. A real-world example of such cross-site information aggregation can be found in Li et al [8]. An attacker associated 5 profiles harvested from various forums and then aggregated the posted information. The identified personal information included laboratory test results, the patient's full name, date of birth, spouse's name, home address, home phone number, cell phone number, 2 email addresses, and occupation.

With the emergence of user-generated Web content, authorship analysis is being increasingly applied to online messages [9,10]. The general task of authorship analysis can mean one of several types of analyses: (1) author attribution in which the system is tasked to assign an unknown text to an author from several authors' writing examples [11], (2) author verification in which the system is tasked to determine if some text was or was not written by an author given an example of the writing of a single author [12], or (3) author profiling in which the system is expected to identify an author's gender, age, personality, cultural background, etc by analyzing given text written by this author [13]. Our focus in this paper is the author attribution.

These studies are characterized by a large number of candidate authors, a small volume of training and test texts, and short messages [14-19]. In Koppel et al [20], 10,000 blogs were used in the task of author detection in which 500-word snippets, one for each author, were considered test examples. Of the texts, 20% to 34% texts were classified with an average accuracy of 80%; the rest of the texts were considered unknown. In a separate study on the same dataset, a 500-word snippet was attributed to 1 of 1000 authors with coverage of 42.2% and precision of 93.2% [21]. The remaining 57.8% of snippets were considered unknown.

None of this previous work, however, dealt with messages posted on medical forums or other online venues that are dedicated to discussions of personal health information. The type of text is important because authorship attribution relies on unique characteristics of an individual's writing style, and it cannot be assumed that one will write the same way when reviewing a fiction novel online as when asking a question about medical treatment or diagnosis.

We chose in vitro fertilization (IVF) forums that host discussions about infertility and attempts to conceive. Such discussions are very personal and it is reasonable to assume that individuals would want to participate anonymously. The website IVF.ca is an infertility outreach resource community created by patients for prospective, existing, and past IVF patients. A number of forums are maintained on the site for messages exchanging emotional support and information [22]. We did not require research ethics review for this study because all the data collected and used was from publically available sources. Our institutional research ethics board confirmed that no review of research on public datasets was necessary.

The most frequent uses of an Internet forum for infertility were sharing personal experience, provision of information or advice, expressions of gratitude/friendship, chat, requests for information, and expressions of universality ("we're all in this together") [23]. We applied Authorship Attributor, a new system to identify messages written by the same author, on the message contents. We used only texts posted by the authors on the forums; no metadata were used in training and testing files.

The choice of text features to analyze is one of the most influential factors in the performance of authorship attribution. The most common features used in the literature are word length [24], sentence length [25], type-token ratio, vocabulary richness [26], word and word n-grams (ie, sequences of n words) frequencies [27], and errors and idiosyncrasies [28]. These features could be obtained by using text analysis tools, such as a tokenizer (breaks a sequence of text into words, phases, etc, called tokens), sentence splitter (breaks text into sentences), lemmatizer (determines the base form for inflected words) or stemmer (reduces inflected words to their base form), and orthographic and synonym dictionaries. Syntactic features, such as parts of speech and part of speech sequences [29], chunks of text [30], syntactic dependencies of words [31], and syntactic structures [32] have been used to a lesser extent, but are still frequently applied. A part of speech tagger (assigns part of speech to each word), chunker (breaks text up into sequences of semantically related words), and syntactic parser (analyzes strings of text into their grammatical elements) are the necessary tools for obtaining these features. Some previous work used semantic features, such as synonyms and semantic dependencies [33]. These features can be obtained through specialized dictionaries and semantic parsers. In some experiments, several application-, content-, or language-specific features were applied as well. In most cases, these features were combined to obtain better results.

In this paper, we describe and evaluate a new system, Authorship Attributor, which has been constructed to crawl through medical forums and identify messages written by the same author.

Methods

Authorship Attribution Task

The task of authorship attribution is to identify who is the author of a text given a list of candidate authors and texts written by these candidates. Its methodology is based on a comparison of a new text to texts knowingly written by the candidates. Koppel [15] compared the accuracy of authorship attribution for a variety of feature sets and learning algorithms for a literature corpus, email, and blog posts corpora. The best accuracy (80%-86%) was obtained by support vector machine (SVM) and Bayesian regression algorithms on the basis of the 1000 most frequent words and the 1000 character trigrams with the highest information gain.

One of the most exhaustive feature sets was used by Abbasi and Chen [34]. It included characters, character bigrams and trigrams, punctuation and special characters, word length, function words, word bigrams and trigrams, vocabulary richness, part of speech tags, part of speech tag bigrams and trigrams, message length and structure, misspelled words, and other features. Experiments with this set of features showed good results: 88% to 96% accuracy (ie, correctly classified texts/all texts) for various datasets including eBay comments, a Java forum, and email and chat corpora.

Narayanan et al [10] reused this feature set but slightly changed it. Frequencies of syntactic category pairs (A, B), where A is the parent of B in the parse tree, were added to the previously described feature set. The overall number of features was approximately 1200. The authors used these features in the experiments with 100,000 blogs with an average length of 7500 words in each blog. As in all such cases, there was a trade-off between precision and recall. With a corpus of texts from 100,000 authors, the classifiers could correctly identify an anonymous author in more than 20% of cases and the correct author was one of the top 20 guesses in approximately 35% of cases. The increase in precision from 20% to more than 80% could be achieved by reducing recall in half.

In Narayanan et al [10], content-specific features (eg, keywords) positively influenced the accuracy of classification if authors were writing texts about different topics. However, many applications seek to identify authors regardless of topic [18]. Other studies have presented good results for gender and age classification [13,15]. The gender- or age-specific differences in writing can help in classification, but hide individual author-specific features. Koppel et al [21] performed a small-scale experiment using 2 authors who had posted on different topics of a listserv collection, but it was pointed out that it is extremely difficult to find writing from the same author on different topics.

Luyckx and Daelemans [17] observed that when a large number of candidate authors were considered, similarity-based methods (ie, an anonymous document is attributed to that author whose known writing is most similar) are more appropriate than classification methods (eg, the known writings of each candidate author are used to construct a classifier which is used to classify anonymous documents). We note, however, that similarity-based methods can be best applied to text within the same medium (eg, messages from medical forums), but might not work as well for text harvested from different mediums (eg, electronic health records vs forum messages).

Character-Based Text Classification Methods

The task of text classification consists of assigning a given text into predetermined categories. Most text classification methods are word-based (eg, they present a text document as a vector of words). In contrast, compression-based classification methods use characters or even bytes as the text representation unit. Researchers have noted that character-based classification methods have a potential advantage over word-based methods because they are able to automatically capture document features other than words. Character-based classification analyzes the text for letter counts, capitalized letters, punctuation and other nonalphabetical character counts, and letter combinations of various lengths [16,35,36]. Other important lexical features include prefixes and suffixes [18], functional words [33], and character n-grams [15]. Experiments demonstrated that letter-based methods yielded more precise results than those based on grammatical information [16].

One classification approach that has been used is compression. Having an anonymous document and several groups of documents representing several classes, a copy of the anonymous document is added to every group of documents. Each of these groups with the added anonymous document is compressed separately. As a result, the anonymous document is compressed differently with different classes of texts because the specific statistical model is created for each class of text. The document is attributed to the class that provides its maximum compression measured in bytes. The maximum compression means that the anonymous document is the most similar to the documents in this class and the created statistical model is the best for it. A relative disadvantage of this algorithm is its comparative slowness.

The most straightforward compression-based method of text categorization using off-the-shelf algorithms was described in Kukushkina et al [16]. The main idea behind this approach is that for every text the compression algorithm creates an individual model adapted to this particular class of texts. Marton et al [37] experimented with 3 compression algorithms, the data compression file format RAR, gzip, and Lempel-Ziv-Welch (LZW) [38], several corpora and types of classification, including the authorship attribution task. The attribution was performed on The Federalist Papers from the Gutenberg Project corpus [39] and a Reuters subcorpus. RAR obtained the best results compared with the other compression algorithms, with 78% overall accuracy for the Reuters corpus, which consisted of smaller texts than the other corpora.

Prediction by Partial Matching

Teahan [40] applied compression-based methods to a multiclass categorization problem to find duplicated documents in large text collections. Comparing several compression algorithms, the author found that the best performance was obtained by the RAR software and the PPMD5 algorithm (84%-89% accuracy for different conditions). Prediction by partial matching (PPM) is an adaptive finite-context method for text compression. It is based on probabilities of the upcoming characters depending on several previous characters. These several previous characters are called “context” of the upcoming character.

Since the algorithm was first presented [41,42], it has been modified and optimized. PPM has set the performance standard for lossless compression of text throughout the past decade. It has been shown that the PPM scheme can predict English text almost as well as humans [40]. The PPM technique blends character context models of varying length to arrive at a final overall probability distribution for predicting upcoming characters in the text. The blending method is similar to the linear interpolation method of n-gram probabilities smoothing. Several methods of interpolation have been proposed [43-46].

An example of the general method of context probability interpolation is provided in Multimedia Appendix 1.

The maximal length of a context equal to 5 in the PPM model was proven to be optimal for text compression [40]. In other experiments, length of character n-grams used for text classification varied from 2 [16] to 4 [21] or a combination of several lengths [34]. Stamatatos [19] pointed out that the best length of character n-grams depends on different conditions and varies for different texts.

The PPM algorithm uses an escape mechanism for blending context probabilities. The algorithm attempts to estimate the probability of an upcoming character by using the maximal context. If this context was not found during training, then the algorithm moves to the shorter context through a so-called escape mechanism in which a probability of escape from the longer context to the shorter one is estimated and added to the final probability. If the probability of the shorter context is equal to zero, the algorithm escapes to the next shorter one and so on. If no one context is found, the algorithm estimates the probability of the upcoming character with the zero context. Given that the maximal context in our experiments is equal to 5, the full name of the method used by Authorship Attributor is PPM5. We provide the specific details of the PPM5 method in Multimedia Appendix 2.

In Bratko and Filipic [38,47], the letter-based PPM models were used for spam detection. In this task, there existed 2 classes only: spam and legitimate email (ham). The created models showed strong performance in a Text Retrieval Conference competition, indicating that data compression models are well-suited to the spam filtering problem.

In Teahan et al [48], a PPM-based text model and minimum cross-entropy as a text classifier were used for various tasks; one of them was an author detection task for The Federalist Papers. The results supported the claim made by historians and other analysts that James Madison had written the disputed papers. The modeling part of the PPM compression algorithm was used to estimate the entropy of text. The entropy provides the estimation of probabilities quality measure; the lower entropy is, the better probabilities are estimated.

In Bobicev and Sokolova [49], the PPM algorithm was applied for text categorization in 2 ways: on the basis of characters and on the basis of words. Character-based methods performed almost as well as SVM, the best method among several machine-learning methods compared in Debole and Sebastiani [50] for the Reuters-21578 Text Categorization Collection corpus.

Comparison With Other Classification Methods

A variety of machine-learning methods have been used for text categorization, including Bayesian classification [6], decision trees [18], cluster classification [15], k-nearest neighbor (k-NN) algorithms [5], and neural nets [20]. Lately, SVM has become the most popular technique [14]. As previously described, words were the most common feature used by these methods in text classification. To put PPM classification in perspective, specifically the PPM5 model used by Authorship Attributor, we compared its performance with the performance of more standard methods.

First, we applied the word-based PPM classification [51]. Here, punctuation marks and other nonalphabetic symbols were eliminated and all letters were converted to lowercase. We used the same set of authors, texts, and other experiment settings to make direct comparison of the results: 10-fold cross-validation, 90 files for training, and 10 files for testing each time.

Next, we applied WEKA's Naïve Bayes and SVM algorithms as the 2 most popular methods in text classification. These algorithms are able to work with various features extracted from texts. The main features in most cases were frequent words. Therefore, we used 3845 words with frequencies of more than 10 in the frequency dictionary of all words appearing in the forum texts. Because we extracted from the text words only without figures and punctuations, we added 24 features with punctuations, and also features with figures and capital letters. We then ran the classification experiment with this set of features on an in vitro fertilization (IVF) support website. The feature set "frequent words + punctuation + figures + capital letters" was built to match PPM features.

Empirical Evaluation

Medical Forums

The IVF.ca website includes 8 forums: Cycle Friends, Expert Panel, Trying to Conceive, Socialize, In Our Hearts, Pregnancy, Parenting, and Administration. Table 1 presents the statistical data about the forums. Each of these forums have subforums; for example, the Cycle Friends forum consists of 6 subforums: Introductions, IVF/FET/IUI Cycle Buddies, IVF Ages 35+, Waiting Lounge, Donor & Surrogacy Buddies, and Adoption Buddies (see the summary of these in Table 1). Each of these subforums consists of a number of topics initiated by one of the participants. For example, the IVF Ages 35+ subforum consists of 506 topics such as "40+ and chances of success," "Over 40 and pregnant or trying to be," etc. Depending on the topic itself and the amount of interest among participants, a different number of replies are associated with each topic. For example, the former topic has 4 replies and the latter topic has 1136 replies.

For our experiments, we wanted to analyze texts authored by many forum users. Ideally, the number of texts written by each author should be as large as possible. We focused on the subforums IVF/FET/IUI Cycle Buddies [52] and IVF Ages 35+ [53] because they have the highest number of posts per author. For IVF Ages 35+ the average number of posts per author was 97.6; for the IVF/FET/IUI Cycle Buddies, the average number of posts per author was 137.8. Another important criterion for the subforum selection was the average number of posts per topic (see Table 1). Analysis showed that a topic was usually discussed through messages posted as responses to other posts on the same topic. We assumed that longer threads of topics were indicators of more posts written by the same author.

Table 1. Statistics on the analyzed subforums on the IVF.ca website at the time of data collection.

Subforum name	Topics, n	Posts, n	Posts per topic, mean
Introduction	1716	13,569	7.91
IVF/FET/IUI Cycle Buddies	2167	116,994	53.99
IVF Ages 35+	506	16,362	32.34
Waiting Lounge	418	3816	9.13
Donor & Surrogacy Buddies	893	7381	8.27
Adoption Buddies	304	4210	13.85

Text Retrieval

We designed a Web crawler to retrieve messages from the Web forums and applied it to the 2 subforums mentioned previously. The 3 main stages in retrieving information using a crawler consist of (1) fetching a website, (2) parsing the HyperText Markup Language (HTML) contents of Web pages within that site, and (3) storing the retrieved data into a database. We used a combination PHP, Apache Server, and MySQL database management system in our design.

Data from each post consisted of forum name, subforum name, topic title, post author name, post author role, post date, and post content. Our code parsed the HTML contents to obtain each of these components corresponding to a given post and placed each component in a corresponding table column in the database. Post content data were used for the subsequent experiments.

Message Preprocessing

We grouped posts by authors to see how much text each author produced. We sorted the data about authors by the number of posts written by each author in descending order. In total, 865 authors posted in the IVF Ages 35+ subforum and 1195 authors posted in the IVF/FET/IUI Cycle Buddies forum. The numbers of posts per author distributions for both subforums are presented in Figure 1.

We wanted to analyze as many authors as possible. At the same time, we wanted these authors to have enough posts for meaningful results. Hence, it was a trade-off between the number of authors and the number of posts, both numbers being as large as possible. For 10-fold cross-validation, 100 posts per author were enough to run machine-learning experiments [54]. In the IVF Ages 35+ subforum, 30 authors posted more than 100 messages. Statistics about the most-prolific 30 authors are presented in Figures 2 and 3, in which the total number and average length of posts for each author were measured in words (mean 126.2 words, SD 47.5). In the IVF/FET/IUI Cycle Buddies subforum, 50 authors had more than 100 posts; hence, their text volumes were larger, but the average length of posts (mean 97.7 words, SD 36) was less than that from the IVF Ages 35+ subforum.

Figure 4 shows the number of posts per topic for each analyzed author in the IVF Ages 35+ subforum. Most authors posted approximately 10 to 20 messages on every topic. At least half of the authors posted on more than 20 topics. Such a large diversity of topics ensures that the author classification would not be influenced by the topic's features.

The average length of posts was also important in our case because it was harder to identify the authors of shorter messages. The average length of posts was approximately 750 characters in the IVF Ages 35+ subforum and approximately 600 characters in the IVF/FET/IUI Cycle Buddies subforum. Given that the average number of characters per word on Wikipedia is 5.2 characters [55], we estimated the average length of the subforum posts to be approximately 100 to 150 words.

After examination of the data, we found that some posts included other posts; for example, “Very hot here today +40°C with the humidity. Summer is finally here! I am soooo jealous-I had my heater on in my office today!” In some cases, there were even 2 inclusions, one inside of another, so the samples of 3 author writings were mixed into 1 message. Such posts can misinform about a writing style of an individual author and should be removed from further consideration. On the IVF Ages 35+ subforum, we removed 1593 of 16,362 posts (9.74%); 14,832 posts remained for further analysis. On the IVF/FET/IUI Cycle Buddies subforum, we removed 5151 posts (15.24%); 28,640 posts remained for further analysis.

No other preprocessing of posts was necessary. Posts did not contain signatures or other personal reference to the post author. Some posts used personal names of the authors, but nicknames were used in most cases.

Figure 1. Number of posts per author distribution for the selected subforums, IVF Ages 35+ (n=865) and IVF/FET/IUI Cycle Buddies (n=1195).

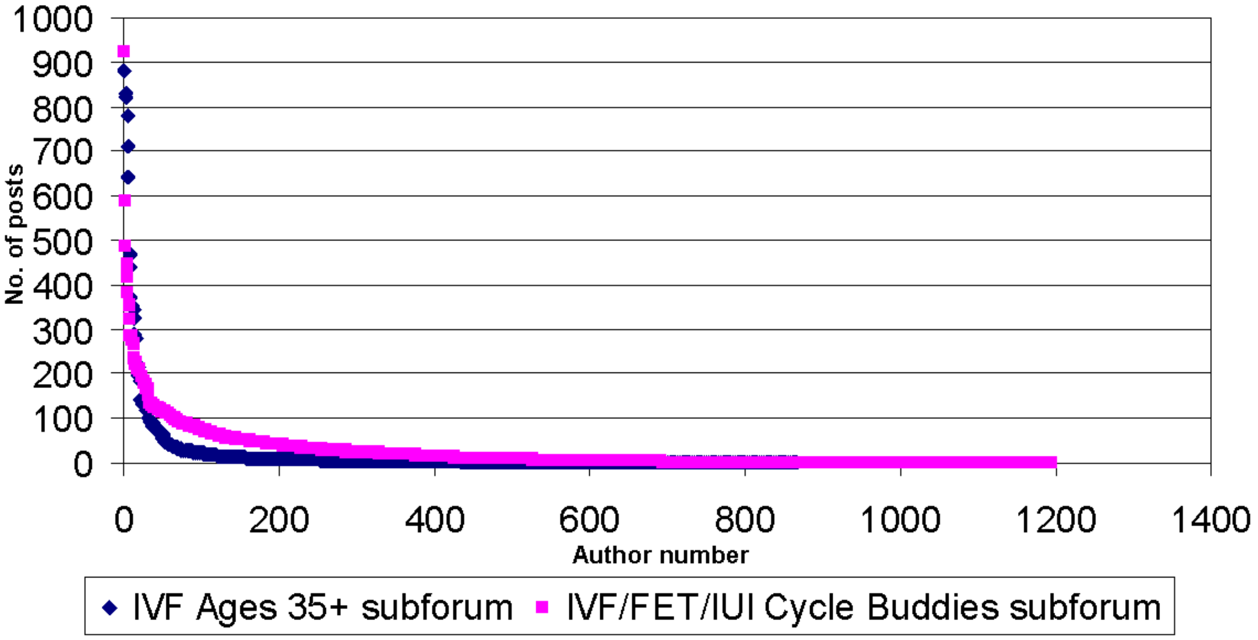


Figure 2. Distribution of the number of posts per author (most prolific) for IVF Ages 35+ subforum (n=30).

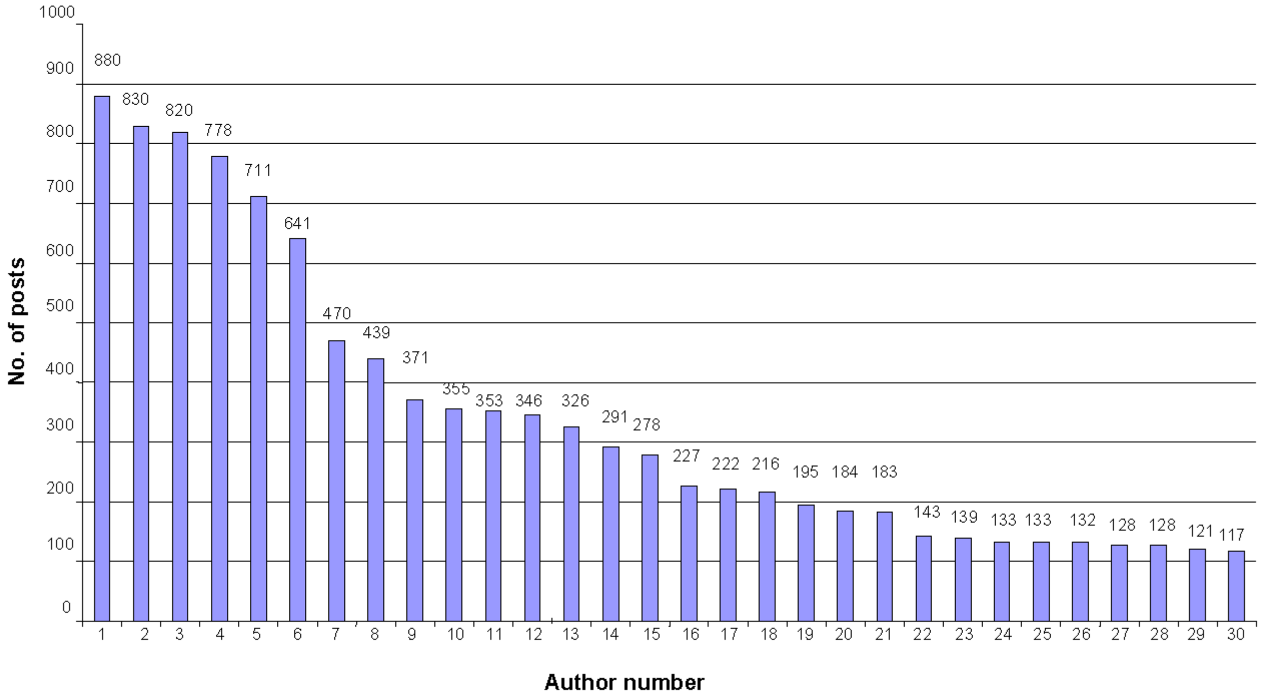


Figure 3. Distribution of the average post length (number of words) for the 30 most-prolific authors in the IVF Ages 35+ subforum.

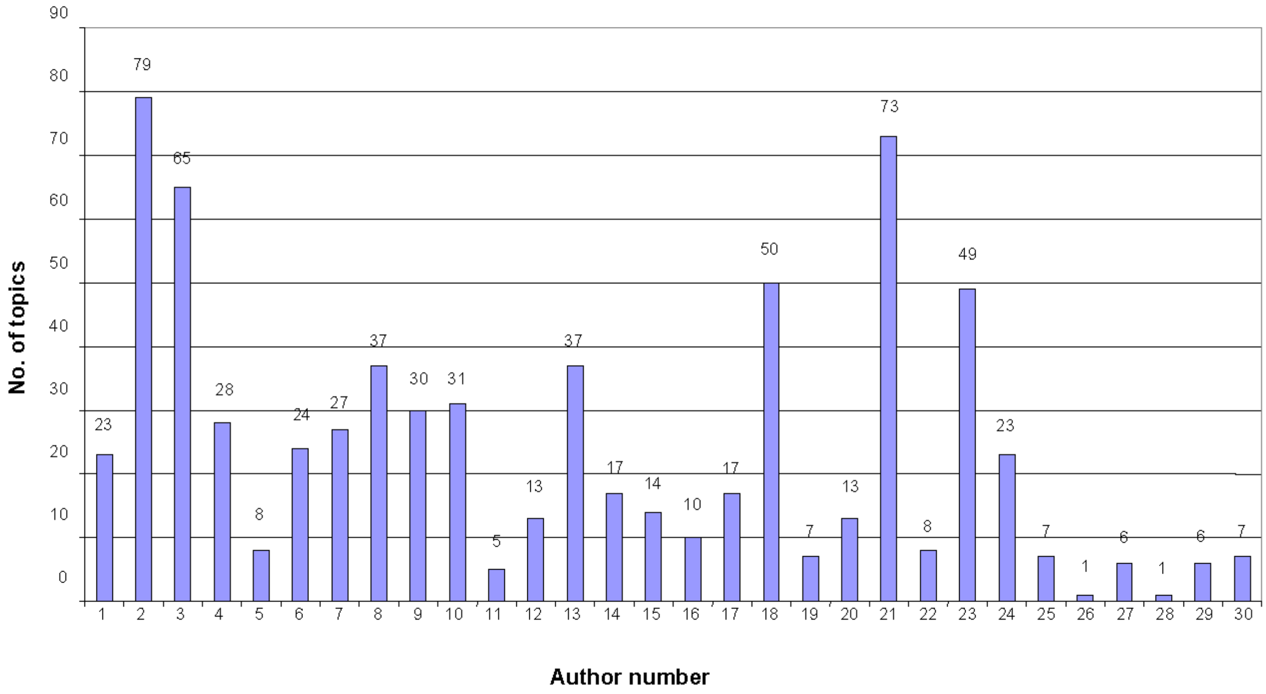
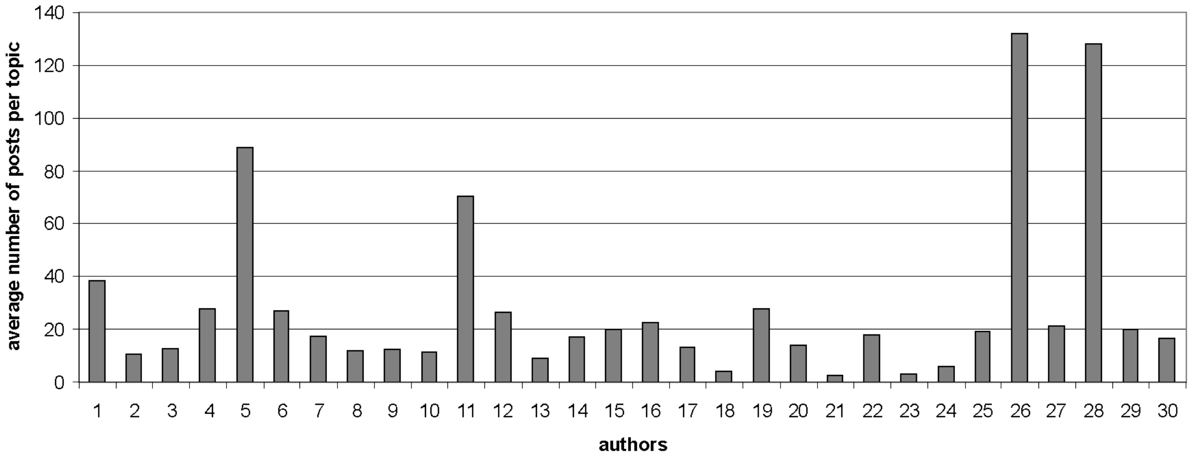


Figure 4. Number of posts per topic for each author (most prolific) in the IVF Ages 35+ subforum (n=30).



Analysis

Experiment 1: Choice of Characters

We tested the PPM method using different sets of characters. We studied whether capitalized letters and nonalphabetic characters (eg, @,#,$,!) hold additional information about an author’s writing style.

To do this analysis, we found 60 authors who posted at least 100 messages: 30 authors from the IVF Ages 35+ subforum and 30 authors from the IVF/FET/IUI Cycle Buddies subforum. The messages from the same author represented 1 class. As a result, we had 3000 messages in the IVF Ages 35+ dataset and 3000 messages in the IVF/FET/IUI Cycle Buddies dataset. On each dataset, we ran the classification experiments by using 10-fold cross-validation. This means 10 runs of the experiment; on each run, 2700 posts were used for training and 300 posts were left for testing. Based on the cross-validation results, the confusion matrix was created and precision, recall, and *F* score were calculated [56].

Figure 2 shows that the volume of texts from the authors differed considerably. The number of posts changed from more than 800 to 100 and the average post length varied from almost 250 words for one author to less than 50 words for another. This imbalance drastically affected the results of the first experiments: the classification was biased toward classes with a larger volume of data for training. Such imbalanced class distribution problems were mentioned in previous studies [10,19,49]. Considering the fact that unbalanced data affected classification results in such a substantial way, we decided to make the data more balanced. We used 10 test texts and 90 training texts for each author, removing the additional texts from the training set.

Even in this case, we obtained an unbalanced class distribution because of different post lengths. Therefore, some normalization was necessary. We used a normalization procedure for balancing entropies of the statistical data models. The normalization procedure goes as follows: In the process of training, statistical models for each class of texts were created and probabilities of text elements were estimated. The next step after training was calculation of entropies of test documents on the basis of each class model. We obtained a matrix of entropies (class statistical models × test documents). The columns were entropies for the class statistical models and rows were entropies for test documents. After this step, the normalization procedure was applied. The procedure consisted of several steps: (1) mean entropy for each class of texts was calculated for each column of the matrix, and (2) each value in the matrix was divided by the mean entropy for this class. Thereby we obtained more balanced values and classification improved considerably. We used normalization in all the PPM5 experiments.

Experiment 2: Attributing Posts From Different Subforums

Machine-learning methods work better on the same types of texts; for example, Koppel et al [21] who analyzed cross-topic author identification. We ran experiments on texts posted by the same author on different subforums. In these experiments, we analyzed all authors who posted in more than 1 subforum. For each author, we extracted training texts from 1 subforum and test texts from the other subforums. We found 9 authors with at least 90 posts in 1 subforum (used for training) and at least 10 posts in other subforums (used in test) and 1 author with 88 posts in the same subforum and more than 10 posts on other subforums. These 10 authors were included in the experiment. In Table 2, we show the statistics for the authors and distribution of their posts per subforum.

Table 2. Statistics for authors and distribution of their posts per subforum.

Author	Subforum, n		
	Introduction	Cycle_Buddies	Age_35+
Author 1	3	6	278
Author 2	35	445	1
Author 3	7	91	3
Author 4	30	11	69
Author 5	6	88	30
Author 6	67	264	4
Author 7	13	16	820
Author 8	54	94	1
Author 9	8	7	355
Author 10	5	130	6

Experiment 3: Important Data Factors

Overview

We tested what data factors affected the accuracy of author recognition. Keeping the method and the post representation constant, we analyzed 3 data factors deemed to be important: the number of authors, the volume of training data, and the volume of test texts.

Number of Candidate Authors

In this set of experiments, we investigated dependence between the number of candidate authors and the accuracy of the authorship identification. We again used 100 posts for each author, splitting them 10 posts for testing and 90 posts for training in 10-fold cross-validation. For both subforums, we repeated the experiments starting with 10 authors and adding 5 authors per iteration. For the IVF Ages 35+ subforum, we had a limit of 30 authors, whereas for the IVF/FET/IUI Cycle Buddies subforum we had a limit of 50 authors.

Volume of Training Data

Training data volume was considered one of the most influential parameters in machine-learning methods. This experiment analyzed how accuracy depended on training data volume. We used mixed candidate authors from both subforums. We selected the first 30 authors [54] from the joint list and used 200 posts for each author. First, 20 posts were used for testing and 180 posts were used for training. Then, for each author, we reduced the number of training posts by 20, repeating the reduction until we reached only 20 training posts per author. The remainder of the settings were the same as in previous experiments: 10-fold cross-validation and PPM5 method using all characters including capitalized letters.

Size of Test Texts

The last critical factor was test text size. As described previously, we considered every post as an independent text author who should be detected. Some posts were really short, containing less than 5 words. Such posts were impossible to classify. Thus, we decided to unify test text sizes. We merged all test texts into 1 text and then split this text in equal fragments measuring their length in words. These experiments were performed with a mixed list of authors from both subforums created for the previous experiment. We only used the first 30 authors with the largest volume of text in their posts. All authors had at least 200 posts. In each of 10 experiments of cross-validation, we used 160 files as a training set and the remaining 40 files of the test set were merged and divided in equal fragments of specified number of words. We repeated the experiments changing test text length starting with 25 words, adding 25 more words each time until the test text reached 500 words per author.

Experiment 4: Comparison With Other Classification Methods

We compared PPM5 results with the results obtained by running Naïve Bayes and SVM algorithms. Both algorithms are often used in text classification and authorship attribution [15,49].

Performance Measures

In text classification, effectiveness is measured by a combination of precision and recall. Precision is the percentage of documents classified into a category that indeed belong in that category, calculated as precision = true positive/(true positive + false positive), where true positive is the number of documents classified into a category that indeed belong to that category and false positive is the number of documents classified into the category that do not belong to that category.

Recall is the percentage of documents belonging to a category that are indeed classified into that category, calculated as recall = true positive/(true positive + false negative), where false negative is the number of documents that indeed belonged to the category but were not classified into the category [57].

The balanced *F* score is the harmonic mean of precision and recall, calculated as *F* score = 2([precision × recall]/[precision + recall]).

When effectiveness is computed for several categories, the results for individual categories can be averaged in several ways [58]: microaveraging (eg, global average of *F* score regardless of topics) or macroaveraging (eg, average of *F* scores of all topics). In our experiments, we calculated the macroaveraged *F* score.

Generalization of Results

We estimated the significance of the PPM5 results (precision, recall, and *F* score) by computing the *t* test against those measures obtained by Naïve Bayes and SVM. Every method comparison was done on the empirical results obtained on the same forum data. Hence, we applied the paired *t* test, which is more rigorous than the unpaired version.

Results

Experiment 1: Choice of Characters

We first report on accuracy of the attribution from IVF Ages 35+ subforum. We used data from 30 authors, 100 posts for each author, and ran 10-fold cross-validation, 90 training and 10 test messages for each fold, to select the best performance. We investigated the impact of letter-based and character-based methods, including original capitalization and conversion to lower case. The results reported in Table 3 show character-based PPM performed better when it worked with all the characters including capitalized letters.

The same experiments were conducted on the base of IVF/FET/IUI Cycle Buddies subforum posts using 100 posts for each of 30 selected authors. The results are presented in Table 4.

Experiment 2: Attributing Posts From Different Subforums

To obtain results on the IVF Ages 35+ subforum using the word-based PPM classification model, we used 1 run of the classifier training and then tested the classifier on the test set. We used 90 training posts from 1 subforum and 10 test texts collected from other subforums. The results are: precision = 0.822, recall = 0.810, *F* score = 0.816. A slight decrease in *F* scores can be explained by the small number of posts. In many cases, the posts were extremely short, especially the test ones, and this affected the results.

Experiment 3: Important Data Factors

Effect of Number of Authors

We used the same dataset as for the rest of our experiments: 100 posts for each author, 10 for testing, 90 for training, 10-fold cross-validation. For both subforums, we repeated the experiments changing the number of authors. Tables 5 and 6 present the results for both subforums. Figure 5 demonstrates the dependencies between the number of authors and the accuracy of the attribution.

Table 3. The IVF 35 Ages + classification results; 10-fold cross-validation, 30 authors, 100 posts per author.

Model	*F* score	Precision	Recall
Letters	0.793	0.803	0.784
Characters lowercase	0.822	0.830	0.831
Original capitalization	0.826	0.836	0.817

Table 4. Classification results for author identification on IVF/FET/IUI Cycle Buddies subforum; 10-fold cross-validation, 30 authors, 100 posts per author.

Features	*F* score	Precision	Recall
Letters	0.836	0.851	0.822
Characters lowercase	0.887	0.896	0.877
Original capitalization	0.902	0.911	0.894

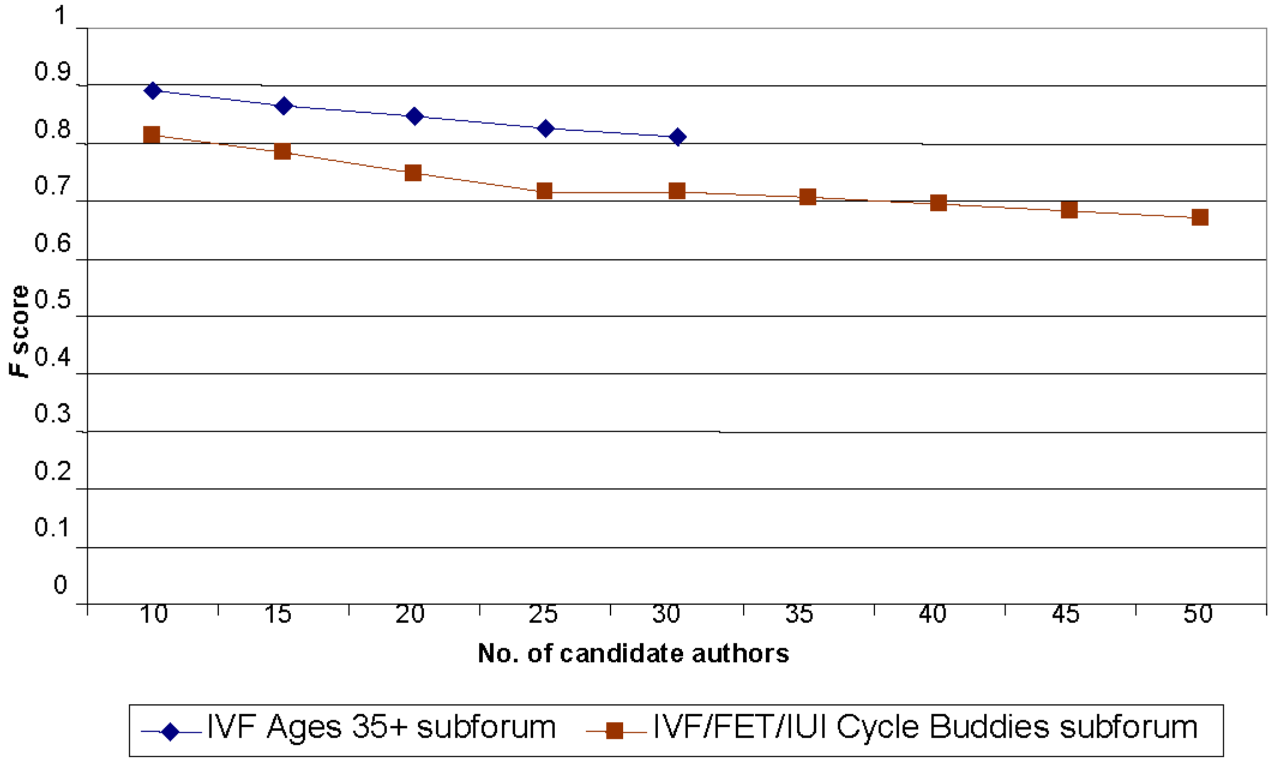
Table 5. Dependency of the accuracy of author detection task on candidate author number on the IVF/FET/IUI Cycle Buddies subforum.

Number of authors	*F* score	Precision	Recall
10	0.965	0.967	0.963
15	0.932	0.937	0.927
20	0.924	0.931	0.917
25	0.912	0.921	0.904
30	0.902	0.911	0.894
35	0.881	0.891	0.872
40	0.845	0.856	0.835
45	0.838	0.849	0.827
50	0.831	0.842	0.820

Table 6. Dependency of the accuracy of author detection task on candidate author number on the IVF Ages 35+ subforum.

Number of authors	*F* score	Precision	Recall
10	0.919	0.921	0.916
15	0.918	0.922	0.914
20	0.885	0.889	0.882
25	0.875	0.882	0.869
30	0.826	0.836	0.817

Figure 5. Dependency of the accuracy on candidate author number for author detection task on the IVF/FET/IUI Cycle Buddies and IVF Ages 35+ subforums.



Effect of Size of Training Data

We analyzed how the attribution accuracy depended on the training data volume. The results of these experiments are presented in Table 7. The *F* score rapidly rose from 0.5 to 0.8 when the number of training texts reached 100 posts. After that, the increase in the training set did not change the *F* score. The graph in Error: Reference source not found Figure 6 visualizes the relationship between the number of training files and the *F* score.

Effect of Test Text Size

We checked the impact of the test size (words) on the author attribution. Table 8 summarizes the results of the experiments. The *F* score rapidly increased with the increase of the text from 25 to 100 words, and then slowly increased until the test text reached 275 words. After that, the *F* score fluctuated, although the overall tendency was still to increase. The relationship between text size and the *F* score is shown in Figure 7.

Table 7. Dependency of the accuracy on training data volume for the author detection task.

Number of training files	*F* score	Precision	Recall
20	0.503	0.496	0.511
40	0.668	0.669	0.667
60	0.765	0.773	0.758
80	0.794	0.800	0.787
100	0.806	0.812	0.800
120	0.815	0.823	0.808
140	0.826	0.834	0.819
160	0.834	0.841	0.827
180	0.837	0.843	0.831

Figure 6. Dependency of the *F* score and the training data volume for the author attribution.

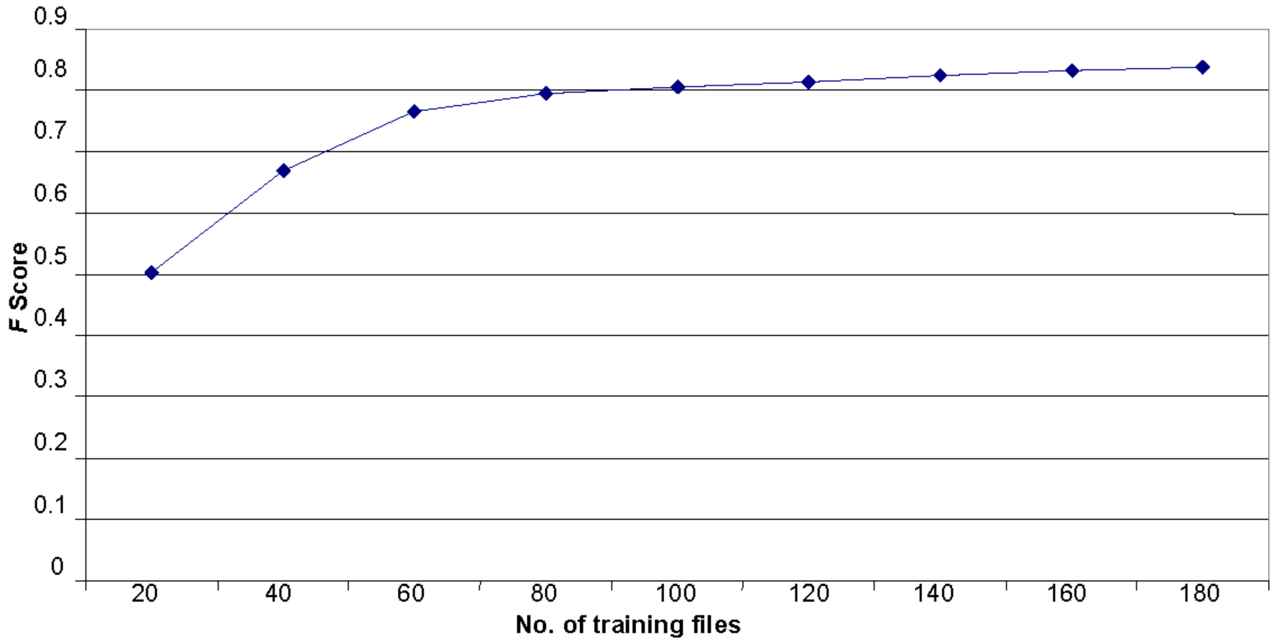
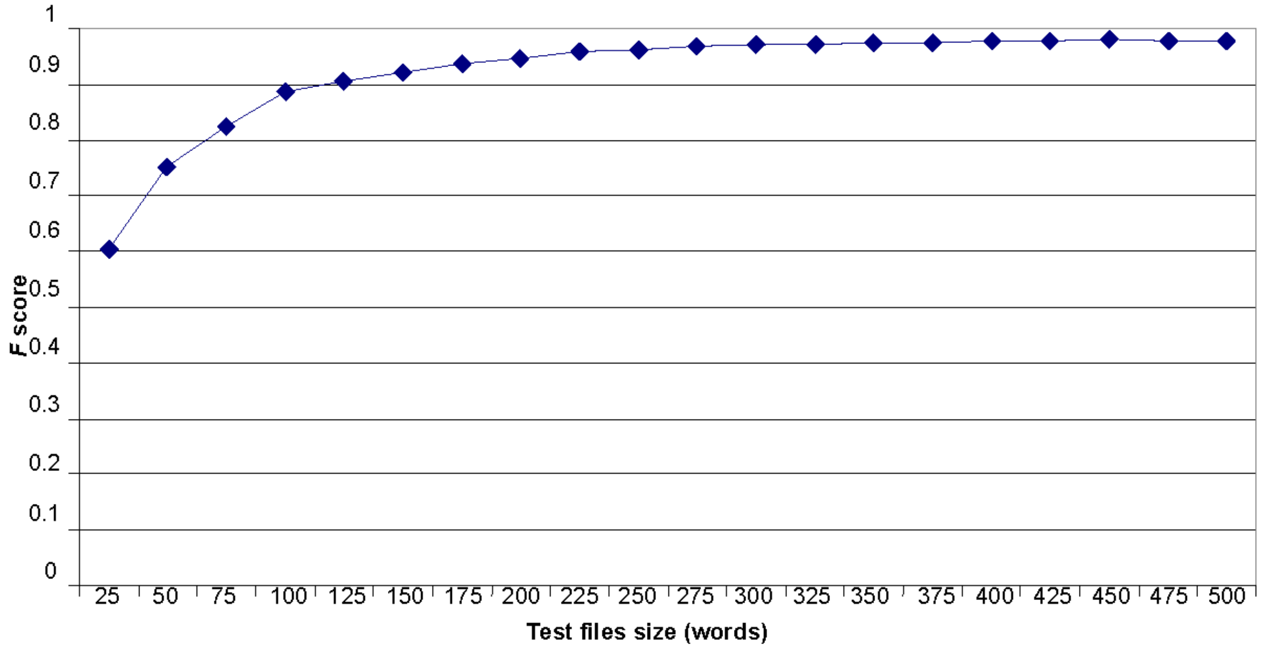


Table 8. Dependency of the results of test files size for author detection task.

Test files size (words)	*F* score	Precision	Recall
25	0.605	0.613	0.599
50	0.752	0.759	0.745
75	0.825	0.833	0.817
100	0.886	0.895	0.877
125	0.907	0.914	0.901
150	0.920	0.926	0.915
175	0.936	0.940	0.933
200	0.948	0.952	0.943
225	0.958	0.963	0.953
250	0.962	0.967	0.957
275	0.970	0.973	0.967
300	0.973	0.976	0.971
325	0.972	0.975	0.969
350	0.976	0.979	0.973
375	0.975	0.978	0.973
400	0.979	0.981	0,976
425	0.977	0.980	0.975
450	0.980	0.982	0.978
475	0.978	0.980	0.975
500	0.979	0.982	0.977

Figure 7. Dependency of the *F* score on the test text size for the author attribution.



Experiment 4: Comparison With Other Classification Methods

When we compared the performance of our method to other classification methods, the results were nonuniform. For the IVF Ages 35+ subforum, SVM on the most complex set of features gave the best result (*F* score=0.766). The Naïve Bayes algorithm performed better on frequent words only, but its *F* score was only 0.636. For the IVF/FET/IUI Cycle Buddies subforum, SVM again was better, but this time on frequent 5-character sequences (*F* score=0.701). The best Naïve Bayes was on frequent words only (*F* score=0.575). The *F* score obtained on different sets of features on both subforums for these methods are presented in Table 9.

The obtained results show that for authorship attribution, word-based classification is not as good as character-based classification. Also, PPM outperformed Naïve Bayes and SVM on the reported experiments for this task.

Statistical Significance of the PPM5 Results

The *t* test results for the IVF35+ subforum show that PPM5 outperformed Naïve Bayes with a significant difference (*P*=.02, standard error of the difference=0.025). PPM5 significantly outperformed SVM (*P*=.001, standard error of the difference=0.002). The *t* test results on the IVF/FET/IUI Cycle Buddies subforum show that PPM5 significantly outperformed Naïve Bayes (*P*=.008, standard error of the difference=0.027). PPM5 significantly outperformed SVM (*P*<.001, standard error of the difference=0.001).

Table 9. Results for author detection task using Naïve Bayes and support vector machine (SVM) classification models implemented in WEKA.

Subforum	Features	*F* score	
		Naïve Bayes	SVM
IVF Ages 35+	Frequent words only	0.636	0.760
IVF Ages 35+	Frequent words + punctuation + figures + capital letters frequency	0.624	0.766
IVF Ages 35+	Frequent 5-character sequences	0.586	0.743
IVF/FET/IUI Cycle Buddies	Frequent words only	0.575	0.690
IVF/FET/IUI Cycle Buddies	Frequent words + punctuation + figures + capital letters frequency	0.567	0.694
IVF/FET/IUI Cycle Buddies	Frequent 5-character sequences	0.550	0.701

Discussion

Principal Findings

In this study, we aimed to empirically examine the accuracy of identifying authors of online posts on a medical forum. Given that individuals may be reluctant to share personal health information on online forums, they may choose to post anonymously. The ability to determine the identity of anonymous posts by analyzing the specific features of the text raises questions about health consumers using anonymous posts as a method to control what is known publicly about them. We measured the accuracy of the direct author matching for a single post that produced an *F* score of 75% to 80% on messages containing 100 to 150 words on average. On messages containing at least 300 words, we obtained an *F* score of 0.979.

The focus of this work was to show that character-based PPM5 can identify authors with a high accuracy. Given the results, we can conclude that our hypothesis was correct. We have shown that the application of PPM5 makes an automated identification of the author of an online post possible. Our method was able to correctly attribute authors with high confidence (ie, *F* score up to 0.979). PPM was demonstrated to create the best statistical text model and to predict it almost as well as humans [40].

It should be noted that the data was very unbalanced. Some authors had hundreds of posts and some had written only tens. In addition, some authors posted long texts with descriptions and discussions and some tended to post just short replies to other posts, for example, “GF - I am so sorry,” “Congrats Lisa!,” and “Saffy - I love you.” As a result, we had to apply the text normalization. The feature set is one of the most important factors in author attribution methods. PPM is character-based because it uses character n-grams as features. Although PPM could be applied on the word-based level, it was demonstrated that it did not perform better than character-based PPM for text classification tasks [51]. A number of researchers used characters and character n-grams for author detection tasks [16,35,36]. Character n-grams captured most of the features used by other methods such as prefixes and suffixes, prepositions, pronouns, conjunctions, abbreviations and other frequent words, errors and idiosyncrasies, punctuations, special symbols (eg, smiles), and others in a natural way without complex preprocessing.

In our experiments, we found evidence that all characters from the text are important for author writing style detection. The results of the experiments demonstrated that the use of different nonalphabetical characters improved the results of character-based PPM experiments.

At the beginning of our experiments, we saw that shorter messages posted on the IVF/FET/IUI Cycle Buddies subforum tended to have poorer classification results. We selected the top 100 long posts for each of the 30 analyzed authors for our experiments. Consequently, the attribution for the IVF/FET/IUI Cycle Buddies subforum improved considerably and was even better than for the IVF Ages 35+ subforum.

Concerns about topic-specific features which helped in classification but did not actually present an author’s specific writing style were expressed in some works dedicated to the authorship attribution problem [10,13,15,18]. To verify the ability of our classification method to work on different topics, we found 10 authors who posted in several subforums. We performed an experiment using training files from 1 subforum and test posts from other ones. The attribution *F* score decreased (from 0.826 to 0.816 for IVF Ages 35+ subforum). This can be explained by short posts that we had to use (eg, “Welcome, glad you found the site!”). In the previous experiments, we were able to delete such short posts; in this one, we did not have enough posts to do this.

Comparisons with the other classification methods demonstrated that the character-based PPM method gives the best results: the *F* score for IVF Ages 35+ subforum was equal to 0.826 with use of nonalphabetical symbols and capitalized letters. The application of word-based PPM, Naïve Bayes, and SVM on the same subforum did not show as good results as the character-based PPM; for example, the best *F* score of 0.766 was obtained by SVM. To evaluate the overall performance of the algorithms, we analyzed the significance of the difference between the PPM5 results and those of Naïve Bayes and the PPM5 results and those of SVM. We applied paired *t* tests and showed that on the data gathered from each subforum and for the all algorithm pairs, the difference is statistically significant.

There were 3 strongly influencing factors in author classification: (1) number of candidate authors, (2) volume of training data, and (3) the size of test text. We analyzed the 3 factors using the data from the 2 subforums.

First, we increased the number of authors from 10 to 30 for the IVF Ages 35+ and from 10 to 50 for the IVF/FET/IUI Cycle Buddies. The main conclusion was that the method was able to handle more authors with a comparatively little loss in accuracy; the author was detected correctly for more than 90% of posts with 10 candidate authors, and we had less than 10% loss of accuracy for 30 authors. Further increase in the number of authors to 50 again decreased the accuracy by less than 10%. The decrease depended on the authors added or removed from the experimental set. Some authors tended to write comparatively long messages and their posts were easier for the method. There were some authors who tended to write a lot of short replies to other posts (10-20 words) for which the accuracy of recognition was considerably lower. Even with 50 candidate authors, the *F* score was approximately 0.83. To compare with previous results for the authorship attribution, in Kukushkina et al [16], 73% accuracy was obtained on 82 literary works in a Russian authors’ corpus, but they worked with much larger volumes of training and testing texts. Luyckx and Daelemans [17] studied dependency of accuracy on number of authors and obtained 82% accuracy on 10 authors, but it had fallen to less than 50% for 50 authors.

Next, as the training data volume was considered the most influential factor in all statistical methods, we tested the relationship between accuracy and the number of files used for training, changing the latter starting with 20 files and adding each time 20 more until we reached 180 files. The *F* score grew fast for the first 100 files—from 0.50 to 0.80—and then the growth slowed down. We hypothesize that to reach *F* score= 0.90, we have to have training data 10 times more than the test data. In practice, this is hard to obtain.

In the experiments with training data volume, our best *F* score was 0.837. It was greater than in the first set of experiments. We can explain this increase by the fact that we mixed texts of 2 subforums. The content of the subforums was different and it helped to categorize messages more precisely. This approach may be helpful when we want to identify authors posting messages on various subforums.

The last factor we tested in the experiments was the test text size. Initially we considered each post as a separate test text and made all our experiments on the basis of these settings. We noticed that some posts were extremely short (3-5 words), as in examples presented previously. Thus, we made experiments with longer fragments of test texts. Even 25-word messages were recognized with an *F* score higher than 0.60 and it grew until message length reached 300 words. The *F* score actually remained the same (approximately 0.97) for messages with lengths from 300 to 500 words. We can conclude that this was the accuracy limit for this method and it was reached for messages with the length at least 300 words.

Based on the reported study and obtained empirical evidence, we have concluded that authorship can be successfully detected in free-form messages posted on medical forums.

Limitations

We focused exclusively on IVF forums in this study; therefore, our results are limited to the IVF context. It is unclear whether these results can be generalized to forums focusing on different topics (eg, smoking cessation, heart disease, cancer). Research on different forum topics should be conducted to expand these results further.

Also, it is unclear whether the results from IVF forums would be useful in identifying anonymous users posting on other forums (eg, smoking cessation forums). Certain text features may be specific to a topic and may not be useful in identifying anonymous authors across forums of varying topics.

Practical Implications

The main implication of our results is that they should caution users from posting sensitive information anonymously. Managers of online properties that encourage user input should also alert their users about the strength of anonymity. Our experiments show that a character-based method can be more effective than word-based methods in authorship attribution. These are novel results for forum analysis because the usual methods of text analysis are based on semantics and analyze the use of words, phrases, and other text segments. We propose that to improve security of forum members, the forum organizers pay more attention to the character-based characteristics of the posts.

Does this mean that posting anonymously is futile and that all consumers should just use their real identity? Moving forward, this is not necessarily the case. Future work can extend tools such as Authorship Attributor to (1) alert anonymous posters about the ease of determining their identity so they can then make a more informed decision about the content of their posts (eg, by informing consumers with many posts on the same topic that they will have a higher chance of being reidentified through their posts than those with fewer posts on many diverse topics), and (2) automatically modify the text to adjust its features to make it correlate less with other text from the same author and, hence, frustrating tools such as Authorship Attributor.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An example of the general method of context probability interpolation.

[PDF File (Adobe PDF File), 63KB - jmir_v15i10e215_app1.pdf]

Multimedia Appendix 2

Details of PPM5 method.

[PDF File (Adobe PDF File), 41KB - jmir_v15i10e215_app2.pdf]

References

1. van der Velden M, El Emam K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. J Am Med Inform Assoc 2013 Jan 1;20(1):16-24. [doi: 10.1136/amiajnl-2012-000949] [Medline: 22771531]
2. Bobicev V, Sokolova M, Jafer Y, Schramm D. Learning sentiments from tweets with personal health information. In: Advances in Artificial Intelligence. 2012 Presented at: 25th Canadian Conference on Artificial Intelligence; May 28-30, 2012; Toronto, Canada p. 37-48. [doi: 10.1007/978-3-642-30353-1_4]
3. Weitzman ER, Cole E, Kaci L, Mandl KD. Social but safe? Quality and safety of diabetes-related online social networks. J Am Med Inform Assoc 2011 May 1;18(3):292-297 [FREE Full text] [doi: 10.1136/jamia.2010.009712] [Medline: 21262920]
4. Balicco L, Paganelli C. Access to health information: going from professional to public practices. In: Information Systems & Economic Intelligence. 2011 Presented at: 4th International Conference - SIIE’2011; Feb 17, 2011; Marrakech, Morocco.
5. Renahy E. Recherche d’information en matière de santé sur Internet: déterminants, pratiques et impact sur la santé et le recours aux soins. Paris: Université Pierre et Marie Curie-école; 2008.
6. 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]
7. Perito D, Castelluccia C, Kaafar M, Manils P. How unique and traceable are usernames? In: Privacy Enhancing Technologies. Heidelberg: Springer; 2011:1-17.

8. Li F, Zou X, Liu P, Chen JY. New threats to health data privacy. BMC Bioinformatics 2011;12 Suppl 12:S7 [FREE Full text] [doi: 10.1186/1471-2105-12-S12-S7] [Medline: 22168526]
9. Zheng R, Li J, Chen H, Huang Z. A framework for authorship identification of online messages: Writing-style features and classification techniques. J Am Soc Inf Sci 2006 Feb 01;57(3):378-393. [doi: 10.1002/asi.20316]
10. Narayanan A, Paskov H, Gong NZ, Bethencourt J, Stefanov E, Shin ECR, et al. On the feasibility of Internet-scale author identification. 2012 Presented at: IEEE Symposium on Security and Privacy (SP); 20-23 May 2012; San Francisco, CA p. 300-314. [doi: 10.1109/SP.2012.46]
11. Popescu M, Dinu L. Kernel methods and string kernels for authorship identification: The Federalist Papers case. In: Recent Advances in Natural Language Processing V. Amsterdam: John Benhamins Publishing Company; 2007:27-29.
12. Koppel M, Schler J. Authorship verification as a one-class classification problem. In: Proceedings of the twenty-first International Conference on Machine learning. 2004 Presented at: 21st International Conference on Machine Learning; 2004; Banff, AB. [doi: 10.1145/1015330.1015430]
13. Argamon S, Koppel M, Pennebaker JW, Schler J. Automatically profiling the author of an anonymous text. Commun ACM 2009 Feb 01;52(2):119-123. [doi: 10.1145/1461928.1461959]
14. Juola P. Authorship Attribution. FNT in Information Retrieval 2007;1(3):233-334. [doi: 10.1561/1500000005]
15. Koppel M, Schler J, Argamon S. Computational methods in authorship attribution. J Am Soc Inf Sci 2009 Jan;60(1):9-26. [doi: 10.1002/asi.20961]
16. Kukushkina OV, Polikarpov AA, Khmelev DV. Using literal and grammatical statistics for authorship attribution. Problems of Information Transmission 2001;37(2):172-184. [doi: 10.1023/A:1010478226705]
17. Luyckx K, Daelemans W. Authorship attribution verification with many authors limited data. In: Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1. 2008 Presented at: International Conference on Computational Linguistics; Aug 18-22, 2008; Manchester, UK.
18. Madigan D, Genkin A, Lewis DD, Lewis EGDD, Argamon S, Fradkin D, et al. Author identification on the large scale. In: Proc of the Meeting of the Classification Society of North America. Heidelberg: Springer; 2005.
19. Stamatatos W. A survey of modern authorship attribution methods. J Am Soc Inf Sci 2009 Mar;60(3):538-556. [doi: 10.1002/asi.v60:3]
20. Koppel M, Schler J, Argamon S, Messeri E. Authorship attribution with thousands of candidate authors. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006 Presented at: SIGIR conference on Research and development in information retrieval; Aug 06-10, 2006; Seattle, WA.
21. Koppel M, Schler J, Argamon S. Authorship attribution in the wild. Lang Resources & Evaluation 2011;45(1). [doi: 10.1007/s10579-009-9111-2]
22. IVF. ca Forums URL: http://ivf.ca/forums/ [accessed 2013-04-19] [WebCite Cache ID 6G0Iv85J5]
23. Malik SH, Coulson NS. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. Patient Educ Couns 2010 Nov;81(2):315-318. [doi: 10.1016/j.pec.2010.01.007] [Medline: 20144521]
24. Brinegar CS. Mark Twain and the Quintus Curtius Snodgrass Letters: A statistical test of authorship. Journal of the American Statistical Association 1963 Mar;58(301):85-96. [doi: 10.1080/01621459.1963.10500834]
25. Morton AQ. The authorship of Greek prose. Journal of the Royal Statistical Society Series A (General) 1965;128(2):169-233.
26. Tweedie FJ, Baayen RH. How variable may a constant be? Measures of lexical richness in perspective. Computers and the Humanities 1998;32(5):323-352. [doi: 10.1023/A:1001749303137]
27. Hoover DL. Frequent collocations and authorial style. Literary and Linguistic Computing 2003 Sep 01;18(3):261-286. [doi: 10.1093/llc/18.3.261]
28. Koppel M, Schler J. Exploiting stylistic idiosyncrasies for authorship attribution. 2003 Presented at: IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis; Aug 2003; Acapulco, Mexico p. 9-15 URL: http://ijcai.org/Past%20Proceedings/IJCAI-2003/content.htm
29. Zhao Y, Zobel J. Searching with style: authorship attribution in classic literature. In: Proceedings of the thirtieth Australasian conference on Computer science, Volume 62. Darlinghurst, Australia: Australian Computer Society; 2007:59-68.
30. Stamatatos E, Fakotakis N, Kokkinakis G. Computer-Based Authorship Attribution without Lexical Measures. Computers and the Humanities 2001;35:193-214. [doi: 10.1023/A:1002681919510]
31. Gerritsen CM. Authorship Attribution Using Lexical Attraction. Cambridge: Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science; 2003.
32. Hirst G, Feiguina O. Bigrams of syntactic labels for authorship discrimination of short texts. Literary and Linguistic Computing 2007 Sep 21;22(4):405-417. [doi: 10.1093/llc/fqm023]
33. Afroz S, Brennan M, Greenstadt R. Detecting hoaxes, frauds, and deception in writing style online. : IEEE Computer Society; 2012 Presented at: IEEE Symposium on Security and Privacy; May 20-23, 2012; San Francisco.
34. Abbasi A, Chen H. Writeprints. ACM Trans Inf Syst 2008 Mar 01;26(2):1-29. [doi: 10.1145/1344411.1344413]
35. Keselj V, Peng F, Cercone N, Thomas C. N-gram-based author profiles for authorship attribution. 2003 Presented at: Proceedings of the Conference of Pacific Association for Computational Linguistics (PACLING '03); Aug 2003; Halifax, NS.

