Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 17 (2015), Issue 6 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Corrigenda and Addendas

Original Paper

Exposure to a Patient-Centered, Web-Based Intervention for Managing Cancer Symptom and Quality of Life Issues: Impact on Symptom Distress

Donna L Berry[1,2*], RN, PhD, FAAN; Traci M Blonquist[3*], MS; Rupa A Patel[4], PhD; Barbara Halpenny[5*], MA; Justin McReynolds[6], BS

[1]Phyllis F. Cantor Center, Dana-Farber Cancer Institute, Boston, MA, United States
[2]Department of Medicine, Harvard Medical School, Boston, MA, United States
[3]Biostatistics & Computational Biology, Dana-Farber Cancer Institute, Boston, MA, United States
[4]Usability Engineering, GoDaddy Group, Kirkland, WA, United States
[5]Phyllis F. Cantor Center, Dana-Farber Cancer Institute, Boston, MA, United States
[6]Biomedical and Health Informatics, University of Washington, Seattle, WA, United States
[*]these authors contributed equally

Corresponding Author:
Donna L Berry, RN, PhD, FAAN
Phyllis F Cantor Center
Dana-Farber Cancer Institute
450 Brookline Ave; LW 518
Boston, MA, 02215
United States
Phone: 1 617 632 1909
Fax: 1 617 394 2877
Email: donna_berry@dfci.harvard.edu

Abstract

Background: Effective eHealth interventions can benefit a large number of patients with content intended to support self-care and management of both chronic and acute conditions. Even though usage statistics are easily logged in most eHealth interventions, usage or exposure has rarely been reported in trials, let alone studied in relationship to effectiveness.

Objective: The intent of the study was to evaluate use of a fully automated, Web-based program, the Electronic Self Report Assessment-Cancer (ESRA-C), and how delivery and total use of the intervention may have affected cancer symptom distress.

Methods: Patients at two cancer centers used ESRA-C to self-report symptom and quality of life (SxQOL) issues during therapy. Participants were randomized to ESRA-C assessment only (control) or the ESRA-C intervention delivered via the Internet to patients' homes or to a tablet at the clinic. The intervention enabled participants to self-monitor SxQOL and receive self-care education and customized coaching on how to report concerns to clinicians. Overall and voluntary intervention use were defined as having ≥2 exposures, and one non-prompted exposure to the intervention, respectively. Factors associated with intervention use were explored with Fisher's exact test. Propensity score matching was used to select a sample of control participants similar to intervention participants who used the intervention. Analysis of covariance (ANCOVA) was used to compare change in Symptom Distress Scale (SDS-15) scores from pre-treatment to end-of-study by groups in the matched sample.

Results: Radiation oncology participants used the intervention, overall and voluntarily, more than medical oncology and transplant participants. Participants who were working and had more than a high school education voluntarily used the intervention more. The SDS-15 score was reduced by an estimated 1.53 points (*P*=.01) in the intervention group users compared to the matched control group.

Conclusions: The intended effects of a Web-based, patient-centered intervention on cancer symptom distress were modified by intervention use frequency. Clinical and personal demographics influenced voluntary use.

Trial Registration: Clinicaltrials.gov NCT00852852; http://clinicaltrials.gov/ct2/show/NCT00852852 (Archived by WebCite at http://www.webcitation.org/6YwAfwWl7).

(J Med Internet Res 2015;17(6):e136) doi:10.2196/jmir.4190

KEYWORDS

patient-centered technology; cancer; symptoms; quality of life; eHealth; Internet

Introduction

Background

Clinicians and researchers have developed eHealth solutions that supplement the limited time for patient report and communication within the confines of the ambulatory care, face-to-face visit [1]. Reported benefits of eHealth solutions for oncology care include improved patient well-being [2,3], better patient-clinician communication [2,4], and lower symptom distress [5]. Effective eHealth interventions can benefit a large number of patients with both generic and tailored content. Even though usage statistics are easily logged in most eHealth interventions, usage or exposure has rarely been reported in trials, let alone studied in relationship to effectiveness. As reviewed by Donkin et al [6], the “dose” of eHealth solutions, comprehensive measures of intervention exposure or patient engagement, have been documented in few trials evaluating health promotion and mental health interventions. Furthermore, eHealth intervention delivery has been studied in only one cancer symptom and quality of life trial, in association with outcomes in breast cancer survivors [7]. The ability and efforts of patients in active cancer treatment to fully utilize such solutions are uncertain.

The Electronic Self Report Assessment for Cancer (ESRA-C) is a patient-centered technology developed with rigorous participatory design methods [8] and evaluated in multi-site randomized trials [4,5]. ESRA-C, a Web-based intervention that supports patients with any cancer diagnosis during treatment, has been shown to significantly increase the frequency of patient-clinician communication about problematic issues [4], reduce symptom distress over the course of active therapy [5], and increase the patient’s unsolicited and specific description of symptoms and quality of life (SxQOL) concerns [9]. However, when we conducted a mediation analysis of the impact of the intervention group’s increased patient verbal reports at one clinic visit during the trial, we found no significant impact on the primary outcome of symptom distress [9]. In other words, another aspect of the intervention was responsible for the reduction of symptom distress.

Objective

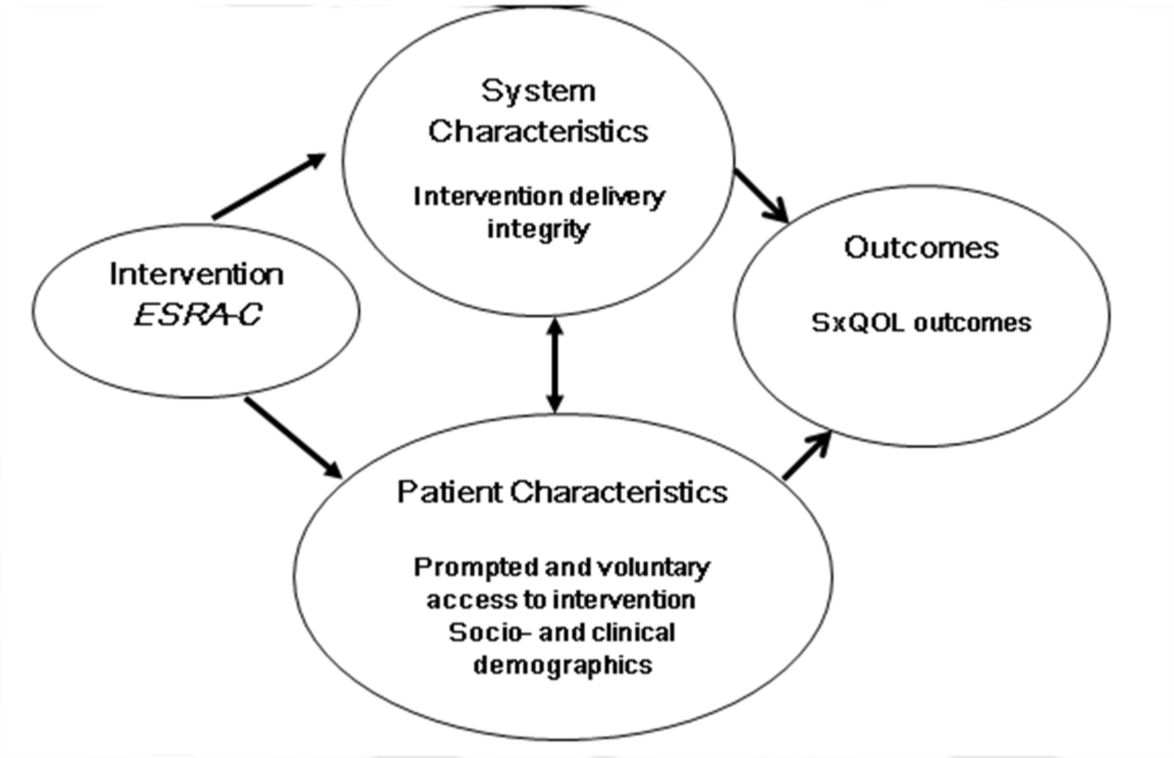
The purpose of this analysis was to determine the impact of the ESRA-C intervention exposure on cancer symptom distress and describe frequency of intervention use by participants in the ESRA-C II trial.

Methods

Overview

This analysis addresses one component of our program of research founded on the Quality Health Outcomes Model, a framework proposed by Mitchell and colleagues [10] to illustrate that patient outcomes are rarely explained only by specific interventions but also by health care system/provider factors and patient-specific factors. The extent of patients’ use of the intervention can be placed in the model (Figure 1) as a patient-specific factor that may influence the impact of the ESRA-C intervention on symptom distress.

Figure 1. Adapted Health Outcomes Model [10].



Design, Sample, Intervention

The ESRA-C II was a randomized trial conducted at two comprehensive cancer centers. Full details of the trial [5] and system development were reported elsewhere [11]. In brief, adult participants with any type or stage of cancer, and about to start a new anti-cancer therapy, used ESRA-C to self-report baseline SxQOL and then were randomized to receive usual education about SxQOL topics or usual education plus the opportunity to self-monitor when not in the clinic, tailored self-care instruction for SxQOL issues, and communication coaching on how to report each SxQOL to clinicians. While in the clinic, the intervention group participants were shown an overview of the ESRA-C intervention and voluntary remote use was encouraged. Participants in the intervention group could access the ESRA-C program from home or in clinic on a touch-screen computer at any time throughout the trial to electronically track SxQOL and view the intervention. Those intervention group participants without Internet access were encouraged to meet the research coordinator during any subsequent visit to the clinic and use ESRA-C on a study tablet. Participants in both groups were asked to report SxQOL using the ESRA-C system from home or clinic at three study time points (T2-T4) throughout the course of therapy, coinciding with clinic visits at which clinicians would receive a printed summary of the patient report for participants in both groups. Home user participants in both groups were prompted by email, 24 hours prior to a scheduled clinic visit, to use the SxQOL report feature of ESRA-C. Clinic users were notified to arrive about 45 minutes prior to scheduled clinic visits corresponding to each study time point in order to use the reporting feature and, if in the intervention group, components. Intervention group participants had access to the ESRA-C intervention *Teaching Tips* and *View My Reports* components.

Following the SxQOL report in prompted T2-T4 sessions, the intervention group participants received pushed teaching tips for those SxQOL issues reported as moderate-to-severe. Within each pushed teaching tip was the option to expand linked text addressing (1) “Why does this happen?”, (2) “What can I do about this?”, and (3) “What do I tell my clinical team?” (Figure 2a). After the SxQOL report and pushed teaching tips, the participant could navigate to the *Teaching Tips* tab or the *View My Reports* tab within the intervention home page (Figure 2b-c) by clicking on the designated tab. A click on the (non-pushed) *Teaching Tips* tab displayed a dropdown list of all 26 SxQOL issues and the option to select and expand any issues. A click on the *View My Reports* tab displayed thumbnail line graphs tracking SxQOL reports over time.

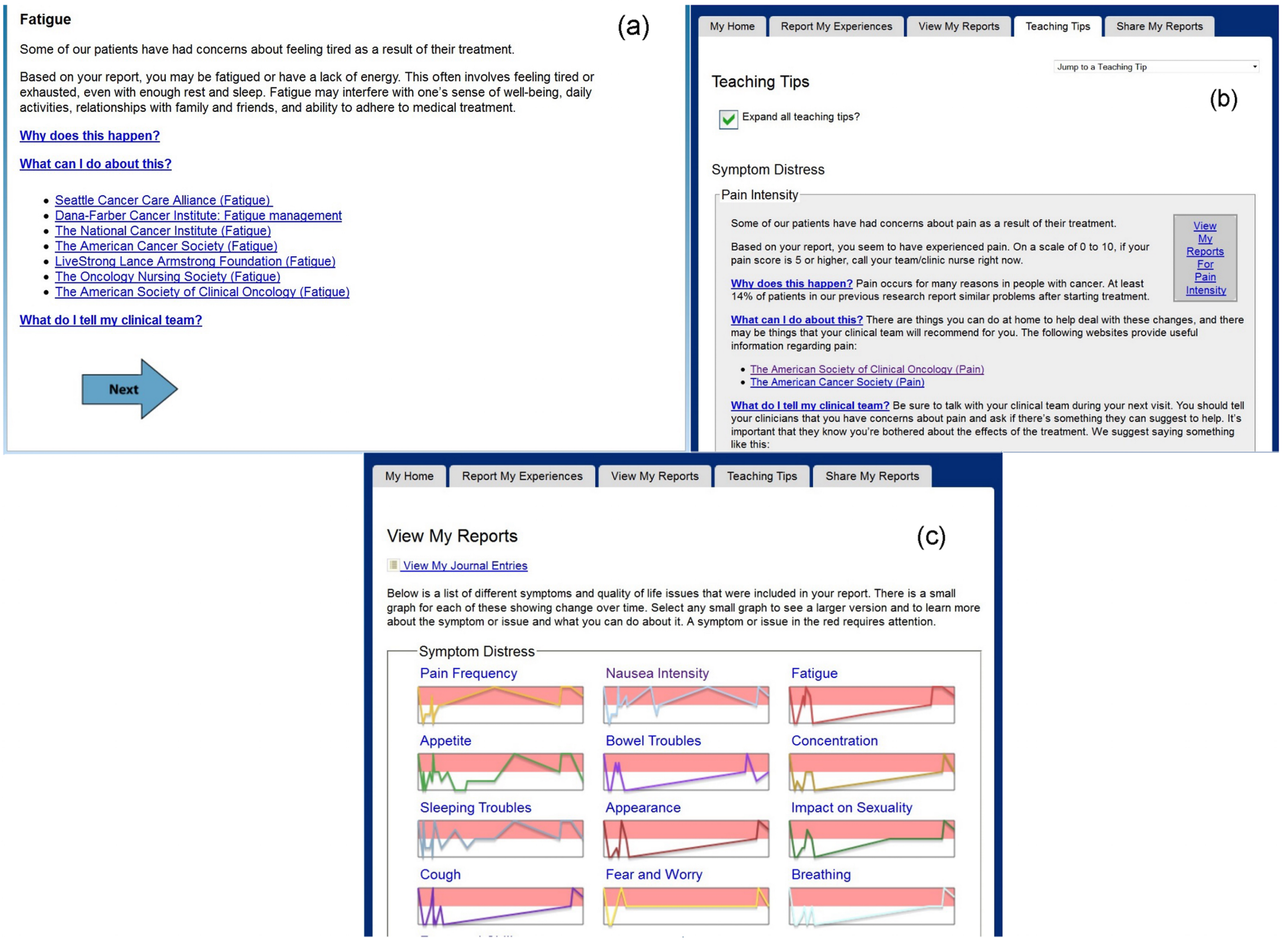
Intervention group participants were invited to access ESRA-C at any time between prompted sessions and clinic visits. These sessions were defined as any intervention use that was not prompted. During voluntary, non-prompted sessions, the participant did not receive pushed *Teaching Tips*, but did have the option to report SxQOL, and click the (non-pushed) *Teaching Tips* tab and the *View My Reports* tab.