36. Koppel M, Schler J, Argamon S, Winter Y. The “Fundamental Problem” of Authorship Attribution. English Studies 2012 May;93(3):284-291. [doi: 10.1080/0013838X.2012.668794]
37. Marton Y, Wu N, Hellerstein L. On compression-based text classification. 2005 Presented at: Proceedings of the 27th European conference on Advances in Information Retrieval Research; Mar 21–23, 2005; Santiago de Compostela, Spain.
38. Welch T. A technique for high-performance data compression. Computer 1984 Jun 01;17(6):8-19. [doi: 10.1109/MC.1984.1659158]
39. Project Gutenberg. URL: http://www.gutenberg.org/ [accessed 2013-04-19] [WebCite Cache ID 6G0HoyQOR]
40. Teahan W. Modelling English Text. Hamilton, New Zealand: University of Waikato; 1998.
41. Cleary J, Witten I. A comparison of enumerative and adaptive codes. IEEE Trans Inform Theory 1984 Mar;30(2):306-315. [doi: 10.1109/TIT.1984.1056889]
42. Cleary J, Witten I. Data compression using adaptive coding and partial string matching. IEEE Trans Commun 1984 Apr;32(4):396-402. [doi: 10.1109/TCOM.1984.1096090]
43. Moffat A. Implementing the PPM data compression scheme. IEEE Transactions on Communications 1990;38(11):1917-1921. [doi: 10.1109/26.61469]
44. Jelinek F, Mercer RL. Interpolated estimation of Markov source parameters from sparse data. 1980 Presented at: Workshop on Pattern Recognition in Practice; May 21-23, 1980; Amsterdam, The Netherlands.
45. Church KW, Gale WA. A comparison of the enhanced Good-Turing and deleted estimation methods for estimating probabilities of English bigrams. Computer Speech & Language 1991 Jan;5(1):19-54. [doi: 10.1016/0885-2308(91)90016-J]
46. Katz S. Estimation of probabilities from sparse data for the language model component of a speech recognizer. IEEE Trans Acoust, Speech, Signal Process 1987 Mar;35(3):400-401. [doi: 10.1109/tassp.1987.1165125]
47. Bratko A, Filipic B. Spam Filtering Using Compression Models; Technical Report IJS-DP-9227. Ljubljana, Slovenia: Department of Intelligent Systems, Jozef Stefan Institute; 2005.
48. Teahan WJ, Wen Y, McNab R, Witten IH. A Compression-based Algorithm for Chinese Word Segmentation. Computational Linguistics 2000 Sep;26(3):375-393. [doi: 10.1162/089120100561746]
49. Bobicev V, Sokolova M. An effective and robust method for short text classification. 2008 Presented at: 23rd National Conference on Artificial intelligence; Jul 13-17, 2008; Chicago.
50. Debole F, Sebastianini F. An analysis of the relative hardness of Reuters-21578 subsets. J Am Soc Inf Sci 2005 Apr;56(6):971-974. [doi: 10.1002/asi.v56:6]
51. Bobicev V. Comparison of Word-based Letter-based Text Classification. In: Recent Advances in Natural Language Processing V. Amsterdam: John Benhamins Publishing Company; 2007.
52. IVF.ca. IVF/FET/IUI Cycle Buddies URL: http://ivf.ca/forums/forum/36-ivffetiui-cycle-buddies/ [accessed 2013-08-02] [WebCite Cache ID 6Ia0p75aq]
53. IVF.ca. IVF Ages 35+ URL: http://ivf.ca/forums/forum/166-ivf-ages-35/ [accessed 2013-08-02] [WebCite Cache ID 6Ia0xPFUR]
54. Mitchell TM. Machine Learning. New York: McGraw-Hill; 1997.
55. WolframAlpha. URL: http://www.wolframalpha.com/ [accessed 2013-08-02] [WebCite Cache ID 6Ia108dS1]
56. Sebastianini F. Text categorization. In: Zanasi A, editor. Text Mining and Its Applications to Intelligence, CRM and Knowledge Management. Southampton, UK: WIT Press; 2005:109-129.
57. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management 2009;45(4). [doi: 10.1016/j.ipm.2009.03.002]
58. Jackson P, Moulinier I. Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization. In: Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization (Natural Language Processing): John Benjamins Pub Co; 2007.

Abbreviations

HTML: HyperText Markup Language
IVF: in vitro fertilization
PPM: prediction by partial matching
SVM: support vector machine

Edited by G Eysenbach; submitted 27.12.12; peer-reviewed by S Argamon, P Juola; comments to author 22.01.13; revised version received 25.06.13; accepted 29.07.13; published 03.10.13.

Please cite as:
Bobicev V, Sokolova M, El Emam K, Jafer Y, Dewar B, Jonker E, Matwin S
Can Anonymous Posters on Medical Forums be Reidentified?
J Med Internet Res 2013;15(10):e215
URL: http://www.jmir.org/2013/10/e215/
doi:10.2196/jmir.2514
PMID:24091380

Original Paper

Can a Novel Web-Based Computer Test Predict Poor Simulated Driving Performance? A Pilot Study With Healthy and Cognitive-Impaired Participants

Tobias Nef[1,2], PhD; René M Müri[1,3], MD; Rahel Bieri[1]; Michael Jäger[1]; Nora Bethencourt[1]; Ioannis Tarnanas[1]; Urs P Mosimann[1,4], MD, PhD

[1]Gerontechnology and Rehabilitation Group, University of Bern, Bern, Switzerland
[2]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland
[3]Division of Cognitive and Restorative Neurology, Department of Neurology, University Hospital Inselspital, University of Bern, Bern, Switzerland
[4]Department of Old Age Psychiatry, University Hospital of Psychiatry, University of Bern, Bern, Switzerland

Corresponding Author:
René M Müri, MD
Gerontechnology and Rehabilitation Group
University of Bern
Murtenstrasse 50
Bern, 3010
Switzerland
Phone: 41 31 632 88 17
Fax: 41 31 632 89 44
Email: Rene.muri@insel.ch

Abstract

Background: Driving a car is a complex instrumental activity of daily living and driving performance is very sensitive to cognitive impairment. The assessment of driving-relevant cognition in older drivers is challenging and requires reliable and valid tests with good sensitivity and specificity to predict safe driving. Driving simulators can be used to test fitness to drive. Several studies have found strong correlation between driving simulator performance and on-the-road driving. However, access to driving simulators is restricted to specialists and simulators are too expensive, large, and complex to allow easy access to older drivers or physicians advising them. An easily accessible, Web-based, cognitive screening test could offer a solution to this problem. The World Wide Web allows easy dissemination of the test software and implementation of the scoring algorithm on a central server, allowing generation of a dynamically growing database with normative values and ensures that all users have access to the same up-to-date normative values.

Objective: In this pilot study, we present the novel Web-based Bern Cognitive Screening Test (wBCST) and investigate whether it can predict poor simulated driving performance in healthy and cognitive-impaired participants.

Methods: The wBCST performance and simulated driving performance have been analyzed in 26 healthy younger and 44 healthy older participants as well as in 10 older participants with cognitive impairment. Correlations between the two tests were calculated. Also, simulated driving performance was used to group the participants into good performers (n=70) and poor performers (n=10). A receiver-operating characteristic analysis was calculated to determine sensitivity and specificity of the wBCST in predicting simulated driving performance.

Results: The mean wBCST score of the participants with poor simulated driving performance was reduced by 52%, compared to participants with good simulated driving performance ($P<.001$). The area under the receiver-operating characteristic curve was 0.80 with a 95% confidence interval 0.68-0.92.

Conclusions: When selecting a 75% test score as the cutoff, the novel test has 83% sensitivity, 70% specificity, and 81% efficiency, which are good values for a screening test. Overall, in this pilot study, the novel Web-based computer test appears to be a promising tool for supporting clinicians in fitness-to-drive assessments of older drivers. The Web-based distribution and scoring on a central computer will facilitate further evaluation of the novel test setup. We expect that in the near future, Web-based computer tests will become a valid and reliable tool for clinicians, for example, when assessing fitness to drive in older drivers.

(J Med Internet Res 2013;15(10):e232) doi:10.2196/jmir.2943

KEYWORDS

cognitive impairment; Web-based cognitive test; computer-based tests; driving simulation

Introduction

Cognition and Driving

Driving a car is a very challenging instrumental activity of daily living that requires the integration of high-level cognition, vision, and motor function [1]. These three domains are usually evaluated when assessing medical fitness to drive in older drivers, as they are commonly affected by age-related diseases [2]. In this article, we focus on the assessment of driving-relevant cognition in older drivers.

Since driving is a complex activity, driving performance is very sensitive to cognitive impairment [3], which is commonly the result of age-related neurodegenerative disorders (eg, Alzheimer's disease and other causes of dementia) [4]. The prevalence of neurodegenerative disorders doubles every five years after the age of 65 years [5]. Therefore, health professionals need easy access to screening tests in order to assess fitness to drive. Due to the ageing population in the Western world and increasing numbers of older drivers, identifying drivers at risk without unnecessarily restricting others is a challenging but important task [6]. This task requires reliable and valid cognitive screening tests with good sensitivity and specificity to identify at-risk drivers.

Testing Driving Performance

On-the-road testing (ORT) has been suggested as being a reasonable proxy measure for naturalistic driving in older adults with a range of cognitive impairments [7]. It is the gold standard for measuring driving performance and several authors suggest using it to assess fitness to drive in older drivers [7-10]. Despite its advantages, ORT has limitations: it is time consuming [11] and may have adverse effects that could lead to dangerous driving situations [12]. In addition, researchers cannot control for environmental conditions such as light, weather, traffic, and pedestrians [13].

That is why more recently, driving simulators (DS) have been recommended as a proxy measure for naturalistic driving and they have been introduced to assess fitness to drive of older drivers with and without cognitive impairment [14,15]. Simulators have the advantage of being intrinsically safe, providing excellent controllability, reproducibility, and standardization. Furthermore, they can be installed in specialist centers, their use is less time consuming, and they require fewer organizational demands than ORT. Several studies demonstrate the validity of DS as a proxy for naturalistic driving [13,16-19]. Disadvantages of DS are that they are expensive and large, both of which reduce their accessibility to primary care physicians and older drivers. Furthermore, there is a lack of standard test protocols and cutoff values. Finally, simulator sickness is a rather common side effect, especially for older female drivers. This interferes with DS driving performance [20].

Web-Based Computer Testing of Driving Performance

In a recent study, Rockwood et al [21] used a Web-based dataset to determine the level of cognitive impairment of older Internet users. They concluded that online tracking of people with cognitive impairment can be used to stage dementia. For review, see [22]. Also, when measuring driving performance, some of the limitations of DS can be resolved by the introduction of computer tests. By using standard personal computers with cheap off-the-shelf interface components similar to those that are used in computer gaming, cheap and easily accessible computer tests can be implemented [23]. Compared to DS, they can be easily integrated into the physician's office. Web-based computer tests use the World Wide Web to distribute the software that can run on local client computers in the physician's office. Moreover, with Web-based computer tests, the scoring of user performance can be conducted on a central server computer. This allows for generation of a dynamically growing database with normative values and ensures that all users have the same up-to-date normative values. This concept has been successfully introduced by Mills et al [24] for the application of a Web-based computer test to assess driving performance under alcohol and drug influence. The authors mention central scoring algorithms and a central database with normative values as the main advantage of Web-based computer tests. That is why we hypothesized that Web-based computer tests would also be helpful to assess fitness to drive of older drivers and we have developed the Web-based Bern Cognitive Screening Test (wBCST) to assess driving-relevant cognitive performance. The novel wBCST is based on a previously developed computer test [25].

In this pilot study, we investigate whether or not the novel wBCST correlates with DS performance and whether it is able to differentiate between participants with poor and good simulated driving performance. To have a broad and diverse test population, we recruited younger and older healthy participants as well as older participants with cognitive impairment for this study. Hence, this paper first describes the novel wBCST and the DS used, followed by a correlation analysis and a receiver operating characteristics (ROC) [26] analysis to calculate sensitivity and specificity of the novel test to predict driving simulator performance. The discussion and conclusion outline advantages and disadvantages of the wBCST and present possible future applications and research directions.

Methods

Participants

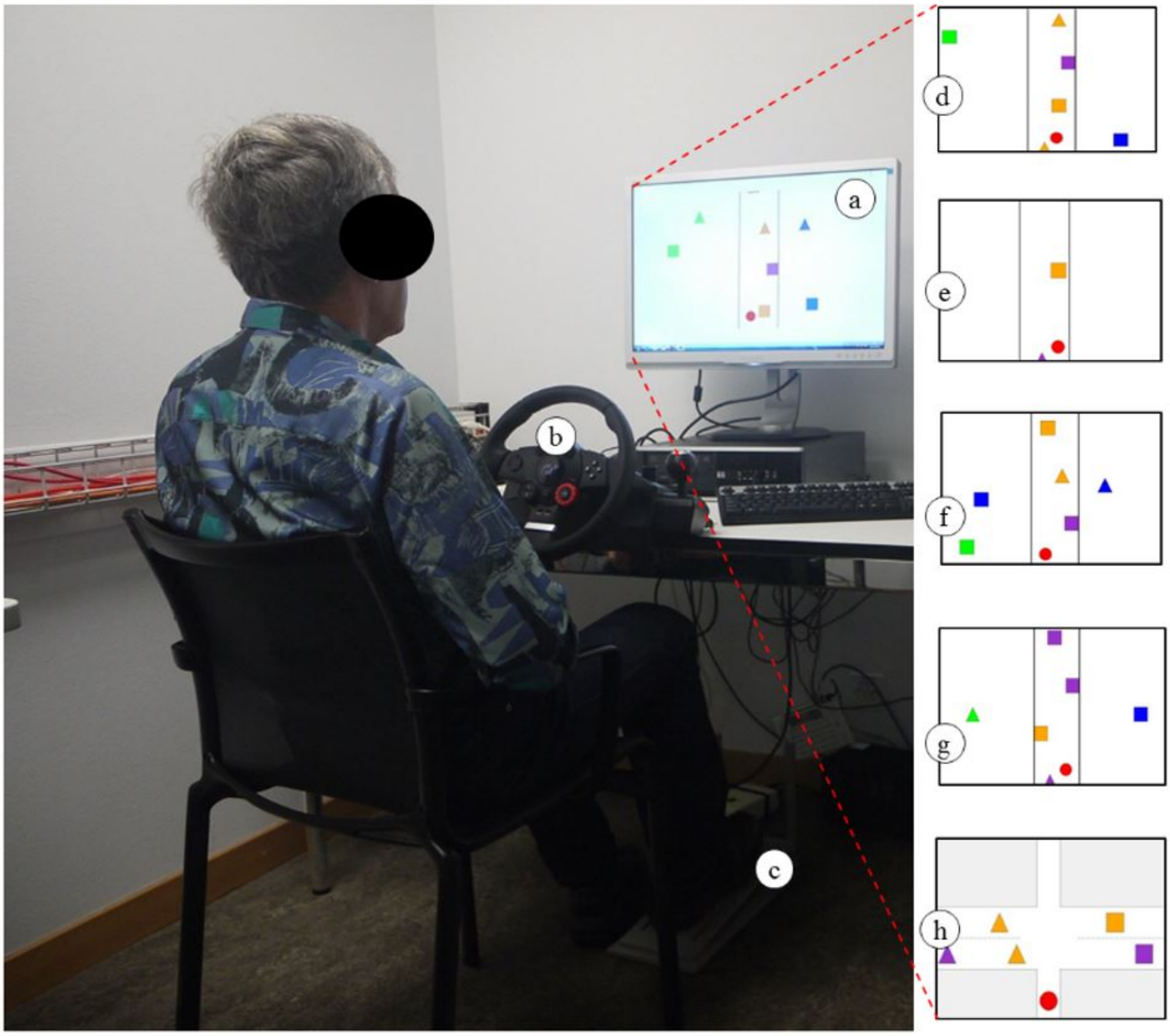
Thirty healthy younger adults (age 22-40 years), 60 healthy older adults (age >50 years), and 15 older (age >50 years) participants with cognitive impairment (Montreal Cognitive Assessment Score [MoCA] <26) [27] were recruited by advertisements in local newspapers and within the local memory clinic. All participants were required to have had a driver's license for at least two years and to have been driving during the last two years. Exclusion criteria for the study were visual impairment (corrected far visual acuity <0.5 degrees, near visual acuity <0.8 degrees) or significant motor impairment (timed-up-and-go-test >12 seconds) [28]. The study was carried out in accordance with the Declaration of Helsinki and was approved by the local ethics board. Written informed consent was obtained from all participants prior to inclusion. No compensation for participation was provided. Six participants were excluded due to visual impairment. Eighteen participants stopped the DS drive due to simulator sickness (four younger, nine older, and five participants with cognitive impairment). Their data were excluded from further analysis. Due to a malfunctioning of the DS, the data of one healthy older test person were not recorded. The data of the remaining 80 participants were included in the analysis. There were 26 young (range 22-39 years, mean 29.4 years, SD 4.7 years), 44 healthy older (range 54-85 years, mean 68.4 years, SD 5.5 years), and 10 impaired older participants (range 55-87 years, mean 72 years, SD 9.6 years). Trail Making Test A (TMT-A) [29], Trail Making Test B (TMT-B) [29], the MoCA score [27], clock-drawing test (CDT) [30], and timed-up-and-go test [28] were assessed to characterize the participants.

Web-Based Computer Test

A literature review of the most important driving-relevant cognitive functions and how they are affected by cognitive impairment was conducted [6]. This analysis was used to develop a novel computer test [25], which we extended in the context of this study to the wBCST. The wBCST measures eye-hand coordination, selective attention, divided attention, executive function, distance judgment, and speed regulation. It is composed of five subtests, each measuring one of the before-mentioned cognitive competencies. The setup of the wBCST comprises a computer screen showing the test scenario (240B1CS/00 24 inc, Philips Inc), a steering wheel (Driving Force GT, Logitec Inc) with foot pedal, and a personal computer with Windows 7 (Microsoft Inc) operating system. Figure 1 shows a healthy test person taking the wBCST and screenshots of the visual representation of the five subtests. The test persons interact with the wBCST via steering wheel and foot pedal.

Subtest 1 measures selective attention with a visual scene consisting of a simplified street in the center of the screen (Figure 1d) with objects moving from the top down. A red dot moves automatically in the horizontal direction to avoid collisions with oncoming objects and with the roadside. The user is instructed to not touch the steering wheel, but to press the foot pedal whenever a visual target (blue square) appears in the periphery (Figure 1d). In subtest 2, eye-hand coordination is measured and the same street is presented. The test person must use the steering wheel to control the horizontal position of the red dot to avoid collisions with the other objects and with the roadside (Figure 1e). There is no peripheral subtask in this test. Subtest 3 measures divided attention and both tasks of subtests 1 and 2 need to be carried out simultaneously; hence, the user must steer (central task) and react to peripheral stimuli (peripheral task) (Figure 1f). Subtest 4 is designed to measure executive functions and the user must react to more complex peripheral stimuli (green triangle and blue square) while ignoring the movement in the center of the screen (Figure 1g). In subtest 5, distance judgment and speed regulation are assessed. The user can control the velocity of a red dot with the foot pedal. The task as shown in Figure 1h is to cross intersections without colliding with the horizontally moving objects. During the tests, the false positive and false negative errors of the peripheral target detection task and the number of collisions with moving objects as well as with the street border are recorded. The test duration is about three minutes per subtest. With instruction, administration of the entire wBCST takes roughly 20 minutes. For videos of the subtests, see Multimedia Appendices 1 and 2.

Figure 1. Older participant during the wBCST evaluation. A 24-inch monitor (a) is used to present the test material and participants interact with the system via a steering wheel (b) and foot pedal (c). Each subtest uses similar graphical objects as shown in the screenshots (d-h).



Drive in the Virtual Reality Driving Simulator

A commercially available high-fidelity fixed-frame driving simulator (F12PI-3/A88, Foerst GmbH) with a custom-built virtual driving circuit was used to measure DS performance. A virtual scene was projected by three projectors (Ultra-Short focus LCD projectors, Sanyo) with 1024x786 pixel resolution onto three projection screens (1.8 x 1.4 m) that were installed in front of the driver. DS components utilized by the participants were steering wheel, brake and gas pedals, rear and side mirrors, and turn signals (Figure 2). The virtual driving scene consisted of a street with two lanes in each direction in a suburban environment. The test drive included two left turns at intersections: one with traffic crossing left to right and one with oncoming traffic. Furthermore, it comprised a construction area with road work on both lanes, a roundabout with an unexpected cyclist, a deer hiding behind trees and suddenly jumping onto the road, a car unexpectedly leaving its parking lot, and a child running into the street after a ball. The instruction was to respect traffic rules and to drive as if in a real-world environment.

Participants performed a familiarization run (three minutes) to get used to the handling of the simulator and a test drive (six minutes) during which data were recorded. Once the end of the track was reached, the car stopped automatically. If a participant felt uncomfortable, the DS was stopped. Primary outcome measure of the driving simulator was the number of errors, E_{DS} (ie, collisions, traffic rule violations, driving in the wrong traffic lane). Secondary outcome measures were mean speed variability, mean lateral acceleration, cumulated time spent on brake, distance to collision, and time to collision. Distance and time to collision are defined by how long (time, respective distance) the virtual car could continue on its current path with constant velocity until a collision would occur.

Figure 2. High-fidelity fixed-frame driving simulator with younger test subject. The steering wheel, cockpit, and parts of two projection screens with the virtual driving screen are shown.

Statistical Analysis

The results of the wBCST were ranked and scored. Thus, the score, S_i, of each subtest was calculated as $S_i= (N-rank(E_i)) / (N-1)$ with N being the total number of test samples and E_i being the number of errors in the subtest. With this formula, 1 corresponds to the best possible test results and 0 the worst. Subtest 3 reveals two results, S_{3p} for the peripheral task and S_{3c} for the central task. The overall test result, S_{wBCST}, was calculated as the mean value of S_1 to S_5. The same formula was also used to score the performance in the DS—namely, the score, S_{DS}, that was calculated out of E_{DS}, as well as the scores for speed variability, lateral acceleration, time on brake, distance to collision, and time to collision.

Pearson product-moment correlations, step-wise regressions, and associated tests of significance were calculated across S_{DS}, S_{wBCST}, and the other secondary measures. The statistical significance of the correlation was computed by transforming the correlation matrix to create a *t* statistic having n-2 degrees of freedom where n was the number of observations.

In a second step, the number of errors E_{DS} in the driving simulator was used as a classification criterion to divide the participants into two groups. As suggested by others [31], the mean value plus 1 SD was used as cutoff for one group with good DS performance and another group with poor DS performance. The mean values of the DS and the wBCST for both groups were calculated and the significance of the differences was calculated using a non-parametric Mann-Whitney U-test. Hypotheses were one-sided. Furthermore, a ROC analysis was conducted to determine the sensitivity and specificity of the wBCST when predicting DS performance. A Matlab (The MathWorks Inc) algorithm based on the work of Hanley et al [32] and implemented by Cardillo et al [33] was used to calculate the ROC curve and the area under the ROC curve (AUC). The Wilcoxon test was used to calculate whether the difference from a random classifier was significant. Also, the sensitivity, specificity, and efficiency (fraction of subjects that are correctly classified) were computed for different cutoff values.

Results

User Statistics

The data of 80 participants were included in the data analysis. Ten were assigned, based on the number of errors in the DS using mean value + 1 SD cutoff, to the poor DS performance group ($E_{DS} \geq 4$) and 70 were assigned to the good DS performance group ($E_{DS} < 4$). With this criterion, all younger participants, 86% (38/44) of the healthy older, and 60% (6/10) of the older participants with cognitive impairment were classified into the good DS group. The mean test performance in MoCA, TMT-A, TMT-B, CDT, and timed-up-and-go-test of the participants with good DS results was better compared to the other group. These differences were statistically significant.

Table 1. User statistics of all participants and the two subgroups.

	All Participants, n=80	Participants with good DS[a] performance, n=70	Participants with poor DS performance, n=10	Significance of group difference
Young (<40 years), n	26	26	0	
Older (>50 years), without cognitive impairment (MoCA[b]≥26), n	44	38	6	
Older (>50 years), with cognitive impairment (MoCA<26), n	10	6	4	
MoCA, score (SD)	28.2 (2.9)	28.5 (2.8)	26.2 (3.1)	$U_{70,10}$=179.00, *P*=.004
TMT[c] A, seconds (SD)	28.0 (14.4)	26.9 (14.2)	35.7 (14.1)	$U_{68,9}$=535.00, *P*<.001
TMT B, seconds (SD) [30]	71.6 (52.5)	68.8 (53.2)	91.2 (45.7)	$U_{68,9}$=479.00, *P*=.003
CDT, score (SD)	6.2 (1.9)	6.5 (1.6)	4.3 (3.1)	$U_{70,10}$=170.00, *P*<.001
Timed-up-and-go-test, seconds (SD)	6.9 (3.2)	6.8 (2.9)	7.5 (4.7)	$U_{63,8}$=420.50, *P*=.001

[a] DS: driving simulator

[b] MoCA: Montreal Cognitive Assessment

[c] TMT: trail making test (A and B)

[d] CDT: clock drawing test

Correlation Analysis of wBCST and DS Performance

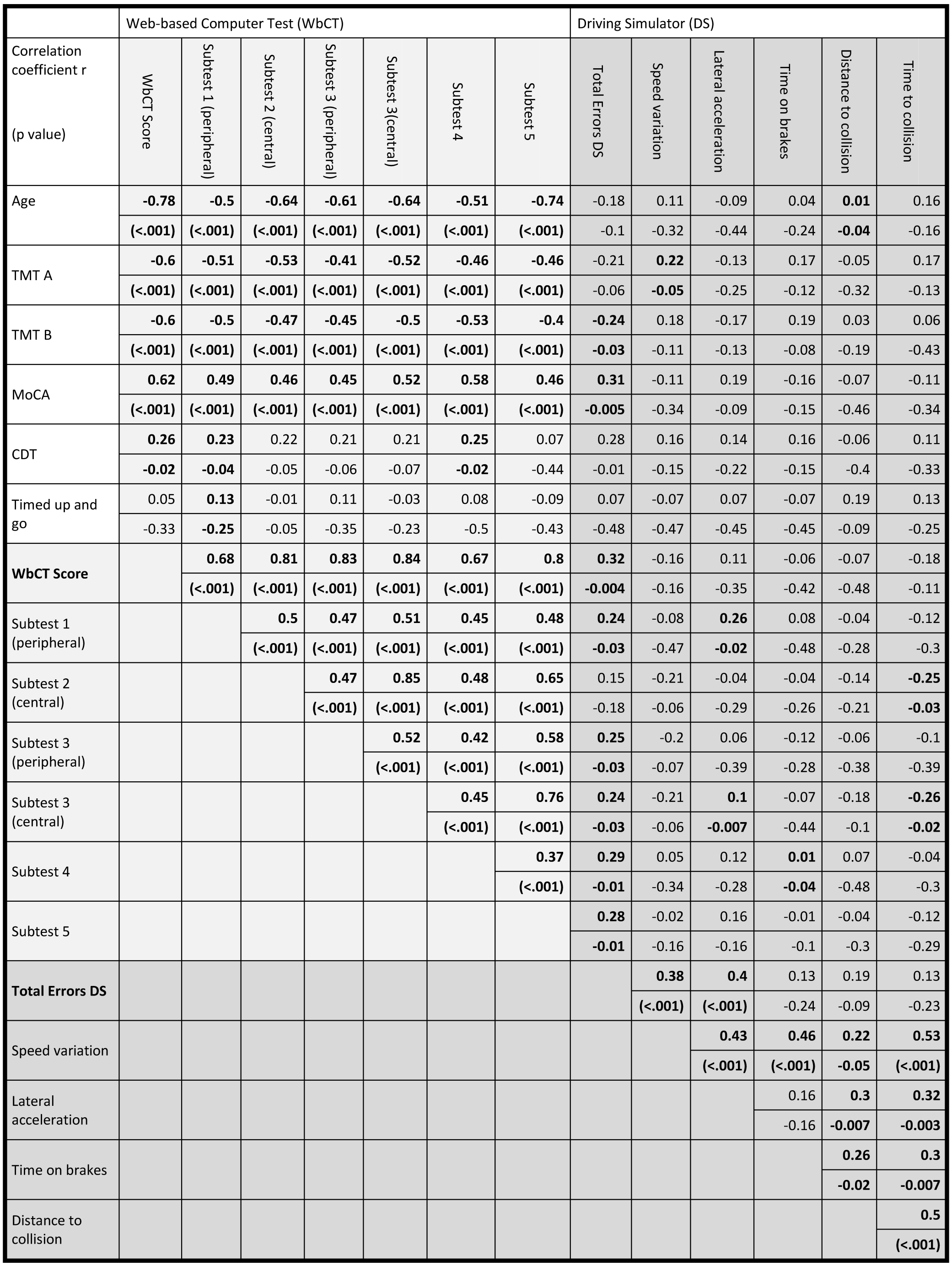
Figure 3 shows the results of the correlation analysis with the correlation coefficient *r* and the associated *P* values. The wBCST score, S_{wBCST}, correlates with the DS score, S_{DS}, with *r*=0.32 (*P*=.004). The individual results of the wBCST subtest, $S_{1...6}$, correlate with the overall score, S_{wBCST}, with *r* values varying between 0.68-0.83 with *P*<.001. In the DS, the overall score S_{DS} correlates with the speed variation with *r*=0.38 (*P*<.001), the lateral acceleration with *r*=0.40 (*P*<.001), the time spent on the brake pedal *r*=0.13 (not significant, *P*=.24), the distance to collision with *r*=0.19 (not significant, *P*=.19), and the time to collision *r*=0.31 (not significant, *P*=.23).

Sensitivity and Specificity of the wBCST in Predicting DS Performance

The ranked normalized wBCST performances of the two groups are represented in Figure 4. The group with poor DS performance performed less well in all tests. The group differences are significant for all subscores and highly significant (*P*<.001) for the overall S_{wBCST} score, subtest 2 (S_2), and subtest 5 (S_5).

The ROC curve for using S_{wBCST} to predict DS performance is shown in Figure 5. The AUC=0.80 is significantly better (*P*<.001) than a random classifier. The 95% confidence interval of the AUC is 0.68-0.92. A selection of possible cutoff scores and the corresponding sensitivity, specificity, and efficiency values is presented in Table 2.

Figure 3. Correlation analysis of the wBCST and the DS. The table shows the Pearson product-moment correlation coefficient *r* and the associated *P* values in parentheses. Correlations with *P*<.05 are indicated in bold font.



	Web-based Computer Test (WbCT)							Driving Simulator (DS)					
Correlation coefficient r (p value)	WbCT Score	Subtest 1 (peripheral)	Subtest 2 (central)	Subtest 3 (peripheral)	Subtest 3(central)	Subtest 4	Subtest 5	Total Errors DS	Speed variation	Lateral acceleration	Time on brakes	Distance to collision	Time to collision
Age	**-0.78**	**-0.5**	**-0.64**	**-0.61**	**-0.64**	**-0.51**	**-0.74**	-0.18	0.11	-0.09	0.04	**0.01**	0.16
	(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	-0.1	-0.32	-0.44	-0.24	**-0.04**	-0.16
TMT A	**-0.6**	**-0.51**	**-0.53**	**-0.41**	**-0.52**	**-0.46**	**-0.46**	-0.21	**0.22**	-0.13	0.17	-0.05	0.17
	(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	-0.06	**-0.05**	-0.25	-0.12	-0.32	-0.13
TMT B	**-0.6**	**-0.5**	**-0.47**	**-0.45**	**-0.5**	**-0.53**	**-0.4**	**-0.24**	0.18	-0.17	0.19	0.03	0.06
	(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**-0.03**	-0.11	-0.13	-0.08	-0.19	-0.43
MoCA	**0.62**	**0.49**	**0.46**	**0.45**	**0.52**	**0.58**	**0.46**	**0.31**	-0.11	0.19	-0.16	-0.07	-0.11
	(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**-0.005**	-0.34	-0.09	-0.15	-0.46	-0.34
CDT	**0.26**	**0.23**	0.22	0.21	0.21	**0.25**	0.07	0.28	0.16	0.14	0.16	-0.06	0.11
	-0.02	**-0.04**	-0.05	-0.06	-0.07	**-0.02**	-0.44	-0.01	-0.15	-0.22	-0.15	-0.4	-0.33
Timed up and go	0.05	**0.13**	-0.01	0.11	-0.03	0.08	-0.09	0.07	-0.07	0.07	-0.07	0.19	0.13
	-0.33	**-0.25**	-0.05	-0.35	-0.23	-0.5	-0.43	-0.48	-0.47	-0.45	-0.45	-0.09	-0.25
WbCT Score		**0.68**	**0.81**	**0.83**	**0.84**	**0.67**	**0.8**	**0.32**	-0.16	0.11	-0.06	-0.07	-0.18
		(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**-0.004**	-0.16	-0.35	-0.42	-0.48	-0.11
Subtest 1 (peripheral)			**0.5**	**0.47**	**0.51**	**0.45**	**0.48**	**0.24**	-0.08	**0.26**	0.08	-0.04	-0.12
			(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	**(<.001)**	**-0.03**	-0.47	**-0.02**	-0.48	-0.28	-0.3
Subtest 2 (central)				**0.47**	**0.85**	**0.48**	**0.65**	0.15	-0.21	-0.04	-0.04	-0.14	**-0.25**
				(<.001)	**(<.001)**	**(<.001)**	**(<.001)**	-0.18	-0.06	-0.29	-0.26	-0.21	**-0.03**
Subtest 3 (peripheral)					**0.52**	**0.42**	**0.58**	**0.25**	-0.2	0.06	-0.12	-0.06	-0.1
					(<.001)	**(<.001)**	**(<.001)**	**-0.03**	-0.07	-0.39	-0.28	-0.38	-0.39
Subtest 3 (central)						**0.45**	**0.76**	**0.24**	-0.21	**0.1**	-0.07	-0.18	**-0.26**
						(<.001)	**(<.001)**	**-0.03**	-0.06	**-0.007**	-0.44	-0.1	**-0.02**
Subtest 4							**0.37**	**0.29**	0.05	0.12	**0.01**	0.07	-0.04
							(<.001)	**-0.01**	-0.34	-0.28	**-0.04**	-0.48	-0.3
Subtest 5								**0.28**	-0.02	0.16	-0.01	-0.04	-0.12
								-0.01	-0.16	-0.16	-0.1	-0.3	-0.29
Total Errors DS									**0.38**	**0.4**	0.13	0.19	0.13
									(<.001)	**(<.001)**	-0.24	-0.09	-0.23
Speed variation										**0.43**	**0.46**	**0.22**	**0.53**
										(<.001)	**(<.001)**	**-0.05**	**(<.001)**
Lateral acceleration											0.16	**0.3**	**0.32**
											-0.16	**-0.007**	**-0.003**
Time on brakes												**0.26**	**0.3**
												-0.02	**-0.007**
Distance to collision													**0.5**
													(<.001)

Table 2. Receiver operating characteristic (ROC) curve analysis: possible cutoff values and corresponding sensitivity, specificity, and efficiency.

Cutoff	Sensitivity	Specificity	Efficiency
0.88	0.94	0.30	0.86
0.84	0.91	0.30	0.84
0.83	0.89	0.30	0.81
0.81	0.86	0.30	0.79
0.77	0.84	0.50	0.80
0.75	0.83	0.70	0.81
0.73	0.80	0.70	0.79
0.68	0.77	0.70	0.76
0.66	0.74	0.70	0.74
0.65	0.71	0.70	0.71
0.64	0.69	0.70	0.69
0.63	0.69	0.90	0.71
0.61	0.66	0.90	0.69
0.57	0.63	0.90	0.66
0.56	0.60	0.90	0.64
0.53	0.57	0.90	0.61
0.52	0.54	0.90	0.59
0.51	0.53	0.90	0.58

Figure 4. Web-based Bern Cognitive Screening Test (wBCST) performance of the group with good simulated driving performance (n=70) and group with poor simulator performance (n=10). All values are normalized and ranked. The score S_wBCST is the mean value of the subscores S (1…5). Subtest 3 is represented with two entries, S_3p for the peripheral recognition task and S_3c for the central steering task. Bars indicate the standard error.

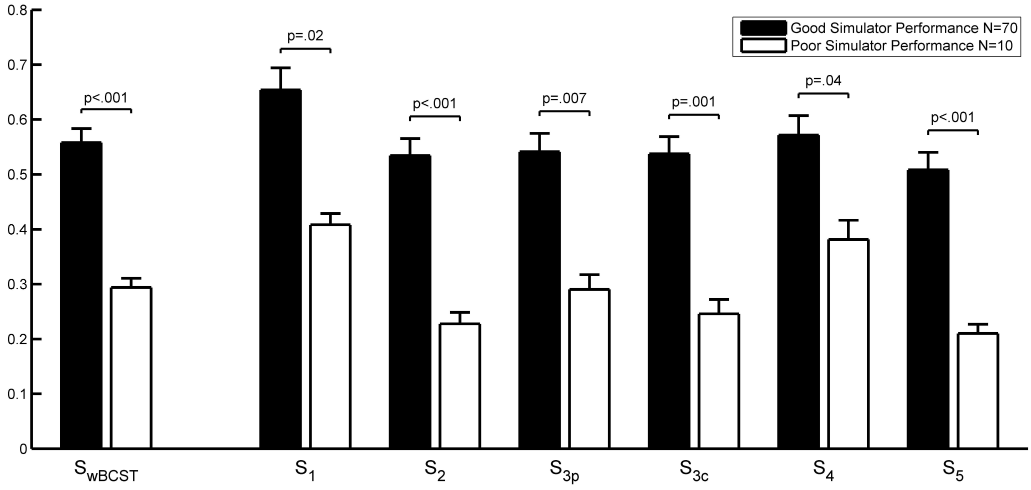
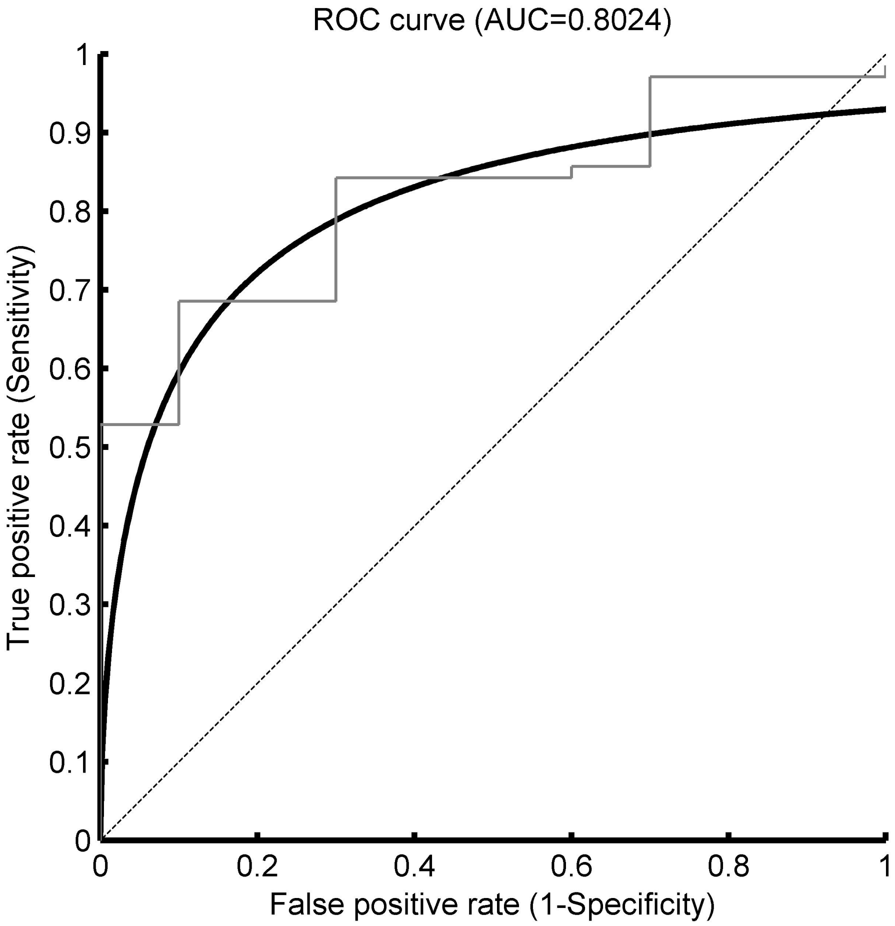


Figure 5. Receiver operating characteristics (ROC) curve for using S_wBCST to predict SD driving performance, respectively to differentiate between subjects with good and poor DS performance. The thin gray line is the empirical curve, the solid black line is the smoothed (Gaussian-based) curve, and the dotted diagonal line indicates no discrimination (50% chance).



Discussion

Principal Results

When using the criteria proposed by Cohen [34], the correlation between S_{DS} and the total score S_{wBCST} has a medium effect size ($r>0.3$) and is significant with $P=.004$, which supports the hypothesis that the wBCST correlates with DS performance. Very good correlations with large effect sizes ($r>0.5$) were found among the five subtests of the wBCST. This is an important result for the novel test since it indicates good consistency among the wBCST subtests. It could be explained by the uniform visual stimulation material that is used in the subtests (Figure 1). The correlation of the main outcome measure, S_{DS}, and the secondary outcome measures is medium and statistically significant for the speed variation and the lateral acceleration, but only small and not significant for the time spent on brakes, distance to collision, and time to collision. This was also reported by others [35]. However, it is worthwhile to emphasize here that the novel tests show a good correlation of S_{wBCST} with existing paper-pencil-based cognitive screening tools, such as the MoCA score, the TMT-A, and the TMT-B, but not with the CDT.

The group differences in the wBCST test and its subscores are all significant, which is a prerequisite for using the wBCST to predict DS performance. The AUC of the ROC curve is slightly larger than 0.80, which is generally considered a good test [33,36]. The ROC curve (Figure 5) and the corresponding table with sensitivity, specificity, and efficiency values allows for selecting a cutoff score that leads to the desired test properties. As mentioned in the introduction, the test should have both a high sensitivity and specificity. When considering Table 2, an appropriate cutoff with a high efficiency could be 0.75, which would lead to 83% sensitivity, 70% specificity, and 81% efficiency. These values are far from perfect, but are within what is to be expected for a test measuring multifaceted characteristics of the cognitive ability to drive safely [37-43].

In the present study, the overall test result, S_{wBCST}, is calculated simply as the mean of the subscores, $S_{1...5}$. It could be that one subtest is more informative than another and, in this case, the subscores should be weighted differently. The sample size of this study is too small to determine the optimal weighting parameters, but this is an interesting question for a future study.

Overall, the novel wBCST was very well accepted by the test population and there were no drop-outs in the wBCST. This is in contrast to the DS with 19 drop-outs (19.1%) due to simulator sickness (18 drop-outs) and technical problems (1 drop-out). Since it is Web-based, the distribution of the test software to different test-sites will be fairly easy and with the central scoring algorithm, data consistency among the different centers can be ensured, which will facilitate validation studies in larger populations.

The online instructions of the test procedure take about 5 minutes. We did not observe difficulties of the test persons to understand the task, except two cognitively impaired test subjects wanted to go through the instructions twice. We concluded that the instructions are clear, but participants should have the option to repeat the instructions.

The steering wheel and the foot pedal are fairly cheap accessories, but to further improve accessibility of the novel test it would be beneficial if the test could also be used with keyboard and mouse. We observed that steering wheel and foot pedal seem to increase the face-validity of the tests, but we would expect that the measured cognitive functions should be independent from the input modality. This could be investigated in a future study.

Limitations

This pilot study has some limitations that need to be mentioned. One is the large number of drop-outs in the DS that might create a selection bias, since it cannot be excluded that people with poor driving performance might be more prone to simulator sickness. Although the current literature on simulator sickness suggests other contributing factors (eg, age, gender), this cannot be excluded [20]. Another limitation is that there are no published and accepted cutoff values for the DS, which makes the selection of the cutoff for the group differentiation difficult. Compared to the DS evaluation, the administration of the wBCST takes 20 minutes, which is quite long. In future studies, one could investigate whether or not test duration could be shortened. Furthermore, the test-retest and inter-rater reliability should be assessed in a future study.

Conclusions

In this pilot study, the novel wBCST looks like a promising test to support clinicians in fitness-to-drive assessments of older drivers. The Web-based distribution and the scoring on a central computer will facilitate further evaluation of the novel test setup. In its current form, the program requires local installation on a client computer in the physician's office. This is currently not considered a disadvantage, but, of course, one could transfer the test program to run within a Web browser, which would not require local installation. The hardware requirements of the wBCST are very minor and include an office-type personal computer with Windows 7 operating system and a simple steering wheel (eg, Driving Force GT, Logitec Inc). Overall, when considering disadvantages of DS (costs, simulator sickness, space requirements), the authors believe that in many clinical environments the wBCST is better suited to support physicians in fitness-to-drive assessments than a DS. That is why we expect that in the near future, Web-based computer tests will become a valuable and usable tool for fitness-to-drive assessment in older drivers.

Acknowledgments

This work has been supported in part by the Swiss foundation for traffic safety (Fonds für Verkehrssicherheit). The authors would like to thank the members of the expert group, namely A Boss-Skupnjak and Prof Dr HP Lindenmann for their valuable input to the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screencast with animated images from the experimental setup (.wmv File, 44 MB).

[WMV File (Windows Media Video), 43MB - jmir_v15i10e232_app1.wmv]

Multimedia Appendix 2

Screencast with animated images from the experimental setup (.mp4 File, 92 MB).

[MP4 File (MP4 Video), 90MB - jmir_v15i10e232_app2.mp4]

References

1. Sherman FT. Driving: the ultimate IADL. Geriatrics 2006 Oct;61(10):9-10. [Medline: 17034267]
2. Marshall S, Demmings EM, Woolnough A, Salim D, Man-Son-Hing M. Determining fitness to drive in older persons: a survey of medical and surgical specialists. Can Geriatr J 2012 Dec;15(4):101-119 [FREE Full text] [doi: 10.5770/cgj.15.30] [Medline: 23259024]
3. Mosimann UP, Bächli-Biétry J, Boll J, Bopp-Kistler I, Donati F, Kressig RW, et al. [Consensus recommendations for the assessment of fitness to drive in cognitively impaired patients]. Praxis (Bern 1994) 2012 Mar 28;101(7):451-464. [doi: 10.1024/1661-8157/a000893] [Medline: 22454307]
4. Chi MJ, Lee CY, Wu SC. The prevalence of chronic conditions and medical expenditures of the elderly by chronic condition indicator (CCI). Arch Gerontol Geriatr 2011;52(3):284-289. [doi: 10.1016/j.archger.2010.04.017] [Medline: 20452688]

5. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Alzheimer's Disease International. Global prevalence of dementia: a Delphi consensus study. Lancet 2005 Dec 17;366(9503):2112-2117 [FREE Full text] [doi: 10.1016/S0140-6736(05)67889-0] [Medline: 16360788]
6. Wagner JT, Müri RM, Nef T, Mosimann UP. Cognition and driving in older persons. Swiss Med Wkly 2011;140:w13136 [FREE Full text] [doi: 10.4414/smw.2011.13136] [Medline: 21240690]
7. Davis JD, Papandonatos GD, Miller LA, Hewitt SD, Festa EK, Heindel WC, et al. Road test and naturalistic driving performance in healthy and cognitively impaired older adults: does environment matter? J Am Geriatr Soc 2012 Nov;60(11):2056-2062. [doi: 10.1111/j.1532-5415.2012.04206.x] [Medline: 23110378]
8. Withaar FK, Brouwer WH, van Zomeren AH. Fitness to drive in older drivers with cognitive impairment. J Int Neuropsychol Soc 2000 May;6(4):480-490. [Medline: 10902417]
9. Ott BR, Heindel WC, Papandonatos GD, Festa EK, Davis JD, Daiello LA, et al. A longitudinal study of drivers with Alzheimer disease. Neurology 2008 Apr 1;70(14):1171-1178. [doi: 10.1212/01.wnl.0000294469.27156.30] [Medline: 18216302]
10. Duchek JM, Carr DB, Hunt L, Roe CM, Xiong C, Shah K, et al. Longitudinal driving performance in early-stage dementia of the Alzheimer type. J Am Geriatr Soc 2003 Oct;51(10):1342-1347. [Medline: 14511152]
11. Di Stefano M, Macdonald W. On-the-road evaluation of driving performance. In: Driver rehabilitation and community mobility: principles and practice. St. Louis, Mo: Elsevier Mosby; 2006:255-274.
12. Lee HC, Lee AH, Cameron D. Validation of a driving simulator by measuring the visual attention skill of older adult drivers. Am J Occup Ther 2003;57(3):324-328. [Medline: 12785671]
13. Shechtman O, Classen S, Awadzi K, Mann W. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. Traffic Inj Prev 2009 Aug;10(4):379-385. [doi: 10.1080/15389580902894989] [Medline: 19593717]
14. Frittelli C, Borghetti D, Iudice G, Bonanni E, Maestri M, Tognoni G, et al. Effects of Alzheimer's disease and mild cognitive impairment on driving ability: a controlled clinical study by simulated driving test. Int J Geriatr Psychiatry 2009 Mar;24(3):232-238. [doi: 10.1002/gps.2095] [Medline: 18615781]
15. de Simone V, Kaplan L, Patronas N, Wassermann EM, Grafman J. Driving abilities in frontotemporal dementia patients. Dement Geriatr Cogn Disord 2007;23(1):1-7. [doi: 10.1159/000096317] [Medline: 17047327]
16. Bella F. Driving simulator for speed research on two-lane rural roads. Accid Anal Prev 2008 May;40(3):1078-1087. [doi: 10.1016/j.aap.2007.10.015] [Medline: 18460376]
17. Bieliauskas LA. Neuropsychological assessment of geriatric driving competence. Brain Inj 2005 Mar;19(3):221-226. [Medline: 15832896]
18. Johnson MJ, Chahal T, Stinchcombe A, Mullen N, Weaver B, Bédard M. Physiological responses to simulated and on-road driving. Int J Psychophysiol 2011 Sep;81(3):203-208. [doi: 10.1016/j.ijpsycho.2011.06.012] [Medline: 21726587]
19. Yan X, Abdel-Aty M, Radwan E, Wang X, Chilakapati P. Validating a driving simulator using surrogate safety measures. Accid Anal Prev 2008 Jan;40(1):274-288. [doi: 10.1016/j.aap.2007.06.007] [Medline: 18215559]
20. Classen S, Bewernitz M, Shechtman O. Driving simulator sickness: an evidence-based review of the literature. Am J Occup Ther 2011;65(2):179-188. [Medline: 21476365]
21. Rockwood K, Richard M, Leibman C, Mucha L, Mitnitski A. Staging dementia from symptom profiles on a care partner website. J Med Internet Res 2013;15(8):e145 [FREE Full text] [doi: 10.2196/jmir.2461] [Medline: 23924608]
22. Eghdam A, Scholl J, Bartfai A, Koch S. Information and communication technology to support self-management of patients with mild acquired cognitive impairments: systematic review. J Med Internet Res 2012;14(6):e159 [FREE Full text] [doi: 10.2196/jmir.2275] [Medline: 23165152]
23. Molnar FJ, Patel A, Marshall SC, Man-Son-Hing M, Wilson KG. Clinical utility of office-based cognitive predictors of fitness to drive in persons with dementia: A systematic review. J Am Geriatr Soc 2006 Dec;54(12):1809-1824. [doi: 10.1111/j.1532-5415.2006.00967.x] [Medline: 17198485]
24. Mills KC, Spruill SE, Walker JM, Lamson M. A clinical trial demonstration of a web-based test for alcohol and drug effects. J Stud Alcohol Drugs 2009 Mar;70(2):308-313. [Medline: 19261244]
25. Bieri R, Gruber N, Nef T, Müri R, Mosimann UP. A novel computer test to assess driving-relevant cognitive functions - a pilot study. International Psychogeriatrics 2013 Dec 01:1 (forthcoming).
26. Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. In: Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Mahwah, N.J: Lawrence Erlbaum Associates; 1996.
27. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
28. Bischoff HA, Stähelin HB, Monsch AU, Iversen MD, Weyh A, von Dechend M, et al. Identifying a cut-off point for normal mobility: a comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women. Age Ageing 2003 May;32(3):315-320 [FREE Full text] [Medline: 12720619]

29. Reitan RM. The relation of the trail making test to organic brain damage. J Consult Psychol 1955 Oct;19(5):393-394. [Medline: 13263471]
30. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry 2000 Jun;15(6):548-561. [Medline: 10861923]
31. Trustram Eve C, de Jager CA. Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: the Cognitive Function Test. Int J Geriatr Psychiatry 2013 Jun 11:1. [doi: 10.1002/gps.3993] [Medline: 23754255]
32. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [Medline: 7063747]
33. Cardillo G. Matlabcentral. 2008 May 16. ROC curve: compute a Receiver Operating Characteristics curve URL: http://www.mathworks.com/matlabcentral/fileexchange/19950-roc-curve [accessed 2013-10-14] [WebCite Cache ID 6KNSXZ9HO]
34. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
35. de Winter JC, de Groot S, Mulder M, Wieringa PA, Dankelman J, Mulder JA. Relationships between driving simulator performance and driving test results. Ergonomics 2009 Feb;52(2):137-153. [doi: 10.1080/00140130802277521] [Medline: 18972239]
36. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988 Jun 03;240(4857):1285-1293. [doi: 10.1126/science.3287615] [Medline: 3287615]
37. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001 May 8;56(9):1133-1142. [Medline: 11342677]
38. Smith T, Gildeh N, Holmes C. The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. Can J Psychiatry 2007 May;52(5):329-332. [Medline: 17542384]
39. Barberger-Gateau P, Commenges D, Gagnon M, Letenneur L, Sauvel C, Dartigues JF. Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. J Am Geriatr Soc 1992 Nov;40(11):1129-1134. [Medline: 1401698]
40. Janke MK. Assessing older drivers. Journal of Safety Research 2001 Mar;32(1):43-74. [doi: 10.1016/S0022-4375(00)00048-7]
41. Ball K. Attentional problems and older drivers. Alzheimer Dis Assoc Disord 1997 Jun;11 Suppl 1:42-47. [Medline: 9194968]
42. De Raedt R, Ponjaert-Kristoffersen I. Predicting at-fault car accidents of older drivers. Accid Anal Prev 2001 Nov;33(6):809-819. [Medline: 11579983]
43. Janke MK, Eberhard JW. Assessing medically impaired older drivers in a licensing agency setting. Accid Anal Prev 1998 May;30(3):347-361. [Medline: 9663294]

Abbreviations

AUC: area under the curve
CDT: clock drawing test
DS: driving simulator
MoCA: Montreal Cognitive Assessment
ORT: on-the-road testing
ROC: receiver operating characteristic curve
TMT A/B: trail making tests A/B
wBCST: Web-based Bern Cognitive Screening Test

Edited by G Eysenbach; submitted 05.09.13; peer-reviewed by D Novak; comments to author 27.09.13; revised version received 04.10.13; accepted 08.10.13; published 18.10.13.

Please cite as:
Nef T, Müri RM, Bieri R, Jäger M, Bethencourt N, Tarnanas I, Mosimann UP
Can a Novel Web-Based Computer Test Predict Poor Simulated Driving Performance? A Pilot Study With Healthy and Cognitive-Impaired Participants
J Med Internet Res 2013;15(10):e232
URL: http://www.jmir.org/2013/10/e232/
doi:10.2196/jmir.2943
PMID:24144946

Original Paper

Pediatric Caregiver Attitudes Toward Email Communication: Survey in an Urban Primary Care Setting

Robert Arthur Dudas[1*], MD; Michael Crocetti[2*], MD, MPH

[1]Johns Hopkins Bayview Medical Center, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, United States
[2]Johns Hopkins Community Physicians, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, United States
[*]all authors contributed equally

Corresponding Author:
Robert Arthur Dudas, MD
Johns Hopkins Bayview Medical Center
Department of Pediatrics
Johns Hopkins University School of Medicine
4940 Eastern Ave
Baltimore, MD, 21224
United States
Phone: 1 410 550 0963
Fax: 1 410 550 8073
Email: rdudas@jhmi.edu

Abstract

Background: Overall usage of email communication between patients and physicians continues to increase, due in part to expanding the adoption of electronic health records and patient portals. Unequal access and acceptance of these technologies has the potential to exacerbate disparities in care. Little is known about the attitudes of pediatric caregivers with regard to their acceptance of email as a means to communicate with their health care providers.

Objective: We conducted a survey to assess pediatric caregiver access to and attitudes toward the use of electronic communication modalities to communicate with health care providers in an urban pediatric primary care clinic.

Methods: Participants were pediatric caregivers recruited from an urban pediatric primary care clinic in Baltimore, Maryland, who completed a 35-item questionnaire in this cross-sectional study.

Results: Of the 229 caregivers who completed the survey (91.2% response rate), 171 (74.6%) reported that they use email to communicate with others. Of the email users, 145 respondents (86.3%) stated that they would like to email doctors, although only 18 (10.7%) actually do so. Among email users, African-American caregivers were much less likely to support the expanded use of email communication with health care providers (adjusted OR 0.34, 95% CI 0.14-0.82) as were those with annual incomes less than US $30,000 (adjusted OR 0.26, 95% CI 0.09-0.74).

Conclusions: Caregivers of children have access to email and many would be interested in communicating with health care providers. However, African-Americans and those in lower socioeconomic groups were much less likely to have positive attitudes toward email.

(J Med Internet Res 2013;15(10):e228) doi:10.2196/jmir.2738

KEYWORDS

electronic mail; email; primary health care; communication; health care disparities; pediatrics

Introduction

It has been more than 10 years since the Institute of Medicine recommended "patients should receive care whenever they need it and in many forms, not just face-to-face visits" and "access to care should be provided over the Internet" [1]. Although electronic methods have the potential to increase communications and the quality and efficiency of care, most research has focused on adult populations [2] and the evidence base is inadequate to assess the effect of email for clinical communication between caregivers and health care professionals [3]. Nevertheless, a study conducted with parents who used email to communicate with their child's pediatrician found that 98% of parents were very satisfied with their experience [4], whereas another study suggested that email enabled physicians to answer medical questions in less time compared with telephone messaging [5]. Surveys from the United States have revealed wide variability in email practices with 16% of physicians using email to communicate with patients in a survey of primary care practitioners to as many as 72% in a large outpatients' department [6,7]. A younger patient population may correlate with increased usage as the majority (79%) of doctors at a student health center in Finland reported email use with their patients [8].

Overall usage of email communication between patient and physician continues to increase, but because of expanding usage of electronic health records and patient portals, there is a concern that unequal access to these technologies could exacerbate disparities in care. Medicaid patients and black patients were less likely to have access to email in pilot studies involving an email intervention [9]. Others have shown that nonwhite or low socioeconomic status patients were less likely to use email in a Web-based secure portal despite access to the necessary technology [10].

We hypothesize that parents bringing their children to a pediatric primary care clinic have access to the Internet and email and would be interested in communicating with their health care providers by these modalities. The aim of this study is to document pediatric caregiver attitudes toward and access to these technologies in an urban pediatric primary care clinic.

Methods

Overview

We conducted a cross-sectional survey involving a convenience sample of 300 caregiver-child dyads with children aged from birth to 21 years presenting for care at an urban pediatric primary care clinic in Baltimore, Maryland. The Institutional Review Board of Johns Hopkins University approved this study.

Study Setting and Participants

This study was undertaken at a community teaching hospital affiliated with a major academic center. The primary care clinic was staffed by 6 pediatric providers and 15 pediatric residents (3 residents each afternoon for continuity clinic).The pediatric clinic has an annual pediatric volume of approximately 10,000.

To be eligible, pediatric patients accompanied by their caregiver had to present to the clinic between November 2010 and January 2011 during clinic hours of 09:00-17:00. Patients were excluded if they were non–English speaking because we lacked the resources to interview them.

Survey Instrument

We developed a questionnaire based upon existing literature [4,11,12] that was piloted on a group of 10 caregivers before study initiation to identify ambiguity. Changes were made to clarify wording before study initiation based upon feedback. We used a final 35-item paper-based questionnaire that included multiple selection and 5-point scale questions. Survey domains included (1) demographic information, including caregiver and child age, sex, race, education, annual family income, and insurance type; (2) caregiver email usage patterns; and (3) caregiver attitudes toward email.

Study Protocol

A 35-item survey instrument was distributed to 300 consecutive English-speaking caregivers presenting to the clinic. The survey was distributed to the caregiver by a registrar during the check-in process. The survey was either returned or self-administered and collected at the conclusion of the health care encounter. No incentives were offered to complete the questionnaire, which took approximately 10 minutes to complete.

Statistical Analysis

Sample size calculations were based upon the determination of the proportion of email users in this cross-sectional study. It was predicted that approximately 80% of the surveyed population would be email users. We assumed a 95% level of confidence and set our precision at .05 to yield a sample size of 246. Data analysis was performed with the use of Stata version 9.2 (StataCorp LP, College Station, TX, USA). Frequencies and simple means were calculated for each variable, where appropriate. For items utilizing a 5-point scale, we considered both "strongly agree" and "agree" as agreement with that item. Unadjusted and adjusted logistic regression was used to assess associations between demographic variables and binary attitudinal variables. The amount of missing data for study variables of interest was minimal with an average missing data rate per variable of 0.8% and the largest missing data occurring for the insurance variable at 6.4%. In the regression analysis, we used listwise deletion to account for missing data. Results are reported as odds ratios (OR) and adjusted odds ratios (adjusted OR) with 95% confidence intervals (CI). A *P* value less than .05 was considered significant.

Results

From 300 consecutive English-speaking caregivers, 229 surveys were available for analysis (Figure 1).

Of the 229 participants, 171 (74.6%) stated that they use email to communicate with others. Table 1 reveals the demographic characteristics of the caregivers based on email usage. Most respondents were mothers (181/229, 79.0%) with an average age of 33.6 years (SD 10). Caregivers who used email were more likely to have a college or greater education (*P*=.003), higher income (*P*<.001), and commercial insurance (*P*=.002) (Table 1).

Of those who did use email, 86.3% (145/168) reported that they would like to communicate with their provider by this method, although only 10.7% (18/168) reported doing so (Table 2).

A large percentage of respondents that use email agree or strongly agree that more doctors should offer email communication to their patients (135/171, 78.9%) (Table 3). Many also stated that email would be a good way to make an appointment (130/170, 76.4%), increase contact with their child's provider (131/170, 77.0%), and improve communication with their provider (125/170, 73.5%).

Email users were asked to select their preferred method to receive test and x-ray results; they were evenly split between email (40.2%, 68/169) and phone (40.2%, 68/169) as preferred method with a minority preferring regular mail (16.6%, 28/169) or text messaging (2.9%, 5/169). Additionally, most caregivers felt that email was an appropriate modality to discuss many of the conditions commonly encountered in pediatric practice, including cold symptoms and sleep and weight issues, although they were slightly less comfortable discussing behavior and development issues by email (Table 4).

In the adjusted logistic regression model, African-American caregivers were less likely to agree with the following statements: more providers should offer email (adjusted OR 0.34, 95% CI 0.14-0.82), email would increase contact with the provider (adjusted OR 0.41, 95% CI 0.19-0.91), email with the provider would be satisfying (adjusted OR 0.32, 95% CI 0.14-0.75), and email would be an easy way to make an appointment (adjusted OR 0.4, 95% CI 0.18-0.88) (Table 5).

Table 1. Caregiver demographics by email usage (N=229).

Variable	Do you ever use email to communicate with others?		*P* value
	Yes (n=171)	No (n=58)	
Mothers surveyed, n (%)	140 (82.3)[a]	41 (70.6)	.06
Age (years), mean (SD)	33.3 (9)	34.1 (12)	.51
Race, n (%)			
African-American	74 (43.2)	31 (53.4)	.22
Caucasian	80 (47.3)	25 (43.1)	.65
Other	17 (9.9)	2 (3.4)	.17
Education, n (%)			
< High school	13 (7.6)[a]	10 (18.1)[b]	.04
High school (GED)	72 (42.3)[a]	37 (67.3)[b]	.44
College or greater	86 (50.6) [a]	8 (14.5) [b]	<.001
Income (US $), n (%)			
≤20,000	34 (20.0)[a]	28 (50.9)[b]	<.001
20,001 - 40,000	44 (25.9)[a]	9 (16.4)[b]	.20
> 40,000	59 (34.7)[a]	1 (1.8)[b]	<.001
Don't know/refuse	33 (19.4)[a]	17 (30.9)[b]	.09
Health insurance, n (%)			
Commercial/private	90 (56.2)[c]	18 (32.7)[b]	.003
Medicaid	61 (38.1)[c]	27 (49.1)[b]	.16
None	1 (0.6)[c]	0 (0.0)[b]	.99
Don't know/refuse	8 (5.0)[c]	10 (18.2)[b]	.004

[a] Percentages calculated based on 170 responses.

[b] Percentages calculated based on 55 responses.

[c] Percentages calculated based on 160 responses.

Table 2. Practices of email users.

Variable	Frequency, n (%)
How often do you check email? [a]	
Daily or more	102 (60.0)
Few times weekly	35 (20.6)
Weekly or less	33 (19.4)
Would like to email child’s doctor (yes)[b]	145 (86.3)
Communicates currently with child’s doctor by email (yes)[b]	18 (10.7)

[a] Percentages based 170 responses.

[b] Percentages based on 168 responses.

Table 3. Email users attitudes toward email communication with doctors (n=171).

Variable	Frequency of agreement, n (%)
More doctors should offer email communication	135 (78.9)
Email would increase contact with my child’s doctor[a]	131 (77.1)
Email would distance us from my child’s doctor	19 (11.1)
Email would improve communication with my child’s doctor[a]	125 (73.5)
Email with my child’s doctor would be satisfying	131 (76.6)
Email would be a good/easy way to make appointment[a]	130 (76.5)
Email hackers are a worry[a]	58 (34.1)

[a] Percentage calculated based on 170 responses.

Table 4. Email communication to discuss common pediatric conditions (N=171).