The ESRA-C intervention was considered delivered if the participant accessed the *Teaching Tips* and/or *View My Reports*. As a conservative measure of exposure to *Teaching Tips*, if at least one pushed tip was delivered during a prompted session, this was considered comparable to a single click on the non-pushed *Teaching Tips*. For example, a participant with at least one pushed teaching tip at each of three prompted sessions would have a total of three pushed teaching tips. Total exposure to the intervention consisted of three components: (1) the number of pushed teaching tips during prompted sessions, (2) the number of clicks on the non-pushed Teaching Tips tab during prompted and non-prompted sessions, and (3) the number of clicks on the *View My Reports* tab during prompted and non-prompted sessions. Voluntary exposure only occurred during non-prompted sessions and consisted of (1) the number of clicks on the (non-pushed) *Teaching Tips* tab, and (2) the number of clicks on the *View My Reports* tab.

At prompted study time points, all participants were presented a set of SxQOL self-report assessments that included the Symptom Distress Scale-15 (SDS-15) [5], cancer quality of life questionnaires EORTC QLQ-C30 [12] and EORTC-CPIN20 [13], the Patient Health Questionnaire (PHQ-9) depression scale [14], a 0-10 pain intensity numerical scale, and a skin problems questionnaire [15]. At unprompted sessions, intervention group participants could choose to access any or all of the questionnaires. These procedures were fully described previously [5].

A total of 752 participants were randomized in the parent trial: 374 intervention and 378 control. Of those, 523 (262 intervention, 261 control) had complete SDS-15 baseline and end-of-study scores. In the primary analysis of covariance (ANCOVA), the average SDS-15 score was reduced by an estimated 1.21 points (95% CI 0.23-2.20; *P*=.02) in the intervention group compared to the control group [5].

Figure 2. Exemplar screen shots from the ESRA-C intervention: a) pushed Teaching Tip, b) Teaching Tips tab, c) View My Reports tab.



Some of our patients have had concerns about feeling tired as a result of their treatment.

Based on your report, you may be fatigued or have a lack of energy. This often involves feeling tired or exhausted, even with enough rest and sleep. Fatigue may interfere with one's sense of well-being, daily activities, relationships with family and friends, and ability to adhere to medical treatment.

☑ Expand all teaching tips?

Some of our patients have had concerns about pain as a result of their treatment.

Based on your report, you seem to have experienced pain. On a scale of 0 to 10, if your pain score is 5 or higher, call your team/clinic nurse right now.

Why does this happen? Pain occurs for many reasons in people with cancer. At least 14% of patients in our previous research report similar problems after starting treatment.

What can I do about this? There are things you can do at home to help deal with these changes, and there may be things that your clinical team will recommend for you. The following websites provide useful information regarding pain:

- The American Society of Clinical Oncology (Pain)
- The American Cancer Society (Pain)

What do I tell my clinical team? Be sure to talk with your clinical team during your next visit. You should tell your clinicians that you have concerns about pain and ask if there's something they can suggest to help. It's important that they know you're bothered about the effects of the treatment. We suggest saying something like this:

Below is a list of different symptoms and quality of life issues that were included in your report. There is a small graph for each of these showing change over time. Select any small graph to see a larger version and to learn more about the symptom or issue and what you can do about it. A symptom or issue in the red requires attention.

Analytic Methods

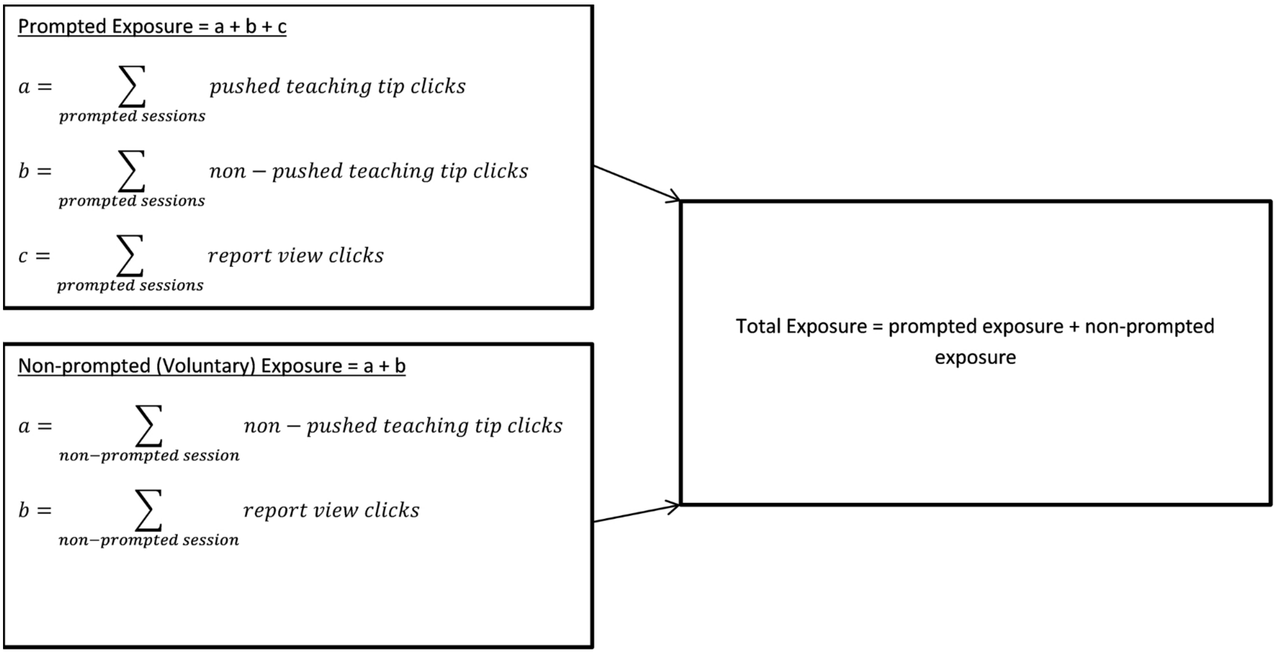
Total exposure was calculated as the sum of the number of clicks on pushed teaching tips, clicks on the non-pushed *Teaching Tips* tab, and the number of clicks on the *View My Reports* tab. Similarly, voluntary exposure was calculated as the sum of the number of clicks on the *Teaching Tips* tab and the number of clicks on the *View My Reports* tab during non-prompted sessions (Figure 3). Descriptive statistics were used to summarize the total and voluntary intervention exposures. The median of total exposure was calculated and used to indicate sufficient exposure to the intervention and defines intervention use.

Factors associated with using the intervention, both overall and voluntarily, were explored with Fisher's exact test. Factors of interest included: age (≥50 years, <50 years), work status, frequent computer user, gender, married/partnered, education (>high school, ≤high school), and service (medical oncology, radiation oncology, and stem cell transplant). Stage of disease was not considered for two reasons: (1) the highly associated relationship of stage and working status, a phenomenon documented in our prior work [16], and (2) the study sample contained participants with hematologic cancers in which the solid tumor staging system was inappropriate. The propensity score [17-19] was used to match a subset of the control group to the exposed intervention group and was defined as the probability of using the intervention given baseline participant characteristics. The following factors, shown in prior work [5] to contribute to outcomes, were used to compute the propensity score: baseline SDS-15 score categories (15-19, 20-23, 24-28, >28), service, gender, frequent computer use, married/coupled, education, minority status, age category, and working status.

The outcome, change in continuous SDS-15 scores from pre-treatment to end of study, was compared by intervention use and selected by propensity score matching using an ANCOVA approach.

Propensity score matching was performed with the complete data method in which no missing information could exist in the covariates used to compute the propensity score. A sensitivity analysis was conducted with the goal to balance the missingness within the two most common missing factors, work status and minority status. The propensity score matched sample was obtained using the R package “MatchIt” [20] and nearest neighbor matching. All analyses were performed in SAS version 9.3 and R version 2.15.2. All tests were two-sided and considered significant at the .05 level.

Figure 3. Components of the ESRA-C intervention with calculation of total exposure.



Results

Intervention Group Analysis

Total and voluntary exposure to the intervention components can only be calculated within the intervention group. The median total exposure was 2 (range 0-29). Intervention use was defined as at least two exposures. Of the 374 intervention participants, 233 (62.3%) used the intervention. A software error precluded exposure to pushed teaching tips for any intervention participants with an appointment date from June 30, 2010 to May 12, 2011. Of the 141 not receiving pushed teaching tips in the intervention, this error precluded three possible exposures in 48 (34.0%), two exposures in 21 (14.9%), and 1 exposure in 15 (10.6%) intervention group participants. There were 55 participants on the intervention arm that required a clinic/assisted point of access; their median total exposure was 2 (range 0-13) and 16 (29%) viewed the intervention voluntarily in the clinic. There were 319 participants on the intervention arm that indicated home/independent access; the median total exposure for these remote users was 3 (range 0-29) and 111 (34.8%) viewed the intervention voluntarily.

There were no statistically significant differences in the proportion of participants using versus not using the intervention based on characteristics, with the exception of clinical service (*P*=.02, Table 1). A total of 70.4% (88/125) of radiation oncology participants, followed by 60.2% (127/211) of medical oncology participants, used the intervention, whereas only 47% (18/38) of transplant participants used the intervention. The median voluntary exposure to the intervention was 0 (range 0-16). Voluntary use was defined as at least one voluntary exposure. Of the 374 participants randomized to the intervention group, 127 (34.0%) voluntarily used the intervention. There were marginally significant differences in the proportion of participants voluntarily using the intervention by work status (*P*=.06) and education (*P*=.05). Participants that used the intervention were working and had more than a high school education. Additionally, there was a significant difference in the proportion of participants voluntarily using the intervention by service (*P*=.001). More radiation oncology participants 58/125 (46.4%) voluntarily used the intervention compared to medical oncology 61/211 (28.9%) and transplant 8/38 (21%) participants (Table 1).

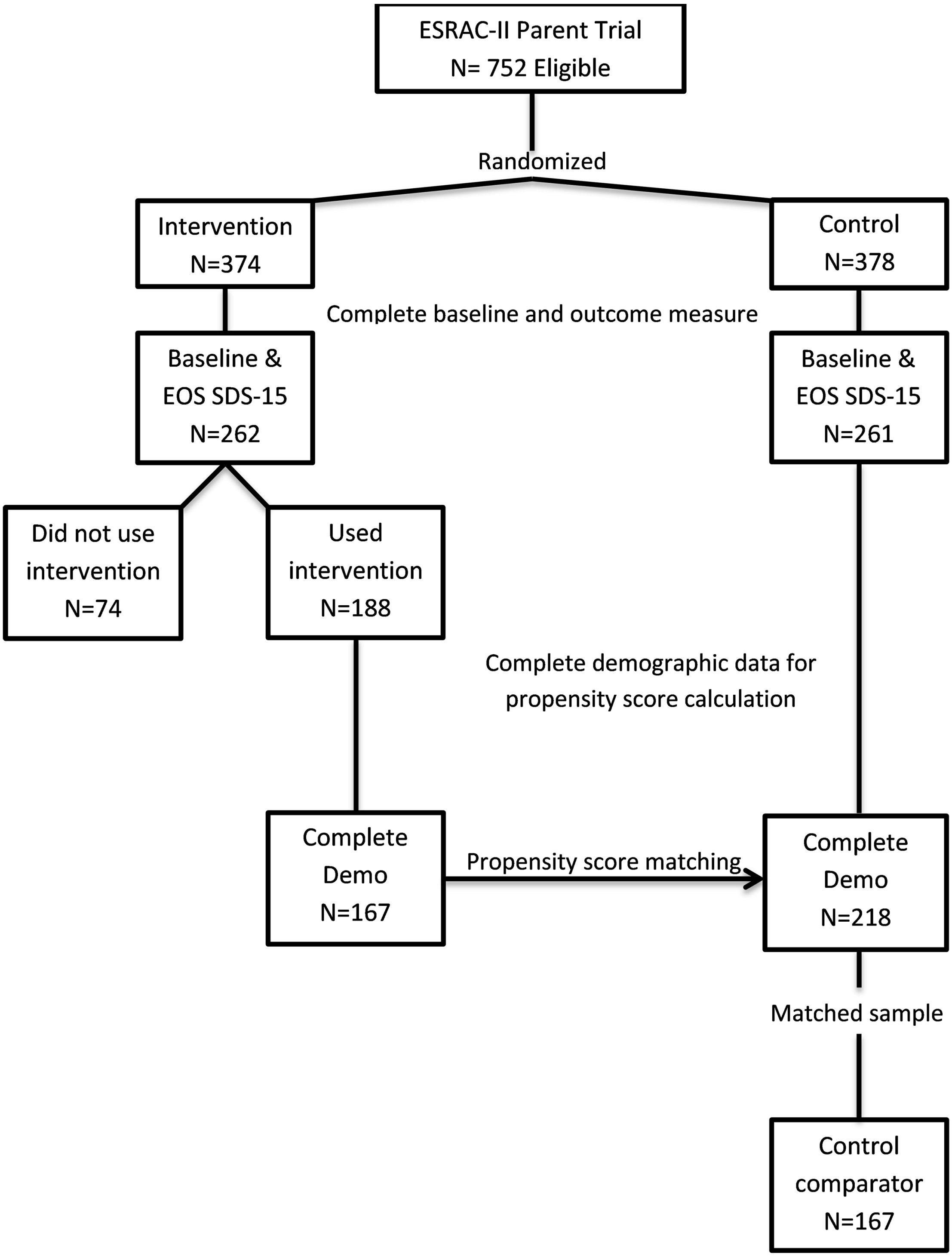
Table 1. Number and frequency of total exposures and voluntary exposures by selected participant characteristic.

Characteristic	Overall n	Total exposure (at least 2) n (%)	*P* value	Voluntary exposure (at least 1) n (%)	*P* value
Total N	374	233 (62.3)	-	127 (34.0)	-
Age			1.00		.82
≥50 years	248	154 (62.1)		83 (33.5)	
<50 years	126	79 (62.7)		44 (34.9)	
Work status			.64		.06
Working	222	141 (63.5)		84 (37.8)	
Not working	123	75 (61.0)		34 (27.6)	
Frequent computer use			.24		.07
No	59	33 (55.9)		14 (23.7)	
Yes	304	195 (64.1)		111 (36.5)	
Gender			.13		.33
Male	185	108 (58.4)		58 (31.3)	
Female	189	125 (66.1)		69 (36.5)	
Minority			.86		1.00
No	304	190 (62.5)		104 (34.2)	
Yes	36	22 (61.1)		12 (33.3)	
Married/Partnered			.79		.14
No	78	48 (61.5)		21 (26.9)	
Yes	293	185 (63.1)		106 (36.2)	
Education			.10		.05
≤High school	71	38 (53.5)		17 (23.9)	
>High school	301	195 (64.8)		110 (36.5)	
Service			.02		.001
Medical oncology	211	127 (60.1)		61 (28.9)	
Radiation oncology	125	88 (70.4)		58 (46.4)	
Hematopoietic stem cell transplant	38	18 (47.4)		8 (21.0)	

Propensity Score Analysis

Figure 4 outlines the sample selection from the parent trial for the propensity score analysis. Of the 262 participants randomized to the intervention group with a baseline and end-of-study SDS-15 score, 188 (71.8%) used the intervention. Complete demographic data were available for 167 (88.8%) of the 188 who used the intervention and 218 (83.5%) of the 261 control participants with baseline and end-of-study SDS-15 score. Using the propensity score and nearest neighbor matching, 167 control participants were selected from the possible 218 as the matched control group. Covariates were confirmed to be balanced (data not shown). Participants who used the intervention had lower symptom distress; mean change in the SDS-15 score was 1.07 (SD 6.55) in the matched control group (higher distress) and −0.57 (SD 5.68) in the intervention group (lower distress). In the ANCOVA analysis, SDS-15 score was reduced by an estimated 1.53 points (95% CI 0.32-2.75; *P*=.01) in the intervention group compared to the matched control group. The sensitivity analysis that balanced the missingness within the work status and minority status factors produced similar results as the complete data analysis (data not shown).

Figure 4. Sample selection for the propensity score analysis. Note: EOS=end of study; SDS=Symptom Distress Scale; Demo=demographics.



Discussion

Principal Findings

The results of this propensity analysis suggested that reduction of cancer symptom distress, the primary outcome of our randomized trial, was associated with use of the ESRA-C intervention components. More than half of participants in the intervention group were exposed to pushed *Teaching Tips*, accessed non-pushed tips, and viewed reports of SxQOL outcomes graphed over time. About a third voluntarily accessed the intervention in between clinic visits. We discovered that use of the intervention significantly reduced the estimated symptom distress score when compared to participants who did not use the intervention. The magnitude of the estimate (1.53) was larger than in the primary outcome analysis when we compared study groups (1.21) [5], indicating that actual use promotes the impact of the intervention. While this may seem intuitive, actual use of psycho-educational or self-administered interventions is not always known to investigators without objective monitoring capability. Our findings are consistent with the Quality Health Outcomes Model [10], illustrating the influence of patients' characteristics.

Comparisons With Prior Work

Use of, and adherence to, Web-based health care interventions have primarily been evaluated in health promotion and chronic disease self-management settings [21]. Few studies have associated changes in physical and/or psychosocial symptom distress with use of an eHealth intervention by patients with cancer and none for patients undergoing active cancer therapies or for those in the United States. Borosund et al [22] analyzed usage of a Web-based, symptom distress self-management system by prostate and breast cancer survivors in Norway over one year post-enrollment. Similar to our analysis, the Norwegian group defined “use” as at least two intervention sessions, but did not analyze symptom outcomes based on use. Van der Berg and colleagues [7] analyzed usage statistics of a Web-based self-management intervention for breast cancer survivors in the Netherlands. Participants were prompted by email to access the intervention. The survivors did not monitor or report symptoms, but were encouraged to read and/or view new educational material provided every 4 weeks over 16 weeks. Active usage was defined by Ven den Berg's group as a log-in to each of the four modules, and was observed in 44% of the 70 women in the trial. Our unprompted voluntary use percentage was lower at 34% of 374 intervention group participants. The explanation for a lower voluntary use percentage may be related to no systematic prompting for voluntary use or the fact that, unlike the group of Dutch survivors, our participants already were receiving the intervention at three time points prior to clinic visits throughout active cancer therapy.

Not surprisingly, ESRA-C was accessed remotely and voluntarily more frequently by those with higher education, who may have been likely to use personal computers or tablets on a regular basis. This is consistent with the finding, while of borderline statistical significance, that working individuals also used ESRA-C more often in between clinic visits. Working when beginning cancer therapy has been shown as a significant variable in two of our earlier analyses, predicting a lower rate of emergency department and unplanned admissions [23], yet predicting depression in participants receiving stem cell transplant [24]. How the fact that a patient is working full- or part-time influences outcomes is not well understood. In this case however, participants who were working when about to start cancer therapy may have had the type of job that enabled easy access to the Internet.

Participants who enrolled in the trial as they were about to undergo radiation treatments also accessed ESRA-C voluntarily significantly more often than those enrolled when beginning medical cancer therapies or stem cell transplant. We are not aware of differences in usual care symptom support between modality services at the cancer centers; yet, if differences existed, patients may have turned to ESRA-C more often in radiation. Alternatively, these participants were reminded of ESRA-C almost every day of the week as they entered the radiation setting where each had consented to the trial.

Limitations

Our findings are limited by the fact that about a third of the intervention group participants never received pushed teaching tips in the assessments prior to on-study clinic visits. Thus, the effect of the intervention may have been different if all had the opportunity to see the tips. Our participant sample was less diverse with regard to race and ethnicity than the rates of cancer diagnoses in minority groups in the United States [25] and all were patients at comprehensive cancer centers, precluding generalization of our findings beyond these parameters.

Implications for Future Research

First, propensity analyses could be replicated in other eHealth trials as a method to investigate the relationship of usage to health outcomes. Although our participants’ raw exposure to the intervention was not high in an absolute sense, we were able to study the association of symptom distress with usage rather than report raw usage. Second, investigators could evaluate whether usage was related to various characteristics, and whether they are characteristics of the intervention, of the user, or of the condition addressed by the intervention. We provided some rationale for the mechanism that triggered usage by certain demographic groups, but this could have been a combination of aspects of the intervention itself in addition to participant demographics. Although some investigators found that educated, older, employed women were the most active users of Web-based, chronic disease [26] and health promotion [27] interventions, other studies have revealed conflicting results with regard to demographic variables [28-29].

Implications for Clinical Practice

Our findings have implications for the many patients treated at institutions that have deployed a patient portal as a component of an electronic medical record system. There may be patients at risk for failed symptom and distress screening and/or failed symptom support delivery if such systems are available only to those Web-savvy, educated patients who regularly use email.

Finally, implications for both future research and practice using patient-centered, Web-based technologies include improved communication of study design and workflow between the design and technical implementation teams and more rigorous quality checks on intervention integrity. Communication of research goals may be facilitated via improved use of shared artifacts such as models of study data and workflows [30]. Software unit-testing and continuous integration goals should be oriented toward detailed research data deliverables [31,32]. Methods to improve quality checks include improved training of software quality assurance staff and making descriptive study data available to the investigators early in the data collection period for interim review.

Conclusions

The intended effects of a Web-based, self-care education, monitoring, and communication coaching intervention on cancer symptom distress were modified by intervention use frequency. The voluntary, remote use of ESRA-C was most frequent in working participants with higher education levels and those receiving radiation therapy.

Acknowledgments

All authors were funded by a grant (R01 NR008726**)** from the National Institute of Nursing Research (NINR), National Institutes of Health, for contributions to the study. No author was paid specifically for manuscript preparation. The NINR did not make any publication decisions. The authors are grateful to the patients, family members, and clinicians who gave time and effort to participate in this trial. The work of many talented research staff and students in collecting and coding data is greatly appreciated.

Conflicts of Interest

None declared.

References

1. Dwamena F, Holmes-Rovner M, Gaulden CM, Jorgenson S, Sadigh G, Sikorskii A, et al. Interventions for providers to promote a patient-centred approach in clinical consultations. Cochrane Database Syst Rev 2012;12:CD003267. [doi: 10.1002/14651858.CD003267.pub2] [Medline: 23235595]
2. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
3. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
4. Berry DL, Blumenstein BA, Halpenny B, Wolpin S, Fann JR, Austin-Seymour M, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol 2011 Mar 10;29(8):1029-1035 [FREE Full text] [doi: 10.1200/JCO.2010.30.3909] [Medline: 21282548]
5. Berry DL, Hong F, Halpenny B, Partridge AH, Fann JR, Wolpin S, et al. Electronic self-report assessment for cancer and self-care support: results of a multicenter randomized trial. J Clin Oncol 2014 Jan 20;32(3):199-205 [FREE Full text] [doi: 10.1200/JCO.2013.48.6662] [Medline: 24344222]
6. 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]
7. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MFM, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. J Med Internet Res 2013;15(8):e170 [FREE Full text] [doi: 10.2196/jmir.2566] [Medline: 23958584]
8. Wolpin SE, Halpenny B, Whitman G, McReynolds J, Stewart M, Lober WB, et al. Development and usability testing of a web-based cancer symptom and quality-of-life support intervention. Health Informatics J 2015 Mar;21(1):10-23. [doi: 10.1177/1460458213495744] [Medline: 24406906]
9. Berry DL, Hong F, Halpenny B, Partridge A, Fox E, Fann JR, et al. The electronic self report assessment and intervention for cancer: promoting patient verbal reporting of symptom and quality of life issues in a randomized controlled trial. BMC Cancer 2014;14:513 [FREE Full text] [doi: 10.1186/1471-2407-14-513] [Medline: 25014995]
10. Mitchell PH, Ferketich S, Jennings BM. Quality health outcomes model. American Academy of Nursing Expert Panel on Quality Health Care. Image J Nurs Sch 1998;30(1):43-46. [Medline: 9549940]
11. Karras BT, Wolpin S, Lober WB, Bush N, Fann JR, Berry DL. Electronic Self-report Assessment--Cancer (ESRA-C): Working towards an integrated survey system. Stud Health Technol Inform 2006;122:514-518. [Medline: 17102311]
12. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, de Haes J C. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [Medline: 8433390]
13. Postma TJ, Aaronson NK, Heimans JJ, Muller MJ, Hildebrand JG, Delattre JY, EORTC Quality of Life Group. The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. Eur J Cancer 2005 May;41(8):1135-1139. [doi: 10.1016/j.ejca.2005.02.012] [Medline: 15911236]
14. 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]
15. Ryan JL, Bole C, Hickok JT, Figueroa-Moseley C, Colman L, Khanna RC, et al. Post-treatment skin reactions reported by cancer patients differ by race, not by treatment or expectations. Br J Cancer 2007 Jul 2;97(1):14-21 [FREE Full text] [doi: 10.1038/sj.bjc.6603842] [Medline: 17565347]
16. Siefert M, Blonquist T, Berry DL, Hong F. Symptom-related emergency department visits and hospital admissions during ambulatory cancer treatment. J Comm Supp Oncol 2015:In-Press (forthcoming)(forthcoming).

17. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
18. Rosenbaum P, Rubin D. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association 1984 Sep;79(387):516-524. [doi: 10.2307/2288398]
19. Jo B, Stuart EA, Mackinnon DP, Vinokur AD. The Use of Propensity Scores in Mediation Analysis. Multivariate Behav Res 2011 May;46(3):425-452 [FREE Full text] [doi: 10.1080/00273171.2011.576624] [Medline: 22399826]
20. Ho D, Imai K, King G, Stuart E. MatchIT: Nonparametric Preprocessing for Parametric Causal Inference. J Stat Software 2011;42(8):1-28 [FREE Full text]
21. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. 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. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
23. Berry DL, Hong F, Blonquist T, Halpenny B, Siefert M, Partridge A. Self report assessment and support for cancer symptoms: Impact on hospital admissions and emergency department visits. In: J Clin Oncol. 2013 Presented at: American Society of Clinical Oncology Annual Meeting; 31 May-4 June, 2013; Chicago p. e20552.
24. Artherholt SB, Hong F, Berry DL, Fann JR. Risk factors for depression in patients undergoing hematopoietic cell transplantation. Biol Blood Marrow Transplant 2014 Jul;20(7):946-950. [doi: 10.1016/j.bbmt.2014.03.010] [Medline: 24650679]
25. American Cancer Society. Cancer Facts & Figures 2014. URL: http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf [accessed 2014-12-29] [WebCite Cache ID 6VBkhx9iT]
26. Wangberg SC, Bergmo TS, Johnsen JAK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]
27. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
28. Couper MP, Alexander GL, Zhang N, Little RJA, 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]
29. 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]
30. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. Best practices for scientific computing. PLoS Biol 2014 Jan;12(1):e1001745 [FREE Full text] [doi: 10.1371/journal.pbio.1001745] [Medline: 24415924]
31. Remmel H, Paech B, Bastian P, Engwer C. System Testing a Scientific Framework Using a Regression-Test Environment. Comput. Sci. Eng 2012 Mar;14(2):38-45. [doi: 10.1109/MCSE.2011.115]
32. Ackroyd KS, Kinder SH, Mant GR, Miller MC, Ramsdale CA, Stephenson PC. Scientific Software Development at a Research Facility. IEEE Softw 2008 Jul;25(4):44-51. [doi: 10.1109/MS.2008.93]

Abbreviations

ANCOVA: analysis of covariance
ESRA-C: Electronic Self Report Assessment for Cancer
SDS-15: Symptom Distress Scale
SxQOL: symptoms and quality of life

Edited by G Eysenbach; submitted 30.12.14; peer-reviewed by E Børøsund, G Juarez; comments to author 14.03.15; revised version received 23.03.15; accepted 07.04.15; published 03.06.15.