Pediatric condition	Agreement, n (%)
Cold symptoms	141 (82.5)
Earache	139 (81.3)
Sleep	136 (79.5)
Weight issues	136 (79.5)
Fever	135 (78.9)
Constipation	133 (77.8)
Diarrhea	133 (77.8)
Vomiting	129 (75.4)
Feeding/diet	134 (78.4)
Safety topics[a]	133 (78.2)
Pink eye	131 (76.7)
Toilet training[b]	129 (76.3)
Immunizations	125 (73.1)
Colic[b]	121 (71.6)
Behavior	121 (70.8)
Development[a]	120 (70.6)

[a] Percentages calculated based on 170 responses.

[b] Percentages calculated based on 169 responses.

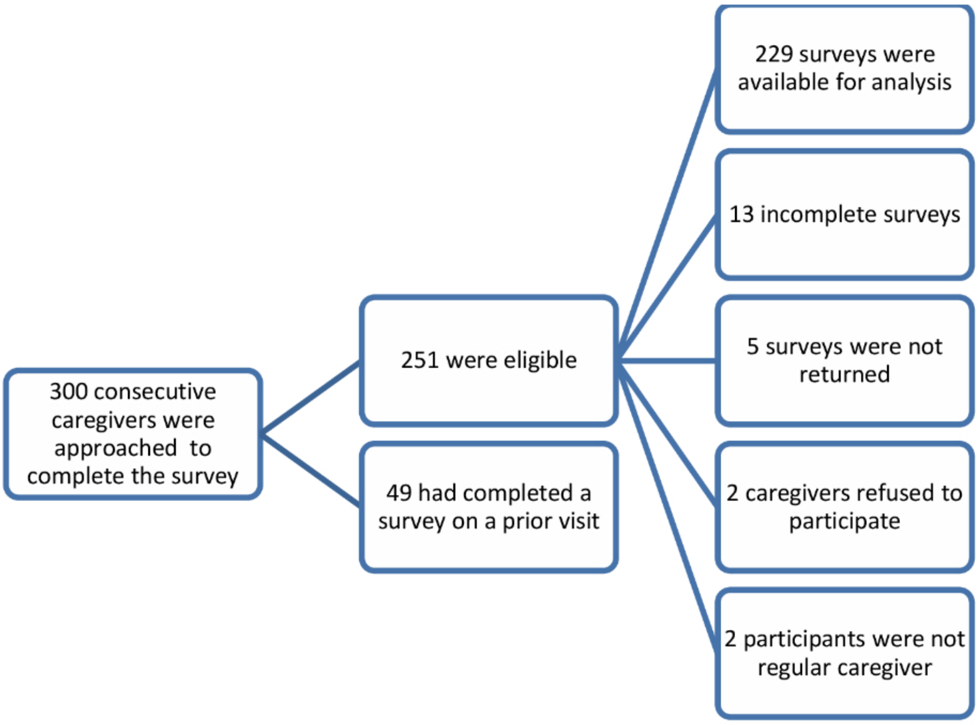
Table 5. Logistic regression of email users' attitudes as a function of age, education level, insurance type, family income, and race.

Covariate (n)	Agreement	
	OR (95% CI)	Adjusted[a] OR (95% CI)
More providers should offer email (171)		
Caregiver age	1.05 (1.00-1.09)	1.00 (0.94-1.08)
Completed > high school	3.45 (1.54-7.72)	2.28 (0.88-5.91)
Medicaid insurance	0.47 (0.22-1.02)	0.75 (0.30-1.9)
Income ≤ US $30,000	0.17 (0.07-0.43)	0.26 (0.09-0.74)
African-American race	0.40 (0.19-0.85)	0.34 (0.14-0.82)
Email would increase contact with provider (170)		
Caregiver age	1.02 (0.98-1.06)	1.02 (0.96-1.09)
Completed > high school	1.92 (0.92-3.99)	1.33 (0.56-3.17)
Medicaid insurance	0.64 (0.30-1.35)	0.73 (0.31-1.72)
Income ≤ US $30,000	0.43 (0.20-0.92)	0.59 (0.24-1.45)
African-American race	0.38 (0.18-0.79)	0.41 (0.19-0.91)
Email would distance us from provider (171)		
Caregiver age	0.95 (0.89-1.01)	0.95 (0.87-1.05)
Completed > high school	0.40 (0.15-1.12)	0.49 (0.13-1.82)
Medicaid insurance	2.16 (0.79-5.9)	1.96 (0.56-6.91)
Income ≤ US $30,000	3.65 (1.15-11.51)	1.51 (0.39-5.88)
African-American race	1.53 (0.59-3.98)	1.40 (0.46-4.27)
Email would improve communication with provider (170)		
Caregiver age	1.03 (0.99-1.07)	1.05 (0.99-1.12)
Completed > high school	1.82 (0.91-3.63)	1.37 (0.59-3.20)
Medicaid insurance	0.83 (0.41-1.68)	1.10 (0.48-2.52)
Income ≤ US $30,000	0.48 (0.24-0.99)	0.57 (0.24-1.38)
African-American race	0.45 (0.23-0.90)	0.49 (0.23-1.04)
Email with provider would be satisfying (171)		
Caregiver age	1.01 (0.97-1.05)	1.00 (0.95-1.07)
Completed > high school	2.34 (1.12-4.88)	1.67 (0.68-4.12)
Medicaid insurance	0.55 (0.26-1.17)	0.79 (0.33-1.91)
Income ≤ $30,000	0.21 (0.09-0.49)	0.25 (0.09-0.66)
African-American race	0.31 (0.15-0.65)	0.32 (0.14-0.75)
Email would be easy way to make appointment (170)		
Caregiver age	1.02 (0.98-1.06)	0.99 (0.94-1.06)
Completed > high school	1.78 (0.86-3.65)	1.90 (0.79-4.58)
Medicaid insurance	0.91 (0.44-1.90)	1.39 (0.59-3.29)
Income ≤ US $30,000	0.47 (0.23-0.99)	0.53 (0.21-1.33)
African-American race	0.41 (0.20-0.85)	0.40 (0.18-0.88)
Email hackers are a worry (170)		
Caregiver age	0.97 (0.94-1.01)	0.98 (0.93-1.04)
Completed > high school	0.53 (0.28-1.01)	0.83 (0.38-1.81)
Medicaid insurance	1.78 (0.92-3.44)	1.35 (0.62-2.95)
Income ≤ US $30,000	3.06 (1.56-6.02)	2.21 (0.97-5.02)

Covariate (n)	Agreement	
	OR (95% CI)	Adjusted[a] OR (95% CI)
African-American race	1.60 (0.85-3.02)	1.34 (0.65-2.73)

[a] Adjusted for all other covariates in a multiple logistic regression model.

Figure 1. Survey response flow diagram.



Discussion

Principal Results

Most caregivers in our urban population have access to email and are interested in communicating with their child's providers by this method, although only 11% currently communicate with their provider via email. Overall, attitudes toward email were favorable with three-quarters of email users reporting that email would improve communication with their provider and be satisfying. Despite the generally favorable disposition toward email, most caregivers preferred not to receive test or x-ray results by email. This may be related to the finding that 34% of caregivers remain concerned about security issues surrounding email, which suggests that there are content areas that caregivers feel less comfortable discussing via email. Regarding discussing specific pediatric conditions commonly encountered in primary care, there was much greater acceptance of such topics with no notable differences among the types of conditions.

Comparison With Prior Work

Our study suggests that caregivers with higher education, higher income, and commercial insurance were more likely to use email to communicate. Furthermore, of those caregivers who did report using email, African-American caregivers and those with incomes less than US $30,000 were less likely to have favorable views toward email as a tool to communicate with their health care providers. Although others have suggested that the digital divide is a function of decreased access to email for these groups [9], our data offer further evidence that disparities persist across racial/ethnic and socioeconomic groups even after controlling for access to email and Internet [10]. There may be factors related to the quality of Internet connection or perhaps cultural preferences pertaining to privacy or other factors as yet not determined. Increased emphasis on the meaningful use of electronic health records has led health care systems to develop patient portals that offer access to health information and frequently allow for 2-way secure messaging between patients and providers. However, recent studies are reporting disparities in adult and pediatric patient portal use along racial and socioeconomic lines [13-15]. A recent review on patient-provider email suggests that email has tremendous potential to improve health care communication between patients and providers and should lead to improved satisfaction and quality of care [16]. Yet, the evidence base to assess the effect of email on clinical communication remains limited and of poor quality [3]. Many questions concerning access, acceptance, privacy, and security issues remain unanswered and established national guidelines are currently lacking.

Limitations

Study limitations include the small sample size and cross-sectional design. However, we surveyed consecutive caregivers and our response rate was high. Other limitations include bias from our convenience sample at a single urban clinic limiting the ability to generalize our results to other populations. Also, we did not make a distinction between personal and professional email usage (or secure patient portals) which could have implications upon caregiver attitudes toward privacy concerns. Lastly, our survey instrument lacks formal testing for reliability and validity.

Further research should continue to closely monitor for exacerbations of existing disparities in pediatrics as the medical community further embraces email and other electronic data communication methods. Text messaging may serve as another alternative communication modality as it too has been shown to be generally accepted by parents [17]. Investigators should help determine which communication modalities are best suited for conveying specific information such as test results, or providing information about medical conditions while taking into consideration the inherent ethical and privacy concerns raised by all forms of communication [18,19].

Conclusions

Caregivers of children in an urban pediatric primary care practice have access to email and would be interested in communicating with health care providers by this method. African-American caregivers and those in lower socioeconomic groups hold less favorable views toward email communication; thus, the use of email may exacerbate existing disparities in health care delivery. Future studies should examine the reasons for these attitudinal differences.

Acknowledgments

Publication of this article was funded in part by the Open Access Promotion Fund of the Johns Hopkins University Libraries.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 14KB - jmir_v15i10e228_app1.pdf]

References

1. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
2. 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]
3. Atherton H, Sawmynaden P, Sheikh A, Majeed A, Car J. Email for clinical communication between patients/caregivers and healthcare professionals. Cochrane Database Syst Rev 2012;11:CD007978. [doi: 10.1002/14651858.CD007978.pub2] [Medline: 23152249]
4. Anand SG, Feldman MJ, Geller DS, Bisbee A, Bauchner H. A content analysis of e-mail communication between primary care providers and parents. Pediatrics 2005 May;115(5):1283-1288 [FREE Full text] [doi: 10.1542/peds.2004-1297] [Medline: 15867036]
5. 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]
6. Brooks RG, Menachemi N. Physicians' use of email with patients: factors influencing electronic communication and adherence to best practices. J Med Internet Res 2006;8(1):e2 [FREE Full text] [doi: 10.2196/jmir.8.1.e2] [Medline: 16585026]
7. Gaster B, Knight CL, DeWitt DE, Sheffield JV, Assefi NP, Buchwald D. Physicians' use of and attitudes toward electronic mail for patient communication. J Gen Intern Med 2003 May;18(5):385-389 [FREE Full text] [Medline: 12795738]
8. Castrén J, Niemi M, Virjo I. Use of email for patient communication in student health care: a cross-sectional study. BMC Med Inform Decis Mak 2005 Jan 27;5:2 [FREE Full text] [doi: 10.1186/1472-6947-5-2] [Medline: 15676077]
9. 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]
10. Hsu J, Huang J, Kinsman J, Fireman B, Miller R, Selby J, et al. Use of e-Health services between 1999 and 2002: a growing digital divide. J Am Med Inform Assoc 2005;12(2):164-171 [FREE Full text] [doi: 10.1197/jamia.M1672] [Medline: 15561786]

11. Kleiner KD, Akers R, Burke BL, Werner EJ. Parent and physician attitudes regarding electronic communication in pediatric practices. Pediatrics 2002 May;109(5):740-744. [Medline: 11986430]
12. 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]
13. 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]
14. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
15. Ketterer T, West DW, Sanders VP, Hossain J, Kondo MC, Sharif I. Correlates of patient portal enrollment and activation in primary care pediatrics. Acad Pediatr 2013;13(3):264-271. [doi: 10.1016/j.acap.2013.02.002] [Medline: 23680344]
16. 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]
17. Dudas RA, Pumilia JN, Crocetti M. Pediatric caregiver attitudes and technologic readiness toward electronic follow-up communication in an urban community emergency department. Telemed J E Health 2013 Jun;19(6):493-496. [doi: 10.1089/tmj.2012.0166] [Medline: 23570276]
18. Davis Giardina T, Singh H. Should patients get direct access to their laboratory test results? An answer with many questions. JAMA 2011 Dec 14;306(22):2502-2503. [doi: 10.1001/jama.2011.1797] [Medline: 22122864]
19. Greene AH. HIPAA compliance for clinician texting. J AHIMA 2012 Apr;83(4):34-36. [Medline: 22567799]

Edited by G Eysenbach; submitted 28.05.13; peer-reviewed by P Altorjai, I Horn, Y Jayaratne; comments to author 22.07.13; revised version received 14.08.13; accepted 15.09.13; published 23.10.13.

Please cite as:
Dudas RA, Crocetti M
Pediatric Caregiver Attitudes Toward Email Communication: Survey in an Urban Primary Care Setting
J Med Internet Res 2013;15(10):e228
URL: http://www.jmir.org/2013/10/e228/
doi:10.2196/jmir.2738
PMID:24152542

Review

Wikis and Collaborative Writing Applications in Health Care: A Scoping Review

Patrick M Archambault[1,2,3,4*], MD, MSc, FRCPC; Tom H van de Belt[5,6*], MSc; Francisco J Grajales III[7,8], BHK, MSc, CSEP, CEP; Marjan J Faber[9], PhD; Craig E Kuziemsky[10], PhD; Susie Gagnon[2], MA; Andrea Bilodeau[2], MSc.; Simon Rioux[2], BA; Willianne LDM Nelen[5], MD, PhD; Marie-Pierre Gagnon[4,11], PhD; Alexis F Turgeon[4,12], MD, MSc, FRCPC; Karine Aubin[11,13], MA; Irving Gold[14], MA, MCA; Julien Poitras[1,15], MD, FRCPC; Gunther Eysenbach[16,17,18], MD, MPH, FACMI; Jan AM Kremer[5,9], MD, PhD; France Légaré[1,4,19], MD, PhD, CCFP, FCFP

[1]Département de médecine familiale et médecine d'urgence, Université Laval, Québec, QC, Canada

[2]Centre de recherche du Centre hospitalier affilié universitaire de Lévis, Centre de santé et de services sociaux Alphonse-Desjardins (Centre hospitalier affilié universitaire de Lévis), Lévis, QC, Canada

[3]Division de soins intensifs, Université Laval, Québec, QC, Canada

[4]Centre de recherche du CHU de Québec, Axe Santé des populations - Pratiques optimales en santé, Traumatologie – Urgence – Soins Intensifs, Québec, QC, Canada

[5]Department of obstetrics and gynecology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands

[6]Radboud REshape and Innovation Center, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands

[7]Social Media Working Group, International Medical Informatics Association, Geneva, Switzerland

[8]eHealth Strategy Office, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[9]Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands

[10]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[11]Faculté des sciences infirmières, Université Laval, Québec, QC, Canada

[12]Division de soins intensifs, Département d'anesthésiologie, Université Laval, Québec, QC, Canada

[13]Université du Québec à Rimouski, Campus de Lévis, Lévis, QC, Canada

[14]Association of Faculties of Medicine of Canada, Ottawa, ON, Canada

[15]Centre de santé et de services sociaux Alphonse-Desjardins (Centre hospitalier affilié universitaire de Lévis), Lévis, QC, Canada

[16]University Health Network, Centre for Global EHealth Innovation & Techna Institute, Toronto, ON, Canada

[17]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[18]JMIR Publications Inc., Toronto, ON, Canada

[19]Canada Research Chair in Implementation of Shared Decision Making in Primary Care, Québec, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Patrick M Archambault, MD, MSc, FRCPC
Centre de recherche du Centre hospitalier affilié universitaire de Lévis
Centre de santé et de services sociaux Alphonse-Desjardins (Centre hospitalier affilié universitaire de Lévis)
143 rue Wolfe
Lévis, QC, G6V 3Z1
Canada
Phone: 1 418 835 7121 ext 3905
Fax: 1 418 835 7276
Email: patrick.m.archambault@gmail.com

Abstract

Background: Collaborative writing applications (eg, wikis and Google Documents) hold the potential to improve the use of evidence in both public health and health care. The rapid rise in their use has created the need for a systematic synthesis of the evidence of their impact as knowledge translation (KT) tools in the health care sector and for an inventory of the factors that affect their use.

Objective: Through the Levac six-stage methodology, a scoping review was undertaken to explore the depth and breadth of evidence about the effective, safe, and ethical use of wikis and collaborative writing applications (CWAs) in health care.

Methods: Multiple strategies were used to locate studies. Seven scientific databases and 6 grey literature sources were queried for articles on wikis and CWAs published between 2001 and September 16, 2011. In total, 4436 citations and 1921 grey literature items were screened. Two reviewers independently reviewed citations, selected eligible studies, and extracted data using a standardized form. We included any paper presenting qualitative or quantitative empirical evidence concerning health care and CWAs. We defined a CWA as any technology that enables the joint and simultaneous editing of a webpage or an online document by many end users. We performed qualitative content analysis to identify the factors that affect the use of CWAs using the Gagnon framework and their effects on health care using the Donabedian framework.

Results: Of the 111 studies included, 4 were experimental, 5 quasi-experimental, 5 observational, 52 case studies, 23 surveys about wiki use, and 22 descriptive studies about the quality of information in wikis. We classified them by theme: patterns of use of CWAs (n=26), quality of information in existing CWAs (n=25), and CWAs as KT tools (n=73). A high prevalence of CWA use (ie, more than 50%) is reported in 58% (7/12) of surveys conducted with health care professionals and students. However, we found only one longitudinal study showing that CWA use is increasing in health care. Moreover, contribution rates remain low and the quality of information contained in different CWAs needs improvement. We identified 48 barriers and 91 facilitators in 4 major themes (factors related to the CWA, users' knowledge and attitude towards CWAs, human environment, and organizational environment). We also found 57 positive and 23 negative effects that we classified into processes and outcomes.

Conclusions: Although we found some experimental and quasi-experimental studies of the effectiveness and safety of CWAs as educational and KT interventions, the vast majority of included studies were observational case studies about CWAs being used by health professionals and patients. More primary research is needed to find ways to address the different barriers to their use and to make these applications more useful for different stakeholders.

(J Med Internet Res 2013;15(10):e210) doi:10.2196/jmir.2787

KEYWORDS

collaborative writing applications; collaborative authoring; knowledge management; crowdsourcing; medical informatics; ehealth; Internet; Wiki; Wikipedia; Google Docs; Google Knol; Web 2.0; knowledge translation; evidence-based medicine; participatory med

Introduction

Health care decision makers—providers, patients, managers, and policy makers—are failing to use research evidence to inform their decisions [1]. By involving knowledge users in the creation and dissemination of knowledge [2], social media—highly accessible, Web-based, interactive vehicles of communication—have the potential to empower users to apply knowledge in practice. Acknowledging this potential and recognizing that social media capitalizes on the free and open access to information, scientists, opinion leaders, and patient advocates have called for research to determine whether social media can equip decision-making constituencies to improve health care delivery [3,4] decrease its costs [2,5,6], accelerate knowledge discovery [7-11], and improve access to knowledge within developing countries [4,12-17].

Collaborative writing applications (CWAs) [18,19] are a category of social media that has surged in popularity in recent years, including within the health care sector [2,6,18,20]. CWAs consist of software that allows users to create online content that anyone who has access can edit or supplement [21]. With these contributions, CWAs can become rich multimodal communication tools enriched with hyperlinks, images, videos, and audio. For example, Internet users have turned to wikis [22,23] to produce a Wikipedia entry on the Global Plan to Stop Tuberculosis [4]; to Google Knol [24,25] to exchange research on influenza at the Public Library of Science [26]; and to Google Docs [19,27] to review the literature on emergency medicine [28,29]. Although now defunct, Google Knol was a Google project that aimed to include user-written articles on a range of topics that could be edited only if the original authors gave access to editing the text. CWAs can also be classified based on who has access. There are open or public CWAs such as Wikipedia, which can be edited by anyone in the world and can also be seen by anyone. There are also partially public CWAs, which can be seen by anyone, but can be edited only by certain members of a restricted community (eg, Ganfyd [30]). There are also closed or private CWAs, part of central knowledge management systems (eg, Intelink [31]) or online learning systems (eg, Blackboard [32]), which are edited by members of the institution and are visible only to members of the institution.

Among the types of CWAs, wikis and its most famous representative—Wikipedia—are perhaps the most popular. Wikipedia is an online encyclopedia whose medical articles are viewed about 150 million times per month and exist in 271 languages [4]. Moreover, readership of Wikipedia's medical content is continuing to increase [33]. New wikis have appeared in all fields of health care [18,28,34-41], and studies of developed countries report 70% of junior physicians using Wikipedia weekly [42]. Patients use wikis to share their experiences [43] and to find information [4]. The Canadian Agency for Drugs and Technologies in Health is exploring the use of wikis to update knowledge syntheses [44-46]; the United States' National Institutes of Health is training its scientists in editing them [47,48]; and the World Health Organization is using a wiki format to update the International Classification of Diseases [49]. In addition, academic institutions have started using wikis to train health professionals [18,22,32,50-54]. Wikis have come to exemplify social media's tremendous promise to enable health professionals, patients, and policy makers to implement evidence-based practice at remarkably low cost [5,28,29,55,56]. In doing so, they could contribute to improving the health of millions of people around the world [4,13].

However, questions remain about the safety [57-59], reliability [60-64], lack of traditional authorship [65,66], and the legal implications for decision making [67,68] regarding the use of CWAs in health care. Researchers question clinicians' intentions to use the applications in their practice [28] and to contribute knowledge collaboratively [4,29,69]. Furthermore, it is unknown how CWAs can enhance the delivery of health care (eg, by empowering patients in decision making [70,71], by improving health care communication and education [18,27,32,72,73,74,75]), and benefiting health in developing countries [4,76]. While researchers have conducted systematic reviews on Internet and communication technologies (ICTs) [77,78] social media in health care [79-84] and research on Wikipedia in general [85], none have specifically focused on wikis and CWAs in health care. Not all social media share the same mechanisms of action [21], therefore examining CWAs in health care is important. The overarching goal of this project was to explore the depth and breadth of evidence about the effective, safe, and ethical use of wikis and CWAs in health care. We conducted a scoping review with the following specific objectives: (1) to map the literature on the use of wikis and other CWAs in health care, (2) to compare the applications' features by investigating how they were used in collaborative writing projects, (3) to synthesize the applications' positive and negative effects as knowledge translation interventions in health care, (4) to inventory the barriers and facilitators that affect how they influence health care delivery, and (5) to produce a research agenda delimiting areas where further knowledge synthesis is needed and where more primary research remains to be done.

Methods

Overview

A detailed description of our peer-reviewed research protocol and conceptual framework can be found elsewhere [86]. This review was planned, conducted, and reported in adherence to standards of quality for scoping reviews [87,88]. A summary of our six-stage methodology follows.

Stage 1: Identifying the Research Question

Our research question was developed by consulting a group of knowledge users to determine their needs and questions about using collaborative writing applications for knowledge translation. We defined "collaborative writing applications" as a category of social media that enables the joint and simultaneous editing of a webpage or an online document by many end users (eg, wikis, Wikipedia, Google Knol, Google Docs, Google Sites) [21]. The participants targeted by this scoping review were health care stakeholders.

Stage 2: Identifying Studies and Grey Literature

Seven scientific databases (Cochrane Library, PubMed, EMBASE, CINAHL, PsycINFO, ERIC, ProQuest Dissertations and Theses) were searched systematically for the period covering January 1, 2001 (Wikipedia's inaugural year), to September 16, 2011. Our search strategy was peer-reviewed using the PRESS criteria [89]. The following keywords were used and adapted to each database: "wiki", "wikis", "Web 2.0", "social media", "Google Knol", "Google Docs", and "collaborative writing applications" (see Table 1).

We did not exclude any citations based on language. In addition, study reference lists; the 2010 and 2011 editions of the Medicine 2.0, WikiSym, and American Medical Informatics Association conference proceedings; clinicaltrials.gov and Open Medicine's websites; expert consultation (eg, the authors of WikiProject Medicine [4]), OpenSIGLE (before 2005), and the Health Technology Assessment international Vortal were searched. Furthermore, environmental scans of the grey literature indexed by Google, Bing, Yahoo, and Mednar were performed. Finally, via email, Twitter, Mendeley, Google Docs, and a health librarianship page (HLWIKI), we called for the crowdsourcing of studies that could potentially fall within the scope of this review.

Table 1. Full search strategy for each database.

Pubmed	Wiki*[All Fields] OR “Web 2.0”[TIAB] OR “Web2.0”[TIAB] OR (google* AND knol) OR (google* AND docs) OR “Social media” [TIAB] OR (Collaborative [tiab] AND writing [tiab]) OR (collaborative technolog*) OR (collaborative software*)
Embase	wiki* OR “collaborative technology” OR “collaborative technologies” OR “collaborative writing” OR “collaborative writings” OR “collaborative software” OR “collaborative softwares” OR “google docs” OR “google knol” OR “ehealth 2.0” OR “health 2.0” OR “e+health 2.0” OR “Web 2.0”
CINAHL	TI (wiki* or “google docs” or “google knol” or “medecine 2.0.” or “Web 2.0” or “collaborative technolog*” or “collaborative writing” or “ehealth” or “e-health” or emedicine or “e-medicine”) OR AB (wiki* or “google docs” or “google knol” or “medecine 2.0.” or “Web 2.0” or “collaborative technolog*” or “collaborative writing” or “ehealth” or “e-health” or emedicine or “e-medicine”)
PsychINFO	(wiki* or “google docs” or “google knol” or “collaborative software” or “collaborative writing” or “collaborative technologies” or “collaborative technology”):Any Field OR (“medicine 2.0” or “emedicine” or e-medicine or “health 2.0” or “ehealth” or e-health or “Web 2.0”):Title OR (“medicine 2.0” or “emedicine” or e-medicine or “health 2.0” or “ehealth” or e-health or “Web 2.0”):Abstract
ERIC	((Keywords:wiki* or Keywords: “Web 2.0” or Keywords: “google docs” or Keywords: “google knol” or Keywords: “collaborative technologies” or Keywords: “collaborative technology” or Keywords: “collaborative software” or Keywords: “collaborative writing” or Keywords: “e-health” or Keywords: ehealth) or (Title: wiki* or Title: “Web 2.0” or Title: “google docs” or Title: “google knol” or Title: “collaborative technologies” or Title: “collaborative technology” or Title: “collaborative software” or Title: “collaborative writing” or Title: “e-health” or Title: ehealth) and (Thesaurus Descriptors: “Health services”))
Dissertation abstract & Thesis	Citation & Abstract (wiki* or “health 2.0” or “Web 2.0” or “e-medicine” or emedicine or “google docs” or “google knol” or “collaborative technologies” or “collaborative technology” or “collaborative writing” or “collaborative software”)
Cochrane Library (n=56)	(wiki* or “Web 2.0” or ehealth or “e-health” or “google docs” or “google knol” or “collaborative writing”) in Title, Abstract or Keywords in All Cochrane Library
Google, Bing, and Yahoo (n=1200 in total)	“wiki in health care”; “Google Knol in health care”; “Google Docs in health care”; “collaborative writing applications in health care”

Stage 3: Selecting Studies

Three teams of 2 reviewers (SR/MF, TB/AB, PA/CK) independently screened titles, abstracts, and grey literature and retained articles that presented empirical data about any CWA applied to the field of health care. In case of disagreements, a third reviewer was consulted (PA, TB, or SR). To reach a high level of agreement, we conducted 4 series of assignments (400 abstracts in total) whereby the screening of a number of studies was followed by a teleconference to reach agreement about which studies to include and to discuss uncertainties. Once consensus was reached for all cases, the remaining studies were coded by the same 3 pairs of screeners (SR/MF, TB/AB, PA/CK). Subsequently, 2 reviewers (TB and PA) conducted another round of screening based on full text studies. As a result, a narrowed definition of health care was applied in order to focus the analysis. Hence, studies that concerned the care of patients were included, and those from the fields of basic medical sciences, the conduct of clinical trials, biomedical library science and medical informatics were excluded.

Stage 4: Charting the Data

A data-charting form was developed and built into EPPI-Reviewer for the extraction of quantitative and qualitative variables and to facilitate data coding. It was tested and refined by 4 reviewers (PA, CN, ME, CF) using the first 50 studies. Three pairs of 2 reviewers (CN/CF, CN/ME, ME/CF) then independently extracted data from the remaining studies. Disagreements were resolved through discussion with a third reviewer (PA or TB). Using EPPI-Reviewer’s inductive coding function, we extracted all the pre-planned variables described in our published protocol [86].

Stage 5: Collating, Summarizing, and Reporting Results

Themes Overview

We summarized the included studies in a table comparing each of the study’s characteristics. Attempting to present an organized description of the current literature on the use of CWAs in health care, we grouped studies based on purpose. Three emergent themes were the use patterns of CWAs (Theme 1), quality of information found in different CWAs (Theme 2), and CWAs used as knowledge translation interventions (Theme 3). We also added a description of each of the applications’ features (the type of CWA and software used) to examine CWA use among studies (Objective 2).

To compare the different CWA applications identified, a Venn diagram was constructed to situate each application in relation to the others depending on two features: their collaborative writing features and their conversational features. To create the most reliable representation of how different CWAs could be represented in relation to each other, each CWA was assessed by 2 reviewers using a scoring system we created based on a classification proposed by Kaplan et al [21]. We attributed a score of 1-5 to characterize the extent of their collaborative writing features and a score of 1-5 to measure the extent of their conversational features. To design our Venn diagram, we plotted each different CWA on a graph presenting the conversational features score on the x axis and the collaborative writing score on the y axis.

Theme 1: Use Patterns of CWAs

Studies whose purpose was to describe the users and the frequency of CWA use were grouped together. We compared each study in a table presenting the population surveyed, the response rate of the population surveyed, the reported results, the prevalence of use, the contribution rate, the time of assessment, and the purpose of CWA use. We also used Eysenbach's Medicine 2.0 map [2] to illustrate the extent to which the different CWAs described in the included studies involve three major stakeholder groups (consumers/patients, professionals, and researchers).

Theme 2: Quality of Information in Different CWAs

We synthesized papers that evaluated the quality of information in CWAs by constructing a table presenting a summary of each evaluation. Three reviewers (PA, TB, SG) assigned a score on a three-point scale based on the original authors' own recommendations about future use of information contained in the different CWAs. When authors concluded that the information contained within the collaborative writing project was of high quality and that it could be used in medical decision making, we gave the paper a score of 1. When the authors concluded that the information reported was not reliable and should never be used in decision making, a score of 3 was attributed. When authors were uncertain and/or suggested that more research was needed, a score of 2 was given. This score was attributed after discussion between the three reviewers until consensus was achieved.

Theme 3: CWAs Used as Knowledge Translation Interventions

Positive/Negative Effects

Three reviewers (PA, TB, SG) performed a mixed inductive and deductive thematic analysis of the content coded in Stage 4 to classify and interpret the perceived positive and negative effects related to the use of a CWA. They began by developing a coding scheme using qualitative content analysis, a method whereby reviewers interpreted the data subjectively by classifying and coding data and identifying patterns [90]. Then, they read the data charted in Stage 4 repeatedly to immerse themselves and obtain a broad perspective [91]. Subsequently, using constant comparison methodology [90], they read the coded content by each reviewer in Stage 4, highlighting words that captured the positive or negative effects. A matrix was created to present any positive or negative effect reported in each study. We then assigned these effects specific codes, organized them into broad categories, and developed a tree diagram to organize the categories into a hierarchical structure [92]. We consolidated codes and categories that expressed the same idea into a comprehensive coding scheme that constituted our taxonomy and guided reviewers' content analysis of the rest of the data. The three reviewers discussed units of text that could not be coded with existing codes and created new codes if necessary.

The Donabedian framework [93] for quality improvement informed the classification of positive and negative effects into processes and outcomes. Elements from the Theoretical Domains Framework [94] were drawn from to classify effects of CWAs on behavior. In order to produce a comprehensive taxonomy for all described positive and negative effects of CWAs in the health care field, we added new items to our taxonomy whenever any unique item was found in a paper. Whenever these items came from a specific theoretical framework, we noted the name of the framework and attempted to label the item using the same terminology as the original source framework.

Barriers/Facilitators

A second thematic content analysis was performed on the data regarding barriers and facilitators to the use of CWAs in health care with the initial coding scheme reflecting an existing framework concerning the determinants of ICT adoption [78]. Many new determinants of social media were inductively added to this framework. Our 3 reviewers created new codes for units of text that could not otherwise be coded using the original framework, thus refining and expanding the list. We also systematically searched each article to determine if a theoretical framework was used to report barriers and facilitators. If so, relevant elements were also added to the existing framework.

Stage 6: Consulting Knowledge Users

As specified in our published protocol [86], we held meetings with representatives from the organizations involved (ie, the Association of Faculties of Medicine of Canada (AFMC), the International Medical Informatics Association (IMIA), the Federation of Patients and Consumer Organization in the Netherlands (NPCF), and the Pan American Health Organization (PAHO)) at the beginning, midway, and draft manuscript stages of this research in order to generate results that were useful for these knowledge users. Knowledge users were selected to represent a broad range of potential stakeholders representing medical education (AFMC), public health (IMIA and PAHO), and patient representatives (NPCF).

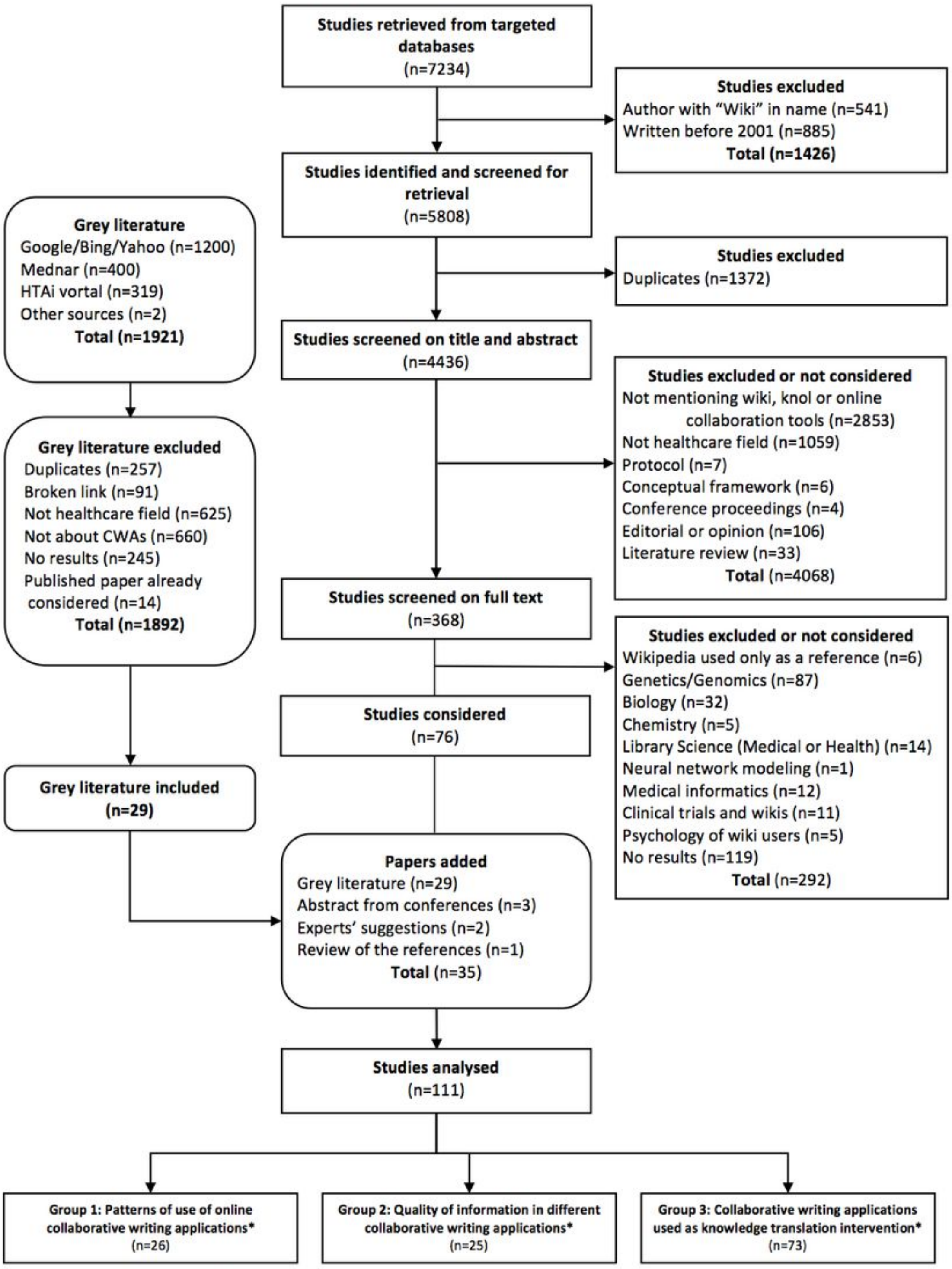
Results

Stages 1, 2, and 3: Mapping of the Literature and Study Selection

After removing duplicates (n=1372), we screened the title and abstract of 4436 citations as well as the studies/abstracts from the grey literature, conference proceedings, expert consultation, and reviewing of reference lists (Figure 1) . All disagreements (n=794) were resolved through discussion.

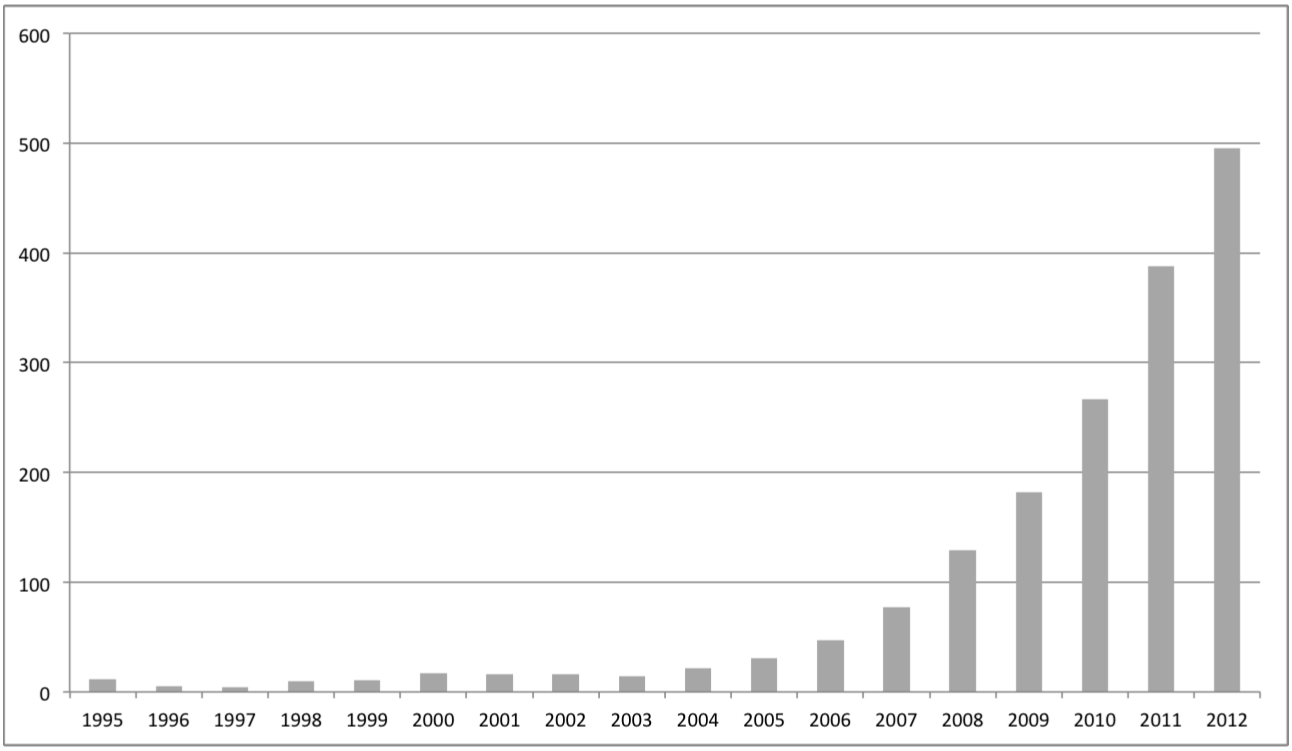
Crowdsourcing identified two studies through Google Docs that were excluded. After review, we included 111 citations. Among these 111 citations, there were 28 abstracts without published full text but with sufficient results to be included. Twenty-six studies were grouped into Theme 1 (use patterns of CWAs), 25 into Theme 2 (quality of information in different CWAs), and 73 into Theme 3 (use of CWAs as a knowledge translation intervention). Figure 2 shows the rapid growth of the number of publications for the period within our search strategy.

Figure 1. Flowchart of our mapping process and study selection.



*10 papers are classified in two groups, and 1 is classified in all three groups

Figure 2. Histogram of the number of publications related to our search strategy per year.



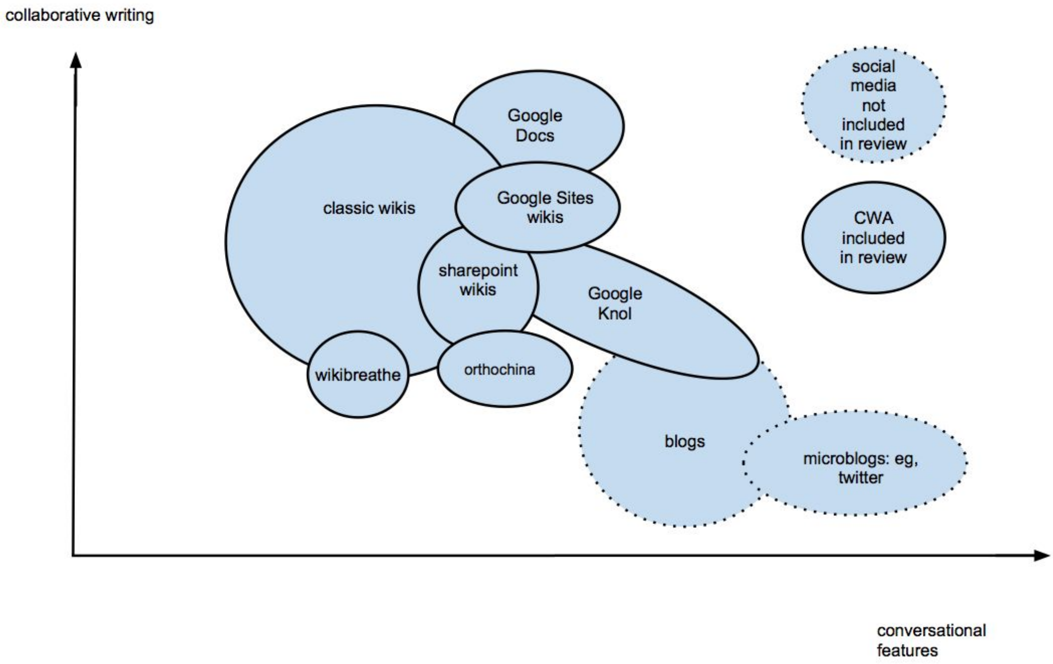
Stages 4 and 5: Charting Data, Collating, Summarizing, and Reporting Results

Study Characteristics

We found 4 experimental studies, 5 quasi-experimental, 5 observational analytic, 52 case studies, 22 describing the quality of wikis, and 23 surveys on wiki use (Multimedia Appendix 1; [27,29-32,38,42,53,54,58,61,63,72,74,76,95-262]). Wikis (n=106) and Google Docs (n=6) are the main types of CWAs used in health care. One grey literature report compared Google Knol to Wikipedia [96]. Wikipedia was the focus of a large number of studies (n=36). The most frequently used wiki software were MediaWiki (n=44), PBworks (n=8), Wikispaces (n=6), Wetpaint (n=6), Microsoft SharePoint (n=3), and Google Sites (n=3). One paper described two wikis using Semantic MediaWiki (WikiEcho [97] and WikiDoc [98,99]). There were studies describing custom-built hybrid wikis (Wikibreathe (n=2) [100,101], Orthochina (n=1) [102], and FreyaWIKI (n=1) [103]; the use of virtual learning environments (eg, Blackboard) to host wikis as aids for supporting educational activities (n=8); and the use of more sophisticated social media platforms (eg, Drupal [104], MijnZorgNet [105], Atlassian [76], and MinJournal [106]) that offer wikis and other social media such as blogs and social networking services. The importance of the collaborative writing features compared to conversational features for each of the CWA studied are presented in a Venn diagram (Figure 3). This diagram shows that wikis and other hybrid wikis are centered more on their collaborative writing features compared to Google Knol, whose conversational features stand out more. Google Docs is different in that it offers both collaborative writing features (eg, real-time online editing) and conversational features (eg, linking documents to authors' email allowing them to discuss a document while it is being created).

Two of the six studies pertaining to Google Docs were experimental [27,107]. The two other experimental studies were conducted with wikis [108,109]. As seen in Multimedia Appendix 1, the types of reported outcomes varied greatly depending on the context, goal, and framework used. Most outcomes concerned intermediate self-reported outcomes (eg, self-efficacy, usability scores, user satisfaction, dialogical communication scores), and some observed process outcomes (eg, wiki usage and contribution statistics, pre/post-test knowledge scores, quality of information, readability scores, number of communications). One study measured patient-oriented outcomes, such as blood pressure, physical activity, and cholesterol levels [107].

Figure 3. Collaborative writing applications Venn diagram.

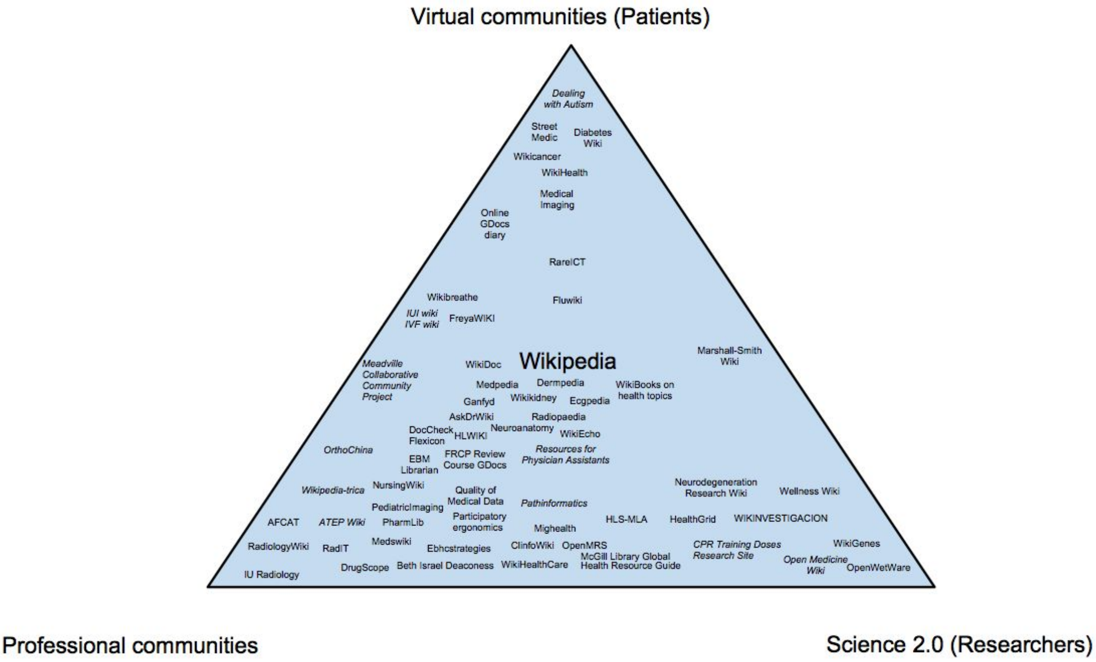


Use Patterns of CWAs

We found a total of 26 studies that presented different patterns of CWA use in health care: who uses the different CWAs, how much, and for what reasons (Multimedia Appendix 2; [29,42,53,110-130,189,263]). Most of these studies were conducted in the United States, United Kingdom, or Australia, and 1 and 3 studies were performed in Spain and in Canada respectively. All studies were published after 2006. Study populations varied widely including health care professionals (n=12), students (n=9), consumers (n=4), teachers/educators (n=2), scholars (n=1), and librarians (n=1). Most recurrent reasons for use were for academic purposes (case-based learning, e-learning, use of Web 2.0 tools for teaching) [110-115,264], for clinical purposes (to support patient care, to obtain drug information, to stay updated) [53,111,116-118], for personal use (by health care professionals and students) [42,118-121] and for seeking health information [122-127] or about specific diseases [128,129]. Other reasons were to update a scoping review [130] and to seek multiple stakeholder input [100,105]. Figure 4 shows that most CWAs described involve peer-to-peer communication between health professionals, followed by CWAs used by patients and researchers respectively.

In general, CWA use varied depending on the training level (eg, 70% or 132/188 first-year medical students using Wikipedia vs 37% or 86/234 third-year medical students [124]), the field of practice (eg, 9% or 4/44 pediatric neurologists used wikis [120] vs 35% or 369/1056 pharmacists [116]), and reason for use (eg, 100% or 51/51 radiology residents using a radiology department wiki [53] vs 15% or 360/2400 first-year psychology students using Wikipedia for personal information needs [121]). We found that a high prevalence of CWA use (ie, more than 50%) was reported in 58% (7/12) of surveys conducted with health care professionals and students (see Multimedia Appendix 2). The only longitudinal study conducted between 2005 and 2009 observed an increase in prevalence of Wikipedia use from 2% to 16% among undergraduate medical and biomedical students [123]. Another study reported higher use among younger medical students (480/593, 81%) compared to older consultants (215/389, 55%) [114]. Studies on the use of Wikipedia by pharmacists report rates of use ranging between 35% using this site for work-related questions in 2009 [116] to 72% using it mainly for personal reasons in 2011 [119]. For consumers, Wikipedia was ranked first when using search engines to find information about rare diseases [125] and to find information on generic drugs [126]. Wikipedia ranked as the second most consulted website both by a group of patients with Crohn's disease [128] as well as by students searching for biomedical information [124]. While CWA rates of use are high, most reports present low rates of contributions to CWAs. From 6%-18% of students contribute to CWAs [114,115,121] while 3%-22% of junior physicians were reported to contribute to a CWA [42,264]. Furthermore, less than 1% of scholars were reported to contribute to a wiki project aiming at updating a scoping review [130]. Rarely, high rates of contribution were found in specific wiki projects [53,100].

Figure 4. Medicine 2.0 map of the different collaborative writing applications (CWAs) and their users described in the included studies.



Quality of the Information in Collaborative Writing Applications

We found 25 papers reporting on the quality of information in CWAs (Multimedia Appendix 3; [54,58,61,63,96,99,104,121,122,124,131-137,182,183,190,195-199]). With the exception of one paper evaluating the quality of information in 52 medical wikis other than Wikipedia [99], all studies focused on evaluating the quality of medical information in Wikipedia (n=24). No studies evaluated the quality of information within projects using Google Docs; however, one did compare the quality of information within Wikipedia and Google Knol [96]. Most studies (64%, 16/25) evaluated information destined to consumers while 32% (8/25) addressed the quality of information for students. Overall, 44% (11/25) of authors concluded that information within wikis and Wikipedia is partially reliable (ie, quality of information needs to be improved or updated) while 28% (7/25) reported that information within wikis and Wikipedia is not reliable and should not be used. Three studies reported no formal conclusion about quality of information [96,121,131]. Three authors concluded that medical information in wikis and Wikipedia was reliable and of high quality [54,104,132], yet only three used a validated quality assessment instrument [99,104,133]. Of the latter, one concluded that expert-moderated wikis could produce higher quality of information [99]. For example, wikis like WikiDoc [98], ECGpedia [234] and WikiKidney [230] were among the top-rated wikis in this study [99]. However, this study also concluded that all the wikis evaluated still needed improvements mainly concerning their completeness before they could safely be used for decision making. Another study concluded that Wikipedia was adequate for clinician and student education [104] while the third study concluded that further improvement of orthognatic surgery information was needed in Wikipedia before referring consumers to the site to support decision making [133]. A recurrent finding about Wikipedia was that its content is accurate, but that it often omits important medical facts and information [58,61].

As an educational tool, Wikipedia was reported to be comprehensive, of high quality, current, and appropriate for learning in gastroenterology and pathology [54,134]. However, variability in the content, accuracy, completeness, and referencing of drug information was reported [135]. Moreover, one study reported that 171 out of 271 (63%) of students do not verify the validity of references in Wikipedia articles [112]. While some think that Wikipedia should not be used by students as a source for referencing [135] or that it is unsuitable as a base for learning [63], others believe that its use by students need not necessarily be discouraged [136] and that it could be an informative and accurate source for education if used in combination with other learning materials [137]. Furthermore, one author considered CWAs to be excellent sources for continuing education and that they could represent the future of medical education as they allow for self-directed and supplementary education as well as for remote access [104].

Online Collaborative Writing Applications as Interventions

We identified four experimental studies in support of CWA use as educational and knowledge translation interventions (Multimedia Appendix 4; [27,107-109]). Three of these studies were conducted in the field of health professions education [27,108,109], and one was in the field of secondary prevention of cardiovascular disease in patients with previous acute coronary syndrome [107]. These studies found that the use of CWAs improved (1) physical activity and blood pressure control, (2) scientific writing skills among health science students, (3) medical student self-confidence and communication skills, and (4) nursing leadership skills. One study found that CWA use worsened diagnostic skills [108].

Taxonomy for Perceived Positive and Negative Effects Associated With CWAs

We classified the perceived positive and negative effects associated with CWAs into a taxonomy, covering eight categories (Table 2; [2,27,32,53,72,76,94,100,102,103,105,107-110,122,130,138-178,265,266]).

In total, 57 positive effects and 23 negative effects were identified. Among the categories of positive effects that we found, the most frequently reported were that CWAs improve collaboration (n=41), positively impact learning (n=30), influence psychological domains (n=28), facilitate knowledge management and accessibility to information (n=30), improve efficiency of health care (n=19), improve quality of health care (n=6), and prevent disease (n=3). Among these effects, the Theoretical Domains Framework [94] was used to label and classify 22 of them into 3 psychological domains (self-efficacy, motivation, emotion) and 2 learning effects (skills and knowledge).

We found 2 studies referring to theoretical frameworks to describe their effects. Among the frameworks, the concept of communities of practice [266] was used to classify 3 studies reporting that CWAs improved the communication of tacit knowledge. The Dialogic Theory of Public Relations [265] was used to describe 5 positive effects wikis could have on public relations between health care organizations and consumers.

The most frequently cited negative effects were that CWAs could have unfavorable impacts on knowledge management (n=14) such as information overload (n=4) and fast dissemination of poorly validated information (n=4), as well as on certain psychological domains (n=6) such as added stress (n=1) and negative emotions (n=5). Some authors stated that CWAs could impede certain aspects of collaborative work (n=4) such as enhancing the perception of unequal work distribution (n=2) and encouraging conversation more than collaborative writing (n=1). Potentially serious negative effects of deletion of important medication information on Wikipedia by pharmaceutical companies (n=1) [177] and breaching of patient confidentiality (n=1) [179] were reported only in the grey literature.

Taxonomy for Barriers and Facilitators to the Use of CWAs in Health Care

A total of 48 barriers and 91 facilitators to the use of CWAs in health care were identified, of which 20 barriers and 69 facilitators were new determinants (Table 3; [32,53,54,76,100-102,106,109,110,114,116,130,141-143,145-149,153-156,159,162-164,166-174,176,178,180,181,267-271]).

Among the latter, some were specific to social media (eg, social aspects of ICT, presence of a moderator, presence of a community of practice) and others were not (eg, information overload, mobile access, lack of proficiency in English). Although we found only 5 studies [101,109,153,155,156] that used a theoretical framework to identify barriers and facilitators, many of these barriers (n=11) and facilitators (n=34) were among those deemed as new.

The five barriers most frequently mentioned, in order of frequency, were unfamiliarity with ICTs (n=8), time constraints and workload (n=6), lack of self-efficacy (belief in one’s competence to use ICT) (n=6), material resources—access to ICT (n=5), worries about the scientific quality of the information (n=5), and the presence of a closed wiki protected by a password (n=5). The five most recurrent facilitators were having had training (n=12), scientific quality of the information (n=10), ease of use (n=8), triability (n=7), presence of a community of practice or a community of learners (n=7), and presence of a moderator (n=7).

Table 2. Positive and negative impacts of collaborative writing applications.

Impacts	Number of papers in which the impacts perceived as positive	Number of papers in which the impacts perceived as negative
Processes (intermediate outcomes) [a]		
1. Effects on psychological domains [b]	28	6
1.1 Beliefs about capabilities (Self-efficacy) [b]		
1.1.1 Self-Efficacy/empowerment: Not further specified	10 [32,108,163,122,138-143]	
1.1.2 Empowering environment	2 [109,139]	
1.1.3 Empowerment of families/relatives	1 [144]	
1.1.4 Patient participation	3 [103,105,110]	
1.2 Motivation [b]		
1.2.1 Engagement	7 [100,145-150]	
1.3 Emotion [b]		
1.3.1 Satisfaction	5 [27,141,145,151,152]	1 [150]
1.3.2 Loss of autonomy/feeling of being monitored		1 [32]
1.3.3 Feeling of working in isolation		1 [153]
1.3.4 Feeling of guilt about not participating		1 [109]
1.3.5 Frustration due to technical issues		1 [154]
1.3.6 Added stress		1 [155]
2. Learning effects	30	1
2.1 Subjective learning improvements: Not further specified	9 [108,140,141,114,145,150,152,156,157]	
2.2 Skills [b]		
2.2.1 Communication skills eg, feedback	2 [138,151]	
2.2.2 Handle fears and feelings	1 [158]	
2.2.3 Adapt to different learning styles	4 [72,109,141,142]	
2.2.4 Information and communication technology skills	1 [154]	
2.2.5 Transfer of knowledge into practice	1 [138]	
2.2.6 More efficient critiquing and evaluating the medical literature	1 [138]	
2.2.7 Development of professionalism on students	1 [32]	
2.2.8 Enhanced understanding of concepts	1 [159]	
2.2.9 Decreased learning of diagnostic skills		1 [108]
2.3 Knowledge [b]		
2.3.1 Knowledge (not further specified)	4 [72,109,154,160]	
2.3.2 Awareness of guidelines	1 [161]	
2.4 Better supervision by teachers	2 [141,154]	
2.5 Better exam preparation	2 [108,110]	
3. Communication	24	2

Impacts	Number of papers in which the impacts perceived as positive	Number of papers in which the impacts perceived as negative
3.1 Communication: Not further specified (impedes/improves)	9 [27,32,76,108,148,153,162-164]	2 [109,141]
3.2 Feedback	2 [151,165]	
3.3 Collegiality	1 [159]	
3.4 Patient/health professionals communication	2 [144,146]	
3.5 Communication of tacit knowledge[b]	3 [76,163,164]	
3.6 Creates a network for families	1 [144]	
3.7 Apomediation (communication process whereby individuals “stand by” to guide consumers to high quality information without being a prerequisite to obtain that information in the first place)[b]	1 [164]	
3.8 Dialogical communication between organizations and individuals [b]		
3.8.1 Mutuality (the recognition of organization–public relationships)[b]	1 [122]	
3.8.2 Propinquity (the temporality and spontaneity of interactions with publics)[b]	1 [122]	
3.8.3 Empathy (the supportiveness and confirmation of public goals and interests)[b]	1 [122]	
3.8.4 Risk (the willingness to interact with individuals and publics on their own terms)[b]	1 [122]	
3.8.5 Commitment (the extent to which an organization gives itself over to dialogue, interpretation, and understanding in its interactions with publics)[b]	1 [122]	
4. Collaboration	41	4
4.1 Collaboration: Not further specified (impedes/improves)	23 [72,76,100,102,110,138-143,145-148,151,154,161,162,166-169]	1 [141]
4.2 Reduces geographical barriers	11 [76,100,138,144,153,154,160,162,163,166,170]	
4.3 Perceived unequal/equal separation of work	3 [100,110,141]	2 [141,154]
4.4 Asynchronous communication	1 [163]	
4.5 Wiki used as a conversational manner without contributing to the same text		1 [141]
4.6 Define team responsibilities	1 [156]	
4.7 Interprofessional collaboration	1 [105]	
4.8 Creation of online presence	1 [156]	
5. Knowledge management and accessibility to information	30	14
5.1 Dissemination of information	8 [110,163,164,167,169,171-173]	
5.2 Fast dissemination of poorly validated information		4 [102,159,164,174]
5.3 Better access to information	8 [138,140,152,163,169,171,175,176]	
5.4 Better exposure to world	1 [168]	
5.5 Better knowledge translation across organizations	2 [146,164]	
5.6 Centralized knowledge management	5 [140,152,156,164,166]	1 [110]

Impacts	Number of papers in which the impacts perceived as positive	Number of papers in which the impacts perceived as negative
5.7 Constantly updated information	1 [169]	
5.8 Facilitates management of various content	1 [172]	
5.9 Privacy issues health related data		1 [146]
5.10 Spam/vandalism		2 [130,177]
5.11 Updating of knowledge synthesis		1 [130]
5.12 Saves paper	1 [175]	
5.13 Information overload		4 [109,164,175,176]
5.14 Wiki allows daily surveillance (looking for spurious edits)	1 [53]	
5.15 Compiling anonymous data	1 [144]	
5.16 Creativity/new ideas	1 [110]	
5.17 Editing wars		1 [167]
Outcomes		
6. Efficiency of health care	19	4
6.1 Efficiency: Not further specified	5 [72,110,146,151,166]	2 [141,164]
6.2 Saves money	1 [166]	
6.3 Saves time/loses time	11 [32,102,146,148,152,155,161,163,166,169,170]	1 [162]
6.4 Decreases/increases duplicate work	1 [164]	1 [155]
6.5 Reduces workload	1 [174]	
7. Quality improvements	6	2
7.1 Quality improvements: Not further specified	5 [27,144,146,151,166]	1 [164]
7.2 Wiki content didn’t meet users’ needs		1 [178]
7.3 Reduces errors	1 [155]	
8. Disease prevention	3 [107,142,146]	

[a] The Donabedian framework [93] for quality improvement was used to describe processes and outcomes.

[b] These items are processes that were taken from other psychological and organizational frameworks for change and used to describe and classify the effects of CWAs found in this review [2,94,265,266].

Table 3. Barriers and facilitators related to the use of collaborative writing applications.