Please cite as:
Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J
Exposure to a Patient-Centered, Web-Based Intervention for Managing Cancer Symptom and Quality of Life Issues: Impact on Symptom Distress
J Med Internet Res 2015;17(6):e136
URL: http://www.jmir.org/2015/6/e136/
doi:10.2196/jmir.4190
PMID:26041682

Review

Computerized Cognitive Behavior Therapy for Anxiety and Depression in Rural Areas: A Systematic Review

Kari Dee Vallury[1], BA, MHlth&IntDev; Martin Jones[1], BSc, RN, PhD; Chloe Oosterbroek[1], BHSc (Hons)

Department of Rural Health (DRH), Division of Health Sciences, University of South Australia, Whyalla Norrie, Australia

Corresponding Author:
Kari Dee Vallury, BA, MHlth&IntDev
Department of Rural Health (DRH)
Division of Health Sciences
University of South Australia
Department of Rural Health
111 Nicolson Avenue
Whyalla Norrie, 5608
Australia
Phone: 61 0433773061
Fax: 61 8 8647 6164
Email: kari.vallury@unisa.edu.au

Abstract

Background: People living in rural and remote communities have greater difficulty accessing mental health services and evidence-based therapies, such as cognitive behavior therapy (CBT), than their urban counterparts. Computerized CBT (CCBT) can be used to effectively treat depression and anxiety and may be particularly useful in rural settings where there are a lack of suitably trained practitioners.

Objective: To systematically review the global evidence regarding the clinical effectiveness and acceptability of CCBT interventions for anxiety and/or depression for people living in rural and remote locations.

Methods: We searched seven online databases: Medline, Embase Classic and Embase, PsycINFO, CINAHL, Web of Science, Scopus, and the Cochrane Library. We also hand searched reference lists, Internet search engines, and trial protocols. Two stages of selection were undertaken. In the first, the three authors screened citations. Studies were retained if they reported the efficacy, effectiveness or acceptability of CCBT for depression and/or anxiety disorders, were peer reviewed, and written in English. The qualitative data analysis software, NVivo 10, was then used to run automated text searches for the word "rural," its synonyms, and stemmed words. All studies identified were read in full and were included in the study if they measured or meaningfully discussed the efficacy or acceptability of CCBT among rural participants.

Results: A total of 2594 studies were identified, of which 11 met the selection criteria and were included in the review. The studies that disaggregated efficacy data by location of participant reported that CCBT was equally effective for rural and urban participants. Rural location was found to both positively and negatively predict adherence across studies. CCBT may be more acceptable among rural than urban participants—studies to date showed that rural participants were less likely to want more face-to-face contact with a practitioner and found that computerized delivery addressed confidentiality concerns.

Conclusions: CCBT can be effective for addressing depression and anxiety and is acceptable among rural participants. Further work is required to confirm these results across a wider range of countries, and to determine the most feasible model of CCBT delivery, in partnership with people who live and work in rural and remote communities.

(J Med Internet Res 2015;17(6):e139) doi:10.2196/jmir.4145

KEYWORDS

eHealth; mHealth; depression; anxiety; cognitive therapy; rural health; mental health

Introduction

Background

In any one year, 10% and 14% of Australian adults experience affective problems and anxiety disorders, respectively [1,2]. These rates of mental illness are in line with global trends—the average 12-month depression prevalence rate of 18 high- and low-income countries is 5.4% [3]. For anxiety disorders, 12-month prevalence rates range from 7% to 15.5% in Euro/Anglo cultures [4].

In Australia, over 30% of the population lives outside major cities, with 11% living in outer regional, remote, and very remote areas [5]. The reported prevalence of mental health disorders is similar across rural and urban areas [6,7]. However, there are certain population groups in rural and remote areas that experience higher levels of mental disorders—men in outer regional and remote areas are significantly more likely to experience higher levels of psychological distress than men in major cities [8], and women in nonmetropolitan areas aged 30 to 44 years also face slightly higher rates of mental health disorders than their urban counterparts [6].

Globally, the prevalence of anxiety disorders is significantly higher among rural versus urban populations [4]. Furthermore, suicide rates are markedly higher in rural areas compared with major cities, as has been documented in Australia, the United States, the United Kingdom, and New Zealand [9]. In Australia, suicide rates increase with level of remoteness and are largely driven by increased suicides among young men [6,10,11].

Treatment, Services, and Access

There is a strong evidence base and subsequently established guidelines for the effective drug and nondrug treatment of depression and anxiety. For example, supportive clinical care, cognitive behavior therapy (CBT), antidepressants, and interpersonal therapy (IPT) are all recommended treatment options for different forms of depression [12]. For anxiety, evidence-based interventions include self-help strategies, group and individual psychoeducational interventions including CBT, and pharmacological treatments for complex conditions [13].

However, despite an understanding of what works in treating depression and anxiety, many people do not receive adequate care. Less than one-quarter of Australians access psychosocial services, even when they are available [14]. Of the people with depression or anxiety who do seek treatment, under half are offered an appropriate treatment option [15].

Accessing health services can be particularly difficult for people living outside metropolitan areas and away from service hubs. Smaller proportions of rural versus urban populations seek or receive professional help for a mental health problem [6,16].

There are numerous factors that prevent people from accessing mental health services. Availability of services and trained mental health professionals are major barriers to access in rural and remote Australia [14,17]. In comparison with the 115 psychologists for every 100,000 persons in major cities, the rate in rural areas declines from 66.5 in inner regional to 29 in very remote areas [18]. Other barriers may include cultural norms around stoicism and not wanting to show vulnerability, denial, poor mental health literacy, stigma around mental illness and mental health service use, and the financial and personal demands required of treatment [6,17,19]. Studies from a number of English-speaking countries have shown that mental health stigma is particularly prevalent in rural areas, is greater among men, and impacts willingness to seek help [16].

The consequences of untreated depression and anxiety are wide ranging and often debilitating. These conditions can lead to reduced quality of life and productivity, increased likelihood of developing substance abuse disorders [20], nonadherence to care and treatment [21,22], increased risk of physical health problems such as cardiovascular disease [23], and increased suicide risk [24]. Stack [24] reported that 87% of suicides involve at least one mental disorder, and that people experiencing major depression are as much as 20 times more likely to commit suicide than people without depression. Accessing appropriate treatment for depression can reduce suicide risk by up to 50%, particularly among young men [12].

Computerized Cognitive Behavior Therapy

Computerized cognitive behavior therapy (CCBT) is an effective treatment option for people with anxiety and/or depression, both as a standalone treatment and as a component of a stepped-care treatment plan. Numerous reviews and meta-analyses have found that CCBT achieves moderate to large effect sizes for depression and anxiety [25-31], similar to those found for therapist-delivered CBT [26,32]. That said, the comparative effectiveness of CCBT and therapist-delivered CBT is somewhat contested. For example, a Cochrane review by Mayo-Wilson and Montgomery [28] found therapist-delivered CBT to be more effective than computerized delivery.

The delivery of evidence-based psychotherapy via personal computers, mobile phones, and tablets provides an opportunity to increase its uptake in rural and remote communities. It may help minimize the impact of inadequate numbers and unequal distribution of appropriately trained therapists, and subsequent long wait times, as well as the financial demands of treatment, travel times [19,33], and stigma associated with accessing mental health services. Computerized CBT as a mechanism to improve evidence-based service provision in rural and remote areas may increase the uptake of evidence-based interventions. However, few studies have explicitly explored the effectiveness and acceptability of CCBT in rural communities.

This review synthesizes the global evidence regarding the clinical effectiveness and acceptability of CCBT interventions for preventing or treating anxiety and depression in people who live in rural and remote areas. It provides recommendations for future research and practice with relevance to rural communities in English-speaking countries around the world.

Methods

Overview

Literature was systematically reviewed in line with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines [34].

Search Strategy

The search strategy was developed by authors KV and CO with input from an academic research librarian. The final strategy included variations of the following terms: anxiety, depression, eHealth, computerized, online, application, health, cognitive behavior therapy, and computerized cognitive behavior therapy. Other terms such as "e-therapy," "Internet-delivered," and "phone-based" did not identify additional citations and were excluded from the strategy. A full copy of the search strategy is provided in Multimedia Appendix 1.

The search was conducted in a number of health and science databases: Medline (1946-2014), Embase Classic and Embase (1947-2014), PsycINFO (1806-2014), CINAHL (1981-2014), Web of Science (1950-2014), Scopus (1960-2014), and the Cochrane Library (all reviews and trials, May 2014). Additional articles were identified through pearling (ie, hand searching) selected reference lists and trial protocols.

Selection Criteria

Two screening phases were conducted. In the initial phase, studies were included if they (1) reported the clinical efficacy, effectiveness, acceptability and/or feasibility of CBT delivered via the Internet, through the use of a computer or other mobile electronic device; (2) had a focus on the prevention or treatment of generalized or social anxiety disorders, multiple forms of anxiety, and/or depression; (3) included participants from any population group in any location; (4) were conducted in any year up until the search date of May 22, 2014; and (5) were written in English. Primary and secondary studies with quantitative and qualitative designs, as well as systematic reviews and meta-analyses, were included.

Studies that focused solely on individual phobias, posttraumatic stress disorder, or postnatal depression were not included in this review, although studies that addressed any combination of anxiety disorders were included. Generalized anxiety disorder (GAD) and social anxiety disorder are two of the most prevalent anxiety disorders, with 12-month prevalence rates in Australia of 2.7% and 4.7%, respectively [15]. In the United Kingdom, approximately 4.4% of the population are experiencing generalized anxiety at any one time, in comparison to panic disorder and obsessive compulsive disorder (OCD) at 1.1% [35]. Furthermore, these forms of anxiety are more likely to be treatable with more generalizable forms of CBT. For these reasons, findings regarding the efficacy of CCBT for these disorders are likely to be more broadly applicable and were, therefore, included as the focus of this review.

In the second phase of screening, studies that measured efficacy or acceptability among rural participants or meaningfully discussed the application of CCBT in rural settings were retained for inclusion.

Study Selection

All three authors were involved in the initial stage of the study selection process. KV conducted an initial review of all citations by title and discarded any that were clearly irrelevant. KV and CO then reviewed the abstracts of all remaining citations (half each), discarding any that did not meet the inclusion criteria. In response to any uncertainty, the other reviewing author was consulted. If both authors were unsure or disagreed, the third author (MJ) was consulted to reach a final decision. Full texts were located for all citations that potentially matched the inclusion criteria. Each text was reviewed by KV and CO independently to decide on the final list of included articles, again with input from MJ when required.

NVivo 10 (QSR International, Cambridge, MA), a software package that supports qualitative data analysis, was used to support the second stage of screening. All studies that met the inclusion criteria at stage one were imported into NVivo. Automated text (word) searches were run to identify studies that included the word “rural,” its stemmed variations, and synonyms. The full texts of studies identified through this process were then assessed by KV to determine the final list of included studies. Where KV was undecided, MJ was consulted.

Data Extraction and Bias Assessment

A structured, but flexible, data extraction table was developed. Data were extracted for a range of outcomes measuring patient experience as well as clinical effect. These included study design, population and intervention characteristics, clinical efficacy, rates of uptake and adherence, qualitative measures of satisfaction, and perceived benefits and disadvantages.

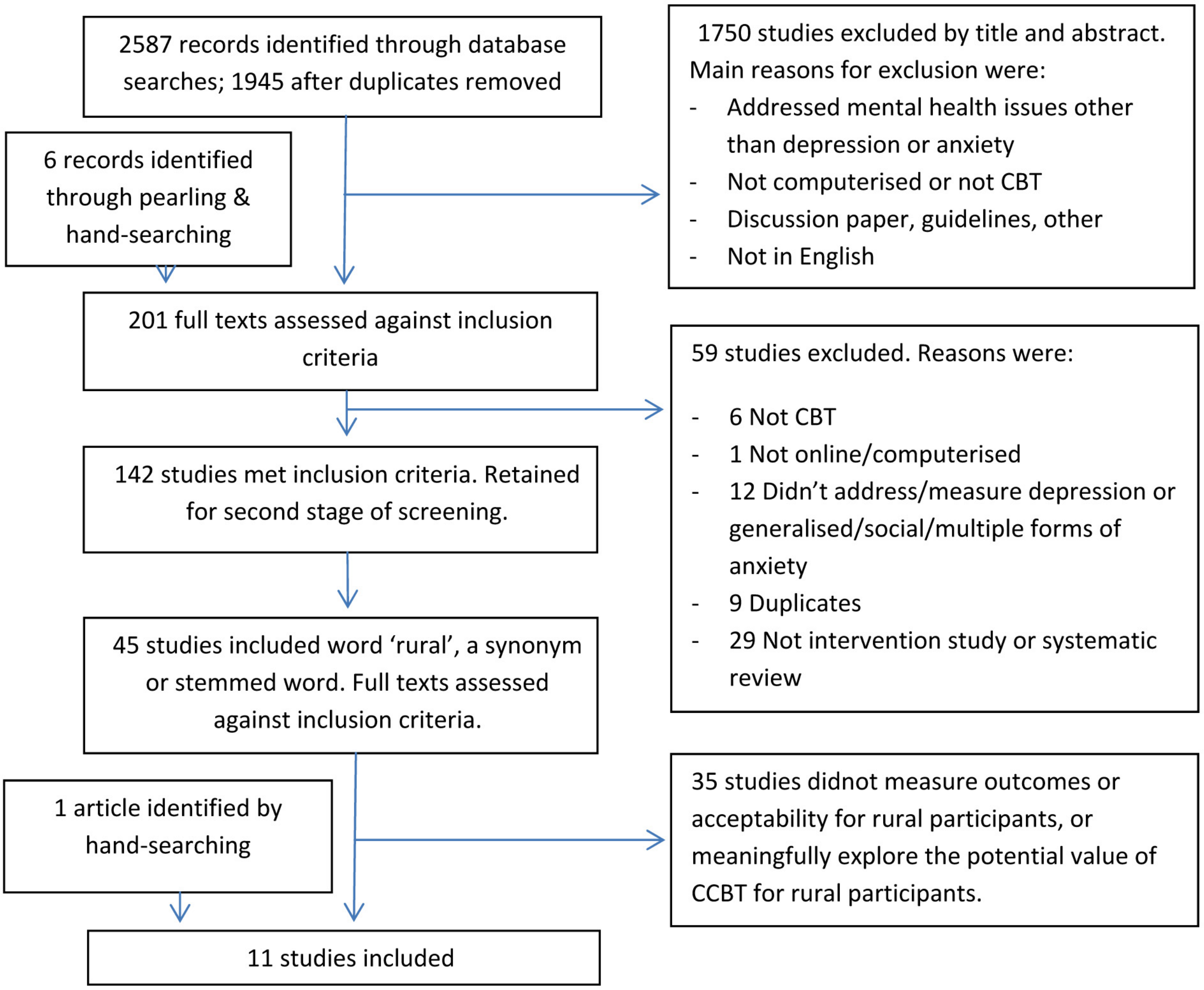
The Cochrane Collaboration’s tool for assessing risk of bias [36] was used to assess bias among randomized controlled trials (RCTs), assessed at study level. Relevant criteria from the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) handbook were used to assess bias in all other study designs [37]. Several authors were contacted with requests for further information regarding study methods to support the accurate completion of these assessments.