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
1. Factors related to ICT (CWA)		
1.1 Design and technical concerns	13	8
1.1.1 Readability of the information[a]		1 [171]
1.1.2 Appearance of wiki (font, etc.)[b]	1 [101]	1 [159]
1.1.3 Organization of information[b]	5 [101,163,169,171,180]	
1.1.4 Immediately available technical information[a]	1 [166]	
1.1.5 Having a sense of continuity and stability[b][267]	1 [109]	
1.1.6 References not intrusive in lay language texts[a]	1 [167]	
1.1.7 Information overload[a]		2 [109,170]
1.1.8 Mobile access[b]	1 [155]	
1.1.9 Spam filter[a]	1 [130]	
1.1.10 System can improve[a]	1 [154]	
1.1.11 Rapid information changes[b]	1 [155]	1 [130]
1.1.12 Design and technical concern – other		3 [109,142,154]
1.2 Characteristics of the innovation	33	5
1.2.1 Ease of use/complexity		
1.2.1.1 Ease of content editing[a]	6 [106,163,166,170,176,180]	
1.2.1.2 Human/computer interactions [b]		
1.2.1.2.1 Consistency (principle of minimum amazement)[b][268]	1 [109]	
1.2.1.2.2 Prevent error messages[b][268]	1 [109]	
1.2.1.2.3 Temporal contiguity (easy mental associations are made between verbal and visual)[b]	1 [109]	
1.2.1.3 Reduce short-term memory load[b][268]	1 [109]	
1.2.1.4 Ease of use/complexity – other	8 [100,109,110,141,146,147,164,166],	4 [109,141,153,172]
1.2.2 Triability		
1.2.2.1 Permit Easy Reversal of Actions[b][268]	3 [106,109,169]	
1.2.2.2 Triability – other	7 [32,102,109,153,154,156,172]	
1.2.3 Relative advantage (usefulness) or lack of		1 [130]
1.3 System reliability	2 [109,169]	
1.4 Interoperability (including Web browser interoperability)	3 [53,146,169]	2 [154,178]
1.5 Legal issues	2	6
1.5.1 Confidentiality - privacy concerns	2 [153,163]	3 [32,109,170]
1.5.2 Liability[a]		1 [172]
1.5.3 Copyright concerns[a]		2 [170,172]
1.6 Validity of the resources	16	9
1.6.1 Scientific quality of the information resources	10 [32,102,142,153,155,159,163,169,170,174]	5 [114,130,171,172,176]

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
1.6.2 Content available (completeness)	2 [169,174]	2 [54,178]
1.6.3 Appropriate for the users (relevance)	2 [53,176]	1 [178]
1.6.4 Content updated frequently[a]		1 [54]
1.6.5 Highly prevalent disease[a]	1 [130]	
1.6.6 Rapidly growing body of research[a]	1 [130]	
1.7 Cost issues: low human and hardware costs	3 [53,146,169]	2 [146,166]
1.8 Social aspects of ICT [a]	28	7
1.8.1 Integrated support tools within wiki (toolbox, FAQ, forum, policies)[b]	6 [149,153,163,164,167,169]	
1.8.2 Open access wiki[b]	1 [53]	5 [109,155,163,169,173]
1.8.3 Good balance between restricted areas within wiki (private info) vs open areas (info for all)[a]	2 [106,130]	
1.8.4 Interface linking content to conversations[b]	2 [109,180]	
1.8.5 Use of template and seed with core set of pages[a]	4 [163,164,167,169]	
1.8.6 Webmetric tool integrated with ICT to measure use (eg, Google Analytics) and contributions/authorship (eg, Wikigenes)[a]	1 [130]	
1.8.7 Simultaneous real-time collaborative editing[a]	1 [109]	
1.8.8 Gives informative feedback[b][268]	1 [109]	
1.8.9 Authorship transparent to increase reliability[a]	3 [130,169,174]	
1.8.10 Socialization tactics (eg, welcome message)[a]	1 [130]	
1.8.11 Controversial content[a]	1 [130]	
1.8.12 Important impact on a large number of health professionals[a]	1 [130]	
1.8.13 Lack of interest in topic[a]		1 [130]
1.8.14 Wiki enabled with an RSS feed or email notifications (reminders)[b]	4 [32,109,159,163]	
1.8.15 Inappropriate automatic computer editing[a]		1 [154]
2. Individual factors or health care professionals characteristics (knowledge and attitude)		
2.1 Knowledge	1	12
2.1.1 Awareness of the existence and/or objectives of the ICT		2 [130,141]
2.1.2 Familiarity with ICT		
2.1.2.1 Skills[b][269]	1 [109]	
2.1.2.2 Familiarity with ICT – other		8 [109,114,116,130,148,153,168,181]
2.1.3 Lack of proficiency in English (the language of the Web)[a]		1 [146]
2.1.4 Lack of knowledge about systematic review methods[a]		1 [130]
2.2 Attitude	17	18

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
2.2.1 Agreement with the particular ICT		
2.2.1.1 Challenge to autonomy		1 [32]
2.2.1.2 Outcome expectancy (use of the ICT leads to desired outcome)	1 [130]	
2.2.1.3 Motivation to use the ICT (readiness)/resistance to use the ICT		4 [109,140,147,149]
2.2.1.4 Motivation to contribute to the wiki (desire to participate and post messages/information) [b] [269]	3 [109,156,174]	1 [130]
2.2.1.4.1 Motivation to contribute needs to be consistent with the person's goals, plans, values, beliefs and interests[b][269]	2 [109,156]	1 [130]
2.2.1.5 Self-efficacy (believes in one's competence to use the ICT)	6 [109,130,141,145,153,168]	6 [32,114,142,153,170,178]
2.2.1.6 Preference for private learning environment compared to open environment[a]	2 [32,162]	
2.2.1.7 Impact on personal life[b][267]	1 [109]	
2.2.1.8 Confidence in ICT developer		1 [116]
2.2.1.9 Agreement with the particular ICT – other	1 [178]	2 [156,170]
2.2.2 Agreement with ICTs in general (welcoming/resistant)	1 [174]	2 [114,168]
3. Human environment		
3.1 Factors associated with patients	3	0
3.1.1 Patient/health professionals interaction		
3.1.1.1 Sharing of information between doctors and patients[a]	1 [174]	
3.1.1.2 Sharing of information between doctors[a]	1 [174]	
3.1.1.3 Sharing of information between patients[a]	1 [174]	
3.2 Factors associated with peers	25	7
3.2.1 Support and promotion of ICT by colleagues		
3.2.1.1 Support by nurses[b]	1 [155]	
3.2.1.2 Support by physicians[b]	1 [155]	
3.2.1.3 Support by trainees[b]	1 [155]	
3.2.1.4 Support and promotion by colleagues (not further specified)	3 [109,153,171]	
3.2.2 Other factors associated with peers (relations between colleagues)		
3.2.2.1 Credential verification[a]		1 [102]
3.2.2.2 Frustration about having someone else edit personal contribution[b]		3 [106,109,141]
3.2.2.3 Reluctance to team work[b]		3 [141,154,156]
3.2.2.4 Using constructivist theoretical framework to setup a wiki is helpful[b][270]	3 [109,153,156]	

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
3.2.2.5 Presence of a community of practice/community of learners [b]		
3.2.2.5.1 Critical mass of scholars[a]	1 [130]	
3.2.2.5.2 Presence of a small group of motivated editors[a]	1 [130]	
3.2.2.5.3 Presence of community of practice/community of learners (not further specified)[b]	7 [76,106,109,149,156,169,174]	
3.2.2.6 Openness, trust and respect[b]	4 [106,109,130,163]	
3.2.2.7 Need for reciprocity (questions answered)[b]	2 [109,156]	
3.2.2.8 Create teams of two collaborators working on same wiki page[a]	1 [162]	
4. Organizational environment		
4.1 Internal environment	69	27
4.1.1 Work (nature of work)		
4.1.1.1 Time constraints and workload		
4.1.1.1.1 Ultra-rapid decision making environment[b]		1 [155]
4.1.1.1.2 Time constraints and workload – other	1 [32]	6 [109,114,141,148,162,170]
4.1.2 Resources availability		
4.1.2.1 Resources available (additional)		1 [116]
4.1.2.2 Material resources (access to ICT)		
4.1.2.2.1 Lack of constant Internet connection/access[b]		2 [146,155]
4.1.2.2.1.2 Material resources (access to ICT) – other	6 [106,109,141,153,166,180]	5 [114,146,153,154,178]
4.1.2.3 Human resources (IT support)	4 [109,154,156,171]	1 [146]
4.1.2.4 Having a single platform[a]	1 [162]	
4.1.3 Organizational factors		
4.1.3.1 Training		
4.1.3.1.1 Face-to-face training[b]	6 [32,76,141,149,153,156]	
4.1.3.1.2 Use smaller groups (n=15-20) for one on one feedback[b]	1 [109]	
4.1.3.1.3 Educators must be aware of human-computer interactions[b]	1 [109]	
4.1.3.1.4 Training medical educators in using Web 2.0 ICTs[a]	1 [114]	
4.1.3.1.5 Need for active learning/constructivist learning[b]		1 [109]
4.1.3.1.6 Training –other	12 [53,76,109,141,143,145,148,153,154,159,163,169]	1 [146]
4.1.3.2 Management (strategic plan to implementing applications)		

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
4.1.3.2.1 Start with pilot project (implementation strategy)[a]	1 [162]	
4.1.3.2.2 Index with Google - use Google Adwords (implementation strategy[a])	1 [167]	
4.1.3.2.3 Monitoring of use with Web metrics[b]	3 [130,156,167]	
4.1.3.2.4 Management – other		2 [109,141]
4.1.3.3 Presence and use of “champions”	1 [54]	
4.1.3.4 Participation of end-users in the design	1 [172]	
4.1.3.5 Communication (includes promotional activities)		
4.1.3.5.1 Work with computer science department to implement a plan to generate traffic to wiki[a]	1 [167]	
4.1.3.5.2 Getting new staff to participate for new look[a]	1 [163]	
4.1.3.5.3 Encourage writers to contribute using their own style[a]	1 [163]	
4.1.3.5.4 Forcing students to edit wiki[a]	1 [130]	
4.1.5.5.5 Participating in a community of wiki editors[a]	1 [130]	
4.1.5.5.6 Communication – other	3 [130,154,167]	
4.1.3.6 Ongoing administrative/organizational support		
4.1.3.6.1 Interactive Web applications permitted and unblocked within the health care institution[b]	1 [109]	
4.1.3.6.2 Administrative/ organizational support – other	3 [109,130,156]	1 [114]
4.1.3.7 Incentive structures		
4.1.3.7.1 Giving continuing medical education (CME) credit[a]	1 [130]	
4.1.3.7.2 New set of scholarly impact metrics[a]	1 [130]	
4.1.3.7.3 Major cultural barrier in academia against participating in social media[a]		1 [130]
4.1.3.7.4 Incentive structures – other	5 [54,102,109,162,169]	2 [130,172]
4.1.3.8 Presence of a moderator[b]	7 [53,102,109,153,156,167,172]	
4.1.3.9 Presence of metacognitive participants and dialogical participants[b][271]	2 [109,156]	
4.1.3.10 Accept that not all will participate and that lurkers will always exists/frustration about the lurkers who don’t contribute[b]	1 [109]	3 [141,149,154]
4.2 External environment	1	1
4.2.1 Financing of ICT/financial support		1 [109]

Factors (Gagnon et al 2012 taxonomy)	Number of papers in which the factor was mentioned as a facilitator	Number of papers in which the factor was mentioned as a barrier
4.2.2 Coupling traditional publications with wiki contributions[a]	1 [130]	

[a] These new determinants did not exist in the Gagnon et al framework

[b] These new determinants were identified in papers using a theoretical framework.

Discussion

Principal Findings

We confirmed that CWAs are currently being used frequently in health care, by a variety of stakeholders including patients, professionals, and researchers, for a large diversity of purposes. Our complete portrait of the literature shows that wikis are by far the most commonly studied type of CWA and that most studies had observational designs. Each type of CWA has different collaborative writing and conversational features that must be considered by decision makers when making a choice about which CWA to use in different collaborative projects. Many positive effects are attributed to the use of CWA in health professions education and knowledge translation. Further systematic synthesis of experimental and quasi-experimental evidence is needed before any clear policy recommendations can be made about implementing these tools in current practice. Moreover, there is an array of potential negative effects and barriers that need to be addressed in future primary research projects.

The Use of CWAs in Health Care

Despite the controversy surrounding the use of information in Wikipedia in clinical decision making [57,65], a high proportion of health professionals and students are already using Wikipedia and other CWAs, with use apparently increasing, especially among younger professionals. Although more research is needed to confirm this trend, these findings are consistent with an overall trend to increased use of social media among health professionals [79,272]. Our systematic mapping of the literature shows that wikis are the most frequently studied type of CWA. Furthermore, the use of Wikipedia by students and professionals represents the focus of many of our included studies. Google Docs studies come second, and we found only one study about Google Knol. This is not surprising since Wikipedia is the sixth most visited website worldwide and appears in top 10 results of search engines concerning health questions [125]. However, as readership of Wikipedia is rapidly changing, it is important to acknowledge that usage percentages depend not only on how you ask the question but also when you ask the question. Moreover, Google terminated the Knol project in 2011 despite interesting health projects using this platform including the PLOS Currents: Influenza project [273,274]. Besides the single publication we found about Google Knol comparing Knol to Wikipedia [96], there are no published accounts of Google's reasons for closing and transferring Knol to the Annotum platform.

Based on the Medicine 2.0 map [2], we demonstrated that current CWAs in use are mainly oriented towards health students and professionals' peer-to-peer interactions. In fact, use of CWAs is a major area of research in health education [275,276]. In particular, of the 4 experimental studies identified, 3 were education studies showing that CWAs positively influenced learning processes and almost half (n=48) of all the studies in this review concerned health professions education. Albeit less common, there are also studies about CWAs involving consumers and professionals to co-create decision-making tools [100,101,105,277]. These four projects seem relevant given that patient-centered care has become a central aspect of knowledge translation and experts have called for new ways of involving patients in the implementation of evidence [278]. Another remarkable finding is that even fewer CWAs involve consumers and researchers in sharing hard to find phenotype information about rare genetic and congenital diseases [106,144].

Researchers are starting to explore the use of CWAs, for example in updating a scoping review [130]. Another expert/researcher driven wiki is the OpenMRS electronic medical record implementation wiki, an example of wikis' full potential for improving health in developing countries. Although the World Health Organization is exploring the use of a wiki to update the 11th International Classification of Disease [49], we did not find any published accounts on their experience, nor did we find any related to the discontinuation of Medpedia [37]. The reasons for ending this ambitious project involving important stakeholders would provide lessons for the future.

CWAs Features and Implications for Health Care

After comparing how each CWA was used in different collaborative writing projects, we found that wikis and certain hybrid custom-built wikis have collaborative writing features that are more prominent compared to their conversational features. These collaborative writing features produce artefacts of synthesized knowledge that lend themselves more readily to daily use than those produced from conversational knowledge. For example, using a wiki to store and update care protocols readily applicable to the care of emergency department patients would be more useful in daily practice than reading the discussion page found in support of the wiki page itself. Conversely, Google Docs, certain knowledge management applications (eg, Google Sites, Microsoft Sharepoint) and other social media platforms (eg, MijnZorgNet, Atlassian Confluence, MinJournal) integrate additional features that favor conversation and deliberation between users. These additional conversational features produce discussions between users about the knowledge being shared and add to users' understanding about the content found on the collaborative writing pages of these applications.

Effects of CWA and Wiki Use in Health Care

Most evidence stemmed from case reports and observational studies demonstrating perceived positive effects of CWA use in health care on behavior change, education, communication, collaboration, knowledge management and access to knowledge, and better quality and efficiency of health care. These findings support claims that CWAs and wikis facilitate that online professional communities create, share, and synthesize knowledge; increase access to health information; and offer opportunity for public participation and citizenship [84,276,279]. Although less frequently reported, we also found a series of perceived negative effects (ie, information overload, fast dissemination of poorly validated information, loss of autonomy, feeling of working in isolation, increased stress, perceived unequal distribution of tasks within teams, biased editing, editing wars, and vandalism/wikispam) that could mask some of the positive effects of CWAs. Innovative developments such as semantic wikis [8,97,98,276,280] and bots [11,281] may decrease some of these negative effects. For example, to reduce the impression of information overload, certain authors are exploring semantic wikis to better organize and structure information based on a logical ontology [97,98]. Semantic wikis could help organize the knowledge being shared [8,276,280], potentially improve its meaningful use [282,283] and eventually allow its integration into intelligent Web-based decision-support tools [280]. Other authors are exploring the use of bots to decrease the risk of vandalism, biased editing, and spam [11,281]. A bot is a computer program that runs automatically and continuously within wikis and can conduct simple tasks like correcting spelling and syntax. Wikipedia contains many different bots that help ensure its quality [281]. More complex bots exist like the one in WikiPathways that surveys the content and identifies potential inconsistencies, redundancies, and incomplete data [11].

Barriers and Facilitators to the Use of CWAs and Wikis in Health Care

The use of CWAs in health care faces barriers that limit their use that are similar to those experienced in other fields: unfamiliarity with ICT [284], time constraints and workload [275], lack of self-efficacy to use CWAs [275], access to CWAs [285], worries about the scientific quality of the information resources [276,281,286,287], readability of information [281], the presence of a closed wiki protected by a password [276,281] and legal concerns [276,286,287].

A recurrent finding about the information in Wikipedia was that it is in large part accurate, free, and easy to access. However, even though Wikipedia does not recommend including medication doses due to concerns about errors [288], it is often incomplete and can lack appropriate referencing of medical information [58,61], thereby possibly indirectly causing patient harm [135]. One observational study demonstrated that involving moderators and experts in the sharing and curation of information within CWAs improves the quality of information [99]. However, as previous authors have demonstrated, finding ways to get these experts to participate remains a challenge [4,130,182,276,289].

Maintaining high-quality information as well as high contribution levels is a heated debate with opposing views (ie, password-protected wiki vs open wiki) [53,105,109,155,163,169,173]. Authors from multiple fields have explored modalities to stimulate participation [276,281,284,285,290-296]. Many facilitators reported from fields other than health care include training [284,296], scientific quality of the information resources [281,286,287], ease of use [291], having access to integrated support tools [296], ease of content editing [297-299], access to CWA [285], self-efficacy [300,301], and the use of incentives [293,294,302-304]. Some propose a set of scholarly metrics that would reward contributions to collaborative projects [130]. The journal *RNA Biology* stimulates contributions to Wikipedia by scholars by requiring that manuscripts be summarized for a Wikipedia page before accepting to publish the article [305]. The WikiGenes project has recognized the importance of authorship [10,36]. Finally, similar to other fields [293,294,297,306], the presence of a community and the sense of community is a frequently reported facilitator that increases contributions by health care stakeholders. Experts suggest that studying CWAs involves looking at both the technology and its community of users [276,285,307]. Thus, understanding the success of a project using a CWA must also include exploring the fundamental elements of communities of practice [266]. Communities of practice can meet online (ie, virtual community) or face to face. Similar to systematic reviews on communities of practice [308,309], our scoping review identified the presence of a moderator and/or a champion as a key factor for a successful collaborative writing project. Related to the concept of community, the success of a collaborative writing project also includes having a critical mass of participants, shared values, openness, trust, and respect.

Clinical Relevance

We believe that our findings are important for consumers, professionals, researchers, and health care organizations around the world that are already using CWAs and/or planning to use a CWA to improve health care. Although we have found some evidence from experimental studies to support the use of CWAs as a health profession’s educational intervention and a large body of observational evidence supporting the use of CWAs as a knowledge translation intervention, a formal systematic review should be conducted to further synthesize the evidence and conduct a formal risk of bias assessment before making practice recommendations. Furthermore, the implementation of CWAs is fraught with barriers and the potential for adverse effects, requiring primary research to assess their safety.

Unfortunately, the breadth and depth of the literature on the use of CWAs specific to public health is scarce. However, based on some ongoing and promising projects [49,76,99,139,146,164], it is clear that the uses of CWAs for public health are vast and far-reaching. Although more research is needed within this specific domain, CWAs improve information access, collaboration, and can improve health education—all tenets of public health. Patients and consumers often experience many barriers in the use of CWAs, with information quality being among the most reported. The readability of articles within Wikipedia is a key area that must be addressed, as it will improve health literacy and knowledge translation [310]. There are also promising projects that may shed light on the effectiveness of involving patients in the development of clinical of guidelines [311]. Evidence from experimental studies about engaging patients with CWAs is still rare and needs to be replicated in robust prospective trials before making recommendations.

Strengths of This Study

This is the first study that has conducted a scoping review to examine the depth and breadth of evidence about the use of CWAs in health care. We rigorously followed scoping review methodology and conducted a systematic and broad search of CWA use in multiple scientific databases and grey literature sources. A scoping review was the ideal methodology to employ for a number of reasons. First, it is an explorative method used when the relevant literature is considered to be broad and diverse [312]. Moreover, the study of these applications is an emerging field that is being examined with diverse methods [28,32,61], with different theoretical frameworks [29] and in different contexts [46,313]. We used a high-quality collaborative Web-based software to manage our review, to import studies, to extract data and to create reports. Every step of our review has been extensively described. By including knowledge users and policy makers, we have produced a relevant synthesis of the evidence targeting their needs. Based on empirical results, this scoping review has also extended an existing taxonomy of adoption determinants to the study of a social media application. The original taxonomy had been developed using a rigorous mixed-methods systematic review methodology [78]. Although our new extended taxonomy is very comprehensive, we believe that this level of detail was important to maintain in order to help future researchers explore the impact of these barriers and facilitators. Moreover, we have also created a new taxonomy of effects based on elements from other sociocognitive and organizational frameworks of change. Our use of the Donebadian framework was very useful because of its generalizability and overarching broad scope. Other more specific frameworks (eg, Theoretical Domains Framework) fit well within this overarching framework. Research should validate our two new taxonomies for future development, assessment, and implementation of other social media applications.

Limitations of This Study

Even though we did everything possible to minimize publication bias by systematically and extensively searching for any sources of the grey literature presenting negative results (eg, including a lay media newspaper article [177]), we believe publication bias is not excluded. For example, we have not found published reports explaining the failed attempts at maintaining Google Knol or Medpedia. Many other CWAs sites have also disappeared over the course of the years without any clear explanations. In 2009, David Rothman had listed 69 medical wikis, many of which are now inactive or simply do not exist anymore [39]. Such reports describing the reasons for CWA failure would help generate important lessons for the advance of the science of collaborative writing.

Second, our scoping review methodology [87,88] did not include formal quality assessment. However, we classified studies based on the strength of their design in order to help us identify areas for primary research and those that produced sufficiently robust evidence for making recommendations.

Third, our scoping review was limited to reviewing CWAs using a definition that excluded related applications like blogs, microblogs, discussion forums, and patient communities (eg, PatientsLikeMe). Even though these social media applications are collaborative as well and share some common features with CWAs, we believe that it is important to study them separately to better understand each application’s impact and interaction with other social media.

Finally, our search strategy is limited to studies published between January 1, 2001, and September 16, 2011, while several more recent studies about CWAs have been published [263,272,314-319].

Unanswered Questions and Future Research

This scoping review has identified a number of research gaps. There is a need to conduct systematic reviews to further synthesize the results of experimental and quasi-experimental studies in the field of health professions education and to further synthesize evidence about implementation strategies addressing the different barriers identified. Given that the majority of the literature presently exists in the form of case reports with self-reported measurements, it is essential that further prospective trials with objective outcomes be conducted. Future trials should identify implementation processes that can be influenced by CWAs and how to measure them (possibly using Web metrics [130,167,276]) as intermediate outcomes of a complex knowledge translation intervention. In this respect, in addition to other frameworks defining evaluation plans of dynamic collaborative applications [320], our taxonomies of CWA adoption determinants and effects will help plan such trials. This will help researchers understand the different mechanisms of action at play leading to improved patient-oriented outcomes (quality of life, morbidity, mortality). Although the feasibility of conducting a randomized clinical trial to study the effectiveness of CWAs seems daunting, other complex interventions have been studied using this methodology [321].

Before conducting such trials, researchers and decision makers must reflect on defining the purpose of using a CWA as a knowledge translation intervention. Researchers must also find ways to adapt CWAs to the particular needs of different stakeholder groups (consumers, professionals, and researchers). Important barriers such as the quality of information contained in different wikis must be better addressed. As previous authors have stated [183,320], measuring the quality of user-generated content and its change over time is a challenging task requiring research [322]. Finding ways of assuring the scientific integrity of evidence within CWAs and recognizing authorship are significant stumbling blocks that need to be addressed for health care [102,114,130,171,176,323]. Studying each specific behavior involved in using CWAs (ie, to use, to contribute, to edit, to delete) with the help of theoretical frameworks will also help inform future interventions.

In addition to other technical considerations [324,325], future studies should explore the impact of collaborative writing and conversational features on information sharing and investigate what kind of knowledge (explicit vs tacit [266]) is shared. This could help knowledge users choose an appropriate CWA. As future communication tools, the impact of using different types of media embedded within CWAs (audio and video recordings) should also be explored. Finally, an important consideration to explore in future studies would be to determine the impact of using a closed vs an open CWA on the quality of the information found within the CWA and on the type of barriers experienced by users.

Conclusion

The prevalence of CWA use is high in various fields of health care, and they are used for a variety of purposes. They present many potential positive and negative effects as knowledge translation tools. Although we found some experimental and quasi-experimental evidence in favor of using CWAs as educational and knowledge translation interventions, the vast majority of included studies were observational case reports about CWAs being used by health professionals and patients. More research is needed to determine which stakeholders benefit the most from using CWAs, to address the barriers to their use, to find ways to ensure the quality of their content, to foster contributions, and to make these tools effective knowledge translation tools for different stakeholders. Answers to these questions are needed before clear policy recommendations can be made about the safe use of CWAs in health care.

Acknowledgments

Funding for this project was provided by a Canadian Institutes for Health Research Knowledge Synthesis Grant (FRN 116632); a Fonds de recherche du Québec—Santé career scientist award (24856); a Fonds de Recherche du Québec—Santé, Establishment of young researchers—Juniors 1 Grant (24856); and a research grant from the Centre de santé et services sociaux Alphonse-Desjardins (CHAU de Lévis). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

We gratefully thank the following people for helping us with importing and reviewing the astonishing number of 7461 abstracts: Michèle Dugas, Cynthia Fournier, Mathieu Édmond, Azadeh Bojmehrani, Catherine Nadeau, and Hugo-Pierre Racine. We gratefully thank Elmie Peters and Monique Clar for their advice on search strategies in scientific databases. We thank Jennifer Petrela, Sandra Owens, and Laura Bégin for editing our manuscript. We also want to extend our gratitude to Bertalan Meskó, Dean Giustini, Lisa Kruesi, Lyndal Trevena, Michaël Laurent, Finn Årup Nielsen, Lynn McCleary, Jacob F. de Wolff, Holly Witteman, Robin Hervé, Linda Hawes Clever, and Tunde Varga-Atkins for providing links to articles, book chapters, and grey literature that were essential to the success of this initiative. We would also like to thank the board of directors of HLWIKI and Dean Giustini for allowing us to use their wiki to crowdsource references. We must also thank Peter Murray (IMIA) for giving us insight into what CWAs we should include in our scoping review. We are also very grateful for the help Ewa Dobrogowska offered in translating an article she wrote in Polish. Finally, we must thank our families for having supported us during the long hours conducting this review.

Conflicts of Interest

One of the authors (GE) is also editor of the Journal of Medical Internet Research (JMIR). Because of his involvement in the conduct of this research and writing of this paper, assessment and peer review have been carried out entirely by an associate editor (Potts), who was not made aware of the fact that GE was a coauthor. GE has not been involved in any editorial decisions related to this paper.

Multimedia Appendix 1

Characteristics of included studies.

[PDF File (Adobe PDF File), 262KB - jmir_v15i10e210_app1.pdf]

Multimedia Appendix 2

Patterns of use of collaborative writing applications.

[PDF File (Adobe PDF File), 124KB - jmir_v15i10e210_app2.pdf]

Multimedia Appendix 3

Quality of information in collaborative writing projects.

[PDF File (Adobe PDF File), 119KB - jmir_v15i10e210_app3.pdf]

Multimedia Appendix 4

Characteristics and results of experimental studies.

[PDF File (Adobe PDF File), 66KB - jmir_v15i10e210_app4.pdf]

References

1. Straus SE, Tetroe J, Graham ID. In: Straus SE, Tetroe J, Graham I, editors. Knowledge Translation in Health Care: Moving from Evidence to Practice. Hoboken, NJ: Wiley-Blackwell; 2009.
2. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
3. The Change Foundation Health Strategy Innovation Cell. Using Social Media to Improve Healthcare Quality: A Guide to Current Practice and Future Promise. In: Part 1, Introduction and Key Issues in the Current Landscape. Toronto, Ontario, Canada: The Change Foundation; 2011.
4. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
5. Giustini D. How Web 2.0 is changing medicine. BMJ 2006 Dec 23;333(7582):1283-1284 [FREE Full text] [doi: 10.1136/bmj.39062.555405.80] [Medline: 17185707]
6. Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med 2008 Apr 17;358(16):1732-1737. [doi: 10.1056/NEJMsb0800220] [Medline: 18420506]
7. Murray S, Choi S, Hoey J, Kendall C, Maskalyk J, Palepu A. Open science, open access and open source software at Open Medicine. Open Medicine 2008;2(1):e1.
8. Giustini D. Web 3.0 and medicine. BMJ 2007 Dec 22;335(7633):1273-1274 [FREE Full text] [doi: 10.1136/bmj.39428.494236.BE] [Medline: 18156223]
9. Huss JW, Orozco C, Goodale J, Wu C, Batalov S, Vickers TJ, et al. A gene wiki for community annotation of gene function. PLoS Biol 2008 Jul 8;6(7):e175 [FREE Full text] [doi: 10.1371/journal.pbio.0060175] [Medline: 18613750]
10. Waldrop M. Big data: Wikiomics. Nature 2008 Sep 4;455(7209):22-25. [doi: 10.1038/455022a] [Medline: 18769412]
11. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. PLoS Biol 2008 Jul 22;6(7):e184 [FREE Full text] [doi: 10.1371/journal.pbio.0060184] [Medline: 18651794]
12. de Silva V, Hanwella R. Why are we copyrighting science? BMJ 2010;341:c4738. [Medline: 20847026]
13. Godlee F, Pakenham-Walsh N, Ncayiyana D, Cohen B, Packer A. Can we achieve health information for all by 2015? Lancet 2004;364(9430):295-300. [doi: 10.1016/S0140-6736(04)16681-6] [Medline: 15262109]
14. Trevena L. WikiProject medicine. BMJ 2011;342:d3387. [Medline: 21653617]
15. Chan L, Kirsop B, Arunachalam S. Towards open and equitable access to research and knowledge for development. PLoS Med 2011 Mar;8(3):e1001016 [FREE Full text] [doi: 10.1371/journal.pmed.1001016] [Medline: 21483470]
16. Chan L, Arunachalam S, Kirsop B. The chain of communication in health science: from researcher to health worker through open access. Open Med 2009;3(3):e111-e119 [FREE Full text] [Medline: 19946398]
17. Bates DW, Wright A. Evaluating eHealth: undertaking robust international cross-cultural eHealth research. PLoS Med 2009 Sep;6(9):e1000105 [FREE Full text] [doi: 10.1371/journal.pmed.1000105] [Medline: 19753106]
18. Chu LF, Young C, Zamora A, Kurup V, Macario A. Anesthesia 2.0: internet-based information resources and Web 2.0 applications in anesthesia education. Curr Opin Anaesthesiol 2010 Apr;23(2):218-227. [doi: 10.1097/ACO.0b013e328337339c] [Medline: 20090518]
19. Eapen BR. Collaborative writing: Tools and tips. Indian J Dermatol Venereol Leprol 2007;73(6):439-441 [FREE Full text] [Medline: 18032878]
20. McLean R, Richards BH, Wardman JI. The effect of Web 2.0 on the future of medical practice and education: Darwikinian evolution or folksonomic revolution? Med J Aust 2007 Aug 6;187(3):174-177. [Medline: 17680746]
21. 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]
22. 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]
23. Leuf B, Cunningham W. The Wiki way: quick collaboration on the Web. Boston, Mass: Addison-Wesley; 2001.
24. Levy S. Wired News. 2008. Google Throws Open Rival for Wikipedia-Anon Authors Discouraged URL: http://www.wired.com/software/coolapps/news/2008/07/google_knol?currentPage=all [WebCite Cache ID 6HLdlfF8c]
25. Manber U. Google. 2007. Encouraging people to contribute URL: http://googleblog.blogspot.ca/2007/12/encouraging-people-to-contribute.html [WebCite Cache ID 6HLfxr94r]
26. Google Knol. Public Library of Science (PLoS) PLoS Currents: Influenza - a Collection by PLoS. 2011. URL: http://knol.google.com/k/plos/aims-and-scope/28qm4w0q65e4w/65 [accessed 2011-10-02]
27. Phadtare A, Bahmani A, Shah A, Pietrobon R. Scientific writing: a randomized controlled trial comparing standard and on-line instruction. BMC Med Educ 2009;9:27 [FREE Full text] [doi: 10.1186/1472-6920-9-27] [Medline: 19473511]

28. Archambault PM, Légaré F, Lavoie A, Gagnon MP, Lapointe J, St-Jacques S, et al. Healthcare professionals' intentions to use wiki-based reminders to promote best practices in trauma care: a survey protocol. Implement Sci 2010;5:45 [FREE Full text] [doi: 10.1186/1748-5908-5-45] [Medline: 20540775]
29. Archambault P, Blouin D, Poitras J, Couture M, Légaré F. Resident participation in an internet-based collaborative teaching tool (Google Docs). Open Medicine 2010;4(3):-.
30. Ganfyd contributors. The Free Medical Knowledge Base. URL: http://www.ganfyd.org [WebCite Cache ID 6GRDURi9z]
31. U.S. Government. Intelink Driven by Mission, Enabled by Innovation. URL: http://www.intelink.gov [WebCite Cache ID 6GQ0dKGZy]
32. Varga-Atkins T, Dangerfield P, Brigden D. Developing professionalism through the use of wikis: A study with first-year undergraduate medical students. Med Teach 2010;32(10):824-829. [doi: 10.3109/01421591003686245] [Medline: 20854158]
33. Wikiproject Medicine contributors. Wikipedia: The Free Encyclopedia. Wikipedia:WikiProject Medicine/Popular pages URL: http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine/Popular_pages [accessed 2013-07-16] [WebCite Cache ID 6IAVYWpS5]
34. Croteau S. Informatisation de l'urgence; AMUQ. 2011. URL: http://www.amuq.qc.ca/amuq/informatisation/ [WebCite Cache ID 5wu4afM0r]
35. de Carvalho EC, Batilana AP, Simkins J, Martins H, Shah J, Rajgor D, et al. Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. PLoS One 2010;5(2):e9314 [FREE Full text] [doi: 10.1371/journal.pone.0009314] [Medline: 20174560]
36. Hoffmann R. A wiki for the life sciences where authorship matters. Nat Genet 2008 Sep;40(9):1047-1051. [doi: 10.1038/ng.f.217] [Medline: 18728691]
37. Martin JB, Hawes Clever L, Omenn GS, Kapor M. Medpedia: An open platform connecting people and information to advance medicine. 2007. URL: http://web.archive.org/web/20130115090302/http://www.medpedia.com/about [accessed 2013-05-05] [WebCite Cache ID 6GOcFjKYG]
38. OpenWetWare Wiki contributors. OpenWetWare: Share your science. 2009. URL: http://openwetware.org/index.php?title=Main_Page&oldid=290994 [WebCite Cache ID 6HHTdjWgE]
39. Rothman D. List of Medical Wikis. 2006. URL: http://davidrothman.net/list-of-medical-wikis/ [WebCite Cache ID 66QDS8K35]
40. Waldrop MM. Science 2.0. Sci Am 2008 May;298(5):68-73. [Medline: 18444327]
41. Wikigenes contributors. Wikigenes. Collaborative publishing URL: http://www.wikigenes.org/ [accessed 2012-03-25] [WebCite Cache ID 66QEVPlul]
42. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
43. Wikia contributors. Wikia. Diabetes Wiki URL: http://diabetes.wikia.com/wiki/Diabetes_Wiki [accessed 2013-05-07] [WebCite Cache ID 6GRBmgh9U]
44. Deshpande A, Khoja S, Lorca J, McKibbon A, Rizos C, Jadad A. Open Medicine. Asynchronous telehealth: a scoping review of analytic studies URL: http://wikisr.openmedicine.ca/index.php/Main_Page [accessed 2013-05-06] [WebCite Cache ID 6GPlPp9xk]
45. Murray S, Giustini D, Loubani T, Choi S, Palepu A. Medical research and social media: Can wikis be used as a publishing platform in medicine? Open Med 2009;3(3):e121-e122 [FREE Full text] [Medline: 21603044]
46. McIntosh B, Cameron C, Singh S, Yu C, Ahuja T, Welton N, et al. Second-line therapy in patients with type 2 diabetes inadequately controlled with metformin monotherapy: A systematic review and mixed treatment comparisons meta-analysis. 2011. URL: http://livewiki.openmedicine.ca/Second-line therapy in patients with type 2 diabetes inadequately controlled with metformin monotherapy: A systematic review and mixed treatment comparisons meta-analysis [WebCite Cache ID 6GObhkM2c]
47. Caputo I. NIH staffers get into the wiki world: scientists learn online etiquette. 2009. URL: http://www.washingtonpost.com/wp-dyn/content/article/2009/07/27/AR2009072701912.html [WebCite Cache ID 6HIwzs52l]
48. National Institute of Health. Guidelines for Participating in Wikipedia from NIH 2011 URL: http://www.nih.gov/icd/od/ocpl/resources/wikipedia/ [WebCite Cache ID 6HTY8oL2G]
49. World Health Organization. ICD11 Beta Draft URL: http://apps.who.int/classifications/icd11/browse/f/en.(Archived [accessed 2013-05-02] [WebCite Cache ID 6FyVwhWC5]
50. McGee JB, Begg M. What medical educators need to know about "Web 2.0". Med Teach 2008;30(2):164-169. [doi: 10.1080/01421590701881673] [Medline: 18464141]
51. Sandars J, Haythornthwaite C. New horizons for e-learning in medical education: ecological and Web 2.0 perspectives. Med Teach 2007 May;29(4):307-310. [doi: 10.1080/01421590601176406] [Medline: 17786742]
52. Sandars J, Homer M, Pell G, Crocker T. Web 2.0 and social software: the medical student way of e-learning. Med Teach 2010 Jun 18:-. [doi: 10.3109/01421590701798729] [Medline: 20560756]
53. Kohli MD, Bradshaw JK. What is a wiki, and how can it be used in resident education? J Digit Imaging 2011 Feb;24(1):170-175 [FREE Full text] [doi: 10.1007/s10278-010-9292-7] [Medline: 20386950]

54. Kim JY, Gudewicz TM, Dighe AS, Gilbertson JR. The pathology informatics curriculum wiki: Harnessing the power of user-generated content. J Pathol Inform 2010;1:- [FREE Full text] [doi: 10.4103/2153-3539.65428] [Medline: 20805963]
55. Schreiber WE, Giustini DM. Pathology in the era of Web 2.0. Am J Clin Pathol 2009 Dec;132(6):824-828 [FREE Full text] [doi: 10.1309/AJCPEC9FZSB4DEDH] [Medline: 19926572]
56. Barsky E, Giustini D. Web 2.0 in physical therapy: a practical overview. Physiother Can 2008;60(3):207-214 [FREE Full text] [doi: 10.3138/physio.60.3.207] [Medline: 20145753]
57. Goodman MJ. Readers' and author's responses to "are traditional peer-reviewed medical articles obsolete?". MedGenMed 2006;8(1):70 [FREE Full text] [Medline: 16915200]
58. Devgan L, Powe N, Blakey B, Makary M. Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference. Journal of the American College of Surgeons 2007 Sep;205(3):S76-S77. [doi: 10.1016/j.jamcollsurg.2007.06.190]
59. Mason F. Medi-wiki worries? Open Med 2009;3(3):e122a [FREE Full text] [Medline: 21603048]
60. Rosenzweig R. Can History Be Open Source? Wikipedia and the Future of the Past. Journal of American History 2006 Jun 01;93(1):117-146. [doi: 10.2307/4486062]
61. Clauson KA, Polen HH, Boulos MN, Dzenowagis JH. Scope, completeness, and accuracy of drug information in Wikipedia. Ann Pharmacother 2008 Dec;42(12):1814-1821. [doi: 10.1345/aph.1L474] [Medline: 19017825]
62. Callis KL, Christ LR, Resasco J, Armitage DW, Ash JD, Caughlin TT, et al. Improving Wikipedia: educational opportunity and professional responsibility. Trends Ecol Evol 2009 Apr;24(4):177-179. [doi: 10.1016/j.tree.2009.01.003] [Medline: 19269059]
63. Pender MP, Lasserre KE, Del Mar C, Kruesi L, Anuradha S. Is Wikipedia unsuitable as a clinical information resource for medical students? Med Teach 2009 Dec;31(12):1095-1096. [Medline: 20050104]
64. Arita M. A pitfall of wiki solution for biological databases. Brief Bioinform 2009 May;10(3):295-296 [FREE Full text] [doi: 10.1093/bib/bbn053] [Medline: 19060305]
65. Giles J. Wikipedia rival calls in the experts. Nature 2006 Oct 5;443(7111):493. [doi: 10.1038/443493a] [Medline: 17024058]
66. Kittur A, Suh B, Pendleton BA, Chi EH. He says, she says: conflict and coordination in Wikipedia. : ACM; 2007 Presented at: SIGCHI conference on Human factors in computing systems; 2007; San Jose, California, USA p. 453-462.
67. Jain SH. Practicing medicine in the age of Facebook. N Engl J Med 2009 Aug 13;361(7):649-651. [doi: 10.1056/NEJMp0901277] [Medline: 19675328]
68. Cohen N. Complaint over doctor who posted inkblot test. 2009. URL: http://www.nytimes.com/2009/08/24/business/24inkblot.html [WebCite Cache ID 5rz9fciJ6]
69. Logan DW, Sandal M, Gardner PP, Manske M, Bateman A. Ten simple rules for editing Wikipedia. PLoS Comput Biol 2010;6(9):- [FREE Full text] [doi: 10.1371/journal.pcbi.1000941] [Medline: 20941386]
70. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": an inter-disciplinary literature and website review. Int J Med Inform 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006] [Medline: 20188623]
71. Vogel L. Dr. YouTube will see you now. CMAJ 2011 Apr 5;183(6):647-648 [FREE Full text] [doi: 10.1503/cmaj.109-3812] [Medline: 21398238]
72. Collier J. Wiki technology in the classroom: building collaboration skills. J Nurs Educ 2010 Dec;49(12):718. [doi: 10.3928/01484834-20101117-02] [Medline: 21117555]
73. Naik AD, Singh H. Electronic health records to coordinate decision making for complex patients: what can we learn from wiki? Med Decis Making 2010 Dec;30(6):722-731. [doi: 10.1177/0272989X10385846] [Medline: 21183759]
74. Wu L, Russel C, Patterson R. Infusing Information Literacy Skills by Researching “Never Events”. 2009 May Presented at: Conference of the Medical Library Association; 2009; Honolulu, Hawaii p. 15-20.
75. Pathology Informatics Curriculum Wiki contributors. Pathology Informatics Curriculum Wiki. URL: http://pathinformatics.wikispaces.com/ [accessed 2013-05-06] [WebCite Cache ID 6GPn6LWlQ]
76. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, OpenMRS Implementers Network. The OpenMRS Implementers Network. Int J Med Inform 2009 Nov;78(11):711-720. [doi: 10.1016/j.ijmedinf.2008.09.005] [Medline: 19157968]
77. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
78. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277. [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
79. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012 Jul;19(5):777-781 [FREE Full text] [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
80. Chou WY, Prestin A, Lyons C, Wen KY. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]

81. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
82. Hollinderbäumer A, Hartz T, Uckert F. Education 2.0 -- how has social media and Web 2.0 been integrated into medical education? A systematical literature review. GMS Z Med Ausbild 2013;30(1):Doc14 [FREE Full text] [doi: 10.3205/zma000857] [Medline: 23467509]
83. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. Acad Med 2013 Jun;88(6):893-901. [doi: 10.1097/ACM.0b013e31828ffc23] [Medline: 23619071]
84. 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]
85. Okoli C, Mehdi M, Mesgari M, Nielsen F, Lanamäki A. The People’s Encyclopedia Under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia. SSRN Journal 2012:-. [doi: 10.2139/ssrn.2021326]
86. Archambault PM, van de Belt TH, Grajales Ill FJ, Eysenbach G, Aubin K, Gold I, et al. Wikis and collaborative writing applications in health care: a scoping review protocol. JMIR Res Protoc 2012;1(1):e1. [doi: 10.2196/resprot.1993] [Medline: 23612481]
87. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
88. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
89. Sampson M, McGowan J, Lefebvre C, Moher D, Grimshaw J. PRESS: Peer Review of Electronic Search Strategies. Ottawa, Ontario, Canada: Canadian Agency for Drugs and Technologies in Health; 2008.
90. 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]
91. Tesch R. Qualitative Research; Analysis Types and Software Tools. New York: Falmer Press; 1990.
92. Morse JM. Qualitative research methods for health professionals. Thousand Oaks: Sage Publications; 1995.
93. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
94. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
95. IU Radiology Wiki contributors. Indiana University. 2013. IU Radiology Wiki URL: http://www.indyradres.org/ [WebCite Cache ID 6GPnOOzPM]
96. Wu HY, Chen KH. A quantitative comparison on Online Encyclopedias - A case study of Wikipedia and Knol. 2010. URL: http://www.lis.ntu.edu.tw/~khchen/writtings/pdf/20101116NTNU02.pdf [WebCite Cache ID 6GRAzlIrr]
97. Sulfi S, Voigt JU. Wiki Echo. 2012. URL: http://www.wikiecho.org [WebCite Cache ID 6GOUXhfEV]
98. Gibson M. WikiDoc: The Living Textbook of Medicine. 2013. URL: http://www.wikidoc.org [WebCite Cache ID 6GOVHXCqR]
99. Dobrogowska-Schlebusch E. Evaluation of the quality of 52 health-related websites, created with wiki technology. Zeszyty Naukowe Ochrony Zdrowia, Zdrowie Publiczne i Zarzadzanie 2009;7(1):102-109.
100. Gupta S, Chignell M, Hall S, Straus SE. Online Collaboration Tool for Asthma Action Plan With Usability (OCTAPUS). Chest 2010 Oct 01;138(4 MeetingAbstracts):172A-172A. [doi: 10.1378/chest.10883]
101. Wan F. Reverse engineering of content as a task for finding usability problems: An evaluative case study using the Wikibreathe tool for online creation of asthma action plans. Canada: University of Toronto (Canada); 2009.
102. Ma ZS, Zhang HJ, Yu T, Ren G, Du GS, Wang YH. Orthochina.org: case-based orthopaedic Wiki project in China. Clin Orthop Relat Res 2008 Oct;466(10):2428-2437 [FREE Full text] [doi: 10.1007/s11999-008-0396-z] [Medline: 18685911]
103. Van Der Schoor-Knijnenburg J, Den Breejen E, Nelen W, Hermens R, Kremer J. Broad collaboration for a new subfertility guideline in the Netherlands: The role of the patient. 2009 Jun 30 Presented at: 25th Annual Meeting of the European Society of Human Reproduction and Embryology; 2009; Amsterdam, Netherlands p. i58.
104. Hanson AH, Krause LK, Simmons RN, Ellis JI, Gamble RG, Jensen JD, et al. Dermatology education and the Internet: traditional and cutting-edge resources. J Am Acad Dermatol 2011 Oct;65(4):836-842. [doi: 10.1016/j.jaad.2010.05.049] [Medline: 21820206]
105. Belt T, Knijnenburg J, Nelen W, Kremer JA. Wikis as an Opportunity to Improve Patient Participation in Developing Information Leaflets: A demonstration project in infertility patients. : Social Media, Mobile Apps,Internet/Web 20 in Health, MedicineBiomedical Research Conference; 2011 Presented at: Medicine 2.0: Social Media, Mobile Apps, and Internet/Web 2.0 in Health, Medicine and Biomedical Research Conference; 2011; Stanford, CA.
106. Moen A, Smørdal O, Sem I. Web-based resources for peer support - opportunities and challenges. Stud Health Technol Inform 2009;150:302-306. [Medline: 19745318]

107. Ioannis Chiotelis I, Giannakopoulos A, Kalafati M, Koutsouradi M, Kallistratos M, Manolis AJ. Secondary prevention with internet support after an acute coronary syndrome in greek patients. European Journal of Cardiovascular Prevention and Rehabilitation 2011;18(1):S11.
108. Moeller S, Spitzer K, Spreckelsen C. How to configure blended problem based learning-results of a randomized trial. Med Teach 2010;32(8):e328-e346. [doi: 10.3109/0142159X.2010.490860] [Medline: 20662568]
109. Stutsky BJ. Empowerment and leadership development in an online story-based learning community. Florida, United States: Nova Southeastern University; 2009.
110. Williams B, Bearman M. Can wikis be used to support case-based learning in paramedic education? Journal of Paramedic Practice 2011;3(7):388-392.
111. Dodson S, Gleason AW. Web 2.0 support for residents' and fellows' patient care and educational needs. Med Ref Serv Q 2011;30(2):95-101. [doi: 10.1080/02763869.2011.562797] [Medline: 21534109]
112. Harris CM, Cameron SL. Displacing Wikipedia: Information literacy for first-year students. In: Dunn DS, Beins BC, McCarthy MA, editors. Best practices for teaching beginnings and endings in the psychology major: Research, cases, and recommendations. New York: Oxford University Press; 2010:125-136.
113. 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]
114. Sandars J, Morrison C. What is the Net Generation? The challenge for future medical education. Med Teach 2007 Mar;29(2-3):85-88. [doi: 10.1080/01421590601176380] [Medline: 17701615]
115. Sandars J, Homer M, Pell G, Croker T. Web 2.0 and social software: the medical student way of e-learning. Med Teach 2008;30(3):308-312. [doi: 10.1080/01421590701798729] [Medline: 18608950]
116. Brokowski L, Sheehan AH. Evaluation of pharmacist use and perception of Wikipedia as a drug information resource. Ann Pharmacother 2009 Nov;43(11):1912-1913. [doi: 10.1345/aph.1M340] [Medline: 19843833]
117. Iyer AK. Drug information-seeking behavior among healthcare professionals within the University of Utah Community Clinics. Salt Lake City, Utah: The University of Utah; 2011.
118. Usher W. Types of social media (Web 2.0) used by Australian allied health professionals to deliver early twenty-first-century practice promotion and health care. Soc Work Health Care 2011;50(4):305-329. [doi: 10.1080/00981389.2010.534317] [Medline: 21512953]
119. Alkhateeb FM, Clauson KA, Latif DA. Pharmacist use of social media. Int J Pharm Pract 2011 Apr;19(2):140-142. [doi: 10.1111/j.2042-7174.2010.00087.x] [Medline: 21385246]
120. González de Dios J, Camino-León R, Ramos-Lizana J. [The use of bibliographic information resources and Web 2.0 by neuropaediatricians]. Rev Neurol 2011 Jun 16;52(12):713-719 [FREE Full text] [Medline: 21594856]
121. Schweitzer NJ. Wikipedia and Psychology: Coverage of Concepts and Its Use by Undergraduate Students. Teaching of Psychology 2008 Apr 15;35(2):81-85. [doi: 10.1080/00986280802004594]
122. Hickerson CA, Thompson SR. Dialogue through wikis: a pilot exploration of dialogic public relations and wiki websites. PRism Online PR Journal 2009;6(1):1-11.
123. Judd T, Kennedy G. A five-year study of on-campus Internet use by undergraduate biomedical students. Computers & Education 2010 Dec;55(4):1564-1571. [doi: 10.1016/j.compedu.2010.06.022]
124. Judd T, Kennedy G. Expediency-Based Practice? Medical Students' Reliance on Google and Wikipedia for Biomedical Inquiries. British Journal of Educational Technology 2011;42(2):351-360. [doi: 10.1111/j.1467-8535.2009.01019.x]
125. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009 Aug;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
126. Law MR, Mintzes B, Morgan SG. The sources and popularity of online drug information: an analysis of top search engine results and web page views. Ann Pharmacother 2011 Mar;45(3):350-356. [doi: 10.1345/aph.1P572] [Medline: 21343404]
127. Martin JR, Kramer S, Slack M. Information literacy skills of first year pharmacy students: Focus group results. 2011 Presented at: 112th Annual Meeting of the American Association of Colleges of Pharmacy; July; 2011; San Antonio, Texas p. 9-13.
128. Limdi JK, Butcher RO. Information resources and inflammatory bowel disease. Inflamm Bowel Dis 2011 Aug;17(8):E89-E90. [doi: 10.1002/ibd.21749] [Medline: 21557399]
129. Santos Arrontes D, García González JI, Martín Muñoz MP, Jiménez Jiménez JI, Paniagua Andrés P. [Internet use in patients attending a hospital urology clinic]. Actas Urol Esp 2007 Dec;31(10):1161-1165. [Medline: 18314655]
130. Bender JL, O'Grady LA, Deshpande A, Cortinois AA, Saffie L, Husereau D, et al. Collaborative authoring: a case study of the use of a wiki as a tool to keep systematic reviews up to date. Open Med 2011;5(4):e201-e208 [FREE Full text] [Medline: 22567076]
131. Johnson PT, Chen JK, Eng J, Makary MA, Fishman EK. A comparison of world wide web resources for identifying medical information. Acad Radiol 2008 Sep;15(9):1165-1172. [doi: 10.1016/j.acra.2008.02.010] [Medline: 18692758]
132. Friedlin J, McDonald CJ. An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database. J Am Med Inform Assoc 2010 Jun;17(3):283-287 [FREE Full text] [doi: 10.1136/jamia.2009.001180] [Medline: 20442145]
133. Aldairy T, Laverick S, McIntyre GT. Orthognathic surgery: is patient information on the Internet valid? Eur J Orthod 2012 Aug;34(4):466-469 [FREE Full text] [doi: 10.1093/ejo/cjr046] [Medline: 21459834]

134. Czarnecka-Kujawa K, Abdalian R, Grover SC. The Quality of Open Access and Open Source Internet Material in Gastroenterology: Is Wikipedia Appropriate for Knowledge Transfer to Patients? Gastroenterology 2008 Apr;134(4):A-325-A-326. [doi: 10.1016/s0016-5085(08)61518-8]
135. Lavsa SM, Corman SL, Culley CM, Pummer TL. Reliability of Wikipedia as a medication information source for pharmacy students. Currents in Pharmacy Teaching and Learning 2011 Apr;3(2):154-158. [doi: 10.1016/j.cptl.2011.01.007]
136. Haigh CA. Wikipedia as an evidence source for nursing and healthcare students. Nurse Educ Today 2011 Feb;31(2):135-139. [doi: 10.1016/j.nedt.2010.05.004] [Medline: 20646799]
137. Wood A, Struthers K. Pathology education, Wikipedia and the Net generation. Med Teach 2010;32(7):618. [doi: 10.3109/0142159X.2010.497719] [Medline: 20653388]
138. Bookstaver PB, Rudisill CN, Bickley AR, McAbee C, Miller AD, Piro CC, et al. An evidence-based medicine elective course to improve student performance in advanced pharmacy practice experiences. Am J Pharm Educ 2011 Feb 10;75(1):9 [FREE Full text] [Medline: 21451761]
139. Cobus L. Using blogs and wikis in a graduate public health course. Med Ref Serv Q 2009;28(1):22-32. [doi: 10.1080/02763860802615922] [Medline: 19197741]
140. Hamilton A, Edney P, Armijo-Olivo S. Mapping our way to the Future of Knowledge Transfer - Development of a Wiki in an Evidence-Based Practice Physical Therapy course. 2008 Presented at: Health Research Transfer Network of Alberta conference; September 28-30, 2008; Banff, Canada.
141. Hulbert-Williams NJ. Facilitating Collaborative Learning Using Online Wikis: evaluation of their application within postgraduate psychology teaching. plat 2010;9(1):45. [doi: 10.2304/plat.2010.9.1.45]
142. Llambí L, Esteves E, Martinez E, Forster T, García S, Miranda N, et al. Teaching tobacco cessation skills to Uruguayan physicians using information and communication technologies. J Contin Educ Health Prof 2011;31(1):43-48. [doi: 10.1002/chp.20100] [Medline: 21425359]
143. Philip CT, Unruh KP, Lachman N, Pawlina W. An explorative learning approach to teaching clinical anatomy using student generated content. Anat Sci Educ 2008 May;1(3):106-110. [doi: 10.1002/ase.26] [Medline: 19177391]
144. Shaw AC, van Balkom ID, Bauer M, Cole TR, Delrue MA, Van Haeringen A, et al. Phenotype and natural history in Marshall-Smith syndrome. Am J Med Genet A 2010 Nov;152A(11):2714-2726. [doi: 10.1002/ajmg.a.33709] [Medline: 20949508]
145. Ciesielka D. Using a wiki to meet graduate nursing education competencies in collaboration and community health. J Nurs Educ 2008 Oct;47(10):473-476. [Medline: 18856102]
146. Cinnamon J, Schuurman N. Injury surveillance in low-resource settings using Geospatial and Social Web technologies. Int J Health Geogr 2010;9:25 [FREE Full text] [doi: 10.1186/1476-072X-9-25] [Medline: 20497570]
147. Hamm KM, Simeonov IM, Heard SE. Using technology to harness and organize expertise in the development of health education materials: How a wiki can help you collaborate. 2009 Presented at: North American Congress of Clinical Toxicology Annual Meeting; Sept. 21-26, 2009; San Antonio, TX, United States p. 09-21.
148. Kardong-Edgren SE, Oermann MH, Ha Y, Tennant MN, Snelson C, Hallmark E, et al. Using a wiki in nursing education and research. Int J Nurs Educ Scholarsh 2009;6:Article6. [doi: 10.2202/1548-923X.1787] [Medline: 19341356]
149. Kitson-Reynolds E. Energizing enquiry based learning through technology advances. British Journal of Midwifery 2009;17(2):118-122.
150. Musil E, Hordesky S. A Technologically Enhanced (Wiki) Activity Series to Integrate Courses Across Curricular Departments. In: The American Journal of Pharmaceutical Education. 2011 Presented at: 112th Annual Meeting of the American Association of Colleges of Pharmacy; July 9-13, 2011; San Antonio, Texas p. 116-117.
151. Andrus LC. Demonstration of interactive wiki website for clinical nursing student documentation. Communicating Nursing Research 2010;43:359.
152. Koerner P, Elliot J, Heasley J. Utilization of a WIKI to Enhance Student Learning within an Elective Pediatric Course. In: The American Journal of Pharmaceutical Education. 2011 Presented at: 112th Annual Meeting of the American Association of Colleges of Pharmacy; July 9-13, 2011; San Antonio, Texas p. 66.
153. Morley DA. Enhancing networking and proactive learning skills in the first year university experience through the use of wikis. Nurse Educ Today 2012 Apr;32(3):261-266. [doi: 10.1016/j.nedt.2011.03.007] [Medline: 21481500]
154. Jones P. Collaboration at a Distance: Using a Wiki to Create a Collaborative Learning Environment for Distance Education and On-Campus Students in a Social Work Course. Journal of Teaching in Social Work 2010 May 18;30(2):225-236. [doi: 10.1080/08841231003705396]
155. Archambault P, Légaré F, Gagnon MP, Bilodeau A, Lavoie A, Lapointe J, et al. Emergency Physicians' Beliefs About Their Intention to Use Wiki-based Reminders to Promote Evidence-based Trauma Care. 2011 Presented at: Medicine 2.0: Social Media, Mobile Apps, and Internet/Web 2.0 in Health, Medicine and Biomedical Research Conference; 2011; Stanford, CA.
156. Culley JM, Polyakova-Norwood V. Synchronous online role play for enhancing community, collaboration, and oral presentation proficiency. Nurs Educ Perspect 2012;33(1):51-54. [Medline: 22416543]

157. Umland E, Klootwyk J, Batool A, Halas C. Wiki Use to Compliment and Add Value to Learning in a Women's Health Elective Course. In: The American Journal of Pharmaceutical Education. 2011 Presented at: 112th Annual Meeting of the American Association of Colleges of Pharmacy; July 9-13, 2011; San Antonio, Texas p. 107-108.
158. Lanning LC, Dadig BA. A strategy for incorporating palliative care and end-of-life instruction into physician assistant education. J Physician Assist Educ 2010;21(4):41-46. [Medline: 21366115]
159. Miller AD, Bookstaver PB, Norris LB. Use of Wikis in advanced pharmacy practice experiences. Am J Pharm Educ 2009 Dec 17;73(8):139 [FREE Full text] [Medline: 20221331]
160. Hawkins J, Campbell A. Wiki travels: using technology to teach global health concepts. Reflections on Nursing Leadership 2010;36(3):1.
161. Dhillon S, Erichsen S, Crowther S, Pianta MJ, Lambert RG. Development of an online teaching resource based on the society of skeletal radiology's musculoskeletal radiology curriculum utilizing open access web material. Skeletal Radiology 2011;40(4):506-507.
162. Mirk SM, Burkiewicz JS, Komperda KE. Student perception of a wiki in a pharmacy elective course. Currents in Pharmacy Teaching and Learning 2010 Mar;2(2):72-78. [doi: 10.1016/j.cptl.2010.01.002]
163. Welsh A. Internal Wikis for procedures and training: from tacit knowledge to self-guided learning. Online 2007;31(6):26-29.
164. Yates D, Paquette S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management 2011 Feb;31(1):6-13. [doi: 10.1016/j.ijinfomgt.2010.10.001]
165. Moser R, Glasgow R, Hesse B. Wiki Approaches to Enhance Reach and Breadth of Stakeholder Involvement in Identification of Practical Patient-Reported Measures for Primary Care. 2011 Presented at: Medicine 2.0: Social Media, Mobile Apps, and Internet/Web 2.0 in Health, Medicine and Biomedical Research Conference; 2011; Stanford, CA.
166. Meenan C, King A, Toland C, Daly M, Nagy P. Use of a wiki as a radiology departmental knowledge management system. J Digit Imaging 2010 Apr;23(2):142-151 [FREE Full text] [doi: 10.1007/s10278-009-9180-1] [Medline: 19184221]
167. Morose T. Using an interactive website to disseminate participatory ergonomics research findings: An exploratory study. Waterloo, Ontario: University of Waterloo; 2007.
168. Muir S, Muir B, Ghahremani S, Hall T, Isidro C, Boechat I. Utilizing a wikispace, dedicated itunes/youtube channels, podcasts and twitter to advance radiology resident education. Pediatric Radiology 2010;40(4):605.
169. Streeter JL, Lu MT, Rybicki FJ. Informatics in radiology: RadiologyWiki.org: the free radiology resource that anyone can edit. Radiographics 2007;27(4):1193-1200 [FREE Full text] [doi: 10.1148/rg.274065090] [Medline: 17620476]
170. Montano BS, Garcia Carretero R, Varela Entrecanales M, Pozuelo PM. Integrating the hospital library with patient care, teaching and research: model and Web 2.0 tools to create a social and collaborative community of clinical research in a hospital setting. Health Info Libr J 2010 Sep;27(3):217-226. [doi: 10.1111/j.1471-1842.2010.00893.x] [Medline: 20712716]
171. Chiarella D, Hasman L, Chiarella DT. Health literacy: using Web 2.0 to create an autism resource. Journal of Consumer Health on the Internet 2009;13(3):281-286.
172. Wright A, Bates DW, Middleton B, Hongsermeier T, Kashyap V, Thomas SM, et al. Creating and sharing clinical decision support content with Web 2.0: Issues and examples. J Biomed Inform 2009 Apr;42(2):334-346. [doi: 10.1016/j.jbi.2008.09.003] [Medline: 18935982]
173. Yu R, Crotty B. Wikis to better manage shared information in a hospitalist group. In: Journal of Hospital Medicine. 2011 Presented at: Society of Hospital Medicine. Research, Innovations, Clinical Vignettes Competition; May 10-13, 2011; Gaylord Texan Resort and Convention Center, Grapevine, TX p. 140-141.
174. Steininger K, Ruckel D, Dannerer E, Roithmayr F. Healthcare knowledge transfer through a web 2.0 portal: an Austrian approach. IJHTM 2010;11(1/2):13. [doi: 10.1504/IJHTM.2010.033272]
175. Felsen U, Kunins H, Jeffers A, Stark R. Implementation of a wiki-based education and resource tool in a primary care residency program. Journal of General Internal Medicine 2010;25:S455-S4S6. [doi: 10.1007/s11606-010-1338-5]
176. Gerber D, Eberle B, Trachsel S. A web-based knowledge database (Wiki platform) for Standard Operational Procedures (SOPS) in cardiac anesthesia. Journal of Clinical Monitoring and Computing 2010;24(1):12-13. [doi: 10.1007/s10877-009-9211-y]
177. Blakely R. The Health Care Blog of the Times Newspaper. 2007. Beyond Wikipedia URL: http://thehealthcareblog.com/blog/2009/05/23/beyond-wikipedia/ [accessed 2012-06-04] [WebCite Cache ID 6JOAdHTwy]
178. Jalali A, Mioduszewski M, Gauthier M, Varpio L. Wiki use and challenges in undergraduate medical education. Med Educ 2009 Nov;43(11):1117. [doi: 10.1111/j.1365-2923.2009.03480.x] [Medline: 19874526]
179. Mosquera M. VA wrestles with physician use of commercial e-tools.: Government Health IT; 2010. URL: http://www.govhealthit.com/news/va-wrestles-physician-use-commercial-e-tools [WebCite Cache ID 6HNDdsoZu]
180. Lauber CA. Using a Wiki for approved clinical instructor training. Athletic Therapy Today 2009;14(6):25-28.
181. Krebs C. Evaluation of a comprehensive neuroanatomy website in a distributed medical curriculum. The FASEB Journal. (S1) 2009;23:476.
182. McInnes N, Haglund BJ. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
183. Lorenz A, Türp JC. [Dentistry in Wikipedia: A quantitative and qualitative analysis]. Schweiz Monatsschr Zahnmed 2010;120(12):1117-1126. [Medline: 21261123]