Results

Overview

The initial database search identified 2587 citations. Of these, 195 were selected for full-text review at the first stage of screening, along with six studies identified through pearling. A total of 142 studies met the inclusion criteria and were retained for the second stage of screening. The automated text search revealed that 45 of these studies included the word “rural,” a synonym, or stemmed word at least once. Of these, 10 met the inclusion criteria and were included in the review. One extra study was identified by hand searching at this stage, resulting in 11 studies being finally included in this review [38-48]. Figure 1 outlines this process and provides the reasons for exclusion at each stage.

Figure 1. Study selection process.



Study Characteristics

Among the 11 included studies were four papers reporting the findings of three RCTs, one systematic review, one qualitative study, and five studies which used quasi-experimental designs. Nine of the studies were conducted in Australia and two—regarding one trial—in Scotland. Characteristics and key findings of the 11 studies are reported in Table 1.

Eight papers regarding six different trials measured rural location as a predictor of outcomes, adherence, or acceptability [38-45]. The qualitative study explored the acceptability of a CCBT package among rural youth [46]. The final two studies discussed the potential application of their results to rural populations [47,48].

Across the nine studies reporting at least some rural participants were a total of 11,260 participants. Between 16% and 100% of study participants lived in rural areas. Four of the 11 studies explored the value of CCBT among children and/or young people, while five studies tested CCBT for adults. The systematic review included five studies with adult participants and four with children and/or young people.

Table 1. Characteristics of included studies.

Citation; program	Study design	Location	Participants: n, % rural; gender, %; age group, age in years, mean or range	Main findings
Calear et al 2013 [38]; MoodGYM, anxiety & depression	RCT[a]	Australia	1477, ~16; female, 56; adolescents, 12-17	Living in a rural location predicted greater adherence.
Neil et al 2009 [43]; MoodGYM, anxiety & depression	Quasi-experimental	Australia	8207, 19; female, 71; adolescents, 13-19	Living in a rural area predicted greater adherence.
Sethi 2013 [47]; MoodGYM, anxiety & depression	RCT	Australia	89, 0; female, 58; youth, 18-25	CCBT[b]may be a viable option for youth, but unsuitable for people with low literacy.
Griffiths & Christensen 2007 [48]; MoodGYM + Blue Pages	Systematic review	International	N/A[c,d]	CCBT may be inconsistent with rural residents' preferred mode of learning—should consider tailoring programs to rural users.
Cheek et al 2014 [46]; SPARX, depression	Qualitative study	Australia	16, 100; male, 75; adolescents, 13-18	New Zealand program acceptable for Australian participants.
Hayward et al 2007 [39]; FearFighter, anxiety & depression	Uncontrolled trial	Scotland	35, 100; female, 66; 16 years and over, 40.2	Participants had significant improvements on measures of depression and anxiety. Patients and GPs[e]were satisfied.
MacGregor et al 2009 [45]; FearFighter, anxiety & depression	Survey & qualitative	Scotland	35, 100; female, 66; 16 years and over, 40.2	Content was generally appropriate for rural dwellers (except for references to city centers, buses, and lifts).
Kay-Lambkin et al 2011; [40] CCBT for comorbid depression & substance use	RCT	Australia	274, 41; male, 57; 16 years and over, 40	Rurality did not affect treatment response (depression). Computerized therapy led to 2.5 times greater reduction in alcohol use than therapist delivered (*P*=.006).
Kay-Lambkin et al 2012 [41][f]	RCT	Australia	163, 33; N/A; 16 years and over[g]	No significant differences between rural and urban regarding preferred treatment method. No effect of rurality on retention or treatment response.
Mewton et al 2012 [42]; CRUfAD clinic, anxiety	Quasi-experimental	Australia	588, 43; female, 71; 16 years and over, 39.5	Those in a nonrural location were 1.8 times more likely to complete the six course components. Need to tailor courses for rural users.
Sunderland et al 2012 [44]; CRUfAD clinic, depression & anxiety	Quasi-experimental	Australia	663, ~45; female, 66; N/A, 43	Rurality did not influence effectiveness of CCBT for anxiety and depression.

[a] Randomized controlled trial (RCT).

[b] Computerized cognitive behavior therapy (CCBT).

[c] Not applicable (N/A).

[d] The review included 12 papers regarding nine studies. Of these nine studies, five were regarding adults, one regarding tertiary students, and three regarding children/secondary school students. Gender breakdown varied across studies.

[e] General practioners (GPs).

[f] Participants from this study were a subset of Kay-Lambkin et al 2011 [40].

[g] Mean age for this subsample was unavailable.

Efficacy

Three papers, regarding two trials, reported clinical efficacy data disaggregated by location [40,41,44]. All found no difference in the treatment response to CCBT for depression and anxiety between rural and urban participants. Another study, conducted entirely with rural participants, found that CCBT led to significant improvements in anxiety and depression [39].

One school-based study found that rurality predicted high adherence to CCBT among adolescents, and that higher adherence led to greater reductions in depression and anxiety [38]. However, of the four trials with children/young people, none disaggregated efficacy data by rural/urban location. Furthermore, while the review by Griffiths and Christensen included studies reporting the efficacy of CCBT for both children/young people and adults, none of their included studies disaggregated the data by location [48]. It is, therefore, not possible to identify patterns regarding the interaction of rural location, age, and efficacy from the studies included in this review.

Uptake and Referral

The included studies did not consistently report rates of uptake, in several cases due to retrospective study designs. The study in rural Scotland found that 24% of people referred to CCBT declined to undertake the treatment [39]. In Kay-Lambkin and colleagues' trial [40], less than 9% (54) of 617 participants assessed for eligibility refused to participate. However, the rate of uptake in this trial was still only 44% of the original participants assessed—40% were excluded as they did not meet inclusion criteria and a further 7% did not attend their first assessment.

The included systematic review [48] found that, of spontaneous users of the CCBT program MoodGYM worldwide, 20.5% were from rural and remote areas. In further support of its acceptability among rural mental health patients, Kay-Lambkin and colleagues reported that almost half of their sample self-referred to CCBT [40].

General practitioners (GPs) have an important role in connecting patients to CCBT in rural communities. Rural participants are more likely to have been referred to CCBT by their GPs than urban participants—23% versus 2% referred by a GP, respectively ($P<.001$) [40]. Hayward and colleagues' trial in rural Scotland also relied on GPs to connect patients to CCBT [39]. The study reported that CCBT was highly acceptable among GPs in regard to suitability for provision to rural and remote patients.

Adherence/Attrition

The included studies reported mixed findings in regard to adherence and attrition rates among rural versus urban participants, though they most commonly reported rural participants to be as likely, if not more likely, to adhere to CCBT treatment.

Two studies with adult participants compared rates of adherence by location. Kay-Lambkin and colleagues found that rurality did not affect retention [41]. In contrast, Mewton and colleagues found that rural participants were significantly less likely to complete CCBT, with urban participants almost twice as likely to complete the full course [42]. The two studies with adolescents that compared adherence outcomes by location both found that rural residence predicted significantly greater adherence to the MoodGYM program [38,43]. Lack of availability of alternative services, greater motivation of supervising staff members, or a preference for health self-management in rural participants are potential explanations for this [38,43,48]. Table 2 shows the efficacy and acceptability outcomes of the studies.

Table 2. Efficacy and acceptability outcomes.

Study	Uptake	Adherence	Other acceptability	Clinical effect
Calear et al 2013 [38]	N/A[a](school based)	Rural had greater adherence (*P*=.01).	N/A	Not disaggregated by location.
Cheek et al 2014 [46]	N/A	N/A	New Zealand program acceptable to rural Australian youth; design important.	N/A
Griffiths & Christensen 2007 [48]	20.5% spontaneous users worldwide rural/remote	N/A	Should consider tailoring content. May not be suitable for learning styles of rural participants.	Both programs examined led to improvements in mental health, knowledge, and attitudes to mental health.
Hayward et al 2007 [39]	89 referred; 13 unsuitable; 21 refused; 55 passwords issued (62%)	26 completed (47% of participants who received passwords)	97% satisfied with help received. GPs feel demos of program could increase referrals by GPs.	Significant improvement in depression and anxiety (*P*<.001).
Kay-Lambkin et al 2011 [40]	617 assessed; 244 unsuitable; 54 refused; 274 randomized (44%); 260 began	86 (33% of starters) received all sessions; 163 (63% of starters) completed 3-month follow-up.	N/A	No significant effect of rurality on effectiveness: depression (*P*=.70) or substance use. Therapist and CCBT equally effective for depression (*P*=.02).
Kay-Lambkin et al 2012 [41]	N/A: 3-month follow-up data	Rurality did not affect attendance or therapeutic alliance.	Rurality did not affect preference for therapist/ computerized delivery. Rural less likely to want more therapist contact—18% vs 48% urban.	Rurality did not influence treatment response.
MacGregor et al 2009 [45]	89 referred; 13 unsuitable; 21 refused; 55 passwords issued (62%)	N/A	Content acceptable to rural/remote participants. Minor changes may be beneficial.	N/A
Mewton et al 2012 [42]	N/A	55.1% completion; rural had poorer adherence (*P*<.05). Urban 1.8 times more likely to complete.	N/A	Significant reduction in anxiety and psychological distress; improved quality of life (WHODAS[b]) (all *P*<.001).
Neil et al 2009 [43]	N/A	Rural had greater adherence: whole sample (*P*=.01), school sample (*P*<.001).	N/A	N/A
Sethi 2013 [47]	103 assessed; 89 eligible and randomized (86%)	100% completed (assume none rural as not reported)	Unsuitable for people with low literacy.	N/A, as location of participants not reported.
Sunderland et al 2012 [44]	N/A: data from completers only	N/A	N/A	Rurality did not influence treatment response: depression (*P*=.83), anxiety (*P*=.77).

[a] Not applicable (N/A).

[b] World Health Organization Disability Assessment Schedule (WHODAS).

Other Measures of Acceptability

There is some evidence to suggest that, on completion, CCBT was considered to be more acceptable to rural than to urban adult participants. In one study, rural participants were more likely to report that CCBT had helped them with their depression and substance use—92% versus 75% of urban participants [41]. Furthermore, fewer rural CCBT participants reported wanting more face-to-face contact as compared with urban participants—18% versus 48%, respectively. Hayward and colleagues found that 97% of their rural (whole) sample was satisfied with the support provided through CCBT [39].

Studies with both young people and adults found that privacy when accessing mental health services was of great importance to rural participants. For example, Cheek et al [46] found that visibility and confidentiality when accessing services, as well as attitudes of health professionals, were barriers to young people accessing mental health services in a small rural town in Australia. They also found that the opportunity to complete CCBT in private was an appealing feature of the treatment. In another study, two-thirds of the rural adult participants missed therapist contact, and yet two-thirds also felt that the benefits of CCBT included increased autonomy and confidentiality [39].

Risk of Bias

A risk of bias assessment was conducted for all studies, with the exception of the systematic review. Overall, the risk of bias was moderate. This is consistent with a large review of the broader CCBT evidence by Grist and Cavanagh [26], which found an overall moderate risk of bias across 49 studies. It also established that risk of bias was unlikely to influence effect sizes in the included studies.

Across the RCTs included in this review, the risk of bias was rated as low for two studies [38,40], and moderate for two studies [41,47]. Baseline differences across treatment groups, and between completers and noncompleters of outcome measures, were the primary sources of potential bias. Of the quasi-experimental studies, the risk in one study was unclear due to insufficient information on several variables [44] and moderate in another [45], due to low numbers of participants completing the outcome measurements. The three other studies were rated as likely to be at low risk of bias [39,42,43].

The qualitative study by Cheek and colleagues [46] was rated as at moderate risk of bias. Due to its size and scope, replication of the study in varying locations would be valuable to further understand the generalizability of the findings.

Across the included studies, strict participant selection criteria in several may limit the generalizability of their findings. However, a number of the studies included groups of participants who are otherwise often excluded, such as youth and people with severe symptoms or comorbidities. We believe this goes some way toward balancing this limitation.

A number of studies in the broader CCBT literature have found evidence of a publication bias. Studies reporting negative findings are less likely to be published [26,49,50]. In this review, publication bias is also a real possibility, given that we identified and included only published data.

Discussion

Overview

Computerized CBT can be clinically effective for the prevention and treatment of anxiety and depression, and offers a valuable alternative to traditional face-to-face delivery. This may be particularly pertinent to the delivery of services in underresourced and otherwise underserved communities.

Efficacy and Acceptability

We located 11 studies that begin to identify the feasibility of CCBT in rural and remote communities. Notwithstanding diversity in study designs, participants, software packages, styles, and locations of delivery, the studies indicate that CCBT has equal effects for urban and rural participants. Furthermore, they support the effectiveness of CCBT in real-world rural clinical practice and community settings, with all included trials conducted in school, university, community, or clinical (ie, online mental health or GP clinic) sites.