184. Deshpande A, Khoja S, Lorca J, McKibbon A, Rizo C, Husereau D, et al. Asynchronous telehealth: a scoping review of analytic studies. Open Med 2009;3(2):e69-e91 [FREE Full text] [Medline: 19946396]
185. Michaud PR. PmWiki. 2013. URL: http://www.pmwiki.org [WebCite Cache ID 6GQ1gyGtH]
186. Planque F. More than a Blog. The EVO factory. 2003. URL: http://b2evolution.net [WebCite Cache ID 6GRO2acbc]
187. LibGuides Community Contributors. LibGuides Community. URL: http://libguides.com [accessed 2013-05-06] [WebCite Cache ID 6GQ1oDJQm]
188. Gohr A. DokuWiki. 2013. URL: https://www.dokuwiki.org [WebCite Cache ID 6GPnc8xQL]
189. Sandars J, Schroter S. Web 2.0 technologies for undergraduate and postgraduate medical education: an online survey. Postgrad Med J 2007 Dec;83(986):759-762 [FREE Full text] [doi: 10.1136/pgmj.2007.063123] [Medline: 18057175]
190. Ayes KB, Bardsley CH. Wikipedia information for toxicologic emergencies: How reliable is it? Clinical Toxicology 2010;48(6):635.
191. Zembowicz A. Dermpedia. URL: http://www.dermpedia.org [WebCite Cache ID 6GQ0y1pZ7]
192. Martin JB, Hawes Clever L, Omenn GS, Kapor M. Medpedia. 2007. The Medpedia Project URL: http://www.medpedia.com/about [WebCite Cache ID 6GOcFjKYG]
193. Buytaert D. Drupal: Come for the software, stay for the community. URL: http://drupal.org [WebCite Cache ID 6GROGOrWu]
194. Pathology Informatics Curriculum Wiki contributors. Path. Pathology Informatics Curriculum Wiki URL: http://pathinformatics.wikispaces.com/ [WebCite Cache ID 6GPn6LWlQ]
195. Leithner A, Maurer-Ertl W, Glehr M, Friesenbichler J, Leithner K, Windhager R. Wikipedia and osteosarcoma: a trustworthy patients' information? J Am Med Inform Assoc 2010;17(4):373-374 [FREE Full text] [doi: 10.1136/jamia.2010.004507] [Medline: 20595302]
196. Mercer J. Wikipedia and open source mental health information. The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work 2007;17(4):373-374.
197. Mühlhauser I, Oser F. Does WIKIPEDIA provide evidence based health care information? A content analysis. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 2008 Jan;102(7):e1-e7. [doi: 10.1016/j.zefq.2008.06.020]
198. Rajagopalan MS, Khanna V, Stott M, Leiter Y, Showalter TN, Dicker A, et al. Accuracy of cancer information on the Internet: A comparison of a Wiki with a professionally maintained database. Journal of Clinical Oncology 2010;28(15):-.
199. Tulbert BH, Snyder CW, Brodell RT. Readability of Patient-oriented Online Dermatology Resources. J Clin Aesthet Dermatol 2011 Mar;4(3):27-33 [FREE Full text] [Medline: 21464884]
200. Buzzi M, Leporini B. Editing Wikipedia content by screen reader: easier interaction with the Accessible Rich Internet Applications suite. Disabil Rehabil Assist Technol 2009 Jul;4(4):264-275. [doi: 10.1080/17483100902903457] [Medline: 19565387]
201. Cousineau LK. Wikipedia-trica: A Library–Department of Pediatrics Collaboration for a Morning Report Wiki. : A Library–Department of Pediatrics Collaboration for a Morning Report Wiki. Conference of the Medical Library Association; 2009 Presented at: Conference of the Medical Library Association; May 15-20; Honolulu, Hawaii.
202. Damani S, Fulton S. Collaborating with Clinical Teams Using Web 2.0 Tools. 2009 Presented at: Conference of the Medical Library Association; May 15-20; Honolulu, Hawaii.
203. Kraft M. Getting wiki with it: using a wiki as a web site for regional health system libraries. 2009 Presented at: Conference of the Medical Library Association; May 15-20; Honolulu, Hawaii.
204. Fetjek M, Krebs C, Wilson J, Findlay T, Chemey N. Neuroanatomy at UBC. 2006. URL: http://www.neuroanatomy.ca [WebCite Cache ID 6GPnqxhl6]
205. Lanning L. Palliative Care Resources for Physician Assistants Home.: Wetpaint; 2009. URL: http://palliativecareforpas.wetpaint.com [WebCite Cache ID 6GPo3S5SS]
206. OrthoChina contributors. 1997. OrthoChina: The Web on Orthopedics in China URL: http://www.orthochina.com/english/index.jsp [WebCite Cache ID 6GPoQhz6K]
207. Matlin T, McGuiness S. Creating an Interactive Resource Guide to Improve Student Researching Skills. 2009 Presented at: Conference of the Medical Library Association; May 15-20; Honolulu, Hawaii.
208. MinJournal contributors. MinJournal. URL: http://goo.gl/07MHm [WebCite Cache ID 6GPoxOMuC]
209. Wikinvestigacion contributors. 2009. Wikinvestigacion URL: http://www.wikinvestigacion.org [WebCite Cache ID 6GPp7aNFi]
210. CRE-MSD Researchers. Centre of Research Expertise for the Prevention of Musculoskeletal Disorders: University of Waterloo. URL: http://www.cre-msd.uwaterloo.ca/ [WebCite Cache ID 6GPpQ2i8y]
211. Muir S. Wikispaces. 2013. PediatricImaging URL: http://pediatricimaging.wikispaces.com [WebCite Cache ID 6GRAzlIrr]
212. Powers E, Skica J, Pierce J, Travis L. “COOL” Libraries Use of Web 2.0 Tools. 2009 Presented at: Conference of the Medical Library Association; May 15-20; Honolulu, Hawaii.
213. OpenMRS Community contributors. Open Source Atlassian Confluence. OpenMRS Community URL: https://wiki.openmrs.org [WebCite Cache ID 6GPpxKbH9]

214. MSS Research Foundation. 2011. Marshall Smith Syndrome Research Foundation URL: http://www.marshallsmith.org [WebCite Cache ID 6GPqV6Mg9]
215. PBwiki inc. 2013. PBWORKS Online Team Collaboration URL: http://pbwiki.com [WebCite Cache ID 6GPzzSyNd]
216. Freya WIKI contributors. FreyaWIKI. URL: http://www.freyawiki.nl [WebCite Cache ID 6GQ0EwA0I]
217. Clinfowiki contributors. 2013. Clinfowiki URL: http://www.clinfowiki.org [WebCite Cache ID 6GQ0SEK17]
218. AskDrWiki contributors. 2012. AskDrWiki: A Medical Wiki with the goal of creating a collective online memory for physicians, nurses and medical students URL: http://askdrwiki.com/ [WebCite Cache ID 6GRAzlIrr]
219. Toronto Public Library Wiki contributors. Canadian Health Information. 2010. URL: http://chis.wikidot.com/about [WebCite Cache ID 6GRBRcpIP]
220. Wikibooks contributors. 2012. Demystifying Depression URL: http://en.wikibooks.org/wiki/Demystifying_Depression [WebCite Cache ID 6GRBcmnBh]
221. Wikia contributors. Wikia. 2006. Diabetes Wiki URL: http://diabetes.wikia.com/wiki/Diabetes_Wiki [WebCite Cache ID 6GRBmgh9U]
222. Wikibooks contributors. 2012. Diagnostic Radiology URL: http://en.wikibooks.org/wiki/Diagnostic_Radiology [WebCite Cache ID 6GRBtKbVE]
223. Antwerpes F, Bartsch S, Beutler B. DocCheck Flexikon: The Cooperative Medical Dictionary. 2013. URL: http://flexikon.doccheck.com/en/Special:Mainpage [WebCite Cache ID 6GRCglSTd]
224. Schardt C, Odato K. EMB Librarian. URL: https://sites.google.com/site/ebmlibrarian/home [WebCite Cache ID 6GRCswo5t]
225. Wikibooks contributors. 2012. Open-Content Textbook of Emergency Medicine URL: http://en.wikibooks.org/wiki/Emergency_Medicine [WebCite Cache ID 6GRD10Xsz]
226. Flu Wiki Forum contributors. 2006. Flu Wiki Forum URL: http://www.newfluwiki2.com [WebCite Cache ID 6GRDII4LH]
227. Wikibooks contributors. 2012. Handbook of Genetic Counseling URL: http://en.wikibooks.org/wiki/Handbook_of_Genetic_Counseling [WebCite Cache ID 6GRDkV3Ab]
228. HealthGrid wiki contributors. 2010. HealthGrid wiki URL: http://www.healthgrid.org/tools/wiki.php [WebCite Cache ID 6HNpn5bDd]
229. Wikibooks contributors. 2012. Immunology URL: http://en.wikibooks.org/wiki/Immunology [WebCite Cache ID 6GRE1SbqP]
230. Fadem SZ. Wikikidney. URL: http://www.wikikidney.org/index.php?title=Main_Page [WebCite Cache ID 6GREEylDG]
231. Wikia contributors. Wikia. 2006. Medical Imaging URL: http://medicalimaging.wikia.com/wiki/Main_Page [WebCite Cache ID 6GREOVnCg]
232. MLA-HLS Wiki contributors. 2013. MLA-HLS URL: http://mla-hls.wikispaces.com [WebCite Cache ID 6GREZPuci]
233. Bhui K, de Jongh B, Ingleby D, Warfa N. MIGHEALTHNET - Information network on good practice in health care for migrants and minorities in Europe. 2011. URL: http://www.mighealth.net/uk/index.php/Main_Page [WebCite Cache ID 6GREuygu1]
234. de Jong JSSG, Postema PG, Kreuger R. de ECGpedia. 2012. URL: http://en.ecgpedia.org/wiki/Main_Page [WebCite Cache ID 6GRFBXN3x]
235. Colosimo A, Duffy E, Gore G. The McGill Library Global Health Resource Guide. 2008. URL: http://wikisites.mcgill.ca/GlobalHealthGuide/index.php/Main_Page [WebCite Cache ID 6GRFMyN44]
236. EBHC Search Strategies Wiki contributors. 2008. EBHC Search Strategies URL: http://ebhcstrategies.wetpaint.com [WebCite Cache ID 6GRFViMmU]
237. Neurodegeneration Research Wiki contributors. Neurodegeneration Research Wiki. URL: http://wiki.iop.kcl.ac.uk/default.aspx/Neurodegeneration/Neurodegeneration%20Research%20Wiki.html [WebCite Cache ID 6GRFjzfyF]
238. PharmLib Wiki contributors. PharmLib Wiki. URL: http://pharmlib.pbworks.com/w/page/16284404/FrontPage [WebCite Cache ID 6GRG24VNu]
239. D'avolio L. The Quality of Medical Data. URL: http://medicaldata.wikia.com/wiki/Main_Page [WebCite Cache ID 6GRGISgXW]
240. Cook B, Dvorak T. Radiation Oncology. 2012. URL: http://en.wikibooks.org/wiki/Radiation_Oncology [WebCite Cache ID 6GRGTLqxU]
241. Gaillard F, Jones J. Radiopaedia. URL: http://radiopaedia.org/ [WebCite Cache ID 6GRGg38OP]
242. Wikibooks contributors. 2012. Rosacea URL: http://en.wikibooks.org/wiki/Rosacea [WebCite Cache ID 6GRGq0A4A]
243. Wikiversity contributors. 2006. School:Medicine URL: http://en.wikiversity.org/wiki/School:Medicine [WebCite Cache ID 6HNr1jSQj]
244. Wikia contributors. Wikia. 2006. Street Medic URL: http://medic.wikia.com/wiki/Main_Page [WebCite Cache ID 6GRGyJ8QW]
245. Wikibooks contributors. 2012. Surgical Procedures URL: http://en.wikibooks.org/wiki/Surgery [WebCite Cache ID 6GRH9fGIM]
246. Giustini D, Ayala AP, Blanchard L. HLWIKI International. URL: http://hlwiki.slais.ubc.ca/index.php/UBC_HealthLib-Wiki_-_A_Knowledge-Base_for_Health_Librarians [WebCite Cache ID 6GRHQW9qs]
247. WebHealth contributors. WebHealth. URL: http://webhealth.com/about/ [WebCite Cache ID 6GRHfkMZc]

248. Beller SE, Monatesti SJ. Quality through Knowledge: Understanding and Curing the Healthcare Crisis. 2013. URL: http://wellness.wikispaces.com [WebCite Cache ID 6GRHtRJOv]
249. WikiCancer contributors. 2008. Cancer: Stories, Support and Information URL: http://www.wikicancer.org [WebCite Cache ID 6GRMKbbWH]
250. WikiHealthCare contributors. WikiHealthCare. 2011. URL: http://wikihealthcare.jointcommission.org/bin/view/Home/WebHome [WebCite Cache ID 6GRMbLgpe]
251. Elsevier's WiserWiki contributors. 2013. Elsevier&#39;s WiserWiki Allows Physicians To Update Evidence-Based Medical Information With Experience-Based Practice Insights URL: http://www.elsevier.com/about/press-releases/health-sciences/elseviers-wiserwiki-allows-physicians-to-update-evidence-based-medical-information-with-experience-based-practice-insights [WebCite Cache ID 6GRMyao4k]
252. Pflege Wiki contributors. 2013. PflegeWiki URL: http://www.pflegewiki.de/wiki/Hauptseite [WebCite Cache ID 6GRN8EPDi]
253. Wikibooks contributors. 2012. Orthopaedic Surgery URL: http://en.wikibooks.org/wiki/Orthopaedic_Surgery [WebCite Cache ID 6GRNLmcCa]
254. Wikibooks contributors. 2012. Human Physiology URL: http://en.wikibooks.org/wiki/Human_Physiology [WebCite Cache ID 6GRNTJ6IG]
255. Wikibooks contributors. 2012. Pharmacology URL: http://en.wikibooks.org/wiki/Pharmacology [WebCite Cache ID 6GRNX7XBp]
256. Oncowiki contributors. 2013. Oncowiki URL: http://oncowiki.info/index.php?title=Main_Page.(Archived [accessed 2013-07-20] [WebCite Cache ID 6IG6pFP9E]
257. Wetpaint Wikis contributors. 2010. Dealing with Autism URL: http://4griffin.wetpaint.com/ [WebCite Cache ID 6GPlsYdAo]
258. MijnZorgnet. URL: https://www.mijnzorgnet.nl [WebCite Cache ID 6GPkwcTwb]
259. Ciesielka D. Meadville Collaborative Community Project.: PBworks; 2010. URL: http://614comm.pbworks.com [WebCite Cache ID 6GPmM06An]
260. WikiHealth contributors. 2013. WikiHealth URL: http://www.wikihealth.com [WebCite Cache ID 6GPmlDvVO]
261. Deshpande A, Khoja S, Lorca J, McKibbon A, Rizos C, Jadad A. Asynchronous telehealth: a scoping review of analytic studies. URL: http://wikisr.openmedicine.ca/index.php/Asynchronous_telehealth:_a_scoping_review_of_analytic_studies [accessed 2013-07-21] [WebCite Cache ID 6IGBfK3IT]
262. Radiologywikicontributors. Radiologywiki. 2013. URL: http://www.radiologywiki.org/w/index.php?title=Radiologywiki:Community_portal [accessed 2013-07-21] [WebCite Cache ID 6IHhdRhe2]
263. Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE. WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building. J Med Internet Res 2011;13(4):e108 [FREE Full text] [doi: 10.2196/jmir.1833] [Medline: 22155694]
264. Archambault PM, Blouin D, Poitras J, Fountain RM, Fleet R, Bilodeau A, et al. Emergency medicine residents' beliefs about contributing to a Google Docs presentation: a survey protocol. Inform Prim Care 2011;19(4):207-216. [Medline: 22828575]
265. Kent ML, Taylor M. Toward a dialogic theory of public relations. Public Relations Review 2002;28(1):21-37. [doi: 10.1080/10668920903527043]
266. Wenger E, McDermott RA, Snyder WM. Cultivating communities of practice: a guide to managing knowledge. Boston, Mass: Harvard Business School Press; 2002.
267. Maloney-Krichmar D, Preece J. A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM Trans. Comput.-Hum. Interact 2005 Jun 01;12(2):201-232. [doi: 10.1145/1067860.1067864]
268. Shneiderman BPC. Designing the user interface: strategies for effective human-computer interaction. Boston: Addison-Wesley; 2010.
269. Bishop J. Increasing participation in online communities: A framework for human–computer interaction. Computers in Human Behavior 2007 Jul;23(4):1881-1893. [doi: 10.1016/j.chb.2005.11.004]
270. Salmon G. E-tivities: the key to active online learning. London and New York: Kogan Page; 2002.
271. Knowlton DS. A taxonomy of learning through asynchronous discussion. Journal of Interactive Learning Research 2005;16(2):155-157. [Medline: 20736728]
272. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
273. Flavall E. PLOS Currents Influenza. 2011. URL: http://currents.plos.org/influenza/about/ [WebCite Cache ID 6GOXVBuVX]
274. Ghosh SS, Klein A, Avants B, Millman KJ. Learning from open source software projects to improve scientific review. Front Comput Neurosci 2012;6:18 [FREE Full text] [doi: 10.3389/fncom.2012.00018] [Medline: 22529798]
275. Karasavvidis I. Wiki uses in higher education: exploring barriers to successful implementation. Interactive Learning Environments 2010 Sep;18(3):219-231. [doi: 10.1080/10494820.2010.500514]

276. Varga-Atkins T, Prescott D, Dangerfield P. Cyber Behavior with Wikis. In: Encyclopedia of Cyber Behavior. Hershey PA: Information Science Reference (an imprint of IGI Global); 2012:164-177.
277. den Breejen EM, Nelen WL, Knijnenburg JM, Burgers JS, Hermens RP, Kremer JA. Feasibility of a wiki as a participatory tool for patients in clinical guideline development. J Med Internet Res 2012;14(5):e138 [FREE Full text] [doi: 10.2196/jmir.2080] [Medline: 23103790]
278. Davies E. Patients: the final frontier. BMJ 2013 Apr 02;346(apr02 2):f2075-f2075. [doi: 10.1136/bmj.f2075]
279. Tapscott D, Williams AD. Wikinomics: how mass collaboration changes everything. New York: Portfolio; 2006.
280. Berners-Lee T, Hendler J, Lassila O. The semantic web. Scientific american 2001;284(5):28-37.
281. Nielsen F. Wikipedia Research and Tools: Review and Comments. SSRN Journal 2012:-. [doi: 10.2139/ssrn.2129874]
282. Lau AS. Implementation of an onto-wiki toolkit using web services to improve the efficiency and effectiveness of medical ontology co-authoring and analysis. Inform Health Soc Care 2009 Jan;34(1):73-80. [doi: 10.1080/17538150902779543] [Medline: 19306201]
283. Technology. Policymaking, Regulation, & Strategy. URL: http://www.healthit.gov/policy-researchers-implementers/meaningful-use [WebCite Cache ID 6GOu0LfWc]
284. Lim W, So H, Tan S. eLearning 2.0 and new literacies: are social practices lagging behind? Interactive Learning Environments 2010 Sep;18(3):203-218. [doi: 10.1080/10494820.2010.500507]
285. Shirky C. Here comes everybody: the power of organizing without organizations. New York: Penguin Press; 2008.
286. Lykourentzou I, Papadaki K, Vergados DJ, Polemi D, Loumos V. CorpWiki: A self-regulating wiki to promote corporate collective intelligence through expert peer matching. Information Sciences 2010 Jan 02;180(1):18-38. [doi: 10.1016/j.ins.2009.08.003]
287. Furtado V, Ayres L, de Oliveira M, Vasconcelos E, Caminha C, D'Orleans J, et al. Collective intelligence in law enforcement – The WikiCrimes system. Information Sciences 2010 Jan 02;180(1):4-17. [doi: 10.1016/j.ins.2009.08.004]
288. Wikipedia's Manual of Style contributors. Wikipedia: The Free Encyclopedia. 2013. Wikipedia:Manual of Style/Medicine-related articles URL: http://en.wikipedia.org/wiki/Wikipedia:MEDMOS#Drugs.2C_medications_and_devices [accessed 2013-07-19] [WebCite Cache ID 6IEY6jWVV]
289. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
290. Lih A. Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource. 2004 Presented at: 5th International Symposium on Online Journalism; April 2004; University of Texas, Austin, USA p. 16-17.
291. Vroom RW, Vossen LE, Geers AM. Aspects to Motivate users of a Design Engineering Wiki to Share their Knowledge. World Academy of Science, Engineering and Technology 2009;57:169-179.
292. Cress U, Kimmerle J. A systemic and cognitive view on collaborative knowledge building with wikis. Computer Supported Learning 2008 Jan 10;3(2):105-122. [doi: 10.1007/s11412-007-9035-z]
293. Cheshire C, Antin J. The social psychological effects of feedback on the production of Internet information pools. Journal of Computer‐Mediated Communication 2008;13(3):705-727.
294. Cheshire C, Antin J. None of us is as lazy as all of us. Information, Communication & Society 2010 Jun;13(4):537-555. [doi: 10.1080/13691181003639858]
295. Yates D, Wagner C, Majchrzak A. Factors affecting shapers of organizational wikis. J Am Soc Inf Sci 2009:n/a-n/a. [doi: 10.1002/asi.21266]
296. Lih A. Wikipedia Revolution, The: How a Bunch of Nobodies Created the World's Greatest Encyclopedia. New York, NY: Hyperion; 2009.
297. Bryant SL, Forte A, Bruckman A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. 2005 Presented at: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work; November 6-9, 2005; Sanibel Island, FL.
298. Ciffolilli A. Phantom authority, self-selective recruitment and retention of members in virtual communities. First Monday 2003;8(12):1.
299. Seitz GM, Reger G. 'Wikipedia, the Free Encyclopedia' as a role model? Lessons for open innovation from an exploratory examination of the supposedly democratic-anarchic nature of Wikipedia. IJTM 2010;52(3/4):457. [doi: 10.1504/IJTM.2010.035985]
300. Yang H, Lai C. Motivations of Wikipedia content contributors. Computers in Human Behavior 2010 Nov;26(6):1377-1383. [doi: 10.1016/j.chb.2010.04.011]
301. Benkler Y, Nissenbaum H. Commons-based Peer Production and Virtue. J Political Philosophy 2006 Dec;14(4):394-419. [doi: 10.1111/j.1467-9760.2006.00235.x]
302. Restivo M, van de Rijt A. Experimental study of informal rewards in peer production. PLoS One 2012 Mar;7(3):e34358 [FREE Full text] [doi: 10.1371/journal.pone.0034358] [Medline: 22479610]
303. Kriplean T, Beschastnikh I, McDonald DW. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. 2008 Presented at: Proceedings of the 2008 ACM conference on Computer supported cooperative work; November 8-12, 2008; San Diego, CA.

304. Forte A, Bruckman A. Why do people write for Wikipedia? Incentives to contribute to open-content publishing. 2005. URL: http://jellis.org/work/group2005/papers/forteBruckmanIncentivesGroup.pdf
305. Butler D. Publish in Wikipedia or perish. Nature 2008 Dec 16:-. [doi: 10.1038/news.2008.1312]
306. George A. Avoiding Tragedy in the Wiki-Commons. SSRN Journal 2007:-. [doi: 10.2139/ssrn.975096]
307. Weinberger D. Everything is miscellaneous: the power of the new digital disorder. New York: Times Books; 2007.
308. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Evolution of Wenger's concept of community of practice. Implement Sci 2009;4:11 [FREE Full text] [doi: 10.1186/1748-5908-4-11] [Medline: 19250556]
309. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Use of communities of practice in business and health care sectors: a systematic review. Implement Sci 2009;4:27 [FREE Full text] [doi: 10.1186/1748-5908-4-27] [Medline: 19445723]
310. Yasseri T, Kornai A, Kertész J. A practical approach to language complexity: a Wikipedia case study. PLoS One 2012;7(11):e48386 [FREE Full text] [doi: 10.1371/journal.pone.0048386] [Medline: 23189130]
311. Brown T, Findlay M, von Dincklage J, Davidson W, Hill J, Isenring E, et al. Using a wiki platform to promote guidelines internationally and maintain their currency: evidence-based guidelines for the nutritional management of adult patients with head and neck cancer. J Hum Nutr Diet 2013 Apr;26(2):182-190. [doi: 10.1111/jhn.12036] [Medline: 23336961]
312. ResearchNet - Canadian Institutes of Health Research. 2012. Funding opportunity details URL: https://www.researchnet-recherchenet.ca/rnr16/vwOpprtntyDtls.do?prog=1425&view=currentOpps&org=CIHR&type=AND&resultCount=25&sort=program&all=1&masterList=true [WebCite Cache ID 65IoN4XLj]
313. Anderson PBJ, Bruell D, Rosenzweig M, Conte M, Song J. An online and social media training curricula to facilitate bench-to-bedside information transfer. 2009 Presented at: Positioning the Profession:the Tenth International Congress on Medical Librarianship; August 31-September 4; Brisbane Australia.
314. McKibbon KA, Lokker C, Keepanasseril A, Colquhoun H, Haynes RB, Wilczynski NL. WhatisKT wiki: a case study of a platform for knowledge translation terms and definitions--descriptive analysis. Implement Sci 2013;8:13 [FREE Full text] [doi: 10.1186/1748-5908-8-13] [Medline: 23347357]
315. Archambault PM. WikiBuild: a new application to support patient and health care professional involvement in the development of patient support tools. J Med Internet Res 2011;13(4):e114 [FREE Full text] [doi: 10.2196/jmir.1961] [Medline: 22155746]
316. Archambault PM, Bilodeau A, Gagnon MP, Aubin K, Lavoie A, Lapointe J, et al. Health care professionals' beliefs about using wiki-based reminders to promote best practices in trauma care. J Med Internet Res 2012;14(2):e49 [FREE Full text] [doi: 10.2196/jmir.1983] [Medline: 22515985]
317. Thomas GR, Eng L, de Wolff JF, Grover SC. An evaluation of Wikipedia as a resource for patient education in nephrology. Semin Dial 2013 Apr;26(2):159-163. [doi: 10.1111/sdi.12059] [Medline: 23432369]
318. Van de Belt TH, Berben SA, Samsom M, Engelen LJ, Schoonhoven L. Use of social media by Western European hospitals: longitudinal study. J Med Internet Res 2012;14(3):e61 [FREE Full text] [doi: 10.2196/jmir.1992] [Medline: 22549016]
319. Corneli J. GravPad. 2010 Presented at: Proceedings of the 6th International Symposium on Wikis and Open Collaboration; July 7-9, 2010; Gdańsk, Poland p. 33.
320. O'Grady L, Witteman H, Bender JL, Urowitz S, Wiljer D, Jadad AR. Measuring the impact of a moving target: towards a dynamic framework for evaluating collaborative adaptive interactive technologies. J Med Internet Res 2009;11(2):e20 [FREE Full text] [doi: 10.2196/jmir.1058] [Medline: 19632973]
321. Légaré F, Labrecque M, Cauchon M, Castel J, Turcotte S, Grimshaw J. Training family physicians in shared decision-making to reduce the overuse of antibiotics in acute respiratory infections: a cluster randomized trial. CMAJ 2012 Sep 18;184(13):E726-E734 [FREE Full text] [doi: 10.1503/cmaj.120568] [Medline: 22847969]
322. Moturu S. Quantifying the trustworthiness of user-generated social media content. Phoenix, Arizona: Arizona State University; 2009.
323. Arazy O, Stroulia E, Ruecker S, Arias C, Fiorentino C, Ganev V, et al. Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. J Am Soc Inf Sci 2010;61(6):1166-1179. [doi: 10.1002/asi.21326]
324. Contributors of Wikipedia: The Free Encyclopedia. 2013. Comparison of wiki software URL: http://en.wikipedia.org/wiki/Comparison_of_wiki_software [WebCite Cache ID 6GOcws35X]
325. Cosmocode. WikiMatrix: Compare them all. URL: http://www.wikimatrix.org/ [WebCite Cache ID 6GOdMHXAY]

Abbreviations

AFMC: Association of Faculties of Medicine of Canada
CWA: collaborative writing application
ICT: information and communication technologies
IMIA: International Medical Informatics Association
NPCF: Federation of Patients and Consumer Organization in the Netherlands
PAHO: Pan American Health Organization

Edited by H Potts; submitted 21.06.13; peer-reviewed by T Varga-Atkins, J Heilman, S Grover; comments to author 05.07.13; accepted 16.08.13; published 08.10.13.

Please cite as:
Archambault PM, van de Belt TH, Grajales III FJ, Faber MJ, Kuziemsky CE, Gagnon S, Bilodeau A, Rioux S, Nelen WLDM, Gagnon MP, Turgeon AF, Aubin K, Gold I, Poitras J, Eysenbach G, Kremer JAM, Légaré F
Wikis and Collaborative Writing Applications in Health Care: A Scoping Review
J Med Internet Res 2013;15(10):e210
URL: http://www.jmir.org/2013/10/e210/
doi:10.2196/jmir.2787
PMID:24103318

Viewpoint

Sociotechnical Challenges and Progress in Using Social Media for Health

Sean A Munson[1], BS, PhD; Hasan Cavusoglu[2], BS, MS, PhD; Larry Frisch[3], MD, MPH; Sidney Fels[4], PhD

[1]Department of Human Centered Design & Engineering, dub group, University of Washington, Seattle, WA, United States

[2]Sauder School of Business, University of British Columbia, Vancouver, BC, Canada

[3]Vancouver Coastal Health Research Institute, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[4]Media and Graphics Interdisciplinary Centre (MAGIC), Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Sean A Munson, BS, PhD
Department of Human Centered Design & Engineering
dub group
University of Washington
423B Sieg Hall, Box 352315
Seattle, WA, 98195
United States
Phone: 1 2062217699
Fax: 1 2065438858
Email: smunson@uw.edu

Abstract

Social media tools that connect patients, caregivers, and health providers offer great potential for helping people access health advice, receive and give social support, manage or cope with chronic conditions, and make day-to-day health decisions. These systems have seen widespread adoption, but often fail to support the goals as fully as designers and users would like. Through Ackerman’s lens of the “sociotechnical gap” and computer supported cooperative work (CSCW) as a science of the artificial, we review contemporary sociotechnical challenges and progress for using social media to support health. These challenges include a tension between privacy and sharing, policy information credibility, accessibility, and tailoring in social spaces. Those studying, building, deploying, and using social media systems to further health goals will benefit from approaching this work by borrowing from Ackerman’s framing of CSCW. In particular, this requires acknowledgment that technical systems will not fully meet our social goals, and then adopting design and educational approaches that are appropriate to fill this gap, building less-nuanced systems as partial solutions and tools for advancing our understanding, and by working with the CSCW research community to develop and pursue key lines of inquiry.

(J Med Internet Res 2013;15(10):e226) doi:10.2196/jmir.2792

KEYWORDS

social media; social computing; privacy; health; sociotechnical systems

Introduction

Advances in technologies that support cheap, ubiquitous sensing and sharing offer great promise for current and future health care. People can now objectively monitor their physical activity and sleep through mobile applications and devices. Mobile applications allow people to log their symptoms, activities, or consumption with relative ease. The basic sensors in mobile phones can support tracking and analysis of symptoms [1,2], and they can share the collected information with peers, their support network, and their health care providers. The last decade has also seen the arrival of “infodemiology” tools, such as Google Flu [3,4], that pool online behavior traces to monitor illness trends.

At a recent Peter Wall Institute for Advanced Studies workshop, our group was tasked with reflecting on contemporary and coming technical challenges for using social media to promote healthy behaviors, communicate health information, and to gather information on current health behaviors or events. We hope to see a continuation and extension of recent technical developments in sensing, connectivity, and large-scale data aggregation and analysis. There are clear areas for improvement—for example, activity inference can be unreliable and drains battery life, and Google Flu is still poor at detecting atypical flu trends, as the most severe often are [5]. We believe, however, that these challenges are being fairly well addressed by current research and market forces, and thus we do not dwell on them here.

Rather, we believe that many of the current grand challenges for the social Web and health, however, are not strictly *technical* challenges but *sociotechnical*. These challenges exist in the gaps between what people want and what is—or ever will be—technically possible [6] or in the complex interactions that emerge between individuals, groups, and technical systems.

In this paper, we provide a background on current trends in social media for health. We then describe one of these challenges: supporting an appropriate balance of privacy and sharing. Using Ackerman's framing of and guidance for CSCW as a science of the artificial [6], we review contemporary work to address this challenge. Before concluding, we highlight additional sociotechnical challenges that will need research attention before social media can better achieve its potential for supporting health information dissemination, sharing, and gathering.

Patient Support

Background

Health researchers have long known that patients receive key support from different people in their lives. Health care providers can provide expert advice and information, while peers can offer "strategies for coping with day-to-day personal health issues gained through trial and error of the lived experience" [7]. Peers are able to offer advice relevant to the health condition and health challenges, while friends' and family's long relationships with a patient make them better suited to offer advice relevant to the patient's personality and context [8] and to offer accountability in everyday life [9].

Increasingly, such support is offered through technology-mediated channels. These channels can allow people to reach each other at scale, to communicate more conveniently and on their own schedules, to reach other patients working with a rare condition, and to share with remote friends, family, providers, and peers. To frame our discussion of technical challenges, we briefly review examples of current research and practice in using technology to connect these different groups.

Patients and Health Care Experts

Health care experts—clinicians and others—are able to offer comprehensive, detailed medical information, delivered in a "prescriptive style and focused on explicit facts and opinions that tied closely to the health care delivery system, biomedical research, and health professionals' work" [10].

A number of mobile applications and ubiquitous health monitoring tools are being studied to help connect patient data to clinicians and to deliver time-sensitive advice from clinicians to the patient [11-13]. Companies such as Numera have sprung up to facilitate the connection between the myriad of consumer sensors and health providers' records systems. Health providers might review transmitted information on a regular basis by the care team, only when it exceeds some defined parameters or during a patient's office visit. Such connections can improve health outcomes (eg, [14]).

Patients and Patients (Peers)

Whether in face-to-face support groups or online interventions, peers can offer important support to people who are working through health issues. Peers who are going through—or who have been through—the same health challenges can draw on their own experiences to offer narratives, coping strategies, and support [8,10]. This shared experience not only makes their support highly relevant for the health challenge, but it also creates a sense of going through a challenge or "being in it together" for the recipients of the support [9].

Several systems help peers share physical activity-related data and have shown improvements in activity levels and retention rates over individual-use applications. For example, during an 8-week Internet-mediated physical activity program at the University of Michigan, participants were more likely to meet weekly physical activity goals if they joined a competitive team than if they participated as individuals [15]. In other studies, sharing physical activity levels, such as step counts, has helped to motivate people to be more active through social support and social pressure. In addition to providing users with individual feedback, the mobile phone application, Houston, facilitated the sharing of step counts and physical activity-related messages among a small group of friends [16]. Participants in the study's sharing condition were more likely to achieve their daily goals than participants without this feature. The Fish'n'Steps study found that sharing with strangers is not always motivating and is sometimes awkward [17]. Nevertheless, interacting with strangers can have benefits. In a 16-week Internet-mediated walking program, subjects with access to a discussion board had a 13% higher retention rate compared to a group without this feature; however, daily step count was not affected [18,19].

In addition to sharing their own data, symptoms, and activities, patients can share their personal trajectory with an illness or medication adherence, as well as experiences with different strategies, medications, and procedures. In some cases, these data can also be used to identify adverse events or poor quality health care (perceived or actual) [20]. These accounts provide both useful information that other patients can use to make decisions, but also provide people with a sense that they are not alone. This can be particularly important in rural areas or for individuals with rare conditions, when there are no physically proximate peers [9]. Peer support can also be enhanced by focusing on connecting a group of people from the same geographic area [21,22] or who share the same health provider [23,24]. When peers share context and constraints, they can offer advice and narratives relevant to the specific health concern and that are more likely to fit into each other's lives.

Patients and Caregivers (Friends and Family)

Technology can also support connections between people and the people in their existing support network who help them manage illness-related challenges, receive emotional support, or help them adopt a new health habit. When a protracted illness or other major health event strikes a patient or their family, friends and extended family often want to pitch in to help with day-to-day tasks. Websites such as CaringBridge and CareCalendar can help patients and families solicit and coordinate that help. Many popular fitness applications, including Daily Mile, RunKeeper, Nike+, Adidas miCoach, FitBit, and LoseIt, also connect users to their existing social networks, including friends and family, on sites such as Twitter and Facebook. These applications typically generate suggested posts and associated data, such as maps of runs or calories burned that users can share as status updates.

Sharing on Facebook can reach friends and family whose opinions matter but who may not be participating in the wellness activity themselves, potentially creating an additional channel for receiving social support and pressure beyond what is available when sharing only with other users of the application [8,9,25,26]. While peers can offer a sense of "going through it together" or advice from their own experiences dealing with a health goal or medical condition, friends and family can offer different support. They know the individual and can give advice that is relevant to their context, and they may be in a better position to understand what sort of support or pressure an individual would benefit from hearing [8]. For health goals related to one's identity and impression management (eg, feeling and being perceived as fit), the opportunity to communicate that identity to friends, family, and even former acquaintances can be an important motivator [9]. Through ongoing relationships in other aspects of life, they can offer accountability and social pressure—even, or especially, when someone stops participating in a health intervention [9].

Health Care Providers and Experts With Other Health Care Providers and Experts

There are also online communities to connect health professionals with each other. For example, the online health community Sermo restricts access to verified MDs and DOs in the United States and has over 125,000 members. In such communities, physicians can share and access recent news and research articles. Members can informally report and share observations or solicit feedback from others through threaded discussion or surveys [27,28].

Spaces

This communication exists and flows across a variety of technology-mediated spaces. It occurs in electronic medical records between caregivers, and now on data that can be inserted from consumer devices. It occurs on social network sites, as posts directly from users or as posts from quantified self-tools like FitBit or RunKeeper, and in all manner of online communities created to support interactions among a single group (physicians, caregivers, patients) or across groups.

This communication, along with other online traces such as search queries and news articles, can also be mined for other purposes. MITRE's MiTAP system monitored newsgroups to detect disease outbreaks such as SARS [29] and to get critical information to medical experts and those involved in relief work [30]. The patient support community PatientsLikeMe aggregates and sells de-identified data to its business partners.

Challenges in Social Media for Health

What issues emerge when we combine these relationship types, spaces, and technical systems? What do we know about how to address them, and where do gaps emerge? While many of the potential benefits of communicating about health through these channels and on these spaces are being achieved even now, they come with costs, barriers, and new challenges. These include privacy and sharing tensions, policy issues, accessibility, and even such fundamentals as the working definition of wellness or what it is to be healthy.

To analyze ways of understanding and designing social media for health, we borrow heavily from Computer Supported Cooperative Work (CSCW). Though CSCW has its origins in workplace and educational settings, many of its primary concerns—adoption and appropriation, designing for groups who have different goals, perspectives and experiences, and remote interaction with varying levels of synchronicity and aggregation—are shared with social media. The value of applying perspectives from CSCW to social media and social computing has not gone unnoticed; the CSCW conference is in the process of being rebranded as a conference on computer supported cooperative work and social computing. This perspective also reminds us that technical progress that is missing a better understanding of people's needs and interactions—with each other and with systems—may not be progress overall.

In the remainder of this paper, we review one of these challenges in some depth—the tension between sharing and privacy in meeting health needs. We show how the challenge emerges from gaps between what designers and users would ideally like in a technical system and what is currently—or will likely ever be—possible, which is what Ackerman terms the "sociotechnical gap". This particular lens has previously been used to examine issues such as electronic voting, how the public perceives risks associated with information technologies, systems and practices to support decision-making, how construction workers adopt and use mobile communication tools, and how people make decisions about managing privacy and communicating identity using information technology. Using Ackerman's proposed ways of moving forward on such challenges, we review current work to improve how people can manage their privacy when sharing to support health. We then briefly highlight additional key challenges of the sociotechnical gap in in social media for health. We identify consistent themes across these challenges and suggest ways forward.

Case Study: Privacy and Sharing

Background

One's health information is often seen as particularly sensitive [31] and often receives unique legal protections. Many patients or caregivers need to share this sensitive information in order to meet their health goals: only by revealing information about their health challenges and about their personal situation can they receive relevant advice and support [8,9,32].

While previous work on social media and health has argued for the potential benefits of using social media to support health goals, for reasons introduced earlier, it has also identified several obstacles and challenges. These include risks associated with others misappropriating or misunderstanding shared information, risks with violating social norms of sharing, and risks of not sharing with the right people to receive the desired type and quality of support. Before reviewing these challenges, we introduce different models and concepts in privacy.

Nissenbaum describes privacy as "contextual integrity" [33]. Nissenbaum notes that all spaces have associated norms about what is and is not appropriate for the information in those spaces. These norms describe both the information that is appropriate for that space and in what ways that information or may or not be reshared, remembered, or further disseminated. Thus, it addresses what an individual discloses, what information is collected about them without their disclosure, and how that information may be used. Privacy violations, then, occur when information is shared or collected that is not appropriate for the given context, or when it is stored or shared (or not stored or shared) in a way that people would not expect for that context. Adams and Sasse propose a model of privacy violations [34] that is largely congruent with Nissenbaum's definition. In their model, individuals have assumptions about information's sensitivity, how it will be used, and who will receive it. When those assumptions turn out to be inaccurate, a privacy violation occurs.

Another model of sharing and disclosure decisions focuses on an individual's privacy concern—their perceived risks and threats—versus the perceived benefits of sharing [35]. More concern reduces their attitude toward sharing and thus their intent to share. This is consistent with major theories in health behavior change, such as the Theory of Planned Behavior [36] and Theory of Reasoned Action [37]. Privacy concerns can be caused by the lack of knowledge about whether the data collected are essential or needed [38], whether the collection process is perceived as fair [39], how the data will be used and disclosed [40], whether the information will be accessed by unauthorized individuals or organizations [38] or used according to its original intent [38], whether the information collected may be subject to deliberate or accidental disclosure errors and whether measures exist to limit such possibilities [35], and the identity of the owner of the data [41].

We next review some specific examples of how norms and system design can influence the choices that individuals make when using social media for health and how these choices can lead to privacy violations or concerns.

The Role of Norms

The norms of any given context can describe both what others share and how they react to what is shared and also more prescriptive information, such as what one should share [36,42]. The norms of spaces can both stifle communications that would be beneficial and encourage sharing that individuals later regret. For example, people may feel uncomfortable asking for health support or sharing successes that potentially appear boastful on general social network sites such as Facebook or Twitter. An individual may also cause a privacy violation by revealing information about themselves that is overly sensitive, and thus inappropriate, for a particular context, making *others* feel uncomfortable. Privacy violations occur not only when sensitive information about oneself is shared or remembered against one's desires, but also when one shares information about oneself against others' desires.

For example, many health applications support regular sharing of physical activity or other health data with one's social network. This information, in this quantity, may not always be appropriate for such spaces. Study participants report concerns about boring their friends with mundane posts or appearing boastful about modest achievements [9,26,43]. A typical Facebook network contains a diverse range of ties [44], and it may not be appropriate to share health-related data with one's entire network, even when a patient would benefit from sharing with a subset of their network.

While this form of violation is common knowledge—the abbreviation "TMI" (too much information) has been coined to describe it—it is often not part of the frame in which system designers and builders approach systems for health information. Coffield and Joiner [45] make a similar observation about this form of violation. They argue that "many people lack common sense about the extent of information that is appropriate to put online". We disagree—for many people to be struggling with this challenge suggests that there is not yet *common* sense to be had. The focus, rather, is often on privacy. One of the authors describes designing health and behavior applications with features to share with one's social network [46]. The team originally designed the applications' interfaces with privacy framed as protecting individuals against disclosing information that they found too sensitive, but not against disclosing information that others found contextually inappropriate. We propose that an alternative perspective, balancing privacy and "appropriate self-casting" might better serve designers and users' needs.

Other norms—real or implied by a system's design—can also have unintended consequences if they induce people to disclose information that is later repurposed by other parties. Each disclosure has associated privacy risks, and some even damage insurance or employment prospects. For example, many people post vacation photos to social sites such as Facebook or Flickr. In one high profile example, however, a woman on sick leave for depression had her insurance benefits cut after the insurer discovered Facebook photos of her on vacation with family and in a bar with friends [47]. Such well-publicized situations where postings led to humiliation, loss of jobs, or loss of insurance coverage have not caused social media enthusiasts to be more reticent, though even many recognize the importance of assuring their own privacy [48].

Even when disclosures of health information in social media channels do not cause others to cringe at their inappropriateness and do not increase the sharer's risk of insurance or employment consequences, such sharing may not get the desired reactions from others. This can occur because others do not know how they are supposed to respond, or because the norms of the space encourage a snarky response rather than the hoped-for, supportive response. Friends and family are also often quite hesitant to hold individuals accountable to health goals unless they have specifically been asked to do so [26].

The Role of Systems Design, Framing, and Defaults

The attributes of system design and perceptions of a system's value can also predict sharing behavior [49-51]. Individuals who report privacy concerns frequently engage in activities that jeopardize the privacy of their personal information [52]. The term "privacy paradox" describes the phenomenon of individuals sharing more information than their privacy positions [53,54]. The privacy paradox is attributed to immediate gratification, bounded rationality, psychological distortion and limited information [53,55], and the value of sharing [56]. In the following paragraphs, we summarize some research that extends our understanding of biases in privacy-related decision making, including the effects of defaults, users' perceived value, perceived control, requested permissions, and framing effects.

The privacy default and suggestions built into a system's interface can be incredibly powerful [49,57,58]. Despite preferences expressed in interviews or surveys, users often share health information according to the default setting [26]. In a study showing the importance of framing and defaults, each participant was asked to select a set of friends with whom they would (or would not) share some personal information [59]. When individuals were asked with whom they would not share (ie, the default was to share with everyone), they shared twice as much as when asked with whom they would share (ie, the default was to share with no one).

Even when flexible privacy controls are available, it is a challenge to help people configure their sharing settings appropriately. Munson and Consolvo designed GoalPost, a physical activity goal-setting and self-monitoring application, to include the ability to configure a "support group" of people with whom to share physical activity goals and progress. Only 25% of participants with access to this feature used it [43]. Industry experts report that privacy controls, when present, often go underutilized. For many, the overhead of configuration may not be worth the extra overhead, or they regard the default as an expression of a norm. Sharing defaults that are not well matched to a space and type of information can lead people to inadvertent privacy violations. For example, the mere presence of affordances such as "share" buttons in fitness can encourage people to do so even when such posts may be off-putting to their friends [46].

Bulgurcu et al investigated an individual's intention to use third-party applications that request access to his or her information on a social network platform [60]. Not surprisingly, the user's perception of an application's value was correlated with their information sharing. There was also an interaction effect between perceived privacy risks and the perceived value of the application—the higher the application's perceived value, the less the perceived privacy risks would affect subjects' sharing behavior. Thus, applications that oversell or overstate their potential benefits, or the value of sharing, can lead individuals to share more than they otherwise would. It is plausible that organizations that stand to benefit from obtaining information about patients, or from patients sharing their use of the organization's product, might misrepresent the benefits of their offerings to bias patients to divulge more. This caution is somewhat balanced by the same research team's investigation of the influence on privacy controls given to, and the permissions requested on, individuals' perceptions of an application's benefit [61]. Requests for more permissions reduced users' perceived value of the system, even when the users were given control over which permissions to grant.

These studies exemplify a key principle of Thaler and Sunstein's recent work, Nudge [62]. Through selection of defaults, by making some actions easier or more available than other actions, or through the particular framing of a choice or decision, all spaces will exert influences on the choices that people make in those spaces. There is no such thing as a neutral choice environment. Thus, it behooves designers to be aware of and carefully consider how the design decisions they make will influence users' choices.

Data Aggregation

Beyond individual acts of sharing and viewing shared data, repurposing aggregate data is also fraught with potential privacy violations. Ideally, de-identified data might be shared broadly with researchers and practitioners who seek to build the next generation of tools like Google Flu. Unfortunately, tools and strategies for re-identifying de-identified data are keeping pace with efforts to make such datasets available anonymously. Some individuals within the "anonymous" search query dataset that AOL released to support academic research were quickly identified [63], Netflix had to cancel their second Netflix prize because they could not assure the anonymity of the users' [64], and multiple public records datasets can be combined to identify mothers' maiden names [65] or predict social security numbers [66].

Unresolved issues around sharing and privacy cause problems for users and designers of systems. When people share too much or too broadly, they expose themselves to risk of others using the information in ways that are harmful to the patient or they risk being perceived as boring or an "oversharer". When they share too little, they may underprovision a social media space with the information that would help them or others meet their health goals. Finally, when they share in an inappropriate channel, they risk both: others misappropriating the data that they do share or social sanction for what they have shared, while also not receiving the health support they might have received in another space.

So What Do We Do?

Seeking Solutions

What can designers and builders do to help people share to support their health goals while reducing potential privacy violations? First, we must acknowledge that perfect privacy and sharing is not going to happen. It is *the* classic example of a sociotechnical gap. Our technological systems cannot fully support users' desires [6]. Finer-grained privacy and sharing controls make for greater configuration challenges; even users confident that they understand controls can make sharing errors in such systems [10]. On the other hand, automated- (or administrator-) configured sharing raises the risk for the technology to introduce mistakes. As Nissenbaum notes, privacy is one of the "enduring social issues associated with information technologies" [33]. With health information, the sharing and privacy needs may be more complex, and the stakes higher, compared to many other types of information. Privacy challenges will endure when trying to use social media to support health goals.

Despite the unobtainability of perfection, researchers and practitioners should continue to seek better solutions. We believe that the broad approaches Ackerman describes for building CSCW as a science of the artificial and working with the sociotechnical gap—palliatives, first order approximations, and fundamental lines of inquiry—can suggest pathways for better handling privacy in social media for health. In most cases, we can borrow from or build on fundamental human computer interaction and CSCW work.

Palliatives

Ackerman noted that ideological, political, and educational efforts were being used to alleviate the sociotechnical gap. Techniques such as stakeholder analyses and participatory design had the value of involving relevant parties to openly produce systems with known characteristics. Through such openness, people can make more informed choices or potentially stop the implementation of systems with particularly problematic consequences. Such approaches are alive and well among HCI and CSCW researchers working on social media for health (eg, [8,67,68]). This is heartening. These approaches can surface and make salient the relevant norms of a space or of the stakeholders as part of the design process. They enable the design of systems that are more responsive to the people and organizations they affect, and with greater awareness of the trade-offs inherent in any system. The application of these methods may be one of the primary contributions that human computer interaction can bring to health communications work overall, whether or not such work is focused on technological artifacts or other forms of sharing.

Educational initiatives, particularly those that inform systems builders and designers, will also prove important. Systems built and released with one set of goals will have further consequences on their users and the organizations in which they are deployed, whether it is by making some choices and actions easier than others or simply through disrupting the existing workflows involving people and artifacts. When systems builders are aware of and attend to these potential effects and the sociotechnical gap, they will hopefully avoid overconfidence that building to the "right" specification can neatly meet any intended goal.

There is also likely a need to better inform users' mental models of how social media systems for health function, and what they do and do not do. For example, the GoalPost system [43] let people share physical activity goals and progress with their Facebook networks. Many study participants were excited, hoping that this would help them get valuable support and accountability from their social networks. When their posts received relatively few likes and comments, however, they reported becoming discouraged or disappointed in their friends. Here, a barrier is the gap between individuals' mental models of how the Facebook feed works (all friends see all of your posts) and how it actually works (some friends may see each post). This misunderstanding can cause them to perceive a lack of comments or likes as being ignored by friends in a time of need rather than a result of their posts just not being seen. Better transparency and understanding of how the feed works might have helped users have more realistic expectations about how many people would see and react to their posts.

First Order Approximations

The second way forward is building first-order approximations: "tractable solutions that partially solve specific problems with known trade-offs" [6]. For Ackerman, these solutions are important tools for exploring the design space of what is possible and for supporting a more detailed understanding of the sociotechnical gap. While such approximations can certainly support these science goals in social media for health, they may also be valuable solutions in and of themselves.

For example, people may want to be able to seamlessly manage all of their different connections for meeting their health goals, with nuanced and well-chosen permissions and disclosure for each piece of data and each relationship. Such a system, however, is not likely to be forthcoming. Instead, people are already using separate, less nuanced channels and spaces for different purposes, even if managing accounts, identity, communication, and relationships across these systems adds overhead. Each space can have its own norms for information sharing, remembering, and dissemination, without the burden of supporting the nuances of a whole range of spaces. The development of these spaces also allows individuals to have "front stage" and "backstage" spaces, which are important for successful impression management [69], including health goals [9]. In front stage social media spaces, such as Twitter or Facebook, individuals can communicate their health successes with friends and family, or give them brief status updates. In backstage spaces, they can let their guard down and reveal weaknesses and struggles so that they can get support and advice from peers or experts. Considered alongside Nissenbaum's definition of privacy and the importance of context, we can see how creating separate spaces, each with their own context, may actually be better than the "convenient" ideal of an integrated, nuanced space to meet all goals.

A related challenge, though, is making each new space sufficiently valuable—especially at first—that individuals will make visits to it part of their routine or tolerate its pushing content into their other channels, such as via push notifications or email. Here, bootstrapping the space with expert or informational content or discussion prompts may help [18].

Designers may also seek to help people better build and shape their networks. Within a peer support community, for example, patients may benefit from being able to identify others with similar circumstances—for example, those who live in the same type of area or who have the same family situation or financial resources—in order to be able to both get and give more contextually relevant advice. Within their own social network, they may benefit from tools that can identify others facing the same health challenges in appropriate ways. Social matching systems have received some attention in the HCI and CSCW communities [70], and determining the most salient issues for applying such systems to health is likely to be a beneficial first-order approximation [71].

Even with separate spaces or channels for meeting different health needs, some spillover will occur. People facing major health events may need to share news and updates with their extended networks, and Facebook or similar tools are key spaces for sharing, even if such updates are not entirely consistent with the normative content for such spaces. Can designers build tools that better select who in one's network will see such updates? Can we design systems that give feedback that helps people craft messages that are more appropriate for the selected channel, and can they help people give helpful responses to a post?

Hansen and Johnson offer one approach that repurposes an existing and broad social channel—Facebook—to deliver sensitive health information [72]. They work with an HPV educational application, called Fact Check: HPV. This application pertains to a stigmatized illness and one that may be contracted through one's social network (sexual partners). They believed that letting people send the application to friends through a semi-anonymous ("veiled") channel (one of your friends—but not which friend—invited you to this application) might make people more willing to invite friends, including past sexual partners, and might make recipients more motivated to access the application. The application's users used both veiled and non-anonymous notifications (1:2 ratio); recipients of veiled invitations were more than five times as likely to access the application.

Fundamental Lines of Inquiry

The final and "most daunting" challenge posed by Ackerman is a set of fundamental inquiries that would further CSCW as a science of the artificial. Work on many of these inquires, such as an understanding of when systems can ignore the need for context, will also advance designers', deployers', and users' abilities to manage privacy and sharing when using social media systems to support health needs. Because these questions are cross-cutting, however, we will return to the question of fundamental inquiries in our discussion.

Summary

Though perfect privacy is an unreasonable goal, technologists and designers are making progress on designing applications and interfaces that help people to better balance their privacy and sharing while meeting health goals. More work in this vein is necessary, as well as work that will address new privacy and sharing challenges that will emerge as people design and adopt new social channels, spaces, and capabilities to support their health needs.

Further Contemporary Sociotechnical Challenges

Policy

While privacy and sharing are the canonical challenge, they are not the only sociotechnical challenge with using social media to support health.

An unfortunate constraint of current health care policy is that it was not written for, or during, the current era of mobile health, electronic health, and social media for health. It is not up to today's challenges and capabilities let alone tomorrow's, and regulatory uncertainties often push health providers to take the most conservative stance with respect to social media.

In the United States, there are many questions about how and when Health Insurance Portability and Accountability Act (HIPAA) applies to social media. Coffield and Joiner highlight several examples in which health professionals posted information about a patient to a social network site, leaving them in a legal grey area and in trouble at work [45]. Sidorov notes that the fit between social media and health care's regulator environment remains unknown and unclear, and that HIPAA's requirements for patient privacy make it difficult for health providers to host participatory communities [73]. There is similar uncertainty about liability when health professionals tweet their expertise or reply informally to an online question about symptoms [45].

The need to meet other HIPAA requirements—such as that information used to make medical decisions be archived—pushes designers and administrators of communication systems to more controlled systems. This can limit their ability to take advantage of a broader ecosystem of tools that may better integrate with patients' lives [74].

In the long term, one can hope for policy reform that better enables health innovation, rather than stifles it or leaves it to those who are willing to take risks and work right up to policy boundaries. There is, though, the risk that policy makers who do not understand the sociotechnical gap will craft policies intended to enable but that impose requirements that cannot technologically be met, and thus further suppress development of systems that are imperfect but would solve real needs. Education of future policy makers and participation in the policy-making process will be essential.

In the short term, palliatives such as better education for health providers about what they can and cannot legally do may help. As of 2010, only 10% of US medical schools had policies or guidelines on social media use, leaving students to navigate its advantages, costs, and limits largely on their own [75].

Information Credibility

Online spaces also create new or expanded challenges for information credibility. This is not a new challenge—there have long been snake oil salesmen and old wives' tales—but new spaces and channels do create new opportunities for incorrect or unverified information to spread, either intentionally or unintentionally.

If the Web is to be used for communication between physicians and patients, it may seem prudent to ensure that someone offering diagnostic or therapeutic advice is truly credentialed to provide these services or expertise. Indeed, some communities have found it beneficial to close their doors to individuals without credentials. Sermo requires new users to verify their credentials as physicians and then lets them post with their real-world identity or anonymously—but readers know that even anonymous posts are coming from credentialed experts.

Requiring credentials, however, is not appropriate for all situations. Such restrictive limits would limit access to peer expertise and support. While such stories may not be rooted in evidenced-based medicine, they are based in lived experiences, and, if taken with appropriate levels of trust, can prove invaluable for both their informational and emotional support. A community designer might be tempted to try to verify that a participant is, indeed, someone who has had to face the medical situation at hand (or a caregiver for someone who has), but such verification is impractical, if not impossible. Overly burdensome verification requirements would stifle contributions to social media spaces: as the cost to contribute goes up, the contributions go down.

Others have argued for online activity to be connected to real-world identity, allowing better evaluation of its credibility and reductions in spiteful remarks made behind a veil of anonymity [76]. We do not, however, believe that this is the right approach for many peer health sites; for people with potentially stigmatizing conditions, anonymity can leave them free to ask questions and seek the help they want.

Instead, designers can build either formal or informal reputation systems [77]. These systems can help surface participation from people whose past posts have proven particularly valuable to the community. They also can give people ways to build up profiles that suggest that they are credible individuals. Unfortunately, once such a system begins to be used broadly, others will have an interest in attacking it to bolster their own reputation.

Researchers are also building first-order approximations that help us understand how spaces can support free participation complemented with material that is known to be credible. For example, Huh et al have been developing an online space that supports peer participation and discussion, with all of the associated potential inaccuracies, while automatically augmenting it with credible information vetted or prepared by experts [78]. Such balanced approaches facilitate patient participation and support, while supplementing patient expertise with health provider expertise.

Accessibility, Exclusion, and Literacy

Leonard Kirsch has described patient engagement as the "blockbuster drug of the century" [79]. Enabling and supporting this engagement at scale, during office visits and in between, in an affordable manner remains difficult. The connections that social media can create between peers, caregivers, and experts may be one way to achieve this goal. Reliance on social media, however, should raise some important questions about access and inclusion. For example, in the United States, patients with chronic illnesses are less likely than others to see health information online (51% vs 66%) [80]. This gap, however, occurs not because they would benefit less from online resources (the number of online communities to support chronic illnesses would suggest otherwise) or because they are unmotivated to see out this information, but because people with chronic illness are less likely to have Internet access at all. 62% of adult Americans with chronic illnesses have Internet access, compared to 81% of those not managing chronic illnesses. If social media is to be a major tool for helping people manage health, then there is a need to ensure that such tools are accessible to all individuals and/or to design other programs to reach those who do not have Internet access.

To address this challenge, we focus on palliatives, including political and ideological stances that advocate for inclusion and for honesty about who may be excluded by a particular solution. Educational efforts to reach out to users to set reasonable expectations for the benefits they can and cannot achieve from a given system or set of systems and to help them best use (or not use) the available tools can also further increase access.

Appropriately Accommodating Different Definitions of Wellness

The term "wellness" is so much a part of our thinking about health and health care that it is easy to forget how relatively recently wellness has come into common use as a health-related idea. From its first recorded written use in the 17th century, *wellness* was most commonly used as an antonym for *illness*, whereas today, it is generally taken to represent a state of healthy that is viewed quite apart from sickness. This transition from antonym to a distinct state of health may have occurred in large part as a result of the Peckham experiment in 20th century England [81]. From about 1926 through 1950, staff at the Pioneer Health Centre in London observed and treated families in a way we would today describe as "holistic". From their observations, they drew four major conclusions:

1. Health is a process that has to be cultivated if it is to thrive.
2. If people are given information about themselves and their families, they will attempt to make decisions that are in the best interests of their families.
3. People thrive when they are given the freedom to make choices about their activities and will choose those that help in their development.
4. When people are given resources in a community to enable them to grow, they will be active in their community for the benefit of that community.

Or, as one of the original Peckham doctors wrote, "Given the opportunity, people can be drawn into a more active lifestyle and greater enjoyment with neighbours" [82]. While these ideas likely seem self-evident today, they appeared quite radical when they were first proposed.

The Pioneer Health Centre closed in 1950, but not long after, the Massachusetts Framingham Heart Study brought the concept of "risk factors" into common usage and identified the heart disease risk factors with which we are familiar today: high blood pressure, smoking, and elevated cholesterol. From these two seeds, “active lifestyle … enjoyment with neighbors” and “risk factors”, the current pervasive wellness movement was born [83]. While the Peckham group focused on families and the healthy development of children, the wellness movement has been more directed to the individual, emphasizing the importance of lifestyle choices that can lead either to illness or good health. In the 1970s, wellness took on many aspects of a quasi-religious movement. For some, it was sufficient to practice a seemingly evidence-based regimen of hygiene, non-smoking, exercise, low animal fat diet, and moderate or no alcohol intake. For others, wellness required adherence to spiritual exercises and/or to strict dietary regimens. We now know that while individual choices do influence health, environmental determinants including social connectedness and satisfying family life are of at least equal importance. More recently, we have come to recognize not only that living in close connection with nature is health-enhancing, but also that many aspects of our current “built environment” have strong adverse effects on our health [84,85].

One challenge facing designers and users of social media to promote health is how to promote the “right” goals. Individuals and social groups may define what it is to be healthy or well differently (for an extreme case, consider pro-anorexia online communities). While some exposure to alternative definitions of wellness, through online social interactions, may be beneficial, large differences in definitions of health may make it hard to reach target populations. Overly strong social rewards (eg, status) or sanctions (eg, stigmatization) may be coercive. Designing a technology-mediated social space that will always offer the optimal support and accountability for appropriate health and wellness goals is unreasonable, though designers of health systems should be sensitive to this issue.

Further increases in measuring, no matter who defines the goal or optimal, may also be harmful. There is an old business adage that “you get what you measure”. As tools for quantitatively tracking health outcomes and health behaviors become increasingly prevalent, and along with explicit and implicit persuasion to optimize those measures, that progress toward more holistic definitions of wellness will be lost. Critics argue that this may already be happening. Purpura et al describe a hypothetical system, Fit4Life, which persistently monitors an individuals’ diet and physical activity and gives feedback as well as shares progress (or lack thereof) with one’s social network [86]. Though Fit4Life includes some technologies that do not yet exist, the overall system and capabilities are not large leaps beyond products currently on the market. While such a system might help someone achieve greater physical health, it is hard to imagine users of such a system feeling more well overall. Indeed, the picture painted is something of a dystopian future.

Here, we again look to palliatives like participatory design and user-centered design to help understand what it means for a given individual or group to be well and for honesty about how a given solution may or may not support that definition. We are also particularly excited by work to build related first-order systems that explore supporting multiple concepts of wellness that emerge from a social group. For example, researchers at Cornell have developed both Vera, a system for new mothers, and Vera+, for a general audience, to support healthy decision-making through open-ended social awareness [87].

Tailoring in Social Spaces

Tailoring health messages has shown promise for increasing individuals’ likelihood of attending to and complying with them [88]. Other work suggests that individual’s responses to different, interactive health behavior change applications, such as those to promote fitness, may be predicted by personality traits [89]. Application features can have different effects based on an individual’s personality [90].

There will likely then be benefits to figuring out how to adapt systems to users’ personalities. This includes automatically reconfiguring the interface or showing different content, as well as sensing (or otherwise collecting) information about users that can be used to inform that tailoring. These problems are not only technical though. For social spaces, people expect a certain shared experience with the other users. Tailoring and personalization of social spaces then may be at odds with this shared experience. Should designers dump different types of people into separate social worlds or applications, such that they experience the best strategy for them but at the cost of diversity and having the broadest possible cross-section of peer expertise? If so, are there ways to identity the best information across the different applications and make sure that all can benefit?

Additionally, there is a danger that people may not choose the system that best meets their health needs. They may, for example, pick applications and health support systems that make them feel the best about the actions they are already taking, and not those that encourage them to make harder choices. If the applications that individuals would choose are, in fact, not the ones that are best suited to helping them make healthy decisions, what can designers of application markets or health experts who suggest applications do about this? What should they ethically do?

Conclusion and the Path Forward

Through several examples, we have demonstrated the importance of attention to the gap between our desires for social media systems to support health and the systems that we can actually build. A failure to acknowledge this gap and account for it in our processes of design, deployment, and evaluation will lead to failures of adoption; to violations of norms, privacy, and users’ expectations; and mismatches between the goals, activities, and tools that systems suggest (or even coerce) and what would be best for individuals or groups.

There is no silver bullet for closing this gap. It is a nuanced and challenging set of problems that we cannot engineer or build our way out of. Instead, social media for the health community of researchers and practitioners must continue to bring together teams representing health experts, those with expertise in human-computer interaction and CSCW, and other stakeholders. From CSCW, we must borrow the palliatives that can improve the gap and know when and how it exists. We must build and study the first-order approximations that will help us better understand the boundaries of the gap and that may serve quite well as partial solutions.