The included studies indicate that rurality is unlikely to have a negative impact on uptake or adherence. Among the wider CCBT evidence base, low uptake has been identified as a key barrier to implementation, with an average of 12% of participants offered CCBT commencing treatment [26,51,52]. The rates of uptake among several studies included in this review were much higher—44% and 56% in studies by Kay-Lambkin et al [40] and Hayward et al [39], respectively. Importantly, these studies included patients with comorbidities and had minimal inclusion criteria, respectively.

Satisfaction and acceptability are generally high among people who undertake CCBT. Acceptability increases significantly once patients have received a demonstration or have undertaken the treatment [19,41,49,52]. Kay-Lambkin et al showed that treatment preference fulfilment—computerized versus therapist delivered—had a greater impact on adherence for rural versus urban participants [41]. Furthermore, changes in depression were significantly associated with treatment preference fulfilment across their whole sample. Fostering understanding and promoting the credibility of CCBT prior to implementation in rural areas may greatly improve its acceptability, uptake, and reach.

This review provides evidence to support a number of the assumed benefits of CCBT for rural populations, including its ability to overcome barriers that have traditionally limited access to mental health services. For example, studies in both Australia and Scotland found that the ability to complete CCBT privately helped minimize confidentiality concerns and stigma regarding accessing mental health services [39,46]. Furthermore, the delivery of CCBT does not rely on a preponderance of trained therapists, and even guided versions require considerably less therapist time than face-to-face CBT [19,53,54]. Staff who are not trained mental health practitioners are able to provide the guidance required by some CCBT programs without significantly reduced clinical effect [32,40,55,56].

Opportunities and Challenges for Rural Implementation

Computerized CBT has the potential to complement the inadequate numbers of qualified mental health professionals in rural communities. Implementing CCBT within the existing service landscape as a "first response" treatment may be appropriate. Within such a model, all patients would first be offered CCBT, with therapist time reserved for those who do not respond well, or require further or more intensive therapies [18,34]. This could alleviate pressure on trained therapists and ensure their services are available to those most in need. A study in the United Kingdom [57] found that 19% of participants required referral to a therapist on completion of CCBT. These patients then required, on average, only 3.5—compared with the usual 15—sessions of CBT with a therapist. Combining therapist and computer-delivered CBT has also been shown to be a particularly effective method for treating anxiety and depression among adolescents and similarly reduces the therapist time required to treat each patient [47,58].

More research is needed into the feasibility of delivering CCBT across varying geographical and demographic sites and groups. Understanding local barriers to uptake and adherence, and solutions to these, will be crucial, as they are likely to vary between towns, regions, and countries.

The extent to which content needs to be tailored to rural users' location and age also requires further study. Computerized CBT packages may benefit from being tailored to more accurately reflect the physical nature of the rural context [42], or in line with different learning styles or education levels [47,48]. However, Cheek and colleagues found that a program developed in New Zealand was acceptable to youth in a rural Australian town [46], suggesting the possibility of translation of CCBT programs across locations without significant alterations.

Limitations

Despite the use of a systematic methodology, it is possible that some studies have been missed. While some hand searching and pearling was conducted, not every reference list of identified reviews and studies was searched. Furthermore, no unpublished findings were included in the review and it is, therefore, at risk of publication bias.

Inclusion criteria were limited to studies that addressed generalized and social anxiety disorders, depression, or several types of anxiety disorders concurrently. CCBT for individual phobias, posttraumatic stress disorder, and postnatal depression were excluded, although such studies could hold valuable insights to inform the wider application of CCBT. Furthermore, CCBT has been used for 25 different clinical disorders [27], not only mental health conditions [59-62]. To ensure that its full potential is realized, a similar review into the efficacy and acceptability of CCBT for other conditions in rural and remote areas would be valuable.

The studies identified were predominantly Australian, with two from Scotland. The conclusions and recommendations drawn are, therefore, particularly relevant to the Australian context. The authors believe, however, that given the similar challenges faced across the world in providing evidence-based mental health to rural communities, these findings can be expected to be relevant to English-speaking countries more broadly.

Conclusions

There is a strong focus on workforce development in rural health research and provision. Yet rural and remote communities, globally, continue to face significant challenges in attracting specialist health professionals, highlighting the need for alternative models of delivering evidence-based care. The studies that we reviewed provide initial evidence that CCBT could be a valuable tool for increasing the accessibility of psychological therapies in rural and remote communities. It is likely to be effective and acceptable among rural participants and practitioners.

In the future, practitioners need to be supported to understand and refer clients with particular needs to appropriate evidence-based CCBT programs. Workforce development programs at university level and beyond need to prepare the workforce to appreciate the potential of CCBT. Demonstration of CCBT packages aimed at both users and practitioners may be an important action to build acceptability and trust in rural communities and to ensure that the therapy is accessed by those who need it.

Future research is required to clarify the findings of this review, given the relatively small number of studies identified and the small number of countries represented. Models of CCBT delivery that work within existing health systems and fill service gaps need to be developed and tested in varied rural and remote environments and countries.

Authors' Contributions

MJ conceptualized the study. KV developed the search strategy with assistance from MJ and Academic Research Librarian, Carole Gibbs. KV conducted the database searches. MJ, KV, and CO conducted the first round of citation screening and extracted data from included studies. KV conducted the second phase of screening with input from MJ. KV wrote the majority of the paper with significant input and revisions by MJ and CO.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 4KB - jmir_v17i6e139_app1.pdf]

References

1. Australian Bureau of Statistics. 2012. 4338.0 - Profiles of health, Australia, 2011-13 URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4338.0main+features12011-13 [accessed 2014-09-15] [WebCite Cache ID 6YpIabSIL]
2. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]
3. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013;34:119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
4. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med 2013 May;43(5):897-910. [doi: 10.1017/S003329171200147X] [Medline: 22781489]
5. Baxter J, Hayes AM, Gray M. Australian Institute of Family Studies, Commonwealth of Australia. 2011. Families in regional, rural and remote Australia URL: https://aifs.gov.au/publications/families-regional-rural-and-remote-australia [accessed 2014-08-28] [WebCite Cache ID 6YpJ6F1aH]

6. Caldwell TM, Jorm AF, Dear KB. Suicide and mental health in rural, remote and metropolitan areas in Australia. Med J Aust 2004 Oct 4;181(7 Suppl):S10-S14. [Medline: 15462636]
7. Judd FK, Jackson HJ, Komiti A, Murray G, Hodgins G, Fraser C. High prevalence disorders in urban and rural communities. Aust N Z J Psychiatry 2002 Feb;36(1):104-113. [Medline: 11929446]
8. AIHW. Rural, Regional and Remote Health: Indicators of Health Status and Determinants of Health. Rural Health Series no. 9. Canberra, Australia: Australian Institute of Health and Welfare; 2008 Mar. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442459831 [accessed 2015-05-26] [WebCite Cache ID 6YpK5ZwAT]
9. Hirsch J, Cukrowicz K. Suicide in rural areas: an updated review of the literature. American Psychological Association 2014;38(2):65-78. [doi: 10.1027/0227-5910.27.4.189]
10. Cheung YT, Spittal MJ, Pirkis J, Yip PS. Spatial analysis of suicide mortality in Australia: investigation of metropolitan-rural-remote differentials of suicide risk across states/territories. Soc Sci Med 2012 Oct;75(8):1460-1468. [doi: 10.1016/j.socscimed.2012.04.008] [Medline: 22771036]
11. Kõlves K, Milner A, McKay K, De Leo D. Suicide in Rural and Remote Areas of Australia. Brisbane, Australia: Australian Institute for Suicide Research and Prevention; 2012. URL: http://www.griffith.edu.au/__data/assets/pdf_file/0007/471985/Suicide-in-Rural-and-Remote-Areas-of-Australia.pdf [accessed 2015-05-26] [WebCite Cache ID 6YpKebQMV]
12. Ellis P, Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines Team for Depression. Australian and New Zealand clinical practice guidelines for the treatment of depression. Aust N Z J Psychiatry 2004 Jun;38(6):389-407. [doi: 10.1111/j.1440-1614.2004.01377.x] [Medline: 15209830]
13. NICE Guidelines. Manchester, UK: National Institute for Health and Care Excellence; 2011. Generalised anxiety disorder and panic disorder (with or without agoraphobia) in adults: Management in primary, secondary and community care URL: http://www.nice.org.uk/guidance/cg113/chapter/guidance [accessed 2015-05-26] [WebCite Cache ID 6YpMxVZLt]
14. Beatty L, Koczwara B, Wade T. Cancer coping online: A pilot trial of a self-guided CBT internet intervention for cancer-related distress. E J Appl Psychol 2011;7(2):17-25.
15. Tiller JWG. Depression and anxiety. Med J Aust 2012 Oct 1;1(4):28-31. [doi: 10.5694/mjao12.10628]
16. Jackson H, Judd F, Komiti A. Mental health problems in rural contexts: What are the barriers to seeking help from professional providers? Aust Psychol 2007;42(2):147-160. [doi: 10.1080/00050060701299532]
17. Fuller J, Edwards J, Procter N, Moss J. How definition of mental health problems can influence help seeking in rural and remote communities. Aust J Rural Health 2000 Jun;8(3):148-153. [Medline: 11249402]
18. HWA. Australia's Health Workforce Series - Psychologists in Focus. Adelaide, Australia: Health Workforce Australia; 2014 Mar. URL: https://www.hwa.gov.au/sites/default/files/HWA_Australia-Health-Workforce-Series_Psychologists%20in%20focus_vF_LR.pdf [accessed 2015-05-26] [WebCite Cache ID 6YpMeTD9d]
19. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
20. Gilman SE, Abraham HD. A longitudinal study of the order of onset of alcohol dependence and major depression. Drug Alcohol Depend 2001 Aug 1;63(3):277-286. [Medline: 11418232]
21. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000 Jul 24;160(14):2101-2107. [Medline: 10904452]
22. Rieckmann N, Gerin W, Kronish IM, Burg MM, Chaplin WF, Kong G, et al. Course of depressive symptoms and medication adherence after acute coronary syndromes: an electronic medication monitoring study. J Am Coll Cardiol 2006 Dec 5;48(11):2218-2222 [FREE Full text] [doi: 10.1016/j.jacc.2006.07.063] [Medline: 17161249]
23. Nemeroff CB, Goldschmidt-Clermont PJ. Heartache and heartbreak--the link between depression and cardiovascular disease. Nat Rev Cardiol 2012 Sep;9(9):526-539. [doi: 10.1038/nrcardio.2012.91] [Medline: 22733213]
24. Stack SJ. Mental illness and suicide. In: The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society. Hoboken, NJ: John Wiley & Sons, Ltd; 2014:1618-1623.
25. Christensen H, Batterham P, Calear A. Online interventions for anxiety disorders. Curr Opin Psychiatry 2014 Jan;27(1):7-13. [doi: 10.1097/YCO.0000000000000019] [Medline: 24257123]
26. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251. [doi: 10.1007/s10879-013-9243-y]
27. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
28. Mayo-Wilson E, Montgomery P. Media-delivered cognitive behavioural therapy and behavioural therapy (self-help) for anxiety disorders in adults. Cochrane Database Syst Rev 2013;9:CD005330. [doi: 10.1002/14651858.CD005330.pub4] [Medline: 24018460]

29. Mureşan V, Montgomery GH, David D. Emotional outcomes and mechanisms of change in online cognitive-behavioral interventions: a quantitative meta-analysis of clinical controlled studies. J Technol Hum Serv 2012 Jan;30(1):1-13. [doi: 10.1080/15228835.2011.653290]
30. 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]
31. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
32. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
33. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct;197(4):297-304 [FREE Full text] [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
34. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. PLOS Med 2009;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097]
35. National Centre for Social Research, Leicester University. Health & Social Care Information Centre. Leeds, UK: The NHS Information Centre for Health and Social Care; 2009 Jan 27. Adult psychiatric morbidity in England - 2007, results of a household survey [NS] URL: http://www.hscic.gov.uk/pubs/psychiatricmorbidity07 [accessed 2015-05-26] [WebCite Cache ID 6YpNmbENd]
36. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
37. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 2011 Apr;64(4):407-415. [doi: 10.1016/j.jclinepi.2010.07.017] [Medline: 21247734]
38. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
39. Hayward L, MacGregor AD, Peck DF, Wilkes P. The feasibility and effectiveness of computer-guided CBT (Fearfighter) in a rural area. Behav Cogn Psychother 2007 Jul;35(4):409-419.
40. Kay-Lambkin FJ, Baker AL, Kelly B, Lewin TJ. Clinician-assisted computerised versus therapist-delivered treatment for depressive and addictive disorders: a randomised controlled trial. Med J Aust 2011 Aug 1;195(3):S44-S50. [Medline: 21806518]
41. Kay-Lambkin FJ, Baker AL, Kelly BJ, Lewin TJ. It's worth a try: the treatment experiences of rural and urban participants in a randomized controlled trial of computerized psychological treatment for comorbid depression and alcohol/other drug use. J Dual Diagn 2012 Nov;8(4):262-276. [doi: 10.1080/15504263.2012.723315]
42. Mewton L, Wong N, Andrews G. The effectiveness of internet cognitive behavioural therapy for generalized anxiety disorder in clinical practice. Depress Anxiety 2012 Oct;29(10):843-849. [doi: 10.1002/da.21995] [Medline: 22949296]
43. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
44. Sunderland M, Wong N, Hilvert-Bruce Z, Andrews G. Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. Behav Res Ther 2012 Jun;50(6):374-380. [doi: 10.1016/j.brat.2012.03.005] [Medline: 22498311]
45. MacGregor AD, Hayward L, Peck DF, Wilkes P. Empirically grounded clinical interventions clients' and referrers' perceptions of computer-guided CBT (FearFighter). Behav Cogn Psychother 2009 Jan;37(1):1-9. [doi: 10.1017/S135246580800492X] [Medline: 19364403]
46. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014 Feb;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
47. Sethi S. Treating youth depression and anxiety: a randomised controlled trial examining the efficacy of computerised versus face-to-face cognitive behaviour therapy. Aust Psychol 2013 Feb 14;48(4):249-257. [doi: 10.1111/ap.12006]
48. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
49. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]

50. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013;13:113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
51. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008 Sep;193(3):181-184 [FREE Full text] [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]
52. 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]
53. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
54. Wright JH, Wright AS, Albano AM, Basco MR, Goldsmith LJ, Raffield T, et al. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry 2005 Jun;162(6):1158-1164. [doi: 10.1176/appi.ajp.162.6.1158] [Medline: 15930065]
55. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
56. Albano AM, Krain AL, Podniesinski E, Ditkowsky KS. Cognitive-behavior therapy with children and adolescents. In: Wright JH, editor. Cognitive Behaviour Therapy: Review of Psychiatry, Volume 23. Washington, DC: American Psychiatric Publishing, Inc; 2004:123-150.
57. Learmonth D, Trosh J, Rai S, Sewell J, Cavanagh K. The role of computer-aided psychotherapy within an NHS CBT specialist service. Counselling and Psychotherapy Research 2008 Jun;8(2):117-123. [doi: 10.1080/14733140801976290]
58. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Hum Serv 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
59. Moss-Morris R, McCrone P, Yardley L, van Kessel K, Wills G, Dennison L. A pilot randomised controlled trial of an Internet-based cognitive behavioural therapy self-management programme (MS Invigor8) for multiple sclerosis fatigue. Behav Res Ther 2012 Jun;50(6):415-421. [doi: 10.1016/j.brat.2012.03.001] [Medline: 22516321]
60. Murray K, Pombo-Carril MG, Bara-Carril N, Grover M, Reid Y, Langham C, et al. Factors determining uptake of a CD-ROM-based CBT self-help treatment for bulimia: patient characteristics and subjective appraisals of self-help treatment. Eur Eat Disord Rev 2003 May;11(3):243-260. [doi: 10.1002/erv.519]
61. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an Internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. Pain 2009 Nov;146(1-2):205-213 [FREE Full text] [doi: 10.1016/j.pain.2009.07.034] [Medline: 19695776]
62. Trockel M, Manber R, Chang V, Thurston A, Taylor CB. An e-mail delivered CBT for sleep-health program for college students: effects on sleep quality and depression symptoms. J Clin Sleep Med 2011 Jun 15;7(3):276-281 [FREE Full text] [doi: 10.5664/JCSM.1072] [Medline: 21677898]

Abbreviations

CBT: cognitive behavior therapy
CCBT: computerized cognitive behavior therapy
GAD: generalized anxiety disorder
GP: general practitioner
GRADE: Grades of Recommendation, Assessment, Development, and Evaluation
IPT: interpersonal therapy
OCD: obsessive compulsive disorder
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
RCT: randomized controlled trial
WHODAS: World Health Organization Disability Assessment

Edited by G Eysenbach; submitted 14.12.14; peer-reviewed by I Mindlis, L Mclellan; comments to author 19.03.15; revised version received 15.04.15; accepted 27.04.15; published 05.06.15.

Please cite as:
Vallury KD, Jones M, Oosterbroek C
Computerized Cognitive Behavior Therapy for Anxiety and Depression in Rural Areas: A Systematic Review
J Med Internet Res 2015;17(6):e139
URL: http://www.jmir.org/2015/6/e139/
doi:10.2196/jmir.4145
PMID:26048193

Original Paper

Web-Based Cognitive Behavioral Therapy for Female Patients With Eating Disorders: Randomized Controlled Trial

Elke D ter Huurne[1,2], MSc; Hein A de Haan[1,2], MD, PhD; Marloes G Postel[1,2,3], PhD; Job van der Palen[4,5], PhD; Joanne EL VanDerNagel[1,2,6], MD; Cornelis AJ DeJong[2,7], MD, PhD

[1]Tactus Addiction Treatment, Enschede, Netherlands
[2]Nijmegen Institute for Scientist-Practitioners in Addiction, Nijmegen, Netherlands
[3]University of Twente, Department of Psychology, Health & Technology, Enschede, Netherlands
[4]Medisch Spectrum Twente, Medical School Twente, Enschede, Netherlands
[5]University of Twente, Department of Research Methodology, Measurement and Data Analysis, Enschede, Netherlands
[6]Aveleijn, Borne, Netherlands
[7]Radboud University, Behavioral Science Institute, Nijmegen, Netherlands

Corresponding Author:
Elke D ter Huurne, MSc
Tactus Addiction Treatment
Raiffeisenstraat 75
Enschede, 7514 AM
Netherlands
Phone: 31 883822887
Fax: 31 570500115
Email: e.terhuurne@tactus.nl

Abstract

Background: Many patients with eating disorders do not receive help for their symptoms, even though these disorders have severe morbidity. The Internet may offer alternative low-threshold treatment interventions.

Objective: This study evaluated the effects of a Web-based cognitive behavioral therapy (CBT) intervention using intensive asynchronous therapeutic support to improve eating disorder psychopathology, and to reduce body dissatisfaction and related health problems among patients with eating disorders.

Methods: A two-arm open randomized controlled trial comparing a Web-based CBT intervention to a waiting list control condition (WL) was carried out among female patients with bulimia nervosa (BN), binge eating disorder (BED), and eating disorders not otherwise specified (EDNOS). The eating disorder diagnosis was in accordance with the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, and was established based on participants' self-report. Participants were recruited from an open-access website, and the intervention consisted of a structured two-part program within a secure Web-based application. The aim of the first part was to analyze participant's eating attitudes and behaviors, while the second part focused on behavioral change. Participants had asynchronous contact with a personal therapist twice a week, solely via the Internet. Self-report measures of eating disorder psychopathology (primary outcome), body dissatisfaction, physical health, mental health, self-esteem, quality of life, and social functioning were completed at baseline and posttest.

Results: A total of 214 participants were randomized to either the Web-based CBT group (n=108) or to the WL group (n=106) stratified by type of eating disorder (BN: n=44; BED: n=85; EDNOS: n=85). Study attrition was low with 94% of the participants completing the posttest assignment. Overall, Web-based CBT showed a significant improvement over time for eating disorder psychopathology (F_{97}=63.07, P<.001, d=.82) and all secondary outcome measures (effect sizes between d=.34 to d=.49), except for Body Mass Index. WL participants also improved on most outcomes; however, effects were smaller in this group with significant between-group effects for eating disorder psychopathology (F_{201}=9.42, P=.002, d=.44), body dissatisfaction (F_{201}=13.16, P<.001, d=.42), physical health (F_{200}=12.55, P<.001, d=.28), mental health (F_{203}=4.88, P=.028, d=.24), self-esteem (F_{202}=5.06, P=.026, d=.20), and social functioning (F_{205}=7.93, P=.005, d=.29). Analyses for the individual subgroups BN, BED, and EDNOS showed that eating disorder psychopathology improved significantly over time among Web-based CBT participants in all three subgroups; however, the between-group effect was significant only for participants with BED (F_{78}=4.25, P=.043, d=.61).

Conclusions: Web-based CBT proved to be effective in improving eating disorder psychopathology and related health among female patients with eating disorders.

Trial Registration: Nederlands Trial Register (NTR): NTR2415; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2415 (Archived by WebCite at http://www.webcitation.org/6T2io3DnJ).

(J Med Internet Res 2015;17(6):e152) doi:10.2196/jmir.3946

KEYWORDS

eating disorders; bulimia nervosa; binge-eating disorder; eating disorders not otherwise specified; randomized controlled trial; eHealth; Web-based treatment; asynchronous therapeutic support; treatment effectiveness; cognitive behavioral therapy

Introduction

In the Netherlands, eating disorders have a lifetime prevalence of 1.74% [1], and these disorders account for severe psychological, physical, and social morbidity. Although early identification and treatment is desired, patients often refrain from seeking or receiving help because of personal barriers, such as feelings of shame and fear of stigmatization, and intervention-related barriers, such as costs, geographical distance, and lack of availability [2-6]. Psychiatric services are challenged to help patients overcome these barriers by providing easily accessible, low-threshold interventions.

The Internet offers many possibilities for these types of interventions because of its relative anonymity, widespread and 24-hour access, and increasing usage. Low-threshold Internet interventions may reach patients with less advanced disorders and prevent their condition from progressing. Moreover, Web-based interventions for psychopathology such as depression, anxiety, and addictive behaviors have already proven successful, with interventions using (intensive) therapeutic support being more effective than self-help or minimal contact interventions [7-9]. In the past few years, several Internet interventions have been developed for patients with eating disorders, and a recent review showed that these treatments can be effective in reducing eating disorder psychopathology, binge eating, and purging, as well as in improving quality of life [10]. However, it should be noted that most studies were conducted among patients with bulimia nervosa (BN) and (to a lesser extent) binge eating disorder (BED) [10], whereas eating disorders not otherwise specified (EDNOS) is the most commonly diagnosed eating disorder [11]. Additionally, studies conducted among patients with EDNOS mostly included interventions aimed at (indicated) prevention or early intervention in eating disorders [12-14]. Although some interventions proved to be effective, most studies included only adolescents and young women [12,13], while the EDNOS subgroup includes a broader population of patients with eating disorder symptoms. Therefore, in 2009 a Web-based cognitive behavioral therapy (CBT) intervention with intensive therapeutic support was developed to treat Dutch patients with all types of eating disorders [15], based on a similar intervention for problem drinkers [16,17], in which patients communicate asynchronously with their therapist twice a week. A before-after study into this intervention showed a reduction in eating disorder psychopathology (*d*=1.14) and body dissatisfaction (*d*=0.86), as well as high patient satisfaction [18]. However, this study had a nonrandomized design and included only those participants who completed the intervention (54% of participants).

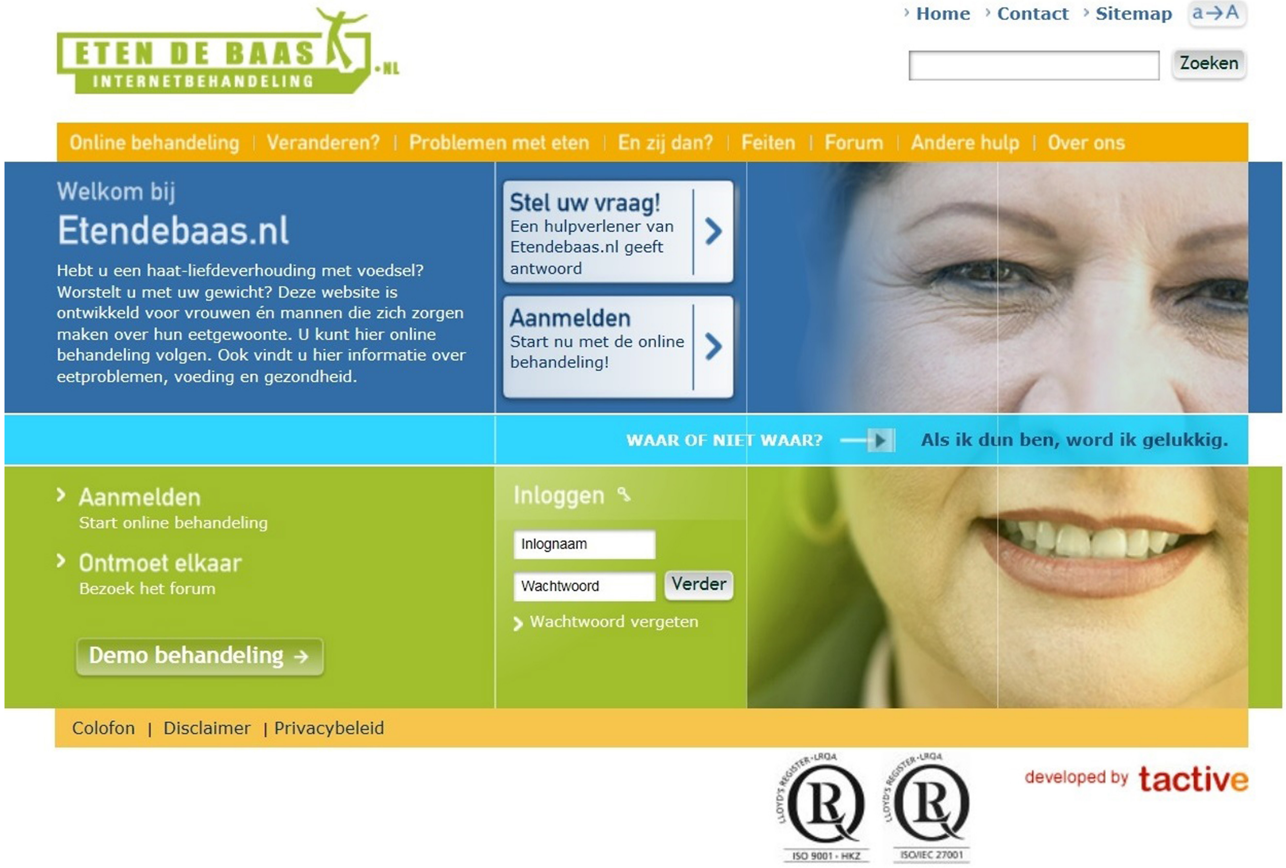
This study, therefore, aimed to explore the effects of the Web-based CBT intervention, compared to a waiting list control group (WL), on eating disorder psychopathology (primary outcome) as well as body dissatisfaction, physical health, body mass index (BMI), mental health, self-esteem, quality of life, and social functioning (secondary outcomes). Furthermore, we were interested in the effects of the Web-based CBT across participants of the specific eating disorder subgroups: BN, BED, and EDNOS.

Methods

Participants

Participants were self-recruited users of the Dutch website “Look at your eating” [19]. This open-access website (see Figure 1) offered general information on eating disorders and related topics, a forum for peer support, as well as information about the Web-based CBT program and the study procedures of this trial. Inclusion criteria for participation were (1) female gender, (2) age ≥18 years, (3) diagnosis of BN, BED, or EDNOS (based on participants’ self-report), (4) written and oral fluency in Dutch language, (5) access to Internet, (6) signed informed consent, and (7) a referral from a general practitioner (GP). Exclusion criteria were (1) severe underweight, (2) suicidal ideation, (3) receiving psychological or pharmaceutical treatment for any eating disorder within the past 6 months, (4) pregnancy, and (5) expected absence for 4 weeks or longer during the intervention period (eg, due to planned vacation). If participation in the intervention was not possible for some reason (eg, lack of Dutch health insurance and therefore funding of the intervention, or patient’s GP did not agree with participation), patients were also not eligible for this study.