Our community must also work to adopt and incorporate the lessons learned from CSCW's fundamental research on sociotechnical systems and to regularly pull new findings and knowledge into the space of building social systems that support health. But we must also identify cross-cutting issues and advocate for study of major themes in health that may not otherwise receive attention from CSCW. Based on the contemporary challenges outlined in this paper, we would add some key questions to the lines of inquiry identified in the original work:

- How can systems balance the competing goals of experts and users, particularly when one dimension may be easier to measure than others? (Here, we suggest that studies of group decision support systems (eg, [91]) have already characterized much of the problem space.)
- Relatedly, what design and deployment processes can help us negotiate issues of individual autonomy and nudging, persuading, or even coercing people toward the actions that experts believe they should take? How do we train system designers and builders to consider the influences they unintentionally create in their systems?
- When is an ecosystem of tools better than attempting to build an integrated tool? When an ecosystem of tools exists, how can systems or other processes guide individuals to the right tool or tools to support their goal (or subgoal)?

Ackerman's description of the sociotechnical gap and of CSCW as a science of the artificial characterizes many of the challenges and predicts many of the failures that we face in designing and building social media systems to support health. Fortunately, it also offers a way forward.

Acknowledgments

The authors wish to thank the participants of the Peter Wall Institute for Advanced Studies Workshop on Social Media for Health for their perspectives and comments, which shaped the direction of this manuscript. We also acknowledge support from the Peter Wall Institute for Advanced Studies for convening this workshop and covering the submission and publication fees.

Conflicts of Interest

None declared.

References

1. Larson EC, Goel M M, Boriello G, Heltshe S, Rosenfeld M, Patel SN. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In: Proceedings of the International Conference on Ubiquitous Computing. New York, New York: ACM Press; 2012 Presented at: UbiComp 2012; 5-8 September 2012; Pittsburgh, Pennsylvania p. 280-289. [doi: 10.1145/2370216.2370261]
2. Larson EC, TienJui L, Liu S, Rosenfeld M, Patel SN. Accurate and Privacy Preserving Cough Sensing using Low-cost Microphone. In: Proceedings of the International Conference on Ubiquitous Computing. New York, New York: ACM Press; 2011 Presented at: UbiComp 2011; 17-21 September 2011; Beijing, China p. 375-384. [doi: 10.1145/2030112.2030163]
3. 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]
4. Watts G. Google watches over flu. BMJ 2008;337:a3076. [Medline: 19118085]
5. Butler D. When Google got flu wrong: US outbreak foxes a leading web-based method for tracking seasonal flu. Nature 2013 Feb 14;494:155-156. [doi: 10.1038/494155a]
6. Ackerman MS. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. Human-Comp. Interaction 2000 Sep 1;15(2):179-203. [doi: 10.1207/S15327051HCI1523_5]
7. Hartzler A, Skeels MM, Mukai M, Powell C, Klasnja P, Pratt W. Sharing is caring, but not error free: transparency of granular controls for sharing personal health information in social networks. AMIA Annu Symp Proc 2011;2011:559-568 [FREE Full text] [Medline: 22195111]
8. Skeels MM, Unruh KT, Powell C, Pratt W. Catalyzing Social Support for Breast Cancer Patients. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, New York: ACM Press; 2010 Presented at: CHI 2010; 10-15 April 2010; Atlanta, Georgia p. 173-182 URL: http://europepmc.org/abstract/MED/21654894 [doi: 10.1145/1753326.1753353]
9. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. "It's not that I don't have problems, I'm just not putting them on Facebook": Challenges and Opportunities in Using Online Social Networks for Health. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work. New York, New York: ACM Press; 2011 Mar Presented at: CSCW 2011; 19-23 March 2011; Hangzhou, China p. 341-350. [doi: 10.1145/1958824.1958876]
10. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
11. Krishna S, Boren SA. Diabetes self-management care via cell phone: a systematic review. J Diabetes Sci Technol 2008 May;2(3):509-517 [FREE Full text] [Medline: 19885219]

12. Morak J, Schindler K, Goerzer E, Kastner P, Toplak H, Ludvik B, et al. A pilot study of mobile phone-based therapy for obese patients. J Telemed Telecare 2008;14(3):147-149. [doi: 10.1258/jtt.2008.003015] [Medline: 18430285]
13. Yoo HJ, Park MS, Kim TN, Yang SJ, Cho GJ, Hwang TG, et al. A Ubiquitous Chronic Disease Care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]
14. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012 Feb;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
15. Buis LR, Poulton TA, Holleman RG, Sen A, Resnick PJ, Goodrich DE, et al. Evaluating Active U: an Internet-mediated physical activity program. BMC Public Health 2009;9:331 [FREE Full text] [doi: 10.1186/1471-2458-9-331] [Medline: 19744311]
16. Consolvo, S; Everitt, K; Smith, I; Landay, JA. Design Requirements for technologies that encourage physical activity. In: Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems. New York, New York: ACM Press; 2006 Apr Presented at: CHI 2006; 22-27 April 2006; Montréal, Canada p. 2006. [doi: 10.1145/1124772.1124840]
17. Lin JJ, Mamykina L, Lindtner S, Delajoux G, Strub HB. Fish’n’Steps: Encouraging Physical Activity with an Interactive Computer Game. In: Proceedings of the 8th international conference on Ubiquitous Computing. Berlin: Springer-Verlag; 2006 Presented at: UbiComp 2006; 17-21 September 2006; Orange County, California p. 261-278. [doi: 10.1007/11853565_16]
18. Resnick PJ, Janney AW, Buis LR, Richardson CR. Adding an online community to an internet-mediated walking program. Part 2: strategies for encouraging community participation. J Med Internet Res 2010;12(4):e72 [FREE Full text] [doi: 10.2196/jmir.1339] [Medline: 21169161]
19. 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]
20. 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]
21. Grimes A, Bednar M, Bolter JD, Grinter RE. EatWell: Sharing Nutrition-Related Memories in a Low-Income Community. In: Proceedings of the ACM Conference on Human Factors in Computing Systems.: ACM Press; 2008 Apr Presented at: CHI 2008; 5-10 April 2008; Florence, Italy. [doi: 10.1145/1460563.1460579]
22. Parker AG, Kantroo V, Lee H, Osornio M, Sharma M, Grinter RE. Health Promotion as Activism: Building Community Capacity to Effect Social Change. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, New York: ACM Press; 2012 Presented at: CHI 2012; 5-10 May 2012; Austin, Texas p. 99-108. [doi: 10.1145/2207676.2207692]
23. Kamal N, Fels S, Blackstock M, Ho K. The VivoSpace Social Network. In: Kranakis E, editor. Advances in Network Analysis and its Applications. Berlin: Springer; 2013:323-348.
24. Kamal N, Fels S. Determining the Determinants of Health Behaviour: Behavior Change through an Online Social Network. In: Lecture Notes in Computer Science Volume 7284. Berlin: Springer; 2012 Presented at: Persuasive Technology 2012; 6-8 June 2012; Linköping, Sweden p. 1-12. [doi: 10.1007/978-3-642-31037-9_1]
25. Khaled R, Barr P, Noble J, Biddle R. Investigating Social Software as Persuasive Technology. In: Lecture Notes in Computer Science 3962. Berlin: Springer; 2006 Presented at: Persuasive 2006; 18-19 May 2006; Eindhoven, The Netherlands p. 104-107. [doi: 10.1007/11755494_15]
26. Munson SA, Lauterbach D, Newman MW, Resnick P. Happier Together: Integrating a Wellness Application Into a Social Network Site. In: Lecture Notes in Computer Science 6137. Berlin: Springer; 2010 Presented at: Persuasive Technology 2010; 7-10 June 2010; Copenhagen, Denmark p. 27-39. [doi: 10.1007/978-3-642-13226-1_5]
27. Bray DA, Croxson K, Dutton WH, Konsynski B. Sermo: A Community-Based, Knowledge Ecosystem. In: SSRN. 2008 Feb Presented at: Distributed Problem-Solving Networks Conference; February 2008; Oxford Internet Institute. [doi: 10.2139/ssrn.1016483]
28. Young P. A physicians-only web site: building an intellectual and activist community online. Community Oncology 2008;5(9):521-524.
29. Damianos LE, Wohlever S, Ponte P, Wilson G, Reeder F, McEntee T, et al. Real users, real data, real problems: the MiTAP system for monitoring bio events. In: Proceedings of the second international conference on Human Language Technology Research. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 2002 Presented at: (HLT '02); 24-27 March 2002; San Diego, California p. 357-362.
30. Damianos LE, Bayer S, Chisholm MA, Henderson J, Hirschman L, Morgan W, et al. MiTAP for SARS detection. In: Demonstration Papers at HLT-NAACL 2004.: Association for Computational Linguistics; 2004 Presented at: HLT-NAACL--Demonstrations '04; 2-7 May 2004; New York, New York p. 13.
31. Cranor LF, Reagle J, Ackerman MS. Beyond Concern: Understanding Net Users’ Attitudes about Online Privacy. In: Vogelsang I, Compaine BM, editors. The Internet upheaval: raising questions, seeking answers in communications policy. Cambridge, Mass: MIT Press; 2000.

32. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Quarterly 2009 Jun;33(2):339-370.
33. Nissenbaum H. Privacy as Contextual Integrity. Washington Law Review 2004;79(1):101-139.
34. Adams A, Sasse MA. Privacy in Multimedia Communications: Protecting Users, Not Just Data. In: Blandford A, Vanderdonckt J, Gray P, editors. People and Computers XV – Interaction without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001. London: Springer; 2001:49-64.
35. Smith HJ. Privacy policies and practices: inside the organizational maze. Commun. ACM 1993;36(12):104-122. [doi: 10.1145/163298.163349]
36. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
37. Ajzen I, Fishbein M. Attitudinal and normative variables as predictors of specific behavior. Journal of Personality and Social Psychology 1973;27(1):41-57. [doi: 10.1037/h0034440]
38. Culnan MJ, Armstrong PK. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. Organization Science 1999 Jan 01;10(1):104-115. [doi: 10.1287/orsc.10.1.104]
39. Malhotra NK, Kim SS, Agarwal J. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. Information Systems Research 2004 Dec;15(4):336-355. [doi: 10.1287/isre.1040.0032]
40. Dinev T, Hart P. An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research 2006 Mar;17(1):61-80. [doi: 10.1287/isre.1060.0080]
41. Croll PR. Determining the privacy policy deficiencies of health ICT applications through semi-formal modelling. Int J Med Inform 2011 Feb;80(2):e32-e38. [doi: 10.1016/j.ijmedinf.2010.10.006] [Medline: 21075675]
42. Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behaviour. In: Zanna MP, editor. Advances in Experimental Social Psychology. CA: Elsevier Inc; 1991:201-234.
43. Munson SA, Consolvo S. Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity, Proceedings of Pervasive Health. In: 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: Pervasive Health 2012; 21-24 May 2012; San Diego, CA p. 25-32.
44. Lampe C, Ellison NB, Steinfield C. Changes in use and perception of Facebook. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. New York, New York: ACM Press; 2008 Nov Presented at: CSCW 2008; 8-12 November 2008; San Diego, California p. 721-730. [doi: 10.1145/1460563.1460675]
45. Coffield RL, Joiner JE. Risky Business: Treating Tweeting the Symptoms of Social Media. AHLA Connections 2010 Mar:10-14.
46. Munson SA. Beyond the Share Button: Making Social Network Sites Work for Health and Wellness. IEEE Potentials 2011 Sep;30(5):42-47. [doi: 10.1109/MPOT.2011.941731]
47. Heussner KM. Tech This Out, ABC News. 2009 Nov 23. Woman Loses Benefits After Posting Facebook Pics URL: http://abcnews.go.com/Technology/AheadoftheCurve/woman-loses-insurance-benefits-facebook-pics/story?id=9154741 [accessed 2013-09-29] [WebCite Cache ID 6K0LZZIFF]
48. van der Velden M, El Emam K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. J Am Med Inform Assoc 2013 Jan 1;20(1):16-24. [doi: 10.1136/amiajnl-2012-000949] [Medline: 22771531]
49. Johnson EJ, Bellman S, Lohse GL. Defaults, Framing and Privacy: Why Opting In-Opting Out. Marketing Letters 2002;13(1):5-15. [doi: 10.1023/A:1015044207315]
50. Tversky A, Kahneman D. Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics 1991;106(4):1039-1061.
51. Xu H, Dinev T, Smith J, Hart P. Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances. Journal of the Association for Information Systems 2011 Dec;12(12):798-824.
52. Sheehan KB, Hoy MG. Flaming, complaining, abstaining: How online users respond to privacy concerns. Journal of Advertising 1999;28(3):37-51.
53. Acquisti A, Grossklags J. Privacy and rationality in individual decision making. IEEE Secur. Privacy Mag 2005 Jan;3(1):26-33. [doi: 10.1109/MSP.2005.22]
54. Madejski M, Johnson M, Bellovin SM. A study of privacy settings errors in an online social network. 2012 Presented at: IEEE Pervasive Computing and Communications Workshops; 24-28 March 2012; Budapest, Hungary p. 340-345. [doi: 10.1109/PerComW.2012.6197507]
55. Shafir E, Simonson I, Tversky A. Reason-based choice. Cognition 1993;49(1-2):11-36. [doi: 10.1016/0010-0277(93)90034-S]
56. Norberg PA, Horne DR, DR DA. The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of Consumer Affairs 2007;41(1):100-126. [doi: 10.1111/j.1745-6606.2006.00070.x]
57. Huber VL, Neale MA, Northcraft GB. Decision bias and personnel selection strategies. Organizational Behavior and Human Decision Processes 1987 Aug;40(1):136-147. [doi: 10.1016/0749-5978(87)90009-4]
58. Levin IP, Jasper JD, Forbes WS. Choosing versus rejecting options at different stages of decision making. J. Behav. Decis. Making 1998 Sep;11(3):193-210. [doi: 10.1002/(SICI)1099-0771(199809)11:3<193::AID-BDM297>3.0.CO;2-G]

59. Aleem U. Master Thesis.: University of British Columbia; 2012. Exploring the influence of privacy management tools on online information sharing decision URL: https://circle.ubc.ca/handle/2429/43340 [accessed 2013-09-30] [WebCite Cache ID 6K1tugIjF]
60. Bulgurcu B. Understanding the information privacy-related perceptions and behaviors of an online social network user (PhD Thesis).: University of British Columbia; 2012. The Role of Primary and Secondary Appraisal in a User's Coping Motivations: An Empirical Study on Facebook Applications URL: https://circle.ubc.ca/handle/2429/43380 [accessed 2013-09-30] [WebCite Cache ID 6K1u8Dihj]
61. Bulgurcu B. Understanding the information privacy-related perceptions and behaviors of an online social network user (PhD Thesis).: University of British Columbia; 2012. The Roles of Permission Requests and Privacy Controls in Shaping a User's Primary Appraisal and Use Intentions: An Empirical Study on Facebook Applications URL: https://circle.ubc.ca/handle/2429/43380 [accessed 2013-09-30] [WebCite Cache ID 6K1uNGQHQ]
62. Thaler RH, Sunstein CR. Nudge: improving decisions about health, wealth, and happiness. New Haven: Yale University Press; 2008.
63. Barbaro M M, Zeller T. The New York Times. 2006 Aug 09. A Face Is Exposed for AOL Searcher No. 4417749 URL: http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0
64. Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. In: Proceedings of the IEEE Symposium on Security & Privacy.: IEEE; 2008 Presented at: SP 2008; 18-22 May 2008; Oakland, CA p. 111-125. [doi: 10.1109/SP.2008.33]
65. Griffith V, Jakobsson M. Messin' with Texas Deriving Mother's Maiden Names Using Public Records. In: Lecture Notes in Computer Science 3531. 2005 Jun Presented at: Applied Cryptography and Network Security; 7-10 June 2005; New York, New York p. 91-103. [doi: 10.1007/11496137_7]
66. Acquisti A, Gross R. Predicting Social Security numbers from public data. PNAS 2009 Jul 06;106(27):10975-10980. [doi: 10.1073/pnas.0904891106]
67. Byrne E, Sahay S. Participatory design for social development: A South African case study on community-based health information systems. Information Technology for Development 2007 Jan;13(1):71-94. [doi: 10.1002/itdj.20052]
68. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
69. Goffman E. 1959. In: The presentation of self in everyday life. Harmondsworth: Penguin Books; 1971.
70. Terveen L, McDonald. Social matching. ACM Trans. Comput.-Hum. Interact 2005 Sep 01;12(3):401-434. [doi: 10.1145/1096737.1096740]
71. Civan A, McDonald DW, Unruh KT, Pratt W. Locating patient expertise in everyday life. In: Proceedings of GROUP 2009. 2009 Presented at: GROUP 2009; 10-13 May 2009; Sanibel Island, Florida p. 291-300. [doi: 10.1145/1531674.1531718]
72. Hansen DL, Johnson C. Veiled viral marketing: disseminating information on stigmatized illnesses via social networking sites. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.: ACM Press; 2012 Presented at: IHI 2012; 28-30 January 2012; Miami, Florida p. 247-254. [doi: 10.1145/2110363.2110393]
73. Sidorov J. Social media and population-based care management. Popul Health Manag 2010 Aug;13(4):175-176. [doi: 10.1089/pop.2010.1341] [Medline: 20735243]
74. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
75. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Med Educ Online 2010;15:- [FREE Full text] [doi: 10.3402/meo.v15i0.5324] [Medline: 20859533]
76. Zhou J. New York Times. 2010 Nov 30. Where Anonymity Breeds Contempt URL: http://www.nytimes.com/2010/11/30/opinion/30zhuo.html?_r=1& [accessed 2013-09-30] [WebCite Cache ID 6K1lvje2E]
77. Resnick P, Kuwabara K, Zeckhauser R, Friedman E. Reputation systems. Commun. ACM 2000;43(12):45-48. [doi: 10.1145/355112.355122]
78. Huh J, Yetisgen-Yildiz M, Hartzler A, McDonald D, Park A, Pratt W. Text Classification to Weave Medical Advice with Patient Experiences. In: Proceedings of the American Medical Informatics Association Fall Symposium. 2012 Presented at: AMIA'12; 3-7 November 2012; Chicago, Illinois.
79. Kirsch L. HL7 Standards. 2012 Aug 28. The Blockbuster Drug of the Century: An Engaged Patient URL: http://www.hl7standards.com/blog/2012/08/28/drug-of-the-century/ [accessed 2013-09-29] [WebCite Cache ID 6K0MaclpO]
80. Fox S, Purcell K. The Pew Internet & American Life Project. 2010 Mar 23. Chronic Disease and the Internet URL: http://www.pewinternet.org/Reports/2010/Chronic-Disease.aspx [WebCite Cache ID 5p0knHRPw]
81. Farber AS. The Peckham Experiment revisited: cultivating health. Health Soc Work 1976 Aug;1(3):27-38. [Medline: 789192]
82. Pioneer Health Foundation. What was the Peckham Experiment? URL: http://thephf.org/index.php/history/what-was-the-peckham-experiment [accessed 2013-09-30] [WebCite Cache ID 6K1l1OmMe]
83. Miller JW. Wellness: the history and development of a concept. Spektrum Freizeit 2005;2005(1):84-102.
84. Hood E. Dwelling disparities: how poor housing leads to poor health. Environ Health Perspect 2005 May;113(5):A310-A317 [FREE Full text] [Medline: 15866753]

85. Largo-Wight E. Cultivating healthy places and communities: evidenced-based nature contact recommendations. Int J Environ Health Res 2011 Feb;21(1):41-61. [doi: 10.1080/09603123.2010.499452] [Medline: 21246432]
86. Purpura S, Schwanda V, Williams K, Stubler W, Sengers P. Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, New York: ACM Press; 2011 Presented at: CHI 2011; 7-12 May 2011; Vancouver, BC, Canada p. 423-432. [doi: 10.1145/1978942.1979003]
87. Baumer EPS, Katz SJ, Adams P, Gonzales AL, Pollak J, Retelny D, et al. Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. New York, New York: ACM Press; 2012 Presented at: CSCW 2012; 11-15 February 2012; Seattle, Washington p. 475-484. [doi: 10.1145/2145204.2145279]
88. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
89. Halko S, Kientz JA. Personality and Persuasive Technology: An Exploratory Study on Health-Promoting Mobile Applications. In: Lecture Notes in Computer Science 6137.: Springer; 2010 Presented at: Persuasive Technology 2010; 7-10 June 2010; Copenhagen, Denmark p. 150-161. [doi: 10.1007/978-3-642-13226-1_16]
90. Nov O, Arazy O, Lopez C, Brusilovsky P. Exploring personality-targeted UI design in online social participation systems. In: Proceedings of the ACM Conference on Human Factors in Computing Systems. New York, New York: ACM Press; 2013 Presented at: CHI 2013; 27 April - 2 May 2013; Paris, France p. 361-370. [doi: 10.1145/2470654.2470707]
91. Aakhus M. Understanding the Socio-Technical Gap: A case of GDSS Facilitation. 2004 Presented at: Second International Conference on Action in Language, Organisations, and Information Systems; 17-18 March 2004; Linköping, Sweden p. 137-148.

Abbreviations

CSCW: computer supported cooperative work
HCI: human-computer interaction
HIPAA: Health Insurance Portability and Accountability Act
MiTAP: MITRE Text and Audio Processing
SARS: Severe Acute Respiratory Syndrome

Edited by G Eysenbach; submitted 26.06.13; peer-reviewed by P Lalli, P Rosen; comments to author 07.08.13; revised version received 27.08.13; accepted 18.09.13; published 22.10.13.

Please cite as:
Munson SA, Cavusoglu H, Frisch L, Fels S
Sociotechnical Challenges and Progress in Using Social Media for Health
J Med Internet Res 2013;15(10):e226
URL: http://www.jmir.org/2013/10/e226/
doi:10.2196/jmir.2792
PMID:24148206

Original Paper

Enabling Community Through Social Media

Anatoliy Gruzd[1], PhD; Caroline Haythornthwaite[2], PhD

[1]School of Information Management, Faculty of Management, Dalhousie University, Halifax, NS, Canada

[2]School of Library, Archival and Information Studies, The iSchool at UBC, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Anatoliy Gruzd, PhD
School of Information Management
Faculty of Management
Dalhousie University
6100 University Ave, Suite 4010
Halifax, NS, B3H 4R2
Canada
Phone: 1 9024946119
Fax: 1 9024942451
Email: gruzd@dal.ca

Abstract

Background: Social network analysis provides a perspective and method for inquiring into the structures that comprise online groups and communities. Traces from interaction via social media provide the opportunity for understanding how a community is formed and maintained online.

Objective: The paper aims to demonstrate how social network analysis provides a vocabulary and set of techniques for examining interaction patterns via social media. Using the case of the #hcsmca online discussion forum, this paper highlights what has been and can be gained by approaching online community from a social network perspective, as well as providing an inside look at the structure of the #hcsmca community.

Methods: Social network analysis was used to examine structures in a 1-month sample of Twitter messages with the hashtag #hcsmca (3871 tweets, 486 unique posters), which is the tag associated with the social media–supported group Health Care Social Media Canada. Network connections were considered present if the individual was mentioned, replied to, or had a post retweeted.

Results: Network analyses revealed patterns of interaction that characterized the community as comprising one component, with a set of core participants prominent in the network due to their connections with others. Analysis showed the social media health content providers were the most influential group based on in-degree centrality. However, there was no preferential attachment among people in the same professional group, indicating that the formation of connections among community members was not constrained by professional status.

Conclusions: Network analysis and visualizations provide techniques and a vocabulary for understanding online interaction, as well as insights that can help in understanding what, and who, comprises and sustains a network, and whether community emerges from a network of online interactions.

(J Med Internet Res 2013;15(10):e248) doi:10.2196/jmir.2796

KEYWORDS

online community; online social networks; information and communication technology; social media; Twitter

Introduction

Background

The use of social media has spread dramatically in the past few years, demonstrated in increasing numbers of users, types of media, mobile applications, and connectivity. This has stimulated growth in applying social media to matters of health and health communities: from work-based communities of practice [1] to forums for patient social and information support (eg, [2-11]). These efforts can be enhanced by taking advantage of the research and experience already existing relating to online communication and community. This research provides a wealth of background theories, studies, and findings that inform the ways that a community is likely to form via newer social media and that can be applied to the development of health communities.

Of the many approaches to community and online communication that have emerged, we highlight a social network perspective. This perspective looks at group or community interactions to determine what kinds of actors and ties make up the network; what exchange of information, social support, socializing, play, or other resources form the basis of the community; and what roles and cliques emerge that provide structure to the community. Social network analysis provides a vocabulary and set of techniques for examining interaction patterns between people and has proven useful for studying health (eg, [12,13]), how relationships are maintained without physical co-presence [14-17], and the development of new, health care-related online networks [3].

In this paper, we first discuss social networks and then illustrate the kind of information that can be revealed about community from a social network perspective through a case study of social media use by the group Health Care Social Media Canada (HCSMCA). This case includes network analyses of the group's structures as shown through a sample of Twitter messages using the hashtag #hcsmca. Results from the network analysis reveal a cohesive group consisting of one major component, including interaction across professional roles. The group founder and participants that are identified as social media health care providers are prominent in posts and in attention from others, and the network is sustained by participation from and recognition of a core set of actors.

The first section below reviews the background on online social networks and describes the HCSMCA group. The following sections describe the analysis of the #hcsmca Twitter networks and then discuss these in relation to previous research on online communities.

Social Network Analysis

Of the many ways to look at the range and effects of social media on interpersonal and collective relations, one that has proved useful for online communities has been a social network perspective. This is not the same as "social networking". It is instead an approach that considers the unit of analysis to be the connections between people and looks at how these connections—social network "relations"—form patterns of interaction that reveal how information and other resources flow in a network, as well as the structures that define the network [18]. Pescosolido [19] has suggested that a network-centered view of health, based on social network principles, can act as a bridge between medical sciences and individual health experience. Her ideas respond to an increasing recognition of the impact of connectivity and experience: "The individual is seen as embedded in an ongoing relational dynamic with sequences of events seen as patterned, contingent and emergent" (p. 196). Her network episode model makes a connection between social context, social support, and illness careers and offers a way to address the complex whole that pertains to health and well-being.

Media use is just one aspect of this complex structure, but it has the potential to set context, add to a social support system, and touch individuals and their closely tied friends and family. From an analytical perspective, one of the advantages of taking a social network perspective is that the focus is on what people do with each other rather than the medium or face-to-face context through which they do it. This allows exploration of the types of interactions that create and define different kinds of relationships and communities [20]. Thus, friendship may be recognized by pairwise exchange of personal information and emotion, discussion of multiple topics, co-participation in events, frequent interaction, and the use of multiple media. Social support emerges as a complex of small and larger exchanges between people, trust in networks to provide services in time of need, and a generalized reciprocity in communities where resources are distributed more generally than in a strictly give-and-take fashion. Analyzing health support networks requires understanding not just what media are used for communication, but also the types of exchanges that constitute support and the roles that start a network, as well as the ones that emerge from networks.

Many years of research on social networks have provided evidence of social network principles as well as statistical and analytical techniques for understanding network behaviors (eg, [21-24]), including health [19,12,25]. The basic principles of social network analysis are derived from graph theory and consider *actors* (eg, people, organizations) as nodes in a network, connected by *relations* (what they do with each other, eg, provide new information, emotional support, resources, and/or services) that form interpersonal *ties*. The nature and variety of relations define the kind of relationship between actors, such as an acquaintanceship, friendship, learning, or work relationship. Research has shown that the closer the relationship, the more different types of exchanges are maintained and the more important these exchanges are for the individuals; close personal relationships also demonstrate a higher level of intimacy and self-disclosure. Such ties are *strong ties*, and pairs who are strongly tied are more motivated to share their resources with each other. These pairs also turn out to be more like each other (more homophilous), with the result that they tend to know and associate with similar others. *Weak ties*, by contrast, are less motivated to share their resources but are more likely to have access to resources different from each other because they do not share similar habits, circles of friends, etc [26]. Pairwise relationships build into the social networks that are recognized as cliques, groups, and communities.

Where bonds are strong, resources are shared generally around the network (generalized reciprocity). This creates the social capital of the network, that is, the accumulated resources held within the network rather than those held by any individual [27]. Where bonds link networks, they connect the network to resources in other networks. Putnam [28] describes these two forms as *bonding* and *bridging social capital*. Both are important means of information and resource access and uptake. Burt [29] identified the important position of the actor who acts as a broker between networks, filling a *structural hole*. Such an actor can choose to control information and resources between these two separate networks, or they can facilitate its transmission. Recently the latter position has come to prominence embodied in the role of the "social entrepreneur", that is, an individual positioned to facilitate the transfer of knowledge or practices to disadvantaged—non-networked—communities (eg, [30]).

The configuration of connections is all-important in social networks. These structures show how actors are connected over the whole network, and thus what paths and obstacles there are for contact, information, and resource flow. Among popular aspects considered for networks are the *positions* of individuals, for example, how prominent or influential they can be based on the ties to and from other actors (creating recognized positions such as network stars, isolates, brokers). For networks as a whole, cliques may be evident as highly interconnected subsets of network actors. Networks may exhibit a high or low density of internal connections, with the former suggesting rapid diffusion of resources and the latter suggesting slow, poor, or long-chain routes for diffusion. Also of interest, particularly when comparing across networks, are similarities in structures and *roles*, for example, as a teacher fills the same role with students no matter what class is examined, or a doctor with a patient no matter what the medical setting.

Our online interactions make these patterns more readily observable, and many examples exist now of how such patterns can be made visible, for example, in social network interaction patterns [31], patterns of text changes in wikis [32,33], and information seeking patterns (eg, Google Flu trends), each of which contributes to understanding emergent community network properties [34]. Social media traces are thus an entry point to describing and later understanding and facilitating community interaction.

In this paper, we examine the social media traces from the #hcsmca Twitter posts. We examine what social network patterns are revealed and the implications these have for #hcsmca as a community. The following section provides background on the #hcsmca group and its operation.

#hcsmca—A Twitter Community

As stated on the HCSMCA website, "#hcsmca is a vibrant community of people interested in exploring social innovation in health care. We share and learn, and together we are making health care more open and connected." It is an example of how those with a common interest can meet and form community online through social media, in this case, in the interests of social innovation in Canadian health care.

The community was founded in September 2010, by Colleen Young, an online community manager and Toronto-based patient advocate and health writer [35]. In her blog [36], she describes the community as follows:

> *Anyone and everyone delivering and receiving health care who is interested in open conversation to help improve quality, access, value and effectiveness of health care. This includes: patients, caregivers, patient advocates, health care professionals, not-for-profit health organizations, educators, health content providers, health institutions, health administrators, health systems and networks, government and health policy makers.*

The community is maintained through four social media: Twitter, a LinkedIn group with 181 members, a Facebook page with 143 "likes" (as of January 8, 2013), and the blog maintained by the founder, Colleen Young. While maintained across these various media, the community relies on Twitter, the popular microblogging site, as their primary communication platform, operating with the hashtag #hcsmca. The community meets weekly on Twitter to discuss various topics relating to health care and social media.

To participate in this group discussion, a participant just needs to post a message on Twitter using this hashtag. At the time of this research, weekly chats on Twitter are scheduled for every Wednesday at 1 pm EST with the last Wednesday of the month being an evening chat at 9 pm EST. Weekly topics and guest moderators are announced in advance and listed in a public Google spreadsheet. For those who miss this real-time meeting, a transcript with messages is available, posted to the community blog by the group moderator. #hcsmca is a great case for study since HCSMCA has been active with this hashtag for over 2 years and generates very active weekly discussions that attract a wide variety of professionals and organizations.

One of the main goals of our research is to gain a better understanding of how social media–based information and communication technologies, such as Twitter, enable a distributed group of people to form and maintain an online community. In particular, we are interested in the following research questions regarding #hcsmca:

1. What accounts for the relative longevity of this particular online community? Is it because of the founder's leadership and continuing involvement, or are there core members who are actively and persistently involved in this community?
2. What is the composition of this community in general? And, more specifically, does their professional role determine a person's centrality within this community? This will allow us to understand generally how professional roles affect online conversational dynamics, and more specifically whether this online community is a welcoming place for a wide range of professionals or is, instead, dominated by professionals from a particular group.

Methods

Study Sample

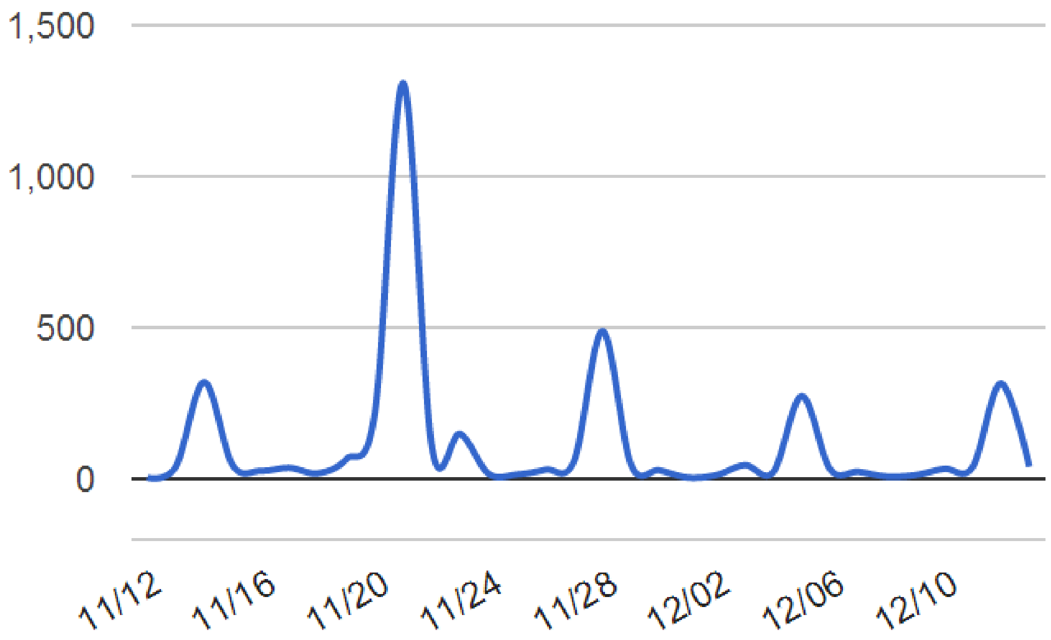
The primary dataset for this research came from Twitter and included all public Twitter messages that included the #hcsmca hashtag, posted between November 12, 2012, and December 13, 2012. The dataset contains a total of 3871 tweets, posted by 486 unique Twitter users. The dataset was collected and analyzed using Netlytic [37] system for automated collection and analysis of social media data. (Netlytic is developed and maintained by author Gruzd).

As noted above, #hcsmca hosts a weekly discussion. Topics covered and the assigned topic moderators for the period studied are shown in Table 1. Figure 1 shows the distribution of the messages over the studied time period; peaks on the chart reflect the weekly live chats.

Table 1. #hcsmca weekly topics (Nov 12-Dec 13, 2012).

Date	Weekly discussion topics	Assigned moderator
Nov 14	Challenge of engaging SM [social media] to inform a research agenda	@QuintePediatric
	Use of innovation, SM, and gamification to encourage uptake of self-care	
Nov 21	Health care blogs should we or shouldn't we, what have we learned, what are the benefits?	@JackieHickeyRN
	Are health care blogs a useful tool for education and knowledge transfer?	
Nov 28	How has social media made you healthier? Unhealthier? Has social media made our health choices more numerous and this overwhelming?	@NaheedD
	What messaging would motivate you to make a positive health change? Who would you listen to?	
Dec 5	What is needed to make cross-organizational collaboration via social networks more effective?	@WillFalk & @MarkCassel-man
	In what settings / sectors are you seeing health care providers and patients interacting via social media?	
Dec 12	How can SM support patient care in an ambulatory care setting?	@CraigTyyz
	How can SM help patients/families navigate a new/unfamiliar hospital/clinic/facility?	

Figure 1. Number of #hcsmca tweets over the studied period.



Twitter Networks

Twitter connections are maintained through the technical means of *usernames*, *following*, and *hashtags*. Twitter usernames identify nodes in the network (eg, author Gruzd is identified as "dalprof"). A direct communication connection can be made person to person by indicating the one recipient by prefacing the message—or tweet—with "@" and the username (eg, @dalprof), or tweets can be sent to the world at large. An indirect communication connection can be made by simply mentioning someone's Twitter username (prefacing it with @) anywhere in a tweet or publicly reposting (retweeting) somebody's else tweet. (While we say "person to person", usernames are also commonly associated with groups or organizations; also, no one-to-one correspondence of person to username is assumed as individuals may have several Twitter usernames.)

Follow and topic hashtags show relational connections between nodes. Searching for someone on Twitter brings up the option to follow that person's postings, with their tweets immediately visible on the user's home Twitter page. *Following* is a node-to-node connection, marking social networks created through the act of designating a *follow* relation in Twitter. A second technical feature for relational connections is the use of hashtags. A hashtag is a microblogging convention that allows users to see others' messages regardless of whether they have chosen to follow that person. When many people tweet with a common hashtag, this creates connections among posts based on a common hashtag relation. For example, the hashtag #med2 was used at the 2013 Medicine 2.0 conference in London, England. Participants both in London and elsewhere could monitor messages with this hashtag to engage with the Twitter conversation regarding the conference. Both following and hashtags provide the infrastructure for social networks, that is, the underpinning structure from which and on which communities grow and prosper.

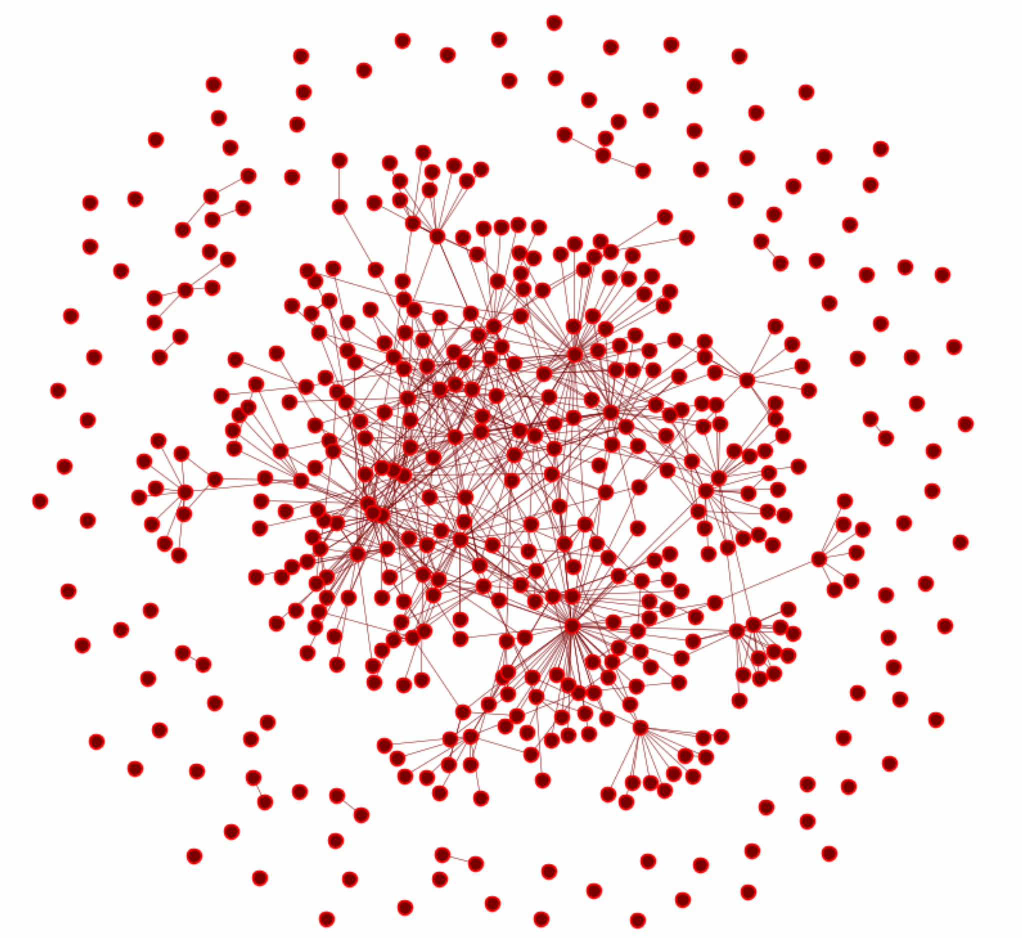
Analyzing Posts for Name Networks

Netlytic was used to discover the communication network among community members. In particular, to discover social connections among community members, the analysis relied on a type of network called "Name Network" [38]. The Name Network technique examines the content of the messages and connects one person to another if they mention, reply, or repost another person's tweet [39,40]. The resulting network generated by Netlytic included 486 nodes and 736 ties. The collected social network dataset was then exported to the network visualization application ORA [41] and to Ucinet [42] for statistical tests.

Figure 2 presents the visualization of the #hcsmca Name Network for the 4-week period. The overall view shows a fairly densely connected, single component of posters who are reading and responding to each others' posts, suggesting an engaged community, paying attention to the topic and actively conversing around the common topic. Isolated nodes (those with no line connecting to others) posted but received no mention, reply, or repost. While there are number of such nodes, their numbers do not overwhelm the number in the central component. Such "legitimate peripheral participation" [43] is a common part of any community and supports learning the way to engage in a community as well as engaging in a partial way that fits individuals' time and needs. Noticeably absent from this figure are subcliques that carry on side conversations with each other. This shows that the #hcsmca community is not fractionated, but rather that participants are all engaged with the single conversational network.

The following sections show results from the analysis of the Twitter posts, with attention to aspects of community. Results address first, the discovery of key actors in the network and their potential influence on others and second, whether and how professional roles affect participation.

Figure 2. Twitter communication network among #hcsmca participants.



Results

Discovering Community Leaders

One way to learn how an online community operates is to find out about key members who have the potential to influence tone, topic, or policy for the whole community. A community organizer may be one such actor, but for a community to operate robustly, actions associated with keeping the community or conversation going need to be distributed to more than one person. Thus, in examining the #hcsmca community, it is of particular interest to see whether more than one individual is leading the discussion.

A brief examination of the community blog shows that, as expected, the founder of the #hcsmca group is heavily involved in planning and running the community. But who else is involved? Are there other members of this group who also take on a leadership role? This is important because the presence of a strong community core with a number of active members suggests a healthy online community that can persist without the presence of particular individuals (eg, as in the failure case described in [44]). For example, if some of the active members cannot participate in a particular weekly discussion, there would be others to carry the conversation.

Three social network measures were used to locate influential individuals in this community: (1) the total number of messages contributed during the studied period, (2) the number of times a person is mentioned or replied to, that is, their @username is used in a post by someone else (*in-degree centrality)*, and (3) the number of times a person mentions or replies to others, that is, an individual uses another person's @username in a post (*out-degree centrality).*

All three measures are important in identifying prominent individuals in the community. An individual posting a high number of messages gains attention for the content they send to others and can add to the social capital of the community by bringing new information to the group as a whole. However, such information needs to be taken up and used by the community. Thus, a high number of posts, by itself, does not mean that the messages contributed are deemed important or interesting by other members of the community—hence the need to look at network structures of message uptake.

Any use of @username signifies a direct connection between the sender and another individual. Being mentioned by others is a case of *in-degree centrality* and signifies the prestige given to that individual by others in the network. A person mentioning or replying to others indicates *out-degree centrality* and signifies the influence that person has as they make their views known to others. To identify people on Twitter with high *in-degree* centrality values, we look for people whose tweets are chosen by others to be retweeted (forwarded) and/or replied to by many others. To identify people with high *out-degree* centrality on Twitter, we measured how often a person mentioned others or replied to others in their tweets. People who have high out-degree centrality tend to have a good awareness of the network and often monitor and retweet messages by others.

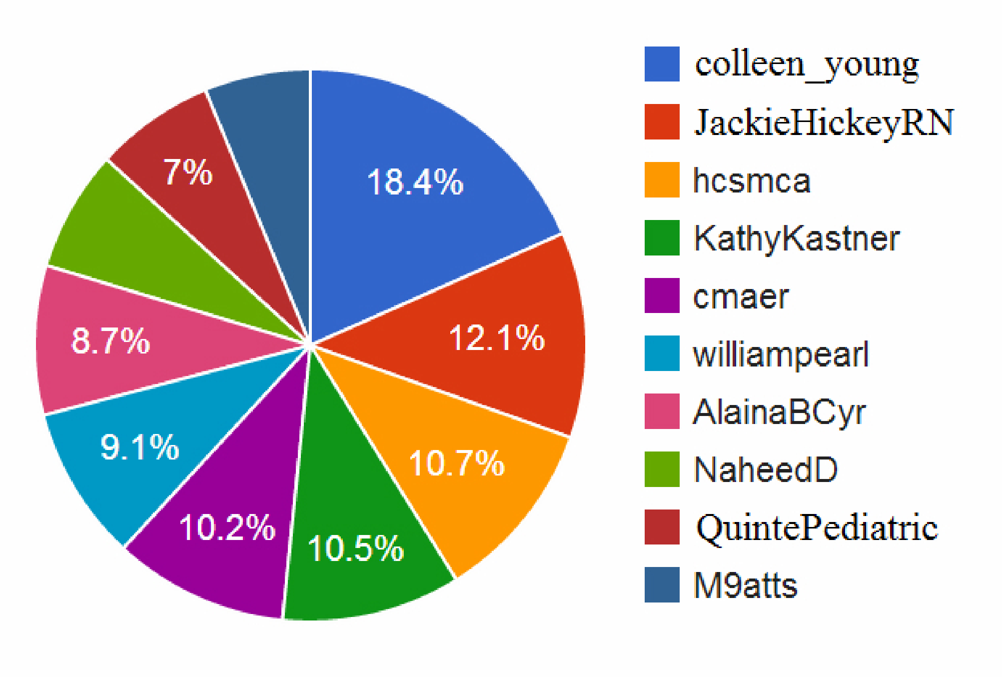
Total Number of Posts

Figure 3 shows the top 10 active members of this community based on the total number of messages posted to this community. Not surprisingly, the group organizer, @colleen_young, posted the most number of messages (18.4% of all messages posted by the top 10 posters). In starting an online community, leaders play a key role by their altruistic or proactive participation, providing more posts to the community than they receive and thus helping create a critical mass of interactions that act as a draw for others.

As well as the founder, there are a few other active participants who contribute heavily to the community, posting about the same number of messages each (approximately 10% of the messages posted by the top 10 posters all together). Among this group are people who moderated weekly chats, such as @JackieHickeyRN, @NaheedD, and @QuintePediatric. Such actors also contribute to the critical mass of the conversation, but the more important result is that there are several people the community can rely on to keep the conversation going, increasing the robustness of the ongoing activity.

However, not all of the moderators are active posters. For example, 3 out of the 6 moderators (see Table 1) do not appear on the top 10 list in Figure 3. One of the possible reasons for this is that some moderators may participate only in their own weekly chat and not be active during other weeks. From the community and knowledge building perspective, it would be beneficial to encourage moderators to participate in discussions moderated by others, especially prior to their own week. This would help the moderator to build authority in this community (encouraging more retweets), get to know what topics are important to this group, and be able to reference and connect to the topics discussed during the prior week(s).

Figure 3. Top 10 most active posters.



Prestige and In-Degree Centrality

As noted above, the total number of posted messages indicates only the engagement level on the part of an individual rather than the uptake of their contributions by the community. To find out whether personal messages influence others and make them reply or retweet, we examined *in-degree centrality* (the number of people who are mentioned or replied to). Table 2 shows the top 10 users based on in-degree centrality. Again, not surprisingly, the group founder is frequently mentioned by others, and her messages were retweeted by 36 people during the collected period of time.

In examining those on this list other than the founder, we noticed that they have something in common. Most of them have a very active online presence in social media in general, not just in this community. They are also very passionate and active commentators on health matters on Twitter. For example, the second most connected account is @cmaer. This username belongs to Pat Rich, who is an online editor for the Canadian Medical Association and has over 1000 followers. Shirley Williams (@williampearl) is a new media enthusiast and advocate at Strategic Leadership Forum and has over 3000 followers. Others on this list also have a considerable number of followers. In fact, there is a weak, monotonic (nonlinear) positive correlation between the number of followers and the in-degree centrality (Spearman rho=0.23, P<.01). In other words, people with more followers on Twitter in general are likely to be more central (based on the in-degree centrality) in this community.

One possible explanation of this could be that by participating in weekly discussions on #hcsmca, these individuals expose their followers to this community through their tweets on this topic (with the #hcsmca hashtag). As a result, their followers may also join #hcsmca chats and retweet or reply to them on this topic, thus increasing their in-degree centrality in this community. Future research is required to confirm or reject this preliminary supposition. If it holds, then one recommendation for growing an online community such as #hcsmca could be to find people who (1) are already actively engaged in online conversations in this area and (2) have a strong base of followers, and invite them to join the discussion. This is a reasonable recommendation in general as it brings in people who can act as bridges between separate networks and communities. People with high in-degree on this list are also good candidates for moderating future discussions as their messages are clearly resonating with this group.

Another observation that we can make about this group is that 6 of 10 people with high in-degree also posted the most number of tweets to this community (see Figure 3). This may just be indicative of their general interest in this topic but may also suggest that because they post more messages, their messages are more likely to be noticed by others on Twitter and thus more likely to be retweeted. Future research is needed to explore this further.

Overall, people with high values of in-degree centrality can be considered as trusted information sources whose opinions and comments are recognized as having value for the community, as evidenced by the frequency with which their messages are retweeted or they are frequently mentioned by others. These people are important for this community as they generate a lot of trusted, “sharable” information that generates discussion, but equally important, by being retweeted, also sustains conversational interaction and the life of the #hcsmca Twitter community.

Table 2. Top 10 users by in-degree centrality.

Twitter handle	Centrality in degree	Centrality out degree	Twitter profile description (as posted by the user)
colleen_young	36	27	Community Manager of Virtual Hospice | Portail en soins palliatifs (@VirtualHospice), Founder of #hcsmca, plain language writer, health literacy advocate
cmaer	33	10	Online editor for the Canadian Medical Association. Views are my own
williampearl	26	4	Facilitating & Finding Pearls in Strategy, SocialMedia & Healthcare, #ROTPt
naheedd	18	12	Medical resident physician. #GlobalHealth+#SDOH advocate. #MedEd+#hcsmca enthusiast. RA @CRICH_StMikes. Writer @HealthyDebate. Humanist. Change agent. Optimist.
infoway	17	1	News & announcements from Canada Health Infoway. Tweets by a team from Infoway. Check out our blog.
rdjfraser	15	1	Nurse & Author. Digital Tool Strategist and Educator. Learn more, help others. Tweets are my own.
anneccpa	14	10	Canadian Certified #PhysicianAssistant, practicing in #Orthopaedics - Sports Medicine & Trauma. Blogger & advocate for the Physician Assistant Profession.
alainabcyr	14	12	Learn, share, create. Grow. Aspiring expert in patient education and health communications on the Web and in print. Opinions are my own.
kathykastner	14	9	love listening learning sharing. appreciate humour happiness and eating chocolate covered almonds. End of life goal: a joyous exit. BestEndings.com
symplur	14	1	Connecting the dots in healthcare social media. Curator of Healthcare Hashtag Project; Social Media Consultancy

Influence and Out-Degree Centrality

Another group of people who are important within any online community are people who monitor and retweet messages from others. To identify these individuals, we used the *out-degree centrality* (a measure of how often a person in the network mentioned or replied to other people in the network). Table 3 shows the top 11 users based on the out-degree centrality (11 users are shown rather than a more conventional “top 10” because of a tie in out-degree centrality for users who ranked 10th and 11th).

There is a strong overlap in who is prominent in both the in-degree and out-degree lists. Accounts such as @colleen_young, @naheedd, @alainabcyr, @cmaer, @anneccpa, and @kathykastner appear in both lists and thus are prominent because of both their in-degree and out-degree network connectivity. This shows their relative importance in this community as their messages resonated within the community (as indicated by their high in-degree centrality) and as they actively engaged others (as indicated by their high out-degree centrality). The remaining individuals on this list also have relative high values of the in-degree centrality (10 or more).

One anomaly is the community’s account @hcsmca. A review of its recent tweets reveals the account primarily posts announcements about upcoming Twitter chats for this community, mentioning Twitter handles of moderators and other special guests, but with little follow-on interaction with others. This suggests a potential method for identifying such accounts in order to exclude them from analyses of social networks: stark differences between in-degree and out-degree centrality may indicate a non-human, or non-community participant within a conversation.

Table 3. Top 11 users ordered by out-degree centrality.

Twitter handle	Centrality in degree	Centrality out degree	Twitter profile description
colleen_young	36	27	Community Manager of Virtual Hospice | Portail en soins palliatifs (@VirtualHospice), Founder of #hcsmca, plain language writer, health literacy advocate
natricer	11	22	retired PSW, family caregiver, passionate about elderly/vulnerable #hcsmca #dwdchat #eolchat #eldercarechat #caregivingchat #HCLDR #theWalkingGallery
naheedd	18	12	Medical resident physician. #GlobalHealth+#SDOH advocate. #MedEd+#hcsmca enthusiast. RA @CRICH_StMikes. Writer @HealthyDebate. Humanist. Change agent. Optimist.
alainabcyr	14	12	Learn, share, create. Grow. Aspiring expert in patient education and health communications on the Web and in print. Opinions are my own
samdunsiger	12	12	Freelance writer. Communicator. Volunteer #PR director for @SOSheadoffice, @stuttersocial. Caffeine and sushi addict.
hcsmca	0	11	Health Care Social Media Canada #hcsmca hosts a tweet chat every Wednesday at 1 pm EST (2 pm AST, noon CST, 11 am MST, 10 am PST).
cmaer	33	10	Online editor for the Canadian Medical Association. Views are my own
anneccpa	14	10	Canadian Certified #PhysicianAssistant, practicing in #Orthopaedics - Sports Medicine & Trauma. Blogger & advocate for the Physician Assistant Profession.
craigtyyz	13	10	Communications professional: digital, online and social media specialist at Women's College Hospital. Views expressed are my own and not those of my employer.
kathykastner	14	9	love listening learning sharing. appreciate humour happiness and eating chocolate covered almonds. End of life goal: a joyous exit. BestEndings.com
quintepediatric	10	9	We provide medical care to infants, children and adolescents. Healthy kids energize our community! Our account is managed by Sara.

Actor Roles

In the second part of the analysis, we were interested in learning more about the professional composition of this community and whether *professional roles* affect an individual's position in the network. To address this, we first manually classified each Twitter user in the dataset into one of 11 roles (see Table 4). The classification was based on information in the user's public Twitter profile. If information provided on Twitter was not sufficient, we followed links to the user's personal website or LinkedIn page (if provided in their Twitter profile). For the purposes of analysis, users with multiple professional roles were listed as whichever they listed first in their own self-description.

Figure 4 shows the distribution of professional job classifications. The majority of participants in #hcsmca fit the category of "Social media health content providers", describing themselves as dedicated to health topics and/or social media groups with a stated purpose of spreading health information. The second largest group was "Communicators", but not those exclusively focused on health. These were mostly social media marketers whose relation to the network seemed topical or client-based. The third largest group was health-related "Communicators". Although the first three groups were providers of social media health content or communicators, generally speaking, the #hcsmca network is relatively diverse as it also includes a number of health professionals, health institutions, advocacy groups along with health students, educators, and others. The smallest group, with only 4 representatives, was "Government and health policy makers".

Absolute counts of the number of members in a particular professional group do not necessarily reflect the importance of any particular professional group in the network. Thus, to see whether any particular group was especially important in this network, an analysis of variance was conducted comparing in-degree centrality by group. We found a statistically significant relationship between professional roles and in-degree centrality (explaining about 7% of the variance, P=.003, using 5000 permutations), indicating that some professional groups are more influential in this community.

Next, we attempted to determine which professional groups were more or less likely to influence discourse in this group. Based on the average in-degree centralities for each of the 11 professional groups (see Table 5), social media health content providers were the most influential group with an average in-degree centrality of 2.89. (Notably, this group is also a clear leader based on the average out-degree centrality.)

The importance of this group in this network can also be visually observed in the graph representation of this community in Figure 4. In this graph, each node represents a Twitter user in this community, and the line connecting any 2 nodes means that there was at least one mention or reply between the 2 users in the network. This network graph shows that social media health content providers (displayed in the light green color) occupy key central positions in this network and often play a bridging role connecting members from other clusters of this network.

Interestingly, although there are fewer health professionals, educators, and health institutions in this community, their average in-degree centrality came in second, third, and fourth (see Table 5), indicating their relative importance in the network. By contrast, communicators, regardless of their strong presence in this network, were not as central as a group as the three groups just mentioned (despite a few nodes that appear to be in a star-configuration in this network). This may suggest that perhaps communicators are participating in this community because it is part of their job description, but they may not have a lot to contribute, and/or they are there to learn more about this subject matter and are tasked with reporting what they find back to their organizations. Future research is needed to explore this.

Another important observation is that although there seems to be a relationship between professional role and in-degree centrality, there is no apparent preferential attachment among people in the same professional group. In other words, the formation of connections among community members is not necessarily constrained by their professional status. This finding was supported by an analysis of variance density test using both the Structural Blockmodel technique (it examines “whether the different classes have significantly different interaction patterns”), and also with the Variable Homophily model (which “assumes that each group or class of actors has a different homophilic tendency” [42]; where homophily is the tendency for connection based on social similarity). Based on this test (run with the 5000 permutations), the professional roles explain only 0.2% of the total variance (*P*=.005) when run with the Structural Blockmodel and only 0.1% (*P*<.001) with the Variable Homophily model. This result indicates connections are more prevalent across members with different professional backgrounds and occupations in this community, which in turn may suggest that this is a welcoming environment that stimulates knowledge exchange and learning across professional boundaries.

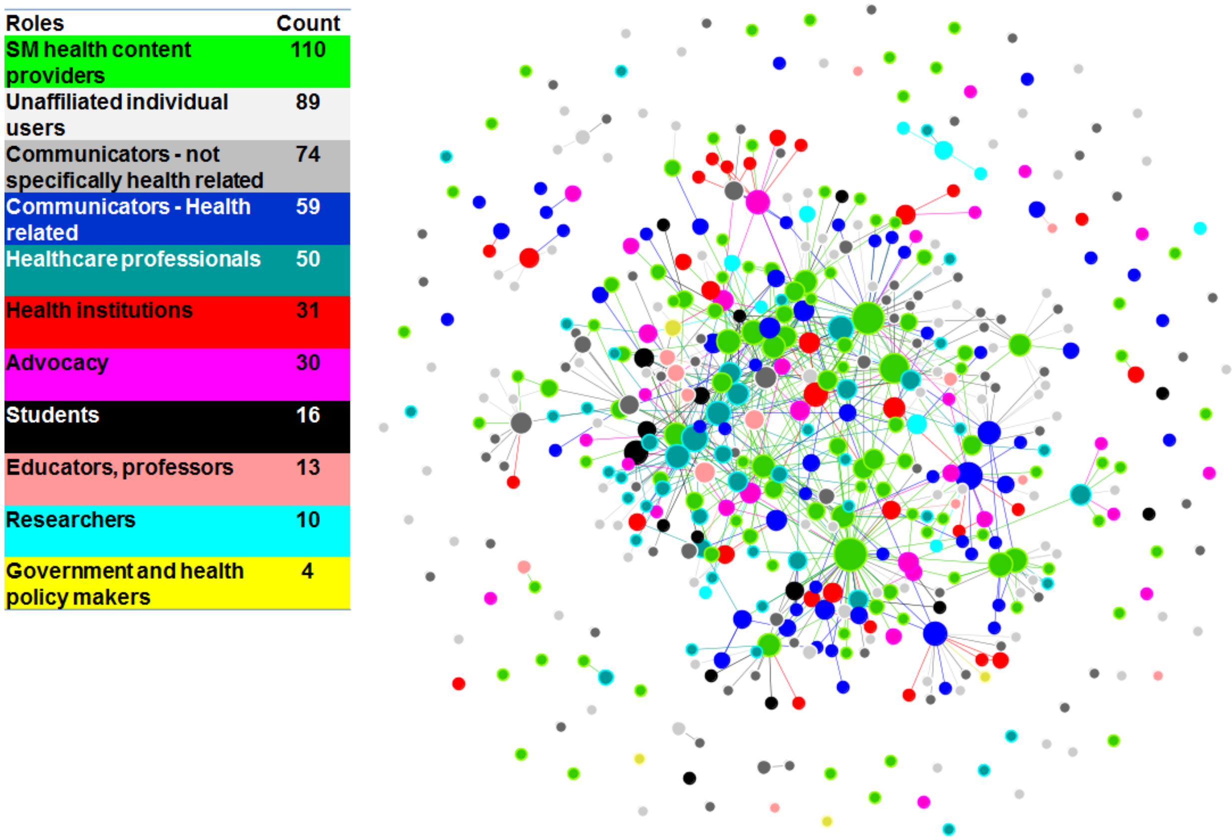
Table 4. Professional roles.

Category	Sample profile of a Twitter user classified under this category
Advocacy	@PatientsAssocCa - The Patients’ Association of Canada promotes the role of the patient in all areas of health care. Follow: Donate:
Communicators—health related	@Infoway - News & announcements from Canada Health Infoway. Tweets by a team from Infoway.
Communicators—not specifically health related	@bobbigreenberg - Dynamic communications & public affairs consultant. Mentor & Coach. Passionate about learning new languages, travel, teaching yoga. Stop, pause and breathe.
Educators, professors	@jendlake - Assistant Professor & Pharmacist. Collaboration/ communication will improve patient-care. Tweets are mine and include primary care, medications and good food
Government and health policy makers	@healthcouncilca - The Health Council of Canada reports on the progress of health care renewal and on innovative practices in Canada.
Health institutions	@QuintePediatric - We provide medical care to infants, children and adolescents. Healthy kids energize our community! Our account is managed by Sara.
Health care professionals	@DrJenGunter - OB/GYN, writer, sex health expert, defender of evidence-based medicine. I wield the lasso of truth. Tweets are not medical advice. I speak for no one but me.
Researchers	@CBoC_HIPE - Independent, leading-edge policy research from the Health Innovation, Policy and Evaluation team at the Conference Board of Canada.
Social media health content providers	@HeartSisters - On women & heart disease from the unique perspective of Carolyn Thomas, a Mayo Clinic-trained heart attack survivor/women’s health advocate. Also
Students	
Unaffiliated individuals	@JEANIESBEACH - music, dance; fashion, women’s rights

Table 5. Average centrality per professional group.

Role	Average in-degree centrality	SD	Average out-degree centrality	SD
Social media health content providers	2.89	6.17	2.21	3.96
Health care professionals	2.48	4.45	1.86	2.28
Educators, professors	2.00	2.97	1.31	1.65
Health institutions	1.65	2.51	1.23	1.71
Advocacy	1.47	2.50	1.10	1.18
Communicators—Health related	1.39	3.03	1.32	1.31
Students	1.38	2.68	1.88	2.03
Researchers	0.90	1.45	0.90	0.99
Government and health policy makers	0.75	1.50	0.50	0.58
Communicators—not specifically health related	0.68	2.18	1.34	1.67
Unaffiliated individual users	0.08	0.34	1.06	1.11

Figure 4. Twitter communication network on #hcsmca colored by professional roles, nodes sized by in-degree centrality.



Roles	Count
SM health content providers	110
Unaffiliated individual users	89
Communicators - not specifically health related	74
Communicators - Health related	59
Healthcare professionals	50
Health institutions	31
Advocacy	30
Students	16
Educators, professors	13
Researchers	10
Government and health policy makers	4

Discussion

Principal Findings

This and previous studies in this area have highlighted how online communication extends the possibilities of community, that is, bringing participants together to form communities of interest for those geographically remote from one another [45]; augmenting geo-community through online information and forums for interaction related to local events and conditions [46]; and extending interaction times and methods through online/offline combinations, as in new forms of online and blended learning, and in the way our communications (text, email, Internet) cross devices (phones, tablets, computers) and contexts (home, work, office). New analytical techniques also push the definition by discovering community on the basis of online interaction, suggesting new definitions and considerations around what constitutes community and what criteria we will accept for identifying it [17].

The current study highlighted some initial observations of the structure of the community formed around the #hcsmca hashtag. As has been asked in the past, how can a group of individuals who meet online, through the lean medium of Twitter, and the constraints of a 140-character posting, sustain and be considered a community? Results from our limited sample set suggest this has happened through a strong core of active participants including the group founder, who lead in posting and prominence in the network. Attention to others is an important aspect of community, and the measures of influence and prominence presented here show that attention to others in the group exists, with key players recognized through mentions and retweeting. The configuration of the community and communication by role suggests one major component connecting all participants, that is, the conversation is not fragmented into isolated cliques. Weekly discussions provide a boost to interaction that stimulates activity and provides a dependable rhythm to interaction patterns and a site to return to each week.

Many studies of community and online community have taken place before ours. The following are some of the observations derived from the results of this and related studies, with commentary on the impact in relation to the #hcsmca community studied here.

Leadership

Some notable attributes of community reported from many studies and associated with both online and offline collectives are local language, shorthands and in-group signifiers [47,48]; group-defined genres, rules of conduct, and policing of conduct [47,49-53]; and interpersonal self-disclosure, emotional support, and shared history (eg, [2,54-58]). In their joint definition of behaviors and practices, there is also attachment to aspirations for a shared future, for example, in group adoption of shared goals and missions, or in the expectation that practices as they exist will be honored and valued in the future. Shared expectations about future commitments enhance trust in the community and its members [59].

In #hcsmca, the very use of Twitter is the community genre, shorthand, and local language. The weekly discussion gives promise of a shared future, as does the general attention to issues relating to health care in Canada and working in this area. While an analysis of tweet content is necessary to discover more about the relations connecting individuals, retweets have provided evidence of attention to others' comments and thus an orientation to community members.

Developing community further depends on continuing attention to the kinds of outcomes that have been found to characterize community, both by design and by emergence from community interaction. Earlier work on online communities and virtual teams has revealed the way rules and norms emerge and evolve with community interaction, with the direction of emergence depending on both technological affordances and the salience to participants of social, informational, and technical features [50,60,61]. Others have noted the need for initial contribution by altruistic or proactive communicators who build the critical mass of participants and participatory interactions [62,63] and create the "safe space" for interaction [49]. However, communities need to move on from these key communicators or risk the demise of the community when such actors leave [44].

Altruistic, proactive use by the #hcsmca founder and by core users remains an important feature for building this community. In looking to the future, the community may face opportunities and challenges in incorporating more and new technology into its repertoire as it expands and as new needs arise that extend the reach and scope of the community. In each round of such expansion, core participants may again have to lead and stimulate contribution and participation as they help develop the character of their community.

Participation

Along with leading a community, there is also the experience of those who lurk, listen, join, participate, and depart from communities. Joining an online community is much like joining any community in the need to learn the norms of behavior, the language used, and who is who among members. Online, this is accomplished through observation and (usually) text-based communications. Joining entails phases. Studies of online learning communities revealed stages of *joining*, *maintaining presence*, and of *disengaging* from the online community [56,64]. Joining can entail learning the norms of the environment, for example, learning how to express oneself in the 140 characters of a tweet. Joining often entails a stage of observation, for example, reading but not posting to online conversations.

It is still a question of how, and why, community can be formed and sustained via text-based communications. Early opposition to the notion of "virtual community" pointed to the lack of nuance of face-to-face interaction and the "leanness" of text as a basis for interaction. Critics noted difficulties in conveying tone, emotion, intimacy, and complex information, and the lack of personal identity and accountability with anonymous participants or the use of pseudonyms (online "handles"). Yet, online communicators found ways around these shortcomings, quickly and easily adopting means of conveying information, enjoying their anonymity, and expressing emotion through texts (eg, with emoticons). Early explorers of these new regions were able to observe the reformation of social and communal ties through online means as these were "uncoupled" from face-to-face interaction (eg, [14-16,45,47,56,65]).

Yet, another effect observed for these lean media has been the reduced inhibitions associated with communicating, for example, the ability to talk through text without face-to-face contact or the need for immediate response. This can be an important feature that encourages new career professionals to communicate (as in #hcsmca), or patients to discuss emotional experiences [2,8,9].

Although our analysis did not focus on newcomers, the results and overall structure of the network suggests that a significant number of isolates, not connected in the Name Networks that signify attention to others (see Figure 2). Such lurkers, while often considered negatively, can also be at the positive stage of what Lave and Wenger termed *legitimate peripheral participation* when new, potential members learn and immerse in the norms and knowledge of the community [43]. However, an overabundance of lurkers can put posters on the spot, inhibit the communal aspect of the site, and fail to create the interactivity necessary for long-term viability [66]. Moreover, while all participants may benefit, individuals may benefit less when lurking rather than participating, as found in a study of a breast cancer support group [5]. Thus, it is important to have new members take up the conversation and participate. Where #hcsmca leaders can become aware of what motivates these isolates, it may help to understand how to support their greater presence in the community.

Online/Offline Synergies

While social media may be considered in their online context only, it has long been recognized that media are not used in isolation from offline interaction and that they are instead embedded in everyday life [67]. This is even more true today as mobile devices, wireless networking, and mobile phone connections afford communication anywhere, anytime [68,69]. We weave and juggle social, learning, and work interactions across media, and across home, school, and work boundaries [46,70-72]. Similarly, while social media may be considered one at a time, relationships, communities, and information behaviors are more often maintained through various media. Several studies have shown that those who maintain closer ties use more media to communicate [73], that is, those who have a greater need or desire to communicate use more of the available media to do so. More forms of interaction, for example, through multiple connections to others, can also increase the value of engagement. This can include using more features of a site: Web access logs of the use of the site PatientsLikeMe, showed those who used more features perceived greater benefit from using the site [7].

Media are not used in isolation but as part of a repertoire that affords connection to resources and to others. This repertoire also includes face-to-face interaction and can support blended learning [70] and blended health [6]. Online interaction provides the opportunity for continuing interaction, learning and care across specializations, disciplines, institutional venues, and structured meeting times. The media then become a tool to facilitate patient-centered, collaborative care [6].

Moreover, it is not just delivery of information that is involved in these collaborations. From considerations of community, attention has expanded along with new forms of social media to consider different forms of interaction, from the friend relationships of Facebook and other social networking sites [11,74,75] to the benefits and interaction patterns associated with participatory culture, peer production [9,76-79], and crowdsourcing for open collaborations and commercial applications [7,8,10,80-85].

Our analysis of HCSMCA did not focus on the multiple platforms that participants use for health care information and conversation, or the way #hcsmca fits with other parts of participants' lives. As such, results about community based solely on Twitter interaction have the potential to underestimate the foundations for community that come from joint and shared interaction across networks and platforms. Again, this is something worthy of further analysis and of interest to community builders as they consider how Twitter works with other venues to help support their community.

Conclusions and Future Directions

We asked at the outset what accounts for the relative longevity of this particular online community and have found that it is based on interaction patterns of participation and prominence of the group's founder and a small core of key participants who are heavily engaged in social media and health care networks. Longevity is supported by the structure of weekly discussions, which creates a communal structure for interaction. We also asked what constitutes the composition of this community and found that it consists of individuals who may be classified as belonging to a number of different roles, but that communication flows across roles and thus the network reveals a community of one major component. The network also shows a large number of individuals who are present in the community but not actively connected to others and who may be benefitting from observation of the conversations and are potential future active participants.

Our purpose has been to show how a social network analysis can reveal such patterns and how past work on online community can help in interpretation of such results for the creation and maintenance of online communities for health. In brief, the implications and recommendations are:

- Leaders and core participants can seed a network by altruistic or proactive use that, initially, provides more benefit to others than they receive in return. However, for long-term sustainability that persists beyond leadership change, the network needs to grow in a way that distributes leadership and participation beyond single leaders.
- Prominence in the network appears to be related to familiarity with individuals, for example, more active participants receive more attention in terms of mentions and retweets. Thus, a recommendation is that moderators of discussions build authority in the network prior to their moderation duties to be able to connect better with ongoing discussions.
- More prominent actors are engaged in multiple networks relating to health matters. As these actors also bridge networks, they are able to carry the message of the network to others. Thus, a recommendation is to engage these types of actors as a way of increasing the reach and prominence of the network itself.
- Peripheral participants represent untapped resources for the network. Finding out what motivates such participants can help identify those who will make contributions in the future and thus how to bring their participation into the community.
- Network analysis and visualizations provide a set of techniques and a vocabulary about network interactions that can help both group leaders and participants to see the size, shape, and configuration of the network in order to gain a better understanding of its operation and the place of individuals in that operation. Attention to roles can reveal both emergent roles (eg, core participants) as well as show the influence of existing roles (eg, different medical or sector roles).

This analysis of one social media site highlights the way social network analysis can be used to gain an understanding of social media use for communication and conversation and how network formations support such communities. However, this example has barely covered the beginnings of potential applications. Some key questions that remain and can form the basis of future work are: How do we implement and measure the impact of social media on health for individual patients and for the general population? What single and/or combination of media provide interaction around health that is effective over the short and long term? What combinations of participation and contribution create interest and sustain communities that discuss and continue to apply better practices for health and well-being? The task is complex as it requires understanding the rapidly changing and expanding media options in relation to changing institutional and societal practices, yet the opportunity is there.

We believe the principles of social networks and the techniques of social network analysis provide a solid foundation for understanding relationships and their formation online and for taking that into social media practice for health. Attention to network relations emphasizes what we do together, rather than what medium or face-to-face venues we use. This approach has proved useful for understanding the societal turn to online communication, relational maintenance, community genesis, and sustainability. There are already studies and models that have addressed health from a social network perspective (eg, [12,13,19]), and there is much that has addressed online interactions from a social network perspective (eg, [17,86]). As we turn to considering their interaction and the specific application of social media for health (eg, [5,9,11,25]), we look forward to combining these to explore further the interplay of social media, social networks, health, and well-being [8,10,87].

Acknowledgments

This paper was first prepared for the workshop *Harnessing the Social Web: Communities for Health and Wellness*, led by Dr Kendall Ho and sponsored by the Peter Wall Institute of Advanced Studies at the University of British Columbia. The authors acknowledge and thank participants at this workshop for their comments and contributions to our thinking. The work on #hcsmca discussed in this paper was supported in part by a grant to author Gruzd from the GRAND (GRaphics, Animation and New meDia) Network of Centres of Excellence, funded by the Canadian government. The authors thank the contributors to #hcsmca for their open discussion and in particular the group founder Colleen Young. The authors would like to thank Thomas Robbins for his help with the manual classification of Twitter profiles and Samantha Fritz for her help with the manuscript preparation.

Conflicts of Interest

None declared.

References

1. Wenger E, McDermott RA, Snyder WM. Cultivating communities of practice: a guide to managing knowledge. Boston, Mass: Harvard Business School Press; 2002.
2. Rubenstein E. 'It's a microcosm of life': weaving a web of information and support in an online breast cancer community. Unpublished Doctoral Dissertation. Champaign, Illinois: University of Illinois; 2011.
3. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
4. Rubenstein EL. Proceedings of the ASIST 2012 Annual Meeting. 2013 Jan 24. “Things my doctor never told me”: Bridging information gaps in an online community URL: https://www.asis.org/asist2012/proceedings/Submissions/126.pdf [accessed 2013-10-29] [WebCite Cache ID 6KjSHFINe]
5. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. J Med Internet Res 2011;13(4):e122 [FREE Full text] [doi: 10.2196/jmir.1696] [Medline: 22204869]
6. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
7. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
8. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. J Med Internet Res 2012 May;14(3):e54 [FREE Full text] [doi: 10.2196/jmir.1824] [Medline: 22555303]
9. Chou WY, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011 Jan;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
10. Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T. Potential benefits and harms of a peer support social network service on the internet for people with depressive tendencies: qualitative content analysis and social network analysis. J Med Internet Res 2009 Jul;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
11. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
12. Smith KP, Christakis NA. Social networks and Health. Annu Rev Sociol 2008 Aug;34(1):405-429. [doi: 10.1146/annurev.soc.34.040507.134601]

13. Valente T. Social networks and health: models, methods, and applications. Oxford: Oxford University Press; 2010.
14. Rheingold H. The virtual community: homesteading on the electronic frontier. Reading, Mass: Addison-Wesley Pub. Co; 1993.
15. Preece J. Online communities: designing usability, supporting sociability. Chichester: John Wiley; 2000.
16. Preece J, Maloney-Krichmar D. Special theme: Online Communities. Journal of Computer Mediated Communication 2005;10(4):1-10. [doi: 10.1111/jcmc.2005.10.issue-4] [Medline: 24156115]
17. Haythornthwaite C. Social networks and online community. In: Joinson A, McKenna K, Reips U, Postmes T, editors. The Oxford handbook of Internet psychology. Oxford: Oxford University Press; 2007:121-136.
18. Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge, MA: Cambridge University Press; 1994.
19. Pescosolido BA. Of Pride and Prejudice: The role of sociology and social networks in integrating the health sciences. Journal of Health and Social Behavior 2006 Sep 01;47(3):189-208. [doi: 10.1177/002214650604700301]
20. Garton L, Haythornthwaite C, Wellman B. Studying online social networks. Journal of Computer-Mediated Communication 1997;3(1):np. [doi: 10.1111/j.1083-6101.1997.tb00062.x]
21. Scott J. Social network analysis: a handbook. 2nd edition. London: SAGE Publications; 2000.
22. Scott J, Carrington PJ. In: Scott J, Carrington PJ, editors. The SAGE Handbook of Social Network Analysis. London: Sage Publications Ltd; 2011.
23. Kadushin C. Understanding Social Networks: Theories, Concepts, and Findings. Oxford: Oxford University Press, USA; 2011.
24. Prell C. Social Network Analysis: History, Theory and Methodology. London: Sage Publications Ltd; 2012.
25. Hackworth BA, Kunz MB. Health care and social media: building relationships through social networks. Academy of Health Care Management Journal 2011;7(2):1.
26. Granovetter MS. The strength of weak ties. American Journal of Sociology 1973;78:1360-1380.
27. Lin N. Social capital: a theory of social structure and action. Cambridge, UK: Cambridge University Press; 2001.
28. Putnam RD. Bowling alone: the collapse and revival of American community. New York: Simon & Schuster; 2000.
29. Burt R. Structural holes: the social structure of competition. Cambridge, Mass: Harvard University Press; 1992.
30. Harrison RT, Leitch CM. Entrepreneurial learning: conceptual frameworks and applications. London: Routledge; 2008.
31. Hansen DL, Shneiderman B, Smith MA. Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Massachusetts: Morgan Kaufmann; 2010.
32. Viegas F. Visualization culture: data literacy for the rest of us. 2013 Jan 8 Presented at: 46th Hawaii International Conference on System Sciences; January 2013; Maui HI.
33. Viegas FB. Democratizing visualization. URL: http://fernandaviegas.com/democratizing_viz.html [accessed 2013-10-29] [WebCite Cache ID 6KjUkE1Wd]
34. Gruzd A, Haythornthwaite C. Networking online: cybercommunities. In: Scott J, Carrington PJ, editors. The SAGE Handbook of Social Network Analysis. London: Sage Publications Ltd; 2011:167-179.
35. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
36. Young C. CY Health Communications. URL: http://cyhealthcommunications.wordpress.com/hcsmca-2/ [accessed 2013-10-29] [WebCite Cache ID 6KjUwA067]
37. Netlytic. URL: http://netlytic.org/ [accessed 2013-10-29] [WebCite Cache ID 6KjVAugjP]
38. Gruzd A. Exploring virtual communities with the Internet Community Text Analyzer (ICTA). In: Daniel B, editor. Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena. Hershey, PA: Information Science Publishing; 2011:205-223.
39. Gruzd A. Studying collaborative learning using name networks. Journal of Education for Library and Information Science 2009;50(4):243-253.
40. Gruzd A, Haythornthwaite C. The analysis of online communities using interactive content-based social. Proc Am Soc Info Sci Tech 2009 Jun 01;45(1):1-5. [doi: 10.1002/meet.2008.1450450318]
41. Carley KM, Pfeffer J, Reminga J, Storrick J, Columbus D. ORA Use's Guide 2012 (No CMU-ISR-12-105). Pittsburgh, PA: Institute for Software Research, Carnegie Mellon University; 2012.
42. Borgatti S, Everett MG, Freeman LC. UCINET for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies; 2002.
43. Lave J, Wenger E. Situated learning: legitimate peripheral participation. Cambridge, UK: Cambridge University Press; 1991.
44. Bruckman A, Jensen C. The mystery of the death of Mediamoo: seven years of evolution of an online community. In: Building virtual communities learning and change in cyberspace. New York: Cambridge University Press; 2002:21-33.
45. Sproull L, Kiesler S. Connections: new ways of working in the networked organization. Cambridge, Mass: MIT Press; 1991.
46. Haythornthwaite C, Kendall L. Internet and community. American Behavioral Scientist 2010 Feb 19;53(8):1083-1094. [doi: 10.1177/0002764209356242]

47. Smith CB, McLaughlin ML, Osborne KK. Conduct control on Usenet. Journal of Computer-Mediated Communication 1996;2(4):-.
48. Herring SC. Computer-mediated communication on the Internet. Annual Review of Information Science and Technology 2002;36:109-168. [doi: 10.1002/aris.1440360104]
49. Bregman A, Haythornthwaite C. Radicals of presentation: visibility, relation, and co-presence in persistent conversation. New Media & Society 2003 Mar 01;5(1):117-140. [doi: 10.1177/14614448030050019l3]
50. DeSanctis G, Poole MS. Capturing the complexity in advanced technology use: adaptive structuration theory. OrgSci 1994 May 01;5(2):121-147. [doi: 10.1287/orsc.5.2.121]
51. Orlikowski WJ, Yates J. Genre repertoire: the structuring of communicative practices in organizations. Administrative Science Quarterly 1994;39:541-574.
52. Miller CR. Genre as social action. Quarterly Journal of Speech 1984 May;70(2):151-167. [doi: 10.1080/00335638409383686]
53. Miller CR. Rhetorical community: the cultural basis of genre. In: Freedman A, Medway P, editors. Genre and the new rhetoric. London: Taylor & Francis; 1994.
54. Baym NK. Tune in, log on: soaps, fandom, and online community. Thousand Oaks, Calif: Sage Publications; 2000.
55. Miller D, Slater D. The Internet: An Ethnographic Approach. Oxford, UK: Berg Publishers; 2000.
56. Haythornthwaite C, Kazmer MM, Robins J, Shoemaker S. Community development among distance learners: temporal and technological dimensions. Journal of Computer-Mediated Communication 2000;6(1):-. [doi: 10.1111/j.1083-6101.2000.tb00114.x]
57. Hearne B, Nielsen A. Catch a cyber by the tale: online orality and the lore of a distributed learning community. In: Haythornthwaite C, Kazmer MM, editors. Learning, culture, and community in online education: research and practice. New York: Peter Lang Publishers; 2004.
58. Rebaza C. The Technological Continuum of Coterie Publication: Fan Fiction Writing Communities. Unpublished Doctoral Dissertation. Champaign, IL: University of Illinois; 2009.
59. Burt RS. The network structure of social capital. Research in Organizational Behavior 2000 Jan;22:345-423. [doi: 10.1016/S0191-3085(00)22009-1]
60. Orlikowski WJ, Yates J, Okamura K, Fujimoto M. Shaping electronic communication: the metastructuring of technology in the context of use. Organization Science 1995 Jul 01;6(4):423-444. [doi: 10.1287/orsc.6.4.423]
61. Spears R, Lea M, Postmes T. Social psychological theories of computer-mediated communication: social pain or social gain? In: Robinson WP, Giles H, editors. The new handbook of language and social psychology. Chichester, England: J. Wiley; 2001:601-623.
62. Markus ML. Toward a 'critical mass' theory of interactive media. In: Fulk J, Steinfield CW, editors. Organizations and communication technology. Newbury Park, Calif: Sage Publications; 1990:194-218.
63. Fulk J, Steinfield CW. Organizations and communication technology. Newbury Park, Calif: Sage Publications; 1990.
64. Kazmer MM. Beyond C U L8R: disengaging from online social worlds. New Media & Society 2007 Feb 01;9(1):111-138. [doi: 10.1177/1461444807072215]
65. Haythornthwaite C, Nielsen A. Revisiting computer-mediated communication for work, community, learning. In: Gackenbach J, editor. Psychology and the internet: intrapersonal, interpersonal, and transpersonal implications. Amsterdam: Elsevier/Academic Press; 2007:167-186.
66. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. Computers in Human Behavior 2004 Mar;20(2):201-223. [doi: 10.1016/j.chb.2003.10.015]
67. Wellman B, Haythornthwaite C. The Internet in everyday life. Malden, MA: Blackwell Pub; 2002.
68. Ling R, Ling RC. New tech, new ties: how mobile communication is reshaping social cohesion. Cambridge, Mass: MIT Press; 2008.
69. Ling R, Stald G. Mobile communities: are we talking about a village, a clan, or a small group? American Behavioral Scientist 2010 Mar 17;53(8):1133-1147. [doi: 10.1177/0002764209356245]
70. Haythornthwaite C, Andrews R. E-learning Theory and Practice. London: Sage Publications Ltd; 2011.
71. Haythornthwaite C, Kazmer MM. Bringing the Internet home: adult distance learners and their Internet, home and work worlds. In: Wellman B, Haythornthwaite C, editors. The internet in everyday life. Malden, MA: Blackwell Pub; 2002:431-463.
72. Allen M, Foth M. Research in action for community informatics: a matter for conversation. Journal of Community Informatics 2011;7(3) [FREE Full text]
73. Haythornthwaite C. Social networks and Internet connectivity effects. Information, Communication & Society 2005 Jun;8(2):125-147. [doi: 10.1080/13691180500146185]
74. Boyd D, Ellison N. Social network sites: definition, history, and scholarship. IEEE Eng Manag Rev 2010;38(3):16-31. [doi: 10.1109/EMR.2010.5559139]
75. Ellison NB, Steinfield C, Lampe C. Connection strategies: social capital implications of Facebook-enabled communication practices. New Media & Society 2011 Jan 27;13(6):873-892. [doi: 10.1177/1461444810385389]
76. Jenkins H, with Clinton K, Purushotma R, Robinson AJ, Weigel M. Confronting the Challenges of Participatory Culture: Media Education for the 21st Century. Chicago, IL: MacArthur Foundation; 2006.

77. Ito M. Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media (John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning). Cambridge, MA: The MIT Press; 2009.
78. Raymond ES. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary (O'Reilly Linux). Cambridge, MA: O'Reilly; 1999.
79. Benkler Y. The wealth of networks: how social production transforms markets and freedom. New Haven [CT]: Yale University Press; 2006.
80. Howe J. Wired. 2006 Jun. The Rise of Crowdsourcing URL: http://www.wired.com/wired/archive/14.06/crowds.html [accessed 2007-05-31] [WebCite Cache ID 6HQBhiyjA]
81. Haythornthwaite C. Crowds and communities: light and heavyweight models of peer production. In: Proceedings of the 42nd Hawaii International Conference on System Sciences. Los Alamitos, CA: IEEE Computer Society; 2009 Presented at: HICSS; Jan 2009; Big Island, HI. [doi: 10.1109/HICSS.2009.137]
82. Budhathoki NR, Haythornthwaite C. Motivation for open collaboration: crowd and community models and the case of OpenStreetMap. American Behavioral Scientist 2012 Dec 28;57(5):548-575. [doi: 10.1177/0002764212469364]
83. Forte A, Lampe C. Defining, understanding, and supporting open collaboration: lessons from the literature. American Behavioral Scientist 2013 Jan 11;57(5):535-547. [doi: 10.1177/0002764212469362]
84. Pedersen J, Kocsis D, Tripathi T, Tarrell A, Weerakoon A, Tahmasbi N, et al. Conceptual foundations of crowdsourcing: a review of IS research. In: Proceedings of the 46th Hawaii International Conference on System Sciences. Los Alamitos, CA: IEEE Computer Society; 2013 Presented at: HICSS; Jan 2013; Maui, HI. [doi: 10.1109/HICSS.2013.143]
85. Erickson LB, Petrick I, Trauth EM. Hanging with the right crowd: matching crowdsourcing need to crowd characteristics. In: Proceedings of the Eighteenth Americas Conference on Information Systems. Atlanta, GA: Association for Information Systems; 2012 Presented at: Eighteenth Americas Conference on Information Systems; Aug 2012; Seattle, Washington p. 9-12.
86. Rainie L, Wellman B. Networked: The New Social Operating System. Cambridge, MA: The MIT Press; 2012.
87. Fernandez-Luque L, Karlsen R, Melton GB. HealthTrust: a social network approach for retrieving online health videos. J Med Internet Res 2012;14(1):e22. [doi: 10.2196/jmir.1985]

Edited by G Eysenbach; submitted 26.06.13; peer-reviewed by S Fels, C Young; comments to author 23.07.13; revised version received 03.10.13; accepted 23.10.13; published 31.10.13.

Please cite as:
Gruzd A, Haythornthwaite C
Enabling Community Through Social Media
J Med Internet Res 2013;15(10):e248
URL: http://www.jmir.org/2013/10/e248/
doi:10.2196/jmir.2796
PMID:24176835

Original Paper

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

[*]these authors contributed equally

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:

This is a corrected version. See correction statement: http://www.jmir.org/2013/11/e268/

Abstract

Background: Three-dimensional virtual worlds are becoming very popular among educators in the medical field. Virtual clinics and patients are already used for case study and role play in both undergraduate and continuing education levels. Dental education can also take advantage of the virtual world’s pedagogical features in order to give students the opportunity to interact with virtual patients (VPs) and practice in treatment planning.

Objective: The objective of this study was to design and evaluate a virtual patient as a supplemental teaching tool for pediatric dentistry.

Methods: A child VP, called Erietta, was created by utilizing the programming and building tools that online virtual worlds offer. The case is about an eight-year old girl visiting the dentist with her mother for the first time. Communication techniques such as Tell-Show-Do and parents’ interference management were the basic elements of the educational scenario on which the VP was based. An evaluation of the simulation was made by 103 dental students in their fourth year of study. Two groups were formed: an experimental group which was exposed to the simulation (n=52) and a control group which did not receive the simulation (n=51). At the end, both groups were asked to complete a knowledge questionnaire and the results were compared.

Results: A statistically significant difference between the two groups was found by applying a *t* test for independent samples ($P<.001$), showing a positive learning effect from the VP. The majority of the participants evaluated the aspects of the simulation very positively while 69% (36/52) of the simulation group expressed their preference for using this module as an additional teaching tool.

Conclusions: This study demonstrated that a pediatric dentistry VP built in a virtual world offers significant learning potential when used as a supplement to the traditional teaching techniques.

(J Med Internet Res 2013;15(10):e240) doi:10.2196/jmir.2651

KEYWORDS

virtual patient; virtual world; pediatric dentistry; simulation; Second Life; OpenSim; communication; tell-show-do; behavior management

Introduction

Three-dimensional (3D) virtual worlds such as “Second Life” (SL) and “OpenSimulator” (OpenSim) are becoming increasingly popular in medical education. The embedded characteristics of social networking, collaboration, constructivism, exploration, 3D building, and programming make them excellent candidates for designing teaching tools, various e-learning activities, and simulations. But what is a virtual world? According to a recent definition, a virtual world is a synchronous, persistent network of people, represented as avatars, facilitated by networked computers [1]. By “persistent”, it is implied that this virtual world, along with the changes the users make, continue to exist and evolve while the user is offline. The users of a virtual world take the form of an avatar, which is the “alter ego” of a human being that is usually represented by a 3D humanoid model. Virtual reality differs from virtual worlds. The latter term has been applied to persistent online social spaces—virtual environments that people experience as ongoing over time and that have large populations which they experience together with others as a world for social interaction [2]. Although virtual reality technology has already been applied to dental education, including operative dentistry [3], endodontics [4], orthodontics [5], and implant surgery [6], and seems to have educational benefits, virtual worlds are relatively new to this area. Virtual worlds are similar to 3D online video games, such as the popular “World of Warcraft” and “The Sims Social” on Facebook. According to the Entertainment Software Association, in 2012, 49% of US households owned a dedicated game console and 31% of game players were between 18-35 years old [7]. Recently, Amer et al [8] developed a video game to teach dentin bonding. Their evaluation concluded that this method of teaching was as good as a passive, non-interactive way of teaching and also dental students preferred it to a lecture. Although virtual worlds are focused on building and socializing, these facts may imply that they may also be used for creating e-learning activities that draw students’ attention.

Second Life is a virtual world, developed by Linden Lab in 2003, consisting of an online 3D environment, the avatars of the users, which are called “residents”, and the objects they create [9]. The users connect to SL from a computer using a program called “Viewer”. Residents can explore the world, communicate with each other, rent their own virtual home, and construct 3D objects that they can share or sell in SL’s “Marketplace”. Communication is possible via text messages, speech, and gestures. SL has its own programming language, LSL (Linden Scripting Language). In 2011, SL’s total virtual area occupied 2060 square kilometres. In order to enter the world, the user must possess an Internet connection. No offline mode is supported. The official SL viewer can be downloaded from the SL website. Third-party viewers are also available.

OpenSim is an open-source virtual world server, very similar to SL and compatible to SL’s viewer and programming language [10]. OpenSim was launched in 2007 and allows users to build their own virtual worlds and operate in online or offline modes. The latter is the main advantage of OpenSim over SL; moreover in SL, the users must own or rent a region of virtual land in order to build interactive objects. Furthermore, that land has a limitation imposed on the number of primitive geometric shapes (called “primitives” or “prims” for short). On the other hand, OpenSim is free of charge and supports any number of prims.

Until today, there have been some good efforts to create medical simulations in virtual worlds. Ohio State University’s “Medical Center” is a virtual building in SL for educational role-play and case studies [11]. “Ann Myers Medical Center” is a complete virtual hospital with detailed interactive apparatuses [12,13]. In the “Respiratory Ward” (Imperial College London), a student can meet virtual patients, listen to their breathing sounds, and make a diagnosis [14]. Auckland University’s “Medical Center” has an emergency department, an ambulance, and virtual classrooms [15]. Creutzfeldt et al [16,17] carried out a scenario-based team training of cardiopulmonary resuscitation using avatars in a virtual world. This activity seemed to be engaging and elicited positive changes in students’ subjective experiences. Wiecha et al [18] designed a postgraduate medical education program in a virtual world and concluded that the virtual world may be used in continuing medical education in order to enhance learning outcomes. Other interesting examples include “Second Health” by Imperial College London [19] and “MUVE” on “Evergreen Islands” [20]. A case which is better related to the interests of dental students is the Kentucky University virtual anatomy lab in SL, illustrating 3D maxillofacial models containing nerves and vessels [21].

Dental educators can also use virtual worlds in order to create clinical scenarios, to allow students to interact with virtual patients and practice in diagnosis and treatment planning [22,23]. There are some interesting examples of dental education applications in virtual worlds. The virtual building of Maryland Dental School in SL was created in 2009. It features a Dental Hygiene clinic, a Pediatric Dentistry room, interactive dental units, lecture rooms, and a small museum [24]. Some dental units are intended for role-play and others for self-assessment, containing case studies with images and dental history. Team-learning activities also take place on the school’s virtual island [25]. Another effort comes from Case Western Reserve University Dental School, pointing to role-play in a virtual treatment room [26].

For playing simulated scenarios with avatars, at least two users must be connected in the virtual world; one plays the patient and the other plays the doctor. Role-play is used mainly for training in interviewing techniques [27] and other nontechnical skills such as communication [28,29]. An instructor and other students can be online too, so that discussions and team activities are feasible. Difficulties arise from the fact that a second user may not always be available nor act successfully as a simulated patient. This is where Virtual Patients (VPs) apply. A VP is a specific type of computer program that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions [30]. By utilizing VPs, students acquire the role of a doctor in a safe and controlled environment where they can develop clinical and communicative skills without the risk of disturbing or hurting people. VPs can be used for practicing interviewing [31], clinical reasoning [32], or even to facilitate the teaching of medical ethics, medical law, or medical professionalism [33]. VPs can also simulate different psychological states of a patient (angry, worried, happy, etc.) and different personalities by proper combination of programming and graphics. This is very important for the development of a student's communication skills. Janda et al [34] found that the use of VPs improved the capability of dental students to take a health history. Another study [35] investigated the use of VPs in dental care for persons with special needs and concluded that dental students demonstrated improved communication skills and became more effective when caring for such patients.

In a recent survey [36], 63.3% of dental schools in the United States and Canada indicated that they are currently using or, at some point in the past, have used VPs in training dental students, while over 80% of the respondents, most of whom were students, seemed to enjoy the use of VPs and considered them advantageous in dental training. In a similar survey [37], 24% of the schools had developed VP case scenarios.

In contrast to adult patients, children are more difficult to simulate due to their complex or unpredictable behavior during dental practice. A dental student or an instructor can easily act as an adult patient during role-play but when it comes to pediatric dentistry, real patients are ideal for teaching behavioral techniques. Unfortunately, this may not be feasible due to practical and ethical limitations. A possible solution is to teach children to act as simulated patients in various clinical conditions [38], but considering the cost and difficulty of this procedure, it may be preferable to follow a more traditional teaching method. Interactive manikin models, multimedia software, virtual patients, or a combination of these three may be helpful in acquiring communication and behavioral skills in pediatric dentistry. Boynton et al [39] developed an Internet-based instruction tool (The Virtual Child) to simulate clinical experience in the dental treatment of a child. This study found that students who had been exposed to this simulation performed significantly better on an examination regarding knowledge of pediatric behavior management than did the control group. Kleinert et al [40] created an offline multimedia-based virtual patient model involving a dental visit for a child with Down syndrome. The study, involving 51 dental students, showed significant changes in both knowledge and perceived difficulty levels for the participants as a result of completing the module.

For our study, a virtual child patient was developed to support training of communication and behavior management in pediatric dentistry. A small 3D clinic was set in a virtual world. Our objective was to examine whether this simulation would result in increased knowledge when used as a supplementary teaching tool compared to the traditional lectures alone and to provide an evaluation of its features by the dental students.

Methods

"Erietta" is a virtual child patient designed by the authors, built in a virtual world combining 3D graphics, LSL programming, educational software principles, and communication management in pediatric dentistry. For this simulation, a small virtual dental clinic was constructed using prims and divided into two rooms. In the Tutorial Room, an interactive presentation board is placed on the wall. The user can watch slides illustrating basic techniques of communication and behavior management in pediatric dentistry. This e-learning content is based on AAPD (American Academy of Pediatric Dentistry) guidelines [41]. The Treatment Room (Figure 1) features two dental units with doctors' seats, a digital radiograph, cabinets, water sink, and a scrubs locker. The latter contains a wearable dental uniform. A box with examination gloves is also available. A mirror and a probe are placed on the unit's tool tray. The humanoid models of Erietta and her mother are placed standing next to a dental unit. All 3D items were designed combining "Blender" (3D software) [42] and the special build tools of SL's viewer. The mother's model was purchased from SL's marketplace. Synth (synthesized) voice and other sounds such as the beeping of the x-ray were incorporated in the models to provide a more realistic simulation experience. Most of the items were animated by using special LSL functions; for example, Erietta can sit, stand, and raise her hand, the radiograph device can be unfolded, the chair can move up/down, and the light can be turned on/off.

According to our educational scenario, Erietta, an eight-year-old girl, with her mother, Mary, is visiting the dental clinic for the first time in order to have Erietta's oral health checked. The child must first sit in the chair and then receive a simple examination. During the visit, the user will have to encounter the child's fear and mother's interference (Figure 2). The goal is to earn Erietta's trust and remove any anxiety by applying basic behavior and communication techniques. The scenario consists of six parts (A: welcome, B: interference, C: examination, D: distraction, E: x-ray, and F: goodbye). Each one of them is implemented using multiple-choice questions and sound/text feedback. Conversation, Tell-Show-Do, word substitution (euphemism), distraction, and positive reinforcement are the basic methods of communication that the user is asked to apply. A sample question is shown in Table 1. Every correct answer is accompanied by a positive reaction from the child, whereas a wrong choice makes Erietta anxious and provides hints to the user. The feedback is provided in the terms of synth speech, short text appearing over the models' heads, and detailed instructions through the chat window. Positive feedback is marked with blue fonts while negative feelings are colored red.

Fifteen multiple-choice questions and their feedback were prepared and imported into Erietta using LSL and SL's notecards. For question reading and answer finding, LSL's API Dataserver was used. The user chooses the correct answer and proceeds to the next question. A diagram of the story is illustrated in Figure 3. There are some prerequisites in order to advance to each next step of the simulation; the user must wash hands and wear gloves prior touching the tools. A bare-hands touch event on the tools will result in a warning voice message advising the user to wear gloves. Similarly, an interaction with the radiograph device is not permitted unless Erietta is seated in the chair and the user has successfully used the mirror and probe. While an avatar is using the simulation, other users can also be online and watch, but only one avatar at a time is permitted to interact with Erietta. In this way, it is ensured that the trainee's learning activity will be uninterrupted while a level of team-learning is provided, via watching, for the rest of the users. Two versions of the simulation were developed: one for SL and one for OpenSim. Both versions are hosted in the "Virtual Islands for Biology Education" (VIBE) [43] in their SL region and OpenSim server respectively [44,45]. For this study, the OpenSim version was preferred due to its advantageous standalone function. OpenSim server v.0.7.3 and SL viewer v.3.2.8 were installed on the PC lab's computers in the AUTh (Aristotle University of Thessaloniki) Dental School.

A class of 103 undergraduate dental students at the Aristotle University of Thessaloniki was included in this study: 82 of the participants (79.6%, 82/103) were 25 years of age or under and the rest (20.4%, 21/103) were between the ages of 26 and 30. Of the 103 students, 53.4% (55/103) were female and 46.4% (48/103) were male. At the time of the study (2012 spring semester), the students were attending their fourth year of the curriculum and had enrolled in the Pediatric Dentistry course during the past winter semester. This course was lecture-based and included behavior guidance and communication techniques.

The class was divided into two groups. The first one, consisting of 52 students, was the experimental (simulation) group. Each student of this group was exposed to the simulation and then completed a knowledge questionnaire. The second group (control group) consisted of 51 students who answered the same questions without having used the simulation. The study was held during the Dental Informatics course and was completed in one week.

The questionnaire was designed to evaluate students' knowledge of behavior and communication management in pediatric dentistry and consisted of seven objective multiple-choice questions. Each one of the questions focused on one of the following seven techniques: (1) welcoming, (2) tell-show-do, (3) euphemism, (4) parent management, (5) communication, (6) distraction, and (7) rewarding. A short clinical scenario describing a problem similar to the ones that the students faced in the simulation and four possible ways of reaction were offered for each question. Students were asked to select the best one. A sample question is shown in Textbox 1. Additionally, both groups answered six 3-scale questions to assess their basic knowledge of computers (duration, frequency, reason for use, 3D video games experience, and subjective level of computer skills). The students' answers were scored individually, totaled, and entered into a spreadsheet.

The students in the simulation group filled in an evaluation questionnaire with three parts. The first part consisted of 11 Likert-scale questions to evaluate the usability and educational characteristics of the simulation and scenario. The rating was between "1" and "5" with lower scores being more positive. In the second part, students had to respond to the hypothesis of choosing whether or not to use Erietta as a teaching tool if they were instructors. The last part consisted of three open-ended questions concerning the ease of use of the virtual world, improvements and suggestions, and subjective definition of the kind of benefit gained as a result of the simulation experience (practical, cognitive, or psychological).

All the answers were scored, totaled, and entered into a spreadsheet. Then, the answers to the open-ended questions were categorized and a list of possible responses was developed for each question. All the data were exported to be analyzed with the Statistical Package for Social Sciences (SPSS).

Table 1. Sample question for users to answer: "After the "Tell" and "Show" steps using the probe on Erietta's nails, what is the best way to ask Erietta open her mouth?"

Answer	Feedback in chat window	Erietta's speech feedback
1. "Erietta, now open your mouth, so we can count your teeth!"	1. You can say that but it's better to combine elements from the "Show" step, try again!	I don't like it!
2. "…so, we have 10 nails on our hands! Let's see how many teeth are in your mouth! If you feel uncomfortable, raise your left hand to stop counting. Open..."	2. Correct! You applied the "Do" step for the probe.	It's fun!
3. "Now it's time to count your teeth! If anything goes wrong, raise your left hand so I can stop counting. Open..."	3. You used the word "wrong". This may make Erietta anxious. Try again.	Mum, I'm afraid!

Textbox 1. Sample question from the knowledge questionnaire.

Dimitris is ten years old and is sitting in the chair, ready to have a primary tooth extracted. You have administered local anesthesia and explained the steps that will follow. Still, Dimitris looks frightened. How would you react?

A. You remain calm and friendly and define the length of the operation by saying: "I will count slowly to 60 and we will have finished." You also try to distract his attention by audiovisual means (TV, music).

B. You ask his mother to help by further explaining the steps of the procedure. This will calm him down.

C. You speak strictly to Dimitris and ask him to stay calm and still, so the extraction can finish quickly. You also tune the TV on a cartoon channel.

D. You say with calm voice: "I understand you feel worried. Take your time and relax. We will start when you feel ready."

Figure 1. "Erietta" simulation: the treatment room.

Figure 2. Erietta raises her hand, while her mother interferes, showing text and speech feedback.


lazaros1982's 'Dental Probe'
-Erietta raised her hand! Touch her to calm down.
Mary is trying to encourage her daughter. What is the best way to behave now?
1. "Please don't speak again, you make Erietta nervous."
2. I say nothing for the moment.
3. "Please, Mary, have a sit and watch quietly."

NEARBY CHAT

Question:4

Your answer:2

Wrong.

If you say nothing, it is possible that the mother will interfere again.

Figure 3. Diagram of the scenario. A: Initial communication (Questions 1-3). B: Talk to Mary (Question 4), C: Tell-Show-Do, word substitution (Questions 5-10). D: Distraction, ask Mary to be quiet (Questions 11-12). E: Tell-Show-Do, diagnosis (Questions 13-14), F: Say goodbye (Question 15).

Results

Students of both groups completed the knowledge (objective) questionnaire. The mean total score for the simulation group was 5.40 (SD 1.40). The mean total score for the control group was 3.61 (SD 1.44). The maximum possible score was 7. A *t* test for independent samples showed a statistically significant difference between the groups, indicating a gain in knowledge for the students who were exposed to the simulation: t_{101}=6.40, *P*<.001 (Table 2). The simulation group had more correct responses than the control group in all seven questions. The greatest difference was noticed in the first question, in which 45 students of the simulation group (87%, 45/52) marked the correct choice, in contrast to only 20 students (39%, 20/51) of the control group. Table 3 shows the scores of the two groups for every question.

In order to assess the potential effect of previous experience with computers and/or 3D games on examination results, data were analyzed to reveal any differences between the two groups. Chi-square analysis of the above-mentioned features found no significant differences between the two groups. However, some interesting results were extracted. In total, 92 of the participants (89.3%, 92/103) had been using computers for more than four years and 51 students (52%, 51/98) subjectively defined their computer skills as “moderate”. Also, 90 of the students (91%, 90/98) were using computers for fun and education and 59 students (60%, 59/98) had played 3D games, of whom, 29 (49%, 29/59) had been playing 3D games for over four years (Table 4).

All 52 students of the simulation group completed the evaluation questionnaire. Students agreed that the simulation, overall, was very comprehensible (rating 1.52 on the scale), very easy to use (1.71), very educative (1.85), very interesting (1.99), very pedagogical (1.92), and original (1.98). Regarding the individual characteristics of the scenario, students rated it as very comprehensible (1.58), very pedagogical (1.88), very educative (1.89), very well-aimed (1.92), and very interesting (2.06). Mean ratings and standard deviations for the Likert-scale questions are illustrated in Tables 5 and 6. Further, 36 students from the simulation group (69%, 36/52) declared that they would use the “Erietta” simulation as a teaching tool if they were instructors. Interestingly, a cumulative 92.3% answered “Yes” or “Maybe” to that question (Table 7).

In the first open-ended question, 50 students (96%, 50/52) responded that the virtual world was very easy to use. In the suggested improvements question, 24 students (46%, 24/52) had no changes to suggest, 5 students (10%, 5/52) would like a variety of selectable scenarios, 3 students (6%, 3/52) asked for better graphics, and 2 students (4%, 2/52) suggested that the dialogs and menus be in Greek (native language). In the last question, 19 students (37%, 19/52) believed that the simulation helped them in all three levels: practical, psychological, and cognitive; 12 students (23%, 12/52) answered “practical” only, 6 students (12%, 6/52) answered “cognitive” only, and 5 students (10%, 5/52) answered “psychological” only (Table 8).

Table 2. Knowledge test scores of the two groups.

Knowledge questionnaire	Simulation (n=52)			Control (n=51)		
	Mean	SD	Std Error Mean	Mean	SD	Std Error Mean
Multiple-choice questions score	5.40	1.40	0.19	3.61	1.44	0.20

Table 3. Correct answers count per question in knowledge questionnaire.

Knowledge questions	Simulation (n=52)		Control (n=51)	
	Frequency	Percent	Frequency	Percent
Q1. Greeting/welcoming the child	45	87	20	39
Q2. Tell-Show-Do	30	58	18	35
Q3. Word substitution	30	58	15	29
Q4. Parent’s interference	47	90	36	71
Q5. Communication	45	87	30	59
Q6. Distraction	47	90	35	69
Q7. Rewarding	37	55	30	45

Table 4. Computer skills profile of the participants (N=103).

Experience in computer usage	Simulation (n=52)		Control (n=51)		Percent of total count
	Frequency	Percent	Frequency	Percent	
Years of computer usage					
0-1	0	0.0	0	0.0	0.0
2-4	6	11.5	5	9.8	10.7
>4	46	88.5	46	90.2	89.3
Subjective level of computer skills					
Extremely good	16	32.7	17	34.7	33.7
Moderate	24	49.0	27	55.1	52.0
Basic	9	18.4	5	10.2	14.3
Primary reason for computer usage					
Education	1	2.0	0	0.0	1.0
Fun	3	6.1	4	8.2	7.1
Both	45	91.8	45	91.8	91.8
Have played 3D games					
Yes	30	61.2	29	59.2	60.2
No	19	38.8	20	40.8	39.8
Years of playing 3D games (If “Yes” was answered to the previous question)					
<2	9	30.0	8	27.6	28.8
2-4	10	33.3	3	10.3	22.0
>4	11	36.7	18	62.1	49.2

Table 5. Students’ evaluation of “Erietta” simulation overall, on a scale of 1 (extremely) to 5 (not at all) (n=52).

Characteristic	Mean	SD
Comprehensible	1.52	0.87
Easy to use	1.71	1.02
Educative	1.85	1.14
Interesting	1.99	1.11
Pedagogical	1.92	1.04
Original	1.98	1.22

Table 6. Students’ evaluation of the simulation scenario on a scale of 1 (extremely) to 5 (not at all) (n=52).

Characteristic	Mean	SD
Comprehensible	1.58	0.94
Educative	1.89	1.02
Aimed	1.92	0.89
Interesting	2.06	1.02
Pedagogical	1.88	1.15

Table 7. Students' answers to the question: "Would you use this simulation as a teaching tool if you were an instructor?" (n=52).

Answer	Frequency	Cumulative percent
Yes	36	69.2
Maybe	12	92.3
No	4	100

Table 8. Students' answers to open-ended questions (n=52).

Categorized answers	Frequency	Percent
Was the virtual world easy to use? Please justify your answer.		
Very easy	50	96
Difficult if you don't have previous computer knowledge	2	4
What changes or improvements would you suggest?		
More scenarios to select from	5	10
More comprehensive dialogs	1	2
Greek menus/dialogs	2	4
No changes	24	46
Better graphics	3	6
Quicker and easier in use	1	2
Other	8	15
Did not respond	8	15
Do you believe that the simulation helped you in a cognitive, practical, or psychological way? Please justify your answer.		
All three ways	19	37
Practical only	12	23
Practical and psychological	5	10
Psychological only	5	10
Cognitive only	6	12
Cognitive and practical	3	6
Other	1	2
Did not respond	1	2

Discussion

Principal Findings

The objective of our study was to measure the efficacy of a child VP in a virtual world, as a supplementary teaching tool for pediatric dentistry. This simulation aimed to support training in communication and behavior techniques. Previous research has shown that VPs can be used for practicing on patient interviewing, communication, and clinical reasoning [30-34]. For this study, a combination of VPs and virtual worlds was attempted. Virtual worlds offer an online programming and 3D building environment in which users can meet and socialize. Previous efforts on creating children VPs, such as the online text-based "Virtual Child" [39] and Kleinert's interactive video-based module [40], resulted in knowledge gain and were evaluated positively by the students. Similarly, our simulation group achieved significantly better results than the control group did in the knowledge questionnaire. This indicates that the simulation acted as a supplement to the lectures of the pediatric dentistry course. More than one-third of the students admitted that this experience helped them in three levels: practical, cognitive, and psychological, while a significant number of answers concentrated in practical level only. This is a very interesting result, taking into account the lack of adequate practical training in communication techniques at the University. The students rated the educational aspects of the simulation highly, such as the comprehensiveness, pedagogical value, ease of use, and originality. The scenario was also accepted very well, although it was simple and linear. Almost half of the participants had no improvements to suggest, some students asked for a variety of selectable scenarios, and a few others suggested better graphics. These changes are easy to implement due to the modular nature of the simulation: Erietta is a VP made of simple 3D graphics, LSL scripts, and notecards. The virtual world offers the capability for expansions and add-ons to be designed online or offline and then uploaded, or purchased by the Marketplace and then programmed to operate at the designer's will. New questions can also be added, as notecards.

Although no difference was found between the two groups regarding to their skills on computers and 3D games, most of the students were found to have been using computers for more than four years. More than half of the participants had played 3D games. It must be noted that the current generation of dental students differs a lot from older ones, in terms of technology knowledge. Today's students could be characterized as aborigines of a "computer era"; most of them have been using computers and/or playing video-games since their childhood. This fact indicates that technologies such as VPs, virtual worlds, serious games, social networks, and the upcoming Web 3.0 are almost a requisite in the education of new healthcare providers. In our study, the majority of the students agreed that they would use Erietta as a teaching tool if they were instructors. This finding proves that such tools are welcome in dental education and will be happily accepted by the students. Being inside a virtual environment as an avatar along with your colleagues and exploring interactive 3D items from your home PC without the need of supervision may have great e-learning potential and also reduce faculty working time. It must be stressed that these new methods of creating educational material are intended to support the traditional learning methods, not to replace them; in this study, a simulation was utilized as a teaching supplement, not a replacement.

Limitations

The evaluation did not include measures of changes in students' skills on behavior management. Ideally, an assessment of communication skills with a child in a real dental operatory should have been made. Although this would require the presence of children, it is planned as a future work, while a new version of the Erietta simulation is currently under development.

All the participants had already completed the Pediatric Dentistry course, so a homogeneous status of knowledge was hypothesized at the beginning of the evaluation. For this reason, the knowledge assessment was not measured at baseline and hence it is difficult to have a clear picture of the significance of the difference between the two groups.

Virtual worlds require computers with good graphics cards and an Internet connection. Although creating simulations with relatively simple graphics such as Erietta is easy, detailed models demand more effort and special software. Programming experience is also required in order to write a script in LSL. For this study, a linear scenario consisting of six steps was written. No branches or alternative ways of achieving the goal were designed because of time limitations. Also, our objective was to examine the efficiency of a pediatric VP in virtual worlds, by utilizing a less sophisticated scenario. Though, as a future improvement, we are planning to expand the current simulation by adding more steps and branches to each selection and creating a pool of stories with various difficulty for the user to select.

Conclusions

This project has indicated that a simulation based on a virtual patient, built in a virtual world, may improve student knowledge in communication management for pediatric dentistry when used as a supplementary teaching tool.

Acknowledgments

We thank Dr Eva Comaroski greatly for contributing her virtual world expertise to developing the Erietta simulation. We thank the VIBE group for hosting the Erietta simulation. VIBE is a collaboration to create learning tools in virtual worlds. What differentiates this effort is that lesson plans and walk-throughs are being created to allow students to navigate through the virtual environment requiring minimal expertise or training in using the virtual world's software. Instructions on accessing the grid and reaching the staff can be found on VIBE's wiki [43]. Finally, we thank Assistant Professor Panagiotis Bamidis, Dr Christina Athanasopoulou, and Maria Karipoglou for their great contribution in the evaluation section.

Conflicts of Interest

None declared.

References

1. Bell M. Toward a definition of virtual worlds. JVWR 2008;1(1):2-5 [FREE Full text]
2. Schroeder R. Defining virtual worlds and virtual environments. JVWR 2008;1(1):2-5 [FREE Full text]
3. Buchanan JA. Experience with virtual reality-based technology in teaching restorative dental procedures. J Dent Educ 2004 Dec;68(12):1258-1265 [FREE Full text] [Medline: 15576814]
4. Marras I, Nikolaidis N, Mikrogeorgis G, Lyroudia K, Pitas I. A virtual system for cavity preparation in endodontics. J Dent Educ 2008 Apr;72(4):494-502 [FREE Full text] [Medline: 18381855]
5. Rodrigues MAF, Silva WB, Barbosa Neto ME, Gillies DF, Ribeiro IM. An interactive simulation system for training and treatment planning in orthodontics. Computers & Graphics 2007 Oct;31(5):688-697 [FREE Full text] [doi: 10.1016/j.cag.2007.04.010]
6. Papadopoulos L. Virtual reality simulators in dental training. Hel Stom Rev 2010;54(3):263-276 [FREE Full text]
7. Entertainment Software Association. Essential Facts about the Computer and Video Game Industry. 2012. URL: http://www.theesa.com/facts/pdfs/ESA_EF_2012.pdf [accessed 2013-02-17] [WebCite Cache ID 6EUhCT3eM]
8. Amer RS, Denehy GE, Cobb DS, Dawson DV, Cunningham-Ford MA, Bergeron C. Development and evaluation of an interactive dental video game to teach dentin bonding. J Dent Educ 2011 Jun;75(6):823-831 [FREE Full text] [Medline: 21642529]

9. Second Life website. URL: http://secondlife.com/whatis/?lang=en-US [accessed 2013-02-17] [WebCite Cache ID 6EUhHTkfQ]
10. OpenSimulator website. URL: http://opensimulator.org/wiki/Main_Page [accessed 2013-02-17] [WebCite Cache ID 6EUhNs2Gz]
11. Ohio State University Medical Center in Second Life. URL: http://slurl.com/secondlife/OSU%20Medicine/83/98/302 [accessed 2013-02-17] [WebCite Cache ID 6EUhSrEyt]
12. Ann Myers Medical Center in Second Life. URL: http://maps.secondlife.com/secondlife/Fashion%20Boulevard%20II/87/78/21 [accessed 2013-02-17] [WebCite Cache ID 6EUhXFUwX]
13. Ann Myers Medical Center blog. URL: http://ammc.wordpress.com/about-us/ [accessed 2013-02-17] [WebCite Cache ID 6EUhZvFDz]
14. Imperial College London Virtual Hospital in Second Life - Respiratory Ward. URL: http://maps.secondlife.com/secondlife/Imperial%20College%20London/150/86/27 [accessed 2013-02-17] [WebCite Cache ID 6EUheUAqA]
15. Diener S, Windsor J, Bodily D. Design and development of clinical simulations in Second Life and OpenSim. 2009 May 03 Presented at: EDUCAUSE Australasia Conference; 2009; Perth Australia URL: https://researchspace.auckland.ac.nz/handle/2292/4305
16. Creutzfeldt J, Hedman L, Medin C, Heinrichs WL, Felländer-Tsai L. Exploring virtual worlds for scenario-based repeated team training of cardiopulmonary resuscitation in medical students. J Med Internet Res 2010;12(3):e38 [FREE Full text] [doi: 10.2196/jmir.1426] [Medline: 20813717]
17. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]
18. Wiecha J, Heyden R, Sternthal E, Merialdi M. Learning in a virtual world: experience with using second life for medical education. J Med Internet Res 2010;12(1):e1 [FREE Full text] [doi: 10.2196/jmir.1337] [Medline: 20097652]
19. Second Health Imperial College London in Second Life. URL: http://maps.secondlife.com/secondlife/National%20Health%20Service/132/121/26 [accessed 2013-02-17] [WebCite Cache ID 6EUhm9W0a]
20. MUVE clinic in Second Life. URL: http://maps.secondlife.com/secondlife/Evergreen%20Island%203/34/171/30 [accessed 2013-02-17] [WebCite Cache ID 6EUhomvzM]
21. University of Kentucky in Second Life. URL: http://slurl.com/secondlife/University%20of%20KY/141/64/27 [accessed 2013-02-17] [WebCite Cache ID 6EUhsQmvF]
22. McAndrew M, Johnston AE. The role of social media in dental education. J Dent Educ 2012 Nov;76(11):1474-1481. [Medline: 23144483]
23. Phillips J, Berge ZL. Second life for dental education. J Dent Educ 2009 Nov;73(11):1260-1264 [FREE Full text] [Medline: 19910474]
24. Maryland Dental School in Second Life. URL: http://slurl.com/secondlife/UMaryland%20Dental%20School/122/103/45 [accessed 2013-02-17] [WebCite Cache ID 6EUhv2wNz]
25. UMDS Virtual Dental School Insider. 2009. Virtual Skull Practical URL: http://dspub.umaryland.edu/vi/ [accessed 2013-02-17] [WebCite Cache ID 6EUhydw14]
26. Griffith S. Case Western Reserve University-News Center. 2008 Jul 28. A virtual Toothache helps student dentists learn patient-side communications URL: http://blog.case.edu/case-news/2008/07/28/virtualdentistry [accessed 2013-02-17] [WebCite Cache ID 6EUi5Ejhr]
27. Fox R, Walker JJ, Draycott TJ. Medical simulation for professional development-science and practice. BJOG 2011 Nov;118 Suppl 3:1-4. [doi: 10.1111/j.1471-0528.2011.03173.x] [Medline: 22039885]
28. Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. Med Educ 2012 Nov;46(11):1042-1054. [doi: 10.1111/j.1365-2923.2012.04343.x] [Medline: 23078681]
29. Bosse HM, Schultz JH, Nickel M, Lutz T, Möltner A, Jünger J, et al. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. Patient Educ Couns 2012 Jun;87(3):300-306. [doi: 10.1016/j.pec.2011.10.007] [Medline: 22137189]
30. AAMC Institute for Improving Medical Education. Effective Use of Educational Technology in Medical Education. Colloquium on Educational Technology: Recommendations and Guidelines for Medical Educators. 2007 Mar. URL: https://members.aamc.org/eweb/upload/Effective%20Use%20of%20Educational.pdf [accessed 2013-02-17] [WebCite Cache ID 6EUiCJtAG]
31. Villaume WA, Berger BA, Barker BN. Learning motivational interviewing: scripting a virtual patient. Am J Pharm Educ 2006 Apr 15;70(2):33 [FREE Full text] [Medline: 17149413]
32. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
33. Hooper CR, Jivram T, Law S, Michell A, Somasunderam A. Using virtual patients to teach medical ethics, medical law and medical professionalism. Med Teach 2012;34(8):674-675. [doi: 10.3109/0142159X.2012.689450] [Medline: 22830327]

34. Schittek Janda M, Mattheos N, Nattestad A, Wagner A, Nebel D, Färbom C, et al. Simulation of patient encounters using a virtual patient in periodontology instruction of dental students: design, usability, and learning effect in history-taking skills. Eur J Dent Educ 2004 Aug;8(3):111-119. [doi: 10.1111/j.1600-0579.2004.00339.x] [Medline: 15233775]
35. Sanders C, Kleinert HL, Boyd SE, Herren C, Theiss L, Mink J. Virtual patient instruction for dental students: can it improve dental care access for persons with special needs? Spec Care Dentist 2008;28(5):205-213. [doi: 10.1111/j.1754-4505.2008.00038.x] [Medline: 18782198]
36. Cederberg RA, Bentley DA, Halpin R, Valenza JA. Use of virtual patients in dental education: a survey of U.S. and Canadian dental schools. J Dent Educ 2012 Oct;76(10):1358-1364 [FREE Full text] [Medline: 23066135]
37. Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med 2007 May;82(5):446-451. [doi: 10.1097/ACM.0b013e31803e8a0a] [Medline: 17457063]
38. Lane JL, Ziv A, Boulet JR. A pediatric clinical skills assessment using children as standardized patients. Arch Pediatr Adolesc Med 1999 Jun;153(6):637-644. [Medline: 10357307]
39. Boynton JR, Green TG, Johnson LA, Nainar SM, Straffon LH. The virtual child: evaluation of an internet-based pediatric behavior management simulation. J Dent Educ 2007 Sep;71(9):1187-1193 [FREE Full text] [Medline: 17761625]
40. Kleinert HL, Sanders C, Mink J, Nash D, Johnson J, Boyd S, et al. Improving student dentist competencies and perception of difficulty in delivering care to children with developmental disabilities using a virtual patient module. J Dent Educ 2007 Feb;71(2):279-286 [FREE Full text] [Medline: 17314390]
41. American Academy of Pediatric Dentistry. Guideline on Behavior Guidance for the Pediatric Dental Patient.: AAPD; 2011. URL: http://www.aapd.org/policies/ [accessed 2013-02-17] [WebCite Cache ID 6EUiFvcoG]
42. Blender website. URL: http://www.blender.org/ [accessed 2013-02-17] [WebCite Cache ID 6EUiIK4Mx]
43. The Virtual Islands for Biology Education website. URL: http://wiki.bio-se.info/doku.php [accessed 2013-02-17] [WebCite Cache ID 6EUiLZarS]
44. Erietta Simulation in Second Life. URL: http://maps.secondlife.com/secondlife/SUNO%20Regents/78/1/1028 [accessed 2013-02-17] [WebCite Cache ID 6EUiVNKsb]
45. Erietta Simulation website. URL: http://wiki.bio-se.info/doku.php?id=pediatric_dentistry [accessed 2013-07-10] [WebCite Cache ID 6I0cL2ecb]

Abbreviations

AAPD: American Academy of Pediatric Dentistry
OpenSim: Open Simulator
SL: Second Life
VP: virtual patient

Edited by P Bamidis, D Giordano, C Pattichis, N Zary; submitted 24.05.13; peer-reviewed by J Wiecha, P Antoniou; comments to author 20.06.13; revised version received 12.07.13; accepted 13.08.13; published 29.10.13.

Please cite as:
Papadopoulos L, Pentzou AE, Louloudiadis K, Tsiatsos TK
Design and Evaluation of a Simulation for Pediatric Dentistry in Virtual Worlds
J Med Internet Res 2013;15(10):e240
URL: http://www.jmir.org/2013/10/e240/
doi:10.2196/jmir.2651
PMID:24168820

Original Paper

Activities on Facebook Reveal the Depressive State of Users

Sungkyu Park[1*]; Sang Won Lee[2*], MD; Jinah Kwak[3]; Meeyoung Cha[3*], PhD; Bumseok Jeong[2*], MD, PhD

[1]Division of Web Science and Technology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, Republic Of
[2]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, Republic Of
[3]Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, Republic Of
[*]these authors contributed equally

Corresponding Author:
Bumseok Jeong, MD, PhD
Graduate School of Medical Science and Engineering
Korea Advanced Institute of Science and Technology (KAIST)
291 Daehak-ro, Yuseong-gu
Daejeon, 305-701
Korea, Republic Of
Phone: 82 42 350 0540
Fax: 82 42 350 7160
Email: bs.jeong@kaist.ac.kr

Abstract

Background: As online social media have become prominent, much effort has been spent on identifying users with depressive symptoms in order to aim at early diagnosis, treatment, and even prevention by using various online social media. In this paper, we focused on Facebook to discern any correlations between the platform's features and users' depressive symptoms. This work may be helpful in trying to reach and detect large numbers of depressed individuals more easily.

Objective: Our goal was to develop a Web application and identify depressive symptom–related features from users of Facebook, a popular social networking platform.

Methods: 55 Facebook users (male=40, female=15, mean age 24.43, SD 3.90) were recruited through advertisement fliers distributed to students in a large university in Korea. Using EmotionDiary, the Facebook application we developed, we evaluated depressive symptoms using the Center for Epidemiological Studies-Depression (CES-D) scale. We also provided tips and facts about depression to participants and measured their responses using EmotionDiary. To identify the Facebook features related to depression, correlation analyses were performed between CES-D and participants' responses to tips and facts or Facebook social features. Last, we interviewed depressed participants (CES-D≥25) to assess their depressive symptoms by a psychiatrist.

Results: Facebook activities had predictive power in distinguishing depressed and nondepressed individuals. Participants' response to tips and facts, which can be explained by the number of app tips viewed and app points, had a positive correlation (P=.04 for both cases), whereas the number of friends and location tags had a negative correlation with the CES-D scale (P=.08 and P=.045 respectively). Furthermore, in finding group differences in Facebook social activities, app tips viewed and app points resulted in significant differences (P=.01 and P=.03 respectively) between probably depressed and nondepressed individuals.

Conclusions: Our results using EmotionDiary demonstrated that the more depressed one is, the more one will read tips and facts about depression. We also confirmed depressed individuals had significantly fewer interactions with others (eg, decreased number of friends and location tagging). Our app, EmotionDiary, can successfully evaluate depressive symptoms as well as provide useful tips and facts to users. These results open the door for examining Facebook activities to identify depressed individuals. We aim to conduct the experiment in multiple cultures as well.

(J Med Internet Res 2013;15(10):e217) doi:10.2196/jmir.2718

KEYWORDS

Facebook; Web application; depressive symptoms; online social network (OSN) activities; mental health; Internet

Introduction

Depression is one of the most common mental disorders. The lifetime prevalence of depressive disorder is 16.2% [1]; it commonly occurs in one's early life and has a chronic course [2]. Depression is also related to reduced individual productivity and functional impairment that can cause a social burden [3]. It is predicted that depression will be the second leading cause of disease burden worldwide by 2020 [4]. The costs associated with depression and mental disorder have grown rapidly, and the National Institute of Mental Health reported in 2008 that major mental disorders cost at least $193 billion in the United States annually in lost earnings alone [5]. Therefore, depression has severe effects on individuals as well as on society.

Early diagnosis and prevention of depression can be an effective way to reduce depression-related problems because the length of the depressive episode is directly related to its recovery rate [6]. Therefore, significant effort has been spent on detecting symptoms of depression earlier in the general population. A number of campaigns have been proposed including National Depression Screening Day [7] and National Anxiety and Depression Awareness Week [8], which include offering free depression screening that can help find participants' depressive symptoms in a prompt and easy way using several questionnaires. While these campaigns are an important step toward the early identification of potential patients, their main limitation lies in their potential bias; participants who have severe depressive symptoms, such as loss of energy or interest, might not attend the campaigns at all since to participate, they need to first go out. Therefore, by adopting our novel strategy, people can more easily access the new method, which actively adopts already widely spread online social media in contrast to conventional campaigns. This new approach may lessen the potential bias of conventional participation-oriented campaigns.

One possible screening method is using the large amount of data on online social networks (OSNs) [9]. OSN sites such as Facebook and Twitter, already used by hundreds of millions of users [10], have large-scale data that can be used to study health-related human behaviors in a cost-effective manner [11-13]. OSN data can be also used to reach and detect a large number of individuals with depression at low cost. Identifying the kinds of online social features that correlate with depression is crucial [9]. With the advent of OSN services, many attempts have been made to detect early symptoms of depression from online large-scale data [13-15].

Most research on finding depressive symptoms on OSNs have used words related to depression. Park et al [16] analyzed short text updates posted on Twitter to characterize the use of language related to depressive moods. The authors found that many online users openly disclose their depressive moods as well as treatment history in a public medium like Twitter. For instance, there was one tweet disclosing a detailed prescription as follows: "My doctor tries to give me birth control for depression, which works for me but I have so many side effects I would rather be moody." Choudhury et al [17] examined linguistic and emotional correlates of postnatal changes in new mothers and built a statistical model to forecast significant postpartum behavioral changes using only prenatal observations. These studies show the potential application of social media in studying depressive symptoms, in particular, to understand the relationship between linguistic markers and mental disorder [18].

Moving beyond text analysis, many other approaches have been used to detect depressive symptoms online. Kotikalapudi et al [19] analyzed the patterns in the Web-browsing activities of college students that could signal depressive symptoms. Moreno et al [20] demonstrated that status updates on Facebook could reveal symptoms of major depressive episodes. Moreover, Rosenquist et al [21] found that levels of depression showed diffusion of up to three degrees of separation in a large social network, suggesting a wide influence of depressive symptoms through social links. It was also found that, in 2008, more than a quarter of Internet users searched for information about depression or mental health issues [22].

In this paper, we build on the abovementioned related work and make an effort to examine social network determinants of depressive symptoms. In doing so, we used data gathered from Facebook, which is currently the most widely used OSN in the world [23]. Even in Korea in 2011, Facebook surpassed the former most famous domestic OSN service, Cyworld, and became the most widely used social network [24]. Facebook contains a wide range of information about users, including demographic features such as age and gender, as well as social features such as friends list, like, interest, and location tagging. Together, these features could represent how a user maintains relationships online as well as offline [25,26]. Our research focus in this paper was to test whether a user's mental health status can be predicted by the wide set of features available on Facebook.

For this study, we developed a mobile Web-based application for Facebook, called EmotionDiary, to recruit participants and seek markers of depressive symptoms on OSNs. The application offers two short self-report scales for measuring depressive symptoms: CES-D (Center for Epidemiologic Studies Depression) [27] and BDI (Beck's Depression Inventory) [28], which are both well-proven approaches for measuring depressive symptomatology in the general population. This work aimed to identify depressive symptom-related features on Facebook that could distinguish depressed individuals from those who are not depressed. To determine whether the test overestimated or underestimated depressive symptoms and to gain a deeper understanding of user behaviors, this study was further assisted by face-to-face interviews of severely depressed individuals by a psychiatrist. Hence this paper provides both quantitative and qualitative findings toward detecting depressive symptoms in OSNs.

Methods

Overview

To demonstrate clearly the overall experiment and evaluation process, a detailed flowchart is provided in Multimedia Appendix 1.

Application

This study is based on a Facebook Web application that the authors developed and named EmotionDiary, which surveys depressive states of users and gathers demographic and social-activity data from Facebook. Upon accessing the app for the first time, users are shown a consent form that asks users for permission to access certain types of data from Facebook (see Figure 1). When users agree, the app becomes available and provides the depression questionnaires. The EmotionDiary app culled user data from Facebook every time a user accessed it. The app terminated for those users who did not consent.

EmotionDiary uses two authorized surveys, CES-D [27] and BDI [28], to screen depression. The CES-D self-report scale contains 20 simple questions, such that each questionnaire is rated 0 to 3 based on the frequency of depressive symptoms. The CES-D was developed to measure the symptoms of depression in community populations; it is commonly used in epidemiological studies. The Korean version of CES-D was standardized in 1993 [29]. The BDI is a self-report scale and consists of 21 questions about various domains of depressive symptoms including emotional, cognitive, physical, and motivational symptoms. Participants rate the severity of each item using a 0 to 3-point scale. The Korean version of the BDI scale used in this study was standardized in 1991, and its reliability and validity were verified [30]. We asked participants to take both surveys to ensure responses were consistent. Only those participants who completed both tests were included in our analysis. To avoid having any questions unanswered, when participants submitted the result with some missing answers, the app directed participants to go back and fill in all the remaining questions, as shown in the error page example in Figure 2.

The CES-D test was the first survey shown to participants, for which we provided three types of feedback (see Figure 2). Given that the score range of CES-D is between 0 and 60, we used the cutoffs of 16 and 25 suggested in previous studies [31,32]: a score of 16 to 24 represents probable depression, and 25 or higher represents definite depression or major depressive disorder. Each participant's feedback included an infographic indicating the participant's level of depression and possible recommendations to reduce their level of depression in everyday life. The precise score was not shown to participants. For instance, users whose CES-D scores fell below 16 received the following feedback: "Your score is in a normal range. However, you need to pay attention to your mental health since depression is a common disease. Taking a walk, exercising for 30 minutes, talking with friends, or sunbathing would be helpful to maintain a good mental state!"

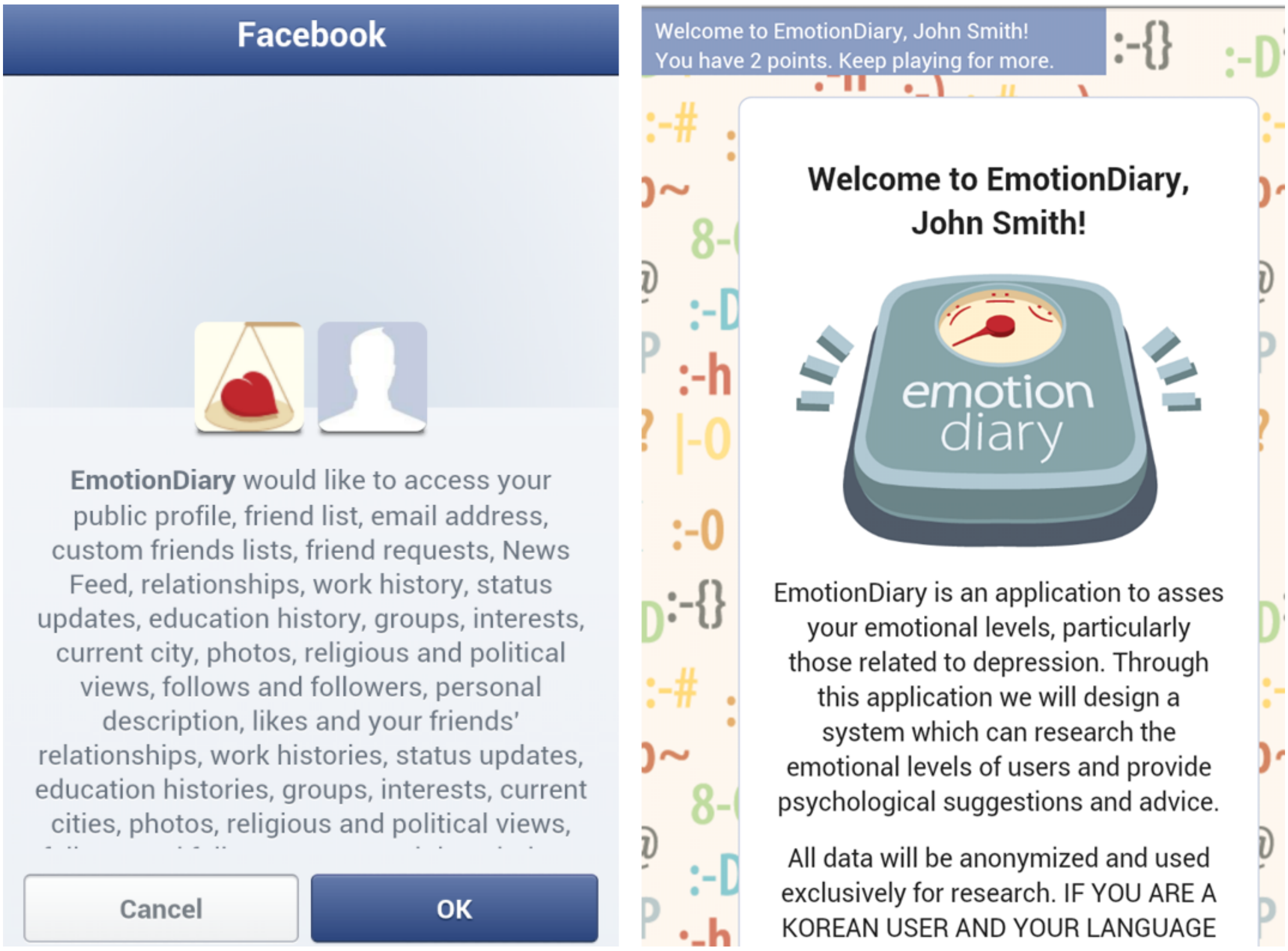
Participants were taken to the BDI test after completing the CES-D test. We did not show any feedback for the BDI to avoid the same participant being judged to have a different depressive state from the CES-D. Once participants finished the BDI survey, they were led to view tips on improving their mental health. EmotionDiary contained 100 tips and facts pages: 40 tips to improve one's mood and 60 general facts about depression (see Figure 3). The helping tips were selected from a self-help book on depression [33], and general facts were from a website entitled Random Facts [34]. Also, the app contained a points system, where participants could gain 1 point for accessing the app each time, 5 points for responding to each survey question, 3 points for viewing each tip, and so on. The point system was designed only to motivate users, and none of the participants were financially rewarded in any ways. Accumulating points and participating in viewing tips, which meant a participant was accessing the app and checking the individual tip and fact pages, can be associated with an acute depressive state rather than a trait. These two activities may better reveal the relationship between depression and OSN activities, because other Facebook social features contain cumulative data that cannot reflect acute behavioral changes.

To check stability of the app and compliance of the participants, before starting the actual research we conducted a pilot study with 28 participants different from those of the main experiment. As a result, we were assured that the app was able to handle data and that participants would be able to comply with the experiment (see Multimedia Appendix 2).

Participants and Recruitment

A total of 115 participants were recruited among undergraduate and graduate students at KAIST (Korea Advanced Institute of Science and Technology) who had Facebook accounts. Participants were randomly recruited through advertisement fliers posted on the major school buildings as well as through an online school BBS (Bulletin Board System) during a 2-week period from April 17-30, 2013 (see Figure 4). All participants joined and answered the questions on a voluntary basis without getting a financial reward. We fully explained the purpose of this study and the individual information that we would gather on Facebook. This study was approved by the KAIST Institutional Review Board (approval number: KH-2012-22).

Figure 1. EmotionDiary screenshots: asking permission for data access and the welcome screen.



EmotionDiary would like to access your public profile, friend list, email address, custom friends lists, friend requests, News Feed, relationships, work history, status updates, education history, groups, interests, current city, photos, religious and political views, follows and followers, personal description, likes and your friends' relationships, work histories, status updates, education histories, groups, interests, current cities, photos, religious and political views,

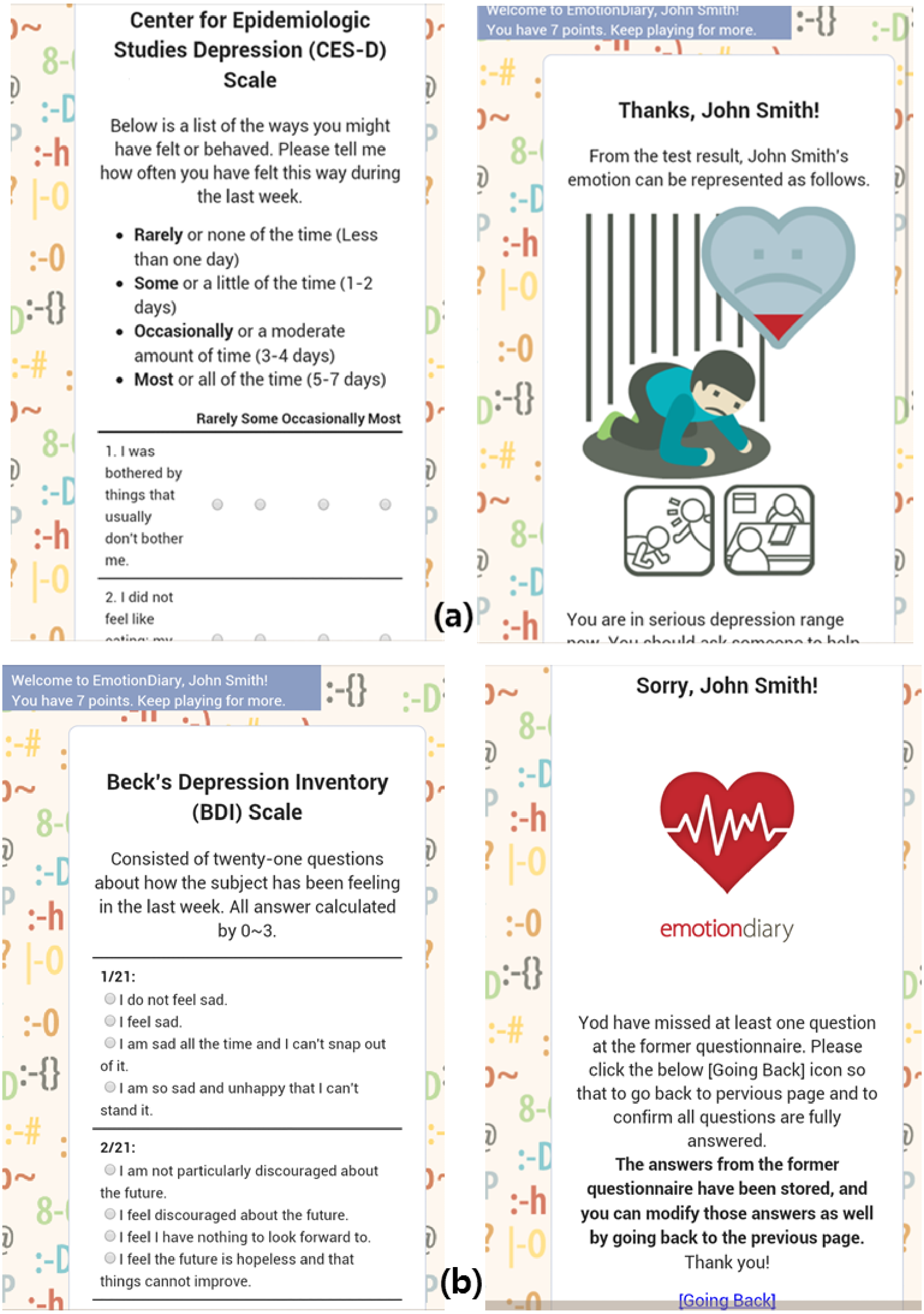
Welcome to EmotionDiary, John Smith!
You have 2 points. Keep playing for more.

Welcome to EmotionDiary, John Smith!

EmotionDiary is an application to asses your emotional levels, particularly those related to depression. Through this application we will design a system which can research the emotional levels of users and provide psychological suggestions and advice.

All data will be anonymized and used exclusively for research. IF YOU ARE A KOREAN USER AND YOUR LANGUAGE

Figure 2. EmotionDiary screenshots: CES-D test and resulting feedback, and the BDI test and an error page when not all questions were answered.



Center for Epidemiologic Studies Depression (CES-D) Scale

Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the last week.

- **Rarely** or none of the time (Less than one day)
- **Some** or a little of the time (1-2 days)
- **Occasionally** or a moderate amount of time (3-4 days)
- **Most** or all of the time (5-7 days)

	Rarely	Some	Occasionally	Most
1. I was bothered by things that usually don't bother me.	○	○	○	○
2. I did not feel like	○	○	○	○

Welcome to EmotionDiary, John Smith!
You have 7 points. Keep playing for more.

Thanks, John Smith!

From the test result, John Smith's emotion can be represented as follows.

You are in serious depression range

Welcome to EmotionDiary, John Smith!
You have 7 points. Keep playing for more.

Beck's Depression Inventory (BDI) Scale

Consisted of twenty-one questions about how the subject has been feeling in the last week. All answer calculated by 0~3.

1/21:
○ I do not feel sad.
○ I feel sad.
○ I am sad all the time and I can't snap out of it.
○ I am so sad and unhappy that I can't stand it.

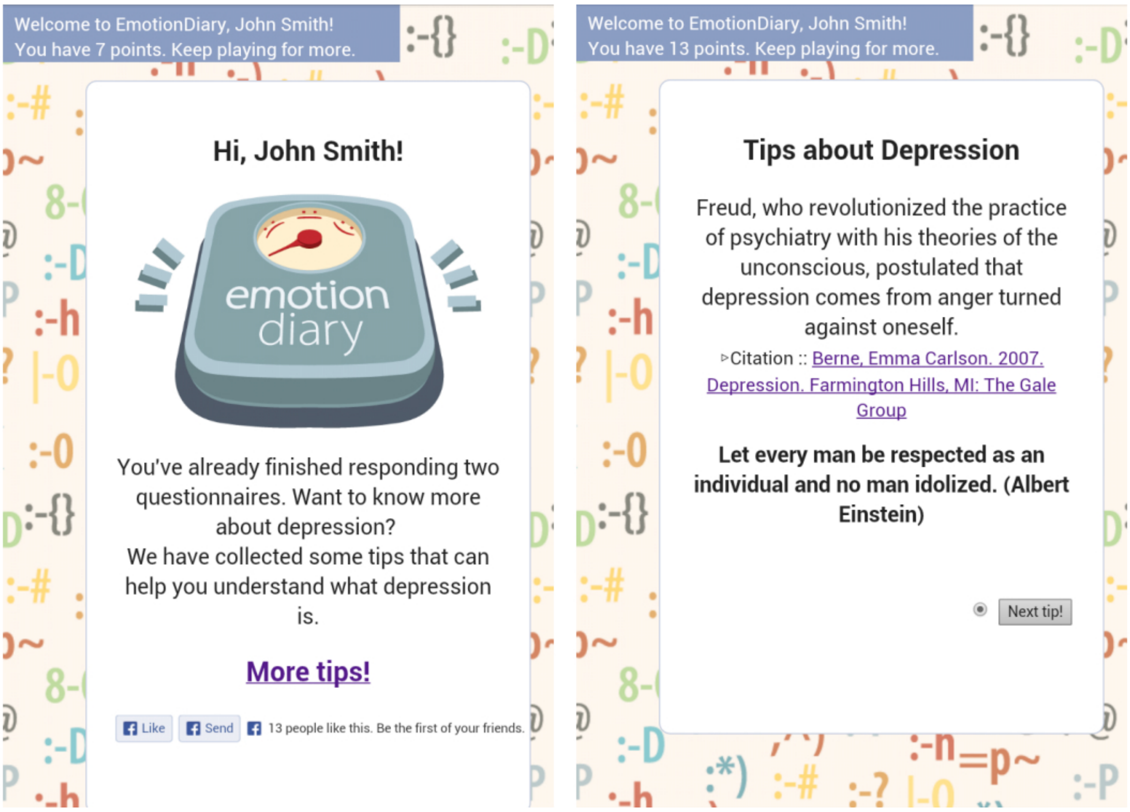
2/21:
○ I am not particularly discouraged about the future.
○ I feel discouraged about the future.
○ I feel I have nothing to look forward to.
○ I feel the future is hopeless and that things cannot improve.

Sorry, John Smith!

Yod have missed at least one question at the former questionnaire. Please click the below [Going Back] icon so that to go back to pervious page and to confirm all questions are fully answered.
The answers from the former questionnaire have been stored, and you can modify those answers as well by going back to the previous page.
Thank you!

[Going Back]

Figure 3. EmotionDiary screenshots: depression tips and facts.



Welcome to EmotionDiary, John Smith!
You have 7 points. Keep playing for more.

Hi, John Smith!

You've already finished responding two questionnaires. Want to know more about depression?
We have collected some tips that can help you understand what depression is.

More tips!

Welcome to EmotionDiary, John Smith!
You have 13 points. Keep playing for more.

Tips about Depression

Freud, who revolutionized the practice of psychiatry with his theories of the unconscious, postulated that depression comes from anger turned against oneself.

▹Citation :: Berne, Emma Carlson. 2007. Depression. Farmington Hills, MI: The Gale Group

Let every man be respected as an individual and no man idolized. (Albert Einstein)

Figure 4. Flier for recruiting participants.

Feeling blue?
You can measure the weight of your mind via simple online tests by scanning the QR code.

Many people have already participated in this Facebook app.
[Web URL: **http://emotiondiary.herokuapp.com**]
Information will be encrypted and used only for research purposes.

Social Computing Lab., KAIST
Clinical Neuroscience & Dev Lab., KAIST

Data Description

The EmotionDiary app culled and stored three broad types of data from each user: (1) demographic information including name, age, and gender, (2) social-activity information such as friends list and likes from Facebook, and (3) app-generated data such as the depression score and the number of tips viewed. Table 1 summarizes the list of data gathered and used for analysis. In collecting data, Facebook's API (Application Programming Interface) was used, similar to other studies on Facebook [35,36]. While the API can be used to gather a wide range of data, we primarily limited our focus to popular activity-related features, as in other studies. As described earlier, the three types of data were culled each time participants accessed the app from April 17-30, 2013.

Statistical Analysis

First, Spearman rank correlation coefficient was used to test any relationship between the various features on Facebook and participants' CES-D scores. Spearman correlation is a nonparametric method that can address outliers appropriately. After the correlation analysis, simple linear regression was applied to evaluate the detailed relationships between Facebook social features and CES-D scores.

Second, a Mann-Whitney *U* test was used to measure the differences in Facebook social features between participants in the probably depressed group and nondepressed group. This analytic method is useful when two groups do not follow normal distributions and when the number of participants is not large.

In dividing users into two such groups (ie, normal and depressed), we jointly used the CES-D and BDI results and adopted the cutoff values proposed for screening from previous studies [37,38], with a CES-D score of 21 and the BDI score of 10. We chose the CES-D score of 21, rather than 16 or 25, because it has been known to be helpful in community-level screening [37]. Therefore, the depressed group consisted of individuals with a CES-D score≥21 and BDI≥10, and the normal group had a CES-D score<21 and BDI score<10.

Face-to-Face Interview

Among participants who voluntarily participated in this study and had completed two screening scales, participants who garnered CES-D scores higher than 25 were invited for a 50-minute (approximately) interview with a psychiatrist at KAIST. This interview was conducted when participants accepted our invitation via a Facebook message. We chose a cutoff score of 25 because it can represent definite depression or major depressive disorder [31,32]. We sent a private invitation message to 7 of the 15 target interviewees through Facebook; two of them paid a visit to the clinic for the interview. The face-to-face interview had three purposes: (1) to evaluate participants' depressive symptoms from a qualitative aspect and to find the relationship between detailed depressive symptoms and Facebook features, (2) to give appropriate feedback to participants with CES-D scores higher than 25, indicating the need for professional help, and (3) to measure the reliability and characteristics of online depression scales.

Table 1. Facebook social activity and app-generated features.

Feature	Feature description (ranges)
User demographic	
Age	User's age
Gender	User's gender information: male (0) and female (1)
Facebook social activity	
Relationships	User's relationship status: no data (excluded), single (coded as 1), in a relationship (2), engaged (3), married (4), widowed (0.5), separated (0.5), divorced (0.5)
Groups	Number of groups to which a user belongs (including groups of which a user is an administrator)
Group admins	Number of groups for which a user is an administrator
Likes	Number of pages a user has marked as like
Pending requests	Number of pending incoming friend requests
Friends	Number of friends a user has
Location tagging	Number of physical locations which a user has tagged
Interests	Number of interest items listed on the user's profile
Activities	Number of activities listed on the user's profile
Events	Number of events a user is attending
App-generated data	
Tips	Number of tip and fact sections list a user checked (total 100 tip and fact sections are offered; we counted the number of tip and fact sections user checked).
Points	Accumulated points by participating in the app
CES-D	Resulting score of the CES-D survey: range from 0 to 60
BDI	Resulting score of the BDI survey: range from 0 to 63

Results

Overview

Of the 115 random participants who accessed the EmotionDiary app, 82 participants completed the CES-D survey (71.3%), and 56 of them also completed the BDI survey (48.7%). We focused only on the 56 participants who completed both questionnaires to select ones who truly participated and to exclude others who dropped off in the middle of the experiment. Furthermore, we set an acceptable range on the number of friends, 0-1000, and groups in which they participated, 0-50, excluding those users who exceeded these limits because such users may be considered “microcelebrities” who exhibit different behavioral patterns from ordinary users on Facebook. Through this screening process, one male participant was excluded from analysis due to having more than 1000 Facebook friends. The remainder of this paper describes the analysis of the 55 participants.

Of the 55 participants retained for further analyses, 40 were males aged 19-36 (mean age 24.89, SD 4.35) and 15 were females aged between 19-28 (mean age 23.33, SD 2.17). Table 2 summarizes the mean and quartile values of participants based on their demographic, social activity, and app features. The table also shows the first quartile (25th) and the third quartile (75th) values to indicate the overall distribution of each feature. Most participants had no “activities” and “events”; hence, we removed these features from the analysis.

Table 2. Mean values of participants’ Facebook social activity and app features (N=55).

Feature	Mean (SD)	1st quartile	3rd quartile
User demographic			
Age	24.43 (3.90)	22	26
Gender	M=40, F=15		
Facebook social activity			
Relationships (n=29)	0.75 (0.79)	1	2
Groups	13.25 (9.02)	7	19
Group admins	0.73 (1.22)	0	1
Likes	46.11 (60.61)	11	54
Pending requests	6.24 (8.62)	0	8
Friends	315.62 (182.80)	179	426
Location tagging	12.8 (10.03)	3	25
Interests	1.13 (2.90)	0	1
App-generated data			
Tips	2.93 (4.02)	0	4
Points	19.96 (12.04)	11	23
CES-D	16.84 (9.56)	10	25
BDI	11.03 (9.29)	3	16

Relationship Between Facebook Social Features and CES-D Scale

To understand the relationship between the various social-network features and one’s depressive state, we examined Spearman rank correlation. Rather than comparing absolute values, we used the rank scale to compensate for the broad range of variations among different features. Table 3 shows the resulting correlation for a representative 9 features. Not all features showed correlation with the CES-D score, such as age (omitted in the table), yet certain features had meaningful correlations. The number of location tags had a negative correlation with the CES-D score (*P*=.045), whereas the accumulated app points and the number of viewed tips had positive correlations (*P*=.04 for both cases). Additionally, the number of friends showed a marginally significant negative correlation (*P*=.08). Some of the features that showed weak trends revealed a stronger tendency when we examined the Spearman correlation among the same features with BDI instead of CES-D scores. For instance, the relationship between BDI scores and the number of friends, app points, and app tips features turned out to be significant (see Multimedia Appendix 3).

Next, we performed a simple linear regression between the BDI and CES-D scores to determine whether participants responded to questionnaires consistently. The BDI and CES-D scores are known to have highly positive correlation [39]. Our linear regression result also confirms the significant relationship between CES-D and BDI (see Multimedia Appendix 4); the Spearman correlation coefficient was 0.839 with *P*<.001, which indicates that participants were consistent in responding to the two surveys. However, we also saw a few outliers whose CES-D and BDI scores were outside the confidence interval in linearity. Hence, in the following group comparison, we excluded these outliers to further increase the level of credibility in data and examine intrinsic traits of individuals in the probably depressed group that can be discriminated from those in the nondepressed group.

Table 3. Spearman rank correlation coefficients between Facebook social features and the CES-D scale.

	Groups	Group admins	Interests	Likes	Pending requests	Location tagging	Friends	App points	App tips
Spearman’s rho	-0.109	-0.104	-0.210	-0.220	-0.074	-0.272	-0.237	0.274	0.278
P value	.43	.45	.12	.11	.59	.045	.08	.04	.04

Difference in Facebook Social Features Between Depressed and Nondepressed Groups

To find differences between depressed and nondepressed individuals, we examined per-group traits by joining the CES-D and BDI scores, as described in the statistical analysis. In doing so, an additional 13 participants were excluded since their CES-D and BDI scores did not match. For instance, some users had a CES-D score greater than or equal to 21 but a BDI score lower than 10. Due to this additional filtering, a total of 42 participants were chosen to be participants for group comparison, where 16 were classified as the probable depression group and 26 as the nondepressed group.

Table 4 shows the results of the Mann-Whitney *U* test between the probably depressed and nondepressed groups. We found that certain Facebook activities had predictive power in distinguishing depressed and nondepressed groups. Among the features, the number of location tags showed marked differences (*P*=.07) in that users in the nondepressed group were far more likely to have a higher number of location tags. In addition, the total accumulated points in EmotionDiary (*P*=.03) and the number of tips viewed (*P*=.01) were also significantly different between the two groups: the depressed group was far more likely to engage in these activities than its nondepressed counterpart.

Table 4. Results of the Mann-Whitney *U* test of Facebook social features between those who were probably depressed[a] (n=16) and nondepressed[b] (n=26).

	Probably depressed[a] mean (SD)	Nondepressed[b] mean (SD)	Z score(two-tailed)	*P* value
Groups	12.56 (8.90)	13.46 (9.35)	-0.207	.84
Group admins	0.75 (1.18)	0.65 (1.26)	-0.365	.72
Likes	34.06 (35.65)	56.46 (79.73)	-0.505	.61
Pending requests	3.50 (6.55)	6.54 (8.67)	-1.122	.26
Friends	253.87 (178.75)	338.42 (183.30)	-1.450	.15
Location tagging	9.62 (9.81)	14.53 (9.78)	-1.790	.07
Interests	0.38 (1.09)	1.62 (3.91)	-1.475	.14
App points	24.69 (14.98)	15.26 (6.43)	-2.229	.03
App tips	4.50 (4.99)	1.38 (2.14)	-2.449	.01

[a] Probably depressed: CES-D≥21 and BDI≥10.

[b] Nondepressed: CES-D≤20 and BDI≤9.

Face-to-Face Interview

One psychiatrist evaluated the depressive symptoms of 2 participants through the Hamilton Depression Rating Scale (HAM-D) [40] within 2 weeks of the EmotionDiary test. All participants showed symptoms of depression including depressed mood, feelings of guilt, insomnia, and anxiety. Their HAM-D scores were higher than 7 points, which is a reliable cutoff point for depression [41]. The participants showed moderate depressive symptoms on both the HAM-D and CES-D; they were judged to be in a chronic depressive state (see Table 5). One participant (Participant A) mainly complained about depressed mood, while the other participant (Participant B) reported a severe loss of interest and exhibited decreased activity in Facebook. This participant used the likes and location-tagging features only once and twice, respectively, and belonged to the 1st quartile in the amount of activities among all Facebook participants. In general, Facebook activities of the 2 participants in location tagging and number of friends, related to depression (shown in Table 3), were relatively minimal. Most values of these features were below the median value.

Table 5. Characteristics of participants.

	Sex	CES-D	HAM-D	Likes	Friends	Location tagging	App tips	Remarks
Participant A	M	32	17	46	19[a]	15	1[b]	Chronic depressive state
Participant B	M	25	17	1[a]	138[a]	2[a]	0[a]	Chronic depressive state

[a] Below 1st quartile value.

[b] Below median value of 55 participants.

Discussion

Principal Findings

OSNs like Facebook have become a primary platform of communication in today's societies. In particular, among young adults, most college students are known to have accounts on OSNs [42] and conduct substantial amounts of interpersonal activities through OSNs [43]. Given their widespread use, this paper explored the idea of using OSNs for a cost-effective and large-scale screening of depression, under the assumption that depressed individuals would exhibit distinguishing behavioral markers online compared to their nondepressed counterparts. Toward this goal, we developed a Facebook Web application called EmotionDiary, which provided users with surveys for depressive symptoms and tips on depression. Based on the analysis driven by data from 55 participants, although preliminary, we found that several Facebook features are associated with depressive symptoms.

First, the number of viewed pages of tips and facts in EmotionDiary is positively correlated with the severity of depression, in that depressed participants read more tips than their nondepressed counterparts. These results suggest that activities in the EmotionDiary app can be more useful and consistent as an acute state marker for depression than other Facebook social features. This may be because many OSN features cannot reflect the acute state of one's mental health since they are formed over a long period of time. Using an app, in contrast, can better reveal the dynamic state of the users. In previous studies, people with psychiatric problems were found to use the Internet to get mental health-related information more frequently than an average person [44,45]. Reading more tips may be related to efforts and interests to overcome their depression. From this finding the app may have the potential to enable successful interaction with people who have depressive symptoms and need some help. One notable observation we made was the difference in behaviors related to chronic quality of depression. From the interviews, although numbers of interviewees were limited, we found that participants with chronic depression (Participants A and B, shown in Table 5) checked tips less frequently than average. Thus, the significant increase in the number of viewed tips for the depression group (see Table 4) indicates that most depressed individuals in our study were experiencing acute stressful events rather than being chronically depressed.

Second, participants who had many Facebook friends showed a low likelihood for depression. Participants with depressive symptoms who suffered from recurrent or chronic depression might not want to increase their number of OSN friends. In previous studies, severely depressed participants did not try to enhance their social network [46]. Also, our observations may be related to findings from other research that shows that Facebook friends can play a role and give social support, which is a key protective factor for depression [47,48]. Previous studies also showed that interpersonal relationships in Facebook are helpful in improving depression [49]. Having many friends on Facebook may allow users to be involved in more communication with other people, thereby reducing depression.

Third, the number of location tags is negatively correlated with one's severity of depression, in that nondepressed individuals were more likely to have location tags. Location tagging is a function commonly used when people visit a new interesting place, for instance, a nice restaurant, park, or concert. Location tagging is the function that requires users to enter their real-world experiences, as one needs to visit a specific place that is notable enough to be worth sharing on Facebook. In fact, loss of interest or pleasure is known to be a key symptom of depression in young adults [50], and this may explain why depressed individuals appear to be less exposed to new experiences (ie, their activities or interests are decreased due to depression) or less likely to “share” their experiences with others. Thus, the decrease in number of location tags might reflect these anhedonia-related symptoms of depression. Additionally, peculiar characteristics related to social withdrawal, such as “Hikikomori”, a Japanese term meaning young adults who withdraw from social life and seek extreme degrees of isolation [51], can affect Facebook features too. Recent findings may show this explanation is possible [52]. The relationships between personal characteristics and Facebook social features should be clarified in future studies. It is possible that the number of location tags can be simply related to total duration of Facebook use, in that the longer the duration of Facebook use, the greater the location tagging. Nonetheless, our findings suggest the possibility that location tagging can be used to mark the status of depressive symptoms.

Interestingly, depressed users show a decreasing tendency to use the like feature, although not to a significant degree, as was seen in the decrease of location tagging. The like feature can show positive empathy, interest, or agreement about someone else's status update. Although the like feature can be related to a decrease in interests in general, this feature is much easier to use than location tagging because location tagging requires physical activity. This difference may explain the weaker correlation between the like feature and depression.

Finally, the qualitative interviews suggested a possibility that differences in individual depression symptoms can affect behaviors on Facebook. For example, participants reporting severe loss of interest showed low activity in likes and location tagging, while participants with chronic depressive symptoms read very few tips. These results may reflect specific types of depressive symptoms or subtypes of depression that can influence which OSNs features users engage with. In fact, depression is a heterogeneous disease and evaluating the subtypes of depression is an important challenge [53,54]. For example, major depressive disorder has several subtypes including melancholic and atypical types, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Different subtypes have differences in symptom domains, as well as epidemiological backgrounds [55], clinical courses [56], endocrine profiles [57], and treatment responses [58]. Therefore, if features of OSNs can give additional information about the specific types of depression more precisely, that can also provide clues toward resolving several debated issues in depression research. We can also consider assessing the answers for each survey question for future studies, since each question is designed to reflect a specific topic such as diet or insomnia, thereby clarifying participants' depressive attributes in more detail.

Results from our interviews could not be generalized to the whole population facing depressive symptoms and could not be evaluated to provide reliability of online screening tests since the number of interviewees was limited (n=2). However, low response rates might reflect the problem of low accessibility to existing face-to-face evaluation. In practice, we were able to contact only 7 of 15 participants who scored greater or equal to 25 in CES-D scores, since the 8 did not want to receive Facebook messages; only 2 of 7 participants who successfully received messages agreed to visit for an interview. Thus, novel approaches, such as EmotionDiary, would be helpful to evaluate and manage depressive symptoms of people who do not want to unveil their depressive symptoms in person.

Limitations and Future Plans

There are several limitations to our study. First, the Facebook social activities we analyzed were accumulated before the CES-D test; we could not measure the patterns of changes in Facebook social features. Therefore, additional prospective studies are needed to mitigate this limitation. Second, the study was conducted with a particular demographic of students in KAIST, which is top-ranked university in South Korea. Also, KAIST is attended mainly by young male students. This limits us from generalizing the findings to the general population and limits the predictive power of Facebook features. Additionally, we think cultural differences could affect our result, as our findings may include Korea-specific trends in the way people use Facebook. However, we could not find ample studies or evidence on how cultural differences affect one's behavior online related to depressive symptoms. To generalize our results, further studies that are focused on transcultural similarities and differences are needed to evaluate depressive mood of groups from various ethnicities and ages. This study is also limited by the relatively small number of participants (N=55) who completed both CES-D and BDI tests. We think much additional research from various perspectives will be necessary to evaluate symptoms and moods appropriately, using online social features.

The number of participants we interviewed is small (n=2), in particular because our aim was to gain quick insights into how participants with depressive symptoms perceived our application. Although the results of interviews could not explain general patterns of groups with depressive symptoms, we could confirm the low accessibility of face-to-face evaluation. We would like to recruit more interviewees or to use other qualitative measurements in future studies. While we did not employ any explicit rewards in EmotionDiary, an appropriate incentive mechanism may be added to facilitate participant recruitment and the successful completion of depression surveys.

Despite the abovementioned limitations, to the best of our knowledge, our study is the first attempt to identify an association between social features on Facebook and users' depressive symptoms. By analyzing Facebook-related depressive traits, we tried to understand human behaviors in a social relationship that could predict depressive moods. Therefore, this study is an important step toward the problem of large-scale screening of depression on OSN platforms. As OSNs are becoming a primary communication platform for more people, we believe mobile and Web-based applications like EmotionDiary can serve an important role in increasing the awareness of depressive symptoms in our society and promoting positive health behaviors in an unintrusive manner.

Acknowledgments

This research was supported by the KUSTAR-KAIST Institute under the R&D program supervised by KAIST, the WCU (World Class University) program under the National Research Foundation of Korea and funded by the Ministry of Education, Science and Technology of Korea (R31-30007), and the Basic Science Research Program through the National Research Foundation of Korea (2011-0012988). The study was also supported by a grant for the KAIST Future Systems Healthcare Project from the Ministry of Education, Science and Technology (N01130009, N01130010).

The application code used in this study will be made available upon request for research purposes [59]. We plan to open our code to the public through an online code-sharing service.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flowchart of the overall experiment and evaluation process.

[PDF File (Adobe PDF File), 171KB - jmir_v15i10e217_app1.pdf]

Multimedia Appendix 2

Pilot study concepts and interviewee characteristics.

[PDF File (Adobe PDF File), 5KB - jmir_v15i10e217_app2.pdf]

Multimedia Appendix 3

Spearman rank correlation coefficients between the online social features and the BDI scores.

[PDF File (Adobe PDF File), 4KB - jmir_v15i10e217_app3.pdf]

Multimedia Appendix 4

Relationship between the CES-D and BDI scores of 55 participants.

[PDF File (Adobe PDF File), 131KB - jmir_v15i10e217_app4.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
2. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]
3. Donohue JM, Pincus HA. Reducing the societal burden of depression: a review of economic costs, quality of care and effects of treatment. Pharmacoeconomics 2007;25(1):7-24. [Medline: 17192115]
4. World Health Organization. An overview of a strategy to improve the mental health of underserved populations. 1999. URL: http://www.who.int/mental_health/media/en/46.pdf [accessed 2013-05-15] [WebCite Cache ID 6Gd8XokRH]
5. National Institute of Mental Health. Mental disorders cost society billions in unearned income. 2008. URL: http://www.nimh.nih.gov/news/science-news/2008/mental-disorders-cost-society-billions-in-unearned-income.shtml [accessed 2013-05-15] [WebCite Cache ID 6GdAvPS6d]
6. Mueller TI, Keller MB, Leon AC, Solomon DA, Shea MT, Coryell W, et al. Recovery after 5 years of unremitting major depressive disorder. Arch Gen Psychiatry 1996 Sep;53(9):794-799. [Medline: 8792756]
7. City of Milwaukee Employee Assistance Program. October 11th is National Depression Screening Day. 2012. URL: http://city.milwaukee.gov/ImageLibrary/User/jkamme/EAP/Info-Library/MentalHealth_NatlDepressionScr.pdf [accessed 2013-07-31] [WebCite Cache ID 6IWS5nHpA]
8. Diane Forrest. National Anxiety and Depression Awareness Week. 2012. URL: http://kisbyto.blogspot.kr/2012/05/national-anxiety-and-depression.html [accessed 2013-07-31] [WebCite Cache ID 6IWSHjFI7]
9. 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]
10. Brodkin J. Facebook vs Twitter. 2010. URL: http://www.networkworld.com/news/2010/060710-tech-argument-facebook-twitter.html [accessed 2013-07-29] [WebCite Cache ID 6ITNVnSR9]
11. Rude S, Valdez C, Odom S, Ebrahimi A. Negative cognitive biases predict subsequent depression. Cognitive Ther Res 2003;27(4):415-429. [doi: 10.1023/A:1025472413805]
12. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011 Mar;16(3):273-281. [doi: 10.1038/mp.2010.13] [Medline: 20231839]
13. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Jul 08 Presented at: 7th International AAAI Conference on Weblogs and Social Media; Jul 8-10, 2013; Boston, MA, USA.
14. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. 2013 May 02 Presented at: 5th ACM International Conference on Web Science; May 2-4, 2013; Paris, France.
15. Sadilek A, Kautz H, Silenzio V. Modeling spread of disease from social interactions. 2012 Jun 04 Presented at: 6th International AAAI Conference on Weblogs and Social Media; June 4-7, 2012; Dublin, Ireland.
16. Park M, Cha C, Cha M. Depressive moods of users portrayed in twitter. 2012 Aug 12 Presented at: ACM SIGKDD Workshop on Healthcare Informatics; Aug 12, 2012; Beijing, China.

17. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 Apr 27 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr 27-May 2, 2013; Paris, France.
18. Neuman Y, Cohen Y, Assaf D, Kedma G. Proactive screening for depression through metaphorical and automatic text analysis. Artif Intell Med 2012 Sep;56(1):19-25. [doi: 10.1016/j.artmed.2012.06.001] [Medline: 22771201]
19. Katikalapudi R, Chellappan S, Montgomery F, Wunsch D, Lutzen K. Associating Internet Usage with Depressive Behavior Among College Students. IEEE Technol Soc Mag 2012 Dec;31(4):73-80. [doi: 10.1109/MTS.2012.2225462]
20. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
21. Rosenquist JN, Fowler JH, Christakis NA. Social network determinants of depression. Mol Psychiatry 2011;16:273-281 [FREE Full text]
22. Fox S, Jones S. The Social Life of Health Information. 2009. URL: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx [accessed 2013-02-05] [WebCite Cache ID 6ECPtTKvb]
23. Carl Jackers. Top 10 Most Visited Social Media Sites in the World. 2013. URL: http://americanlivewire.com/social-media-sites/ [accessed 2013-07-29] [WebCite Cache ID 6ITleDr9u]
24. Yoon Ja-young. Is homegrown Cyworld giving in to Facebook?. 2011. URL: http://www.koreatimes.co.kr/www/news/special/2011/10/182_96267.html [accessed 2013-07-29] [WebCite Cache ID 6ITUse2G3]
25. Danah M, Nichol BE. Social network sites: Definition, history, and scholarship. J Comput-Mediat Comm 2007;13:210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
26. Ellison N, Steinfield C, Lampe C. The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites. J Comput-Mediat Comm 2007;12:1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
27. Radloff LS. The CES-D scale: A self-report symptom scale to depression scale in the general population. Appl Psych Meas 1977;1(3):385-401. [doi: 10.1111/j.1083-6101.2007.00367.x]
28. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
29. Cho MJ, Kim KH. Diagnostic validity of the CES-D (Korean version) in the assessment of DSM-III-R major depression. J Korean Neuropsychiaty Assoc 1993;32:381-399.
30. Lee YH, Song JY. A study of the reliabilitythe validity of the BDI, SDS. and MMPI-D scales. Korean J of Clin Psychol 1991;10:98-113.
31. Cho MJ, Nam JJ, Suh GH. Prevalence of symptoms of depression in a nationwide sample of Korean adults. Psychiatry Res 1998 Dec 14;81(3):341-352. [Medline: 9925185]
32. Haringsma R, Engels GI, Beekman AT, Spinhoven P. The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. Int J Geriatr Psychiatry 2004 Jun;19(6):558-563. [doi: 10.1002/gps.1130] [Medline: 15211536]
33. Ilardi SS. The Depression Cure: The 6-Step Program to Beat Depression without Drugs. Cambridge, MA: Da Capo Lifelong Books; 2009.
34. RandomHistory. 2009. URL: http://facts.randomhistory.com/random-facts-about-depression.html [accessed 2013-07-26] [WebCite Cache ID 6IOM3hRGK]
35. Ferebee SS, Davis JW. Factors that persuade continued use of Facebook among new members. New York: ACM; 2009 Apr 26 Presented at: 4th International Conference on Persuasive Technology; Apr 26–29, 2009; Claremont, CA, USA.
36. Hsu Y. Facebook as international eMarketing strategy of Taiwan hotels. International Journal of Hospitality Management 2012 Sep;31(3):972-980. [doi: 10.1016/j.ijhm.2011.11.005]
37. Cho MJ, Kim KH. Use of the Center for Epidemiologic Studies Depression (CES-D) Scale in Korea. J Nerv Ment Dis 1998 May;186(5):304-310. [Medline: 9612448]
38. Beck AT, Beamesderfer A. Assessment of depression: the depression inventory. Mod Probl Pharmacopsychiatry 1974;7:151-169. [Medline: 4412100]
39. Wilcox H, Field T, Prodromidis M, Scafidi F. Correlations between the BDI and CES-D in a sample of adolescent mothers. Adolescence 1998;33(131):565-574. [Medline: 9831873]
40. Hamilton M. Rating depressive patients. J Clin Psychiatry 1980 Dec;41(12 Pt 2):21-24. [Medline: 7440521]
41. Riedel M, Möller HJ, Obermeier M, Schennach-Wolff R, Bauer M, Adli M, et al. Response and remission criteria in major depression--a validation of current practice. J Psychiatr Res 2010 Nov;44(15):1063-1068. [doi: 10.1016/j.jpsychires.2010.03.006] [Medline: 20447651]
42. Lee EB. Young, black, and connected: Facebook usage among African American college students. J Black Stud 2012;43(3):336-354. [Medline: 22536626]
43. Ljepava N, Orr RR, Locke S, Ross C. Personality and social characteristics of Facebook non-users and frequent users. Computers in Human Behavior 2013 Jul;29(4):1602-1607. [doi: 10.1016/j.chb.2013.01.026]
44. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry 2006 Sep;189:273-277 [FREE Full text] [doi: 10.1192/bjp.bp.105.017319] [Medline: 16946364]

45. Haviland MG, Pincus HA, Dial TH. Type of illness and use of the Internet for health information. Psychiatr Serv 2003 Sep;54(9):1198. [Medline: 12954932]
46. Amann G. Social network and social support deficits in depressed patients: a result of distorted perception? Eur Arch Psychiatry Clin Neurosci 1991;241(1):49-56. [Medline: 1832306]
47. Garber J. Depression in children and adolescents: linking risk research and prevention. Am J Prev Med 2006 Dec;31(6 Suppl 1):S104-S125. [doi: 10.1016/j.amepre.2006.07.007] [Medline: 17175406]
48. Prince MJ, Harwood RH, Blizard RA, Thomas A, Mann AH. Social support deficits, loneliness and life events as risk factors for depression in old age. The Gospel Oak Project VI. Psychol Med 1997 Mar;27(2):323-332. [Medline: 9089825]
49. Wright KB, Rosenberg J, Egbert N, Ploeger NA, Bernard DR, King S. Communication competence, social support, and depression among college students: a model of facebook and face-to-face support network influence. J Health Commun 2013;18(1):41-57. [doi: 10.1080/10810730.2012.688250] [Medline: 23030518]
50. Friedman RC, Hurt SW, Clarkin JF, Corn R, Aronoff MS. Symptoms of depression among adolescents and young adults. J Affect Disord 1983 Feb;5(1):37-43. [Medline: 6220042]
51. Furlong A. The Japanese hikikomori phenomenon: acute social withdrawal among young people. The sociological review 2008;56:309-325 [FREE Full text]
52. Martin EA, Bailey DH, Cicero DC, Kerns JG. Social networking profile correlates of schizotypy. Psychiatry Res 2012 Dec 30;200(2-3):641-646. [doi: 10.1016/j.psychres.2012.06.031] [Medline: 22796101]
53. Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiatr Scand 2011 Dec;124(6):495-496. [doi: 10.1111/j.1600-0447.2011.01744.x] [Medline: 21838736]
54. Parker G. Classifying clinical depression: an operational proposal. Acta Psychiatr Scand 2011 Apr;123(4):314-316. [doi: 10.1111/j.1600-0447.2011.01681.x] [Medline: 21320078]
55. Benazzi F. Prevalence and clinical features of atypical depression in depressed outpatients: a 467-case study. Psychiatry Res 1999 Jun 30;86(3):259-265. [Medline: 10482345]
56. Chen L, Eaton WW, Gallo JJ, Nestadt G. Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. J Affect Disord 2000 Jul;59(1):1-11. [Medline: 10814765]
57. Gold PW, Chrousos GP. The endocrinology of melancholic and atypical depression: relation to neurocircuitry and somatic consequences. Proc Assoc Am Physicians 1999;111(1):22-34. [Medline: 9893154]
58. Henkel V, Mergl R, Allgaier AK, Kohnen R, Möller HJ, Hegerl U. Treatment of depression with atypical features: a meta-analytic approach. Psychiatry Res 2006 Jan 30;141(1):89-101. [doi: 10.1016/j.psychres.2005.07.012] [Medline: 16321446]
59. The EmotionDiary project webpage (application code available). URL: http://mia.kaist.ac.kr/publications/emotiondiary/ [accessed 2013-09-25] [WebCite Cache ID 6JtbDWtof]

Abbreviations

API: application programming interface
App: application
BBS: bulletin board system
BDI: Beck's Depression Inventory
CES-D: Center for Epidemiologic Studies-Depression
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders
HAM-D: Hamilton Depression Rating Scale
KAIST: Korea Advanced Institute of Science and Technology
OSN: online social network

Edited by G Eysenbach; submitted 15.05.13; peer-reviewed by M Conway, S Langrial; comments to author 18.06.13; revised version received 31.07.13; accepted 29.08.13; published 01.10.13.

Please cite as:
Park S, Lee SW, Kwak J, Cha M, Jeong B
Activities on Facebook Reveal the Depressive State of Users
J Med Internet Res 2013;15(10):e217
URL: http://www.jmir.org/2013/10/e217/
doi:10.2196/jmir.2718
PMID:24084314

Letter to the Editor

Security Concerns to Be Considered When Downloading Human Immunodeficiency Virus/Sexually Transmitted Disease Related Smartphone Applications

Savita Lorreta Brito-Mutunayagam[1], BmedSci, MBChB; Imali Fernando[1], Dip HIV, MRCP, MB ChB

[1] Chalmers Sexual Health Centre, Edinburgh, United Kingdom

Corresponding Author:
Savita Lorreta Brito-Mutunayagam, BmedSci, MBChB
Chalmers Sexual Health Centre
2a Chalmers Street
Edinburgh, EH3 9ES
United Kingdom
Phone: 44 447833497767
Fax: 44 0131 5361609
Email: savita.brito.mutunayagam@gmail.com

(J Med Internet Res 2013;15(10):e222) doi:10.2196/jmir.2650

KEYWORDS

HIV/STD; smartphone applications; security

“Mobile Phone Applications for the Care and Prevention of HIV and Other Sexually Transmitted Diseases: A Review” by Muessig et al is an excellent review that succinctly summarizes currently available mobile phone applications (apps) related to the prevention and care of human immunodeficiency virus (HIV) and other human immunodeficiency virus/sexually transmitted disease (STDs) [1]. The authors have comprehensively reviewed apps related to HIV/STDs and identified the need for health care professionals to work closely with app developers to provide accurate evidence based advice, and design effective risk reduction interventions.

While Muessig et al rightly raised concerns regarding the accuracy and reliability of the content of these apps, an important area not discussed in their review is the security of the apps. When downloading an app, the user is asked to authorize the “permissions” requested by the application. These “permissions” enable the optimum performance of the app on a smartphone [2]. There are over 100 different “ permissions” requested by smartphone applications. While some of the “permissions” requested are harmless, many raise serious concerns regarding the confidentiality and security of the apps requesting them [2]. These include permissions that request;

The above mentioned permissions that an app may require for optimum functioning involve access to and control of sensitive personal data. Applications often have legitimate reasons for accessing this sensitive and private data. Permission to obtain the exact GPS location of the app user is necessary if the app is designed to provide information on the nearest HIV/ STD testing center. If the app is designed as a personal assistant for those living with HIV, access to the user’s calendar is important to remind them of their next hospital appointment.

However, the concern arises when the app is not developed by a named professional health care body/organization and there is no assurance of confidentiality. Today’s smartphone applications often fail to provide users with visibility into where their private data is being stored and how it is being used. There are often significant social implications associated with a diagnosis of HIV and the secure storage of their personal information is of immense importance to those living with the condition. Even individuals simply looking for information on the topic, or calculating their risk of contracting a STD after unprotected sexual intercourse, may be concerned if an unverified smartphone application had access to their personal information including precise location.

Muessig et al reviewed HIV and STD related apps that matched their search criteria in the Apple iTunes Store and the Android Google Play Store, as combined these two companies account for over 86% of the global app market [1]. In the android store this information is readily available within the app details. Apple does not explicitly specify permissions required in the app details, but this information is available on download. Apple states that all their apps are pre-screened prior to making them available for download. However, the recent controversy surrounding Apple, for enabling the download of malicious apps that stole their users’ address books, show that this screening process is not infallible [3].

Providing HIV and STD prevention and care services via smartphone applications is an area of rapid and immense growth. If provided by trusted and professional organizations which guarantee the security of their users’ personal information, they can be a powerful and rapidly accessible resource. However, it is essential that users are aware of the potential confidentiality and security breaches when downloading these apps. They must be encouraged to pay attention to the app developer in order to ascertain if it is a reputed body Furthermore, they should note the permissions requested by the apps and only proceed with the download if they are comfortable with these requests. Whilst more vigilance amongst app users is essential, it is also the responsibility of the companies that offer these apps to ensure their products are not malicious and employ the highest levels of data protection software.

Conflicts of Interest

None declared.

References

1. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
2. Chia HP, Yamamoto Y, Asokan N. International World Wide Web Conference. 2012 Apr. Is this App Safe? A Large Scale Study on Application Permissions and Risk Signals URL: http://www2012.wwwconference.org/proceedings/proceedings/p311.pdf [accessed 2013-04-04] [WebCite Cache ID 6Fd53y4iK]
3. Gilbert J. The Huffington Post. 2012 Feb. iPhone App Privacy: Path, Facebook, Twitter and Apple Under Scrutiny for Address Book Controversy URL: http://www.huffingtonpost.com/2012/02/15/iphone-privacy-app-path-facebook-twitter-apple_n_1279497.html [accessed 2013-04-04] [WebCite Cache ID 6Fd6eINX7]

Edited by G Eysenbach; submitted 04.04.13; this is a non–peer-reviewed article;accepted 01.10.13; published 07.10.13.

Please cite as:
Brito-Mutunayagam SL, Fernando I
Security Concerns to Be Considered When Downloading Human Immunodeficiency Virus/Sexually Transmitted Disease Related Smartphone Applications
J Med Internet Res 2013;15(10):e222
URL: http://www.jmir.org/2013/10/e222/
doi:10.2196/jmir.2650
PMID:24100134

Letter to the Editor

Online Survey on Twitter: A Urological Experience

Fabrizio Dal Moro[1], MD

[1] Urology, Department of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy

Corresponding Author:
Fabrizio Dal Moro, MD
Urology
Department of Surgical, Oncological and Gastroenterological Sciences
University of Padova
Via Giustiniani 2
Padova, 35126
Italy
Phone: 39 0498212720
Fax: 39 0498218757
Email: fabrizio.dalmoro@unipd.it

Related Articles:

Comment on: http://www.jmir.org/2012/6/e176/

Comment on: http://www.jmir.org/2013/4/e62/

(J Med Internet Res 2013;15(10):e238) doi:10.2196/jmir.2719

KEYWORDS

Twitter; social media; survey

After my return from the Annual Congress of the American Urological Association in San Diego (May 3-8, 2013), I would like to congratulate McKendrick et al [1] and Hanson et al [2] on their recent papers on the impact of Twitter, one of the most commonly used social media, in the medical and non-medical fields. In their elegant communications, both groups of authors stressed the role of the Internet and the Web as new methods of publicizing scientific data. Starting from the use of Twitter during a medical conference, McKendrick et al reported their first experience of using social media as support for conference organizers, highlighting the use of a Twitter stream as an integral part of the communication structure of a conference on anesthetics. Hanson et al reported a Twitter-based surveillance method for monitoring public health regarding the use and abuse of a psychostimulant drug, emphasizing the potential role of this social media in collecting data for a survey.

To confirm these findings and encourage the use of these novel tools “to do science”, I would like to share a personal pioneering experiment I carried out during the Annual Congress of the European Association of Urology (March 2013, in Milan) and the recent American Congress in San Diego. During the meetings, I launched an online survey using Twitter, and posted an interactive specific urological question concerning the choice of preferred approach to robotic radical prostatectomy, the most recent and widespread surgical intervention for prostate cancer. The tweet was: “ONLINE-SURVEY: Do you perform ONLY Transperitoneal (T), ONLY Extraperitoneal (E) or BOTH approaches to dVP? (dVP= Da Vinci robotic prostatectomy) ReTweet T, E or TE” (Figure 1).

I used two specific hashtags (#EAU2013 for Milan, and #AUA2013 for San Diego).

In the course of both congresses, I received no fewer than 326 answers. Considering the absolute number of participants (approximately 18,000 in both congresses) the percentage of Tweets received (about 2%) may seem very low, but considering that it was a first "urological" experience, I think it may be significant.

Evaluating the survey results, 81 out of the 326 urologists (24.9% of those who used Twitter and retweeted me their answers!) perform both procedures. In 165 cases (50.6%), the preferred approach to robotic prostatectomy is transperitoneal.

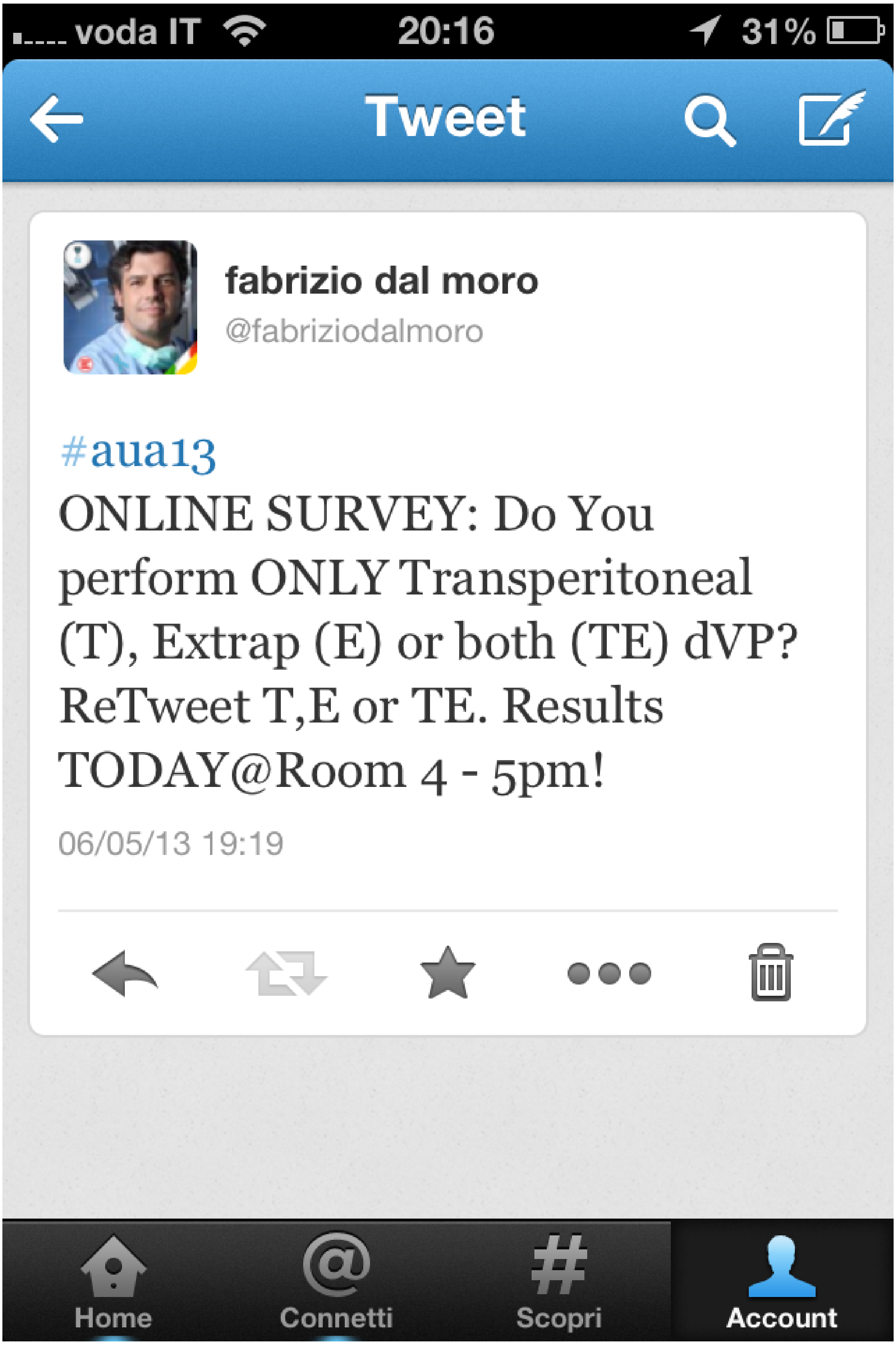
I presented the results of this survey during my podium lecture in San Diego.

Obviously, my study is not a scientific survey without biases, but it did reveal the spread of social media, to the scientific community too.

In my opinion, this simple experiment not only confirms considerations about the use of social media by these authors, but also helps all of us to appreciate the fact that Twitter, like

Facebook, Google+, Linkedin and all the other weblinked social tools, is not limited to young people "chatting about love or friends", but are novel instruments which can be used to even greater scientific advantage.

Figure 1. Screenshot of the Tweet.



fabrizio dal moro
@fabriziodalmoro

#aua13
ONLINE SURVEY: Do You perform ONLY Transperitoneal (T), Extrap (E) or both (TE) dVP? ReTweet T,E or TE. Results TODAY@Room 4 - 5pm!

06/05/13 19:19

Conflicts of Interest

None declared.

References

1. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res 2012;14(6):e176 [FREE Full text] [doi: 10.2196/jmir.2144] [Medline: 23232765]
2. 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]

Edited by G Eysenbach; submitted 15.05.13; peer-reviewed by C Giraud-Carrier, J West; comments to author 06.10.13; revised version received 15.10.13; accepted 22.10.13; published 25.10.13.

Please cite as:
Dal Moro F
Online Survey on Twitter: A Urological Experience
J Med Internet Res 2013;15(10):e238
URL: http://www.jmir.org/2013/10/e238/
doi:10.2196/jmir.2719
PMID:24164710

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org