Figure 1. Website homepage.



Welkom bij Etendebaas.nl

Hebt u een haat-liefdeverhouding met voedsel? Worstelt u met uw gewicht? Deze website is ontwikkeld voor vrouwen én mannen die zich zorgen maken over hun eetgewoonte. U kunt hier online behandeling volgen. Ook vindt u hier informatie over eetproblemen, voeding en gezondheid.

Stel uw vraag! Een hulpverlener van Etendebaas.nl geeft antwoord

Aanmelden Start nu met de online behandeling!

WAAR OF NIET WAAR? Als ik dun ben, word ik gelukkig.

Study Design and Procedure

This study was a randomized controlled trial with two groups: Web-based cognitive behavioral therapy (Web-based CBT) and waiting list control (WL). Figure 2 presents a flow chart of the design and timeline of the study. Recruitment took place from March 2011 until December 2013. Information about the study was disseminated through announcements on eating disorder–related websites and forums, and newspaper advertisements. Website visitors were invited to read the information about the study explicitly, provide their email addresses and telephone numbers, and agree with the conditions of the Web-based CBT protocol. Furthermore, they had to provide written informed consent, personal data, and data of their GP. The GP was informed about the patients' participation in the study (as covered by the Ethics Committee approval) and requested to sign and return the referral form. To assess eligibility and to obtain baseline data, participants completed an online self-report questionnaire during the sign-up procedure. Based on this questionnaire, the *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition (DSM-IV) eating disorder diagnosis was assessed and randomization took place. Participants not eligible for this study were offered participation in the regular Web-based CBT intervention (outside this study). This was possible only by logging in with a personal code that individuals received by mail after providing personal data. In case of urgent medical risks, no funding of the intervention, or disagreement of the GP for participation in the intervention, participants were referred to their GP or advised on more appropriate treatment.

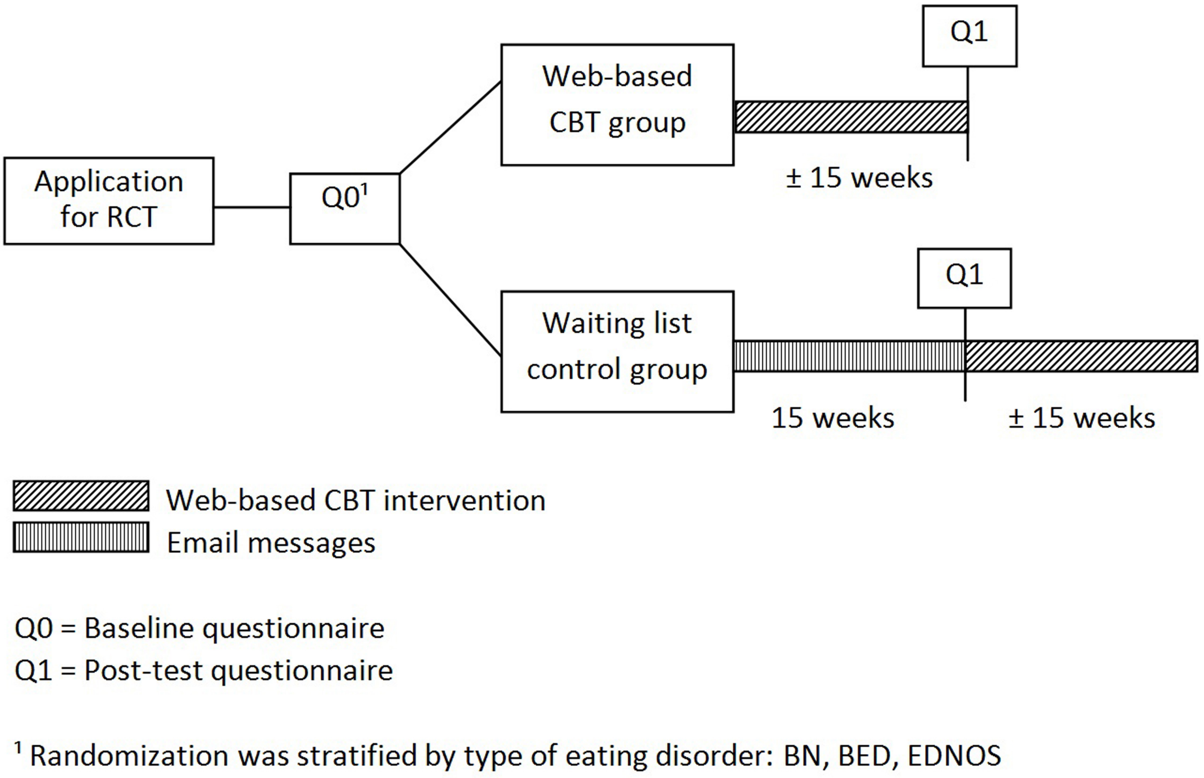
Participants were randomized to the Web-based CBT or WL through computer-generated randomly varying block sizes (2, 4, or 8), stratified by type of eating disorder (BN, BED, EDNOS). Randomization was performed at a 1:1 ratio. The allocation schedule was prepared by an independent researcher not involved in data collection. The assignment of participants to the conditions was not dependent on the participants' characteristics. Participants assigned to the Web-based CBT started the intervention immediately while participants of the WL had to wait 15 weeks, during which they received supportive email messages once every 2 weeks to keep them involved. All WL participants were guaranteed treatment after the waiting period, and they were advised to contact their GP in case earlier treatment was needed.

To measure efficacy, the WL group completed the posttest questionnaire after the 15-week waiting period, and the Web-based CBT group completed this questionnaire after completing or prematurely ending the intervention, or in case of a longer treatment duration, 18 weeks after randomization (to keep the time frame between the first and second assessments as close as possible in both groups). The posttest questionnaire measured all primary and secondary outcomes and included evaluation questions about the Web-based CBT intervention and treatment non-completion when applicable. Participants received a €10 digital voucher for an online store for each completed questionnaire, except for the baseline questionnaire.

The study protocol was approved by the Ethics Committee of Medical Spectrum Twente in March 2011 (reference number NL31717.044.010, P10-31) and was registered on the Dutch

Trial Registry (NTR2415). Details of this protocol have been published previously [20]. See Multimedia Appendix 1 for the CONSORT-EHEALTH checklist [21].

Figure 2. Flow chart of study design and timeline.



[1] Randomization was stratified by type of eating disorder: BN, BED, EDNOS

Web-Based Cognitive Behavioral Therapy Intervention

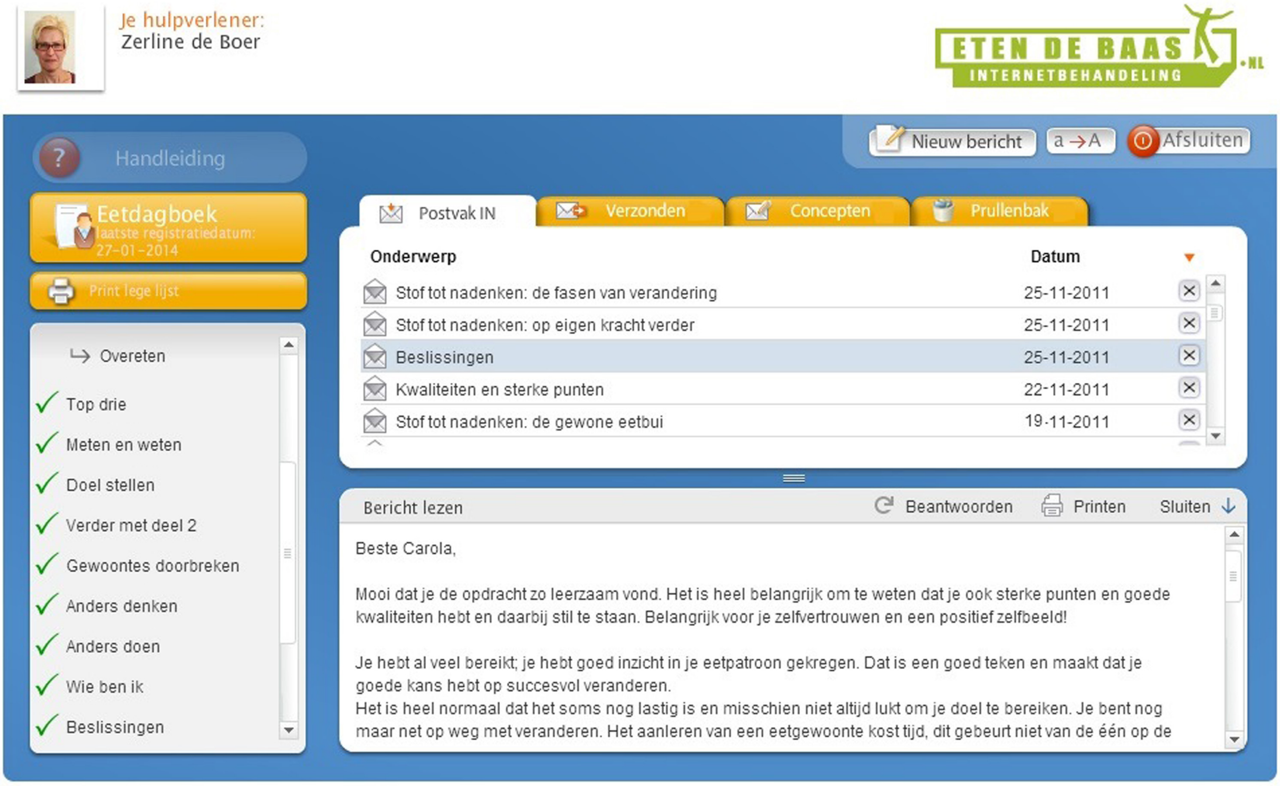
The Web-based CBT intervention [15,18] “Look at your eating” (in Dutch: Etendebaas) was developed at Tactus Addiction Treatment by health care professionals (psychologist, addiction medicine physician, psychotherapist, psychiatrist, dietician, registered nurses, and social workers), a software development team (The Factor-E), and patients and members of a Dutch organization for people with eating disorders. Development of the intervention was an interactive and iterative process, involving patients providing input and feedback on different versions of the content, layout, visual features, and ease of navigation of the Web-based CBT intervention. The intervention included a structured two-part program with at least 21 contact moments and 10 assignments within a secure Web-based application. Multimedia Appendix 2 presents an overview of the content of the intervention. The first part aimed to analyze participants’ eating attitudes and behaviors, while the second part focused on behavioral change. The modules were organized in a pre-specified order, and participants had access only to the information and assignments that were sent by the therapist at a specific point. CBT [22-24] and motivational interviewing (MI) [25,26] were the fundamentals of the intervention including techniques such as psycho-education, self-monitoring, thought restructuring, problem solving, and relapse prevention.

Participants had to log in to the application via the website [19] with a personal username and password, secured by Secure Sockets Layer, to have access to their personal file (see Figure 3). All data transferred between the participant’s computer and the application was encrypted and sent via the Hypertext Transfer Protocol Secure (https) protocol. The application was entirely server-based, and all information gathered was stored on an encrypted database. Daily backups of the server on an offsite location were made to ensure further data security.

During the Web-based CBT, participants normally had asynchronous contact twice a week with their personal therapists, solely via the Internet, unless participants specifically requested an additional telephone contact. Participation in the intervention took approximately 20 minutes per day, and participants were instructed about completing home assignments and registering eating behaviors daily in their online eating diary. Participants could access the intervention in their personal environment at any time, and in their private file they could read the messages from their therapist or complete assignments. Accordingly, therapists responded within 3 working days on the participants’ messages or assignments. The therapists’ messages were personalized to the participants’ situation but also comprised pre-programmed text fragments, for example, explaining the assignments. For each module, a format was available including the topics that should be addressed and the information that should be given by the therapist. Therapists could also include hyperlinks to information on the website [19] in their messages. These formats ensured consistency in the therapists’ messages. However, as all messages were adjusted to the situation of participants, with differences in complexity of personal issues, their content and length varied. Therapists always responded on participants' completed eating diary, assignments, and questionnaires, but the content of these texts was dependent on what participants had filled out. The responses of the therapists were supportive and included also CBT and MI techniques. Moreover, their communication primarily focused on providing accurate and objective information, hopeful writing, reinforcement, and relabeling of demotivating statements. The progress of the participants was monitored by the therapists. When participants did not respond to the messages of the therapists within the next week, they received a reminder with a request to keep in touch regularly. When participants did not respond for 4 weeks, the intervention was terminated by the therapist. Participants were considered treatment completers if they had completed all 10 assignments and attended at least 21 sessions. In case participants stopped the intervention prematurely and still needed help, the therapist discussed with the participant what kind of treatment would be more appropriate. Possible options were face-to-face treatment with a professional (therapist or dietician), day care, or hospitalization in a specialist eating disorder institution. If preferred by the participant, the therapist also initiated the first contact with the other professional or institution.

A total of 17 therapists with either a bachelor's degree in nursing or social work or a master's degree in psychology were involved in this study. All therapists completed 2 days of training including theoretical information and practice-oriented assignments focusing on the design and implementation of the Web-based application, on technical aspects of delivering this intervention, and on using different strategies to apply the CBT and MI techniques (eg, restructuring of non-helpful thoughts into helpful thoughts, enhancing self-efficacy, expression of empathy, and evoking cognitive dissonance). Additionally, they also received a 1-day training session about eating disorders and related issues, and about the treatment content and protocol. Subsequently, all therapists completed a full treatment program with a test patient, and 3 months of intensive supervision by experienced coaches. A comprehensive manual to the Web-based CBT intervention was available for all therapists, which included a detailed description of the different modules of the intervention. Also safety protocols were described in this manual, covering the criteria for admission to part 2, as well as guidelines about what to do in case of severe eating problems, relapse or suicidal ideations, and when to inform the participant's GP. Safety protocols were also included in the formats of the different modules. Besides the use of formats for all modules and the intensive training of therapists, several other methods were used to ensure quality and consistency in the treatment of participants. For example, all messages of the therapists were checked retrospectively (and adjusted if necessary) by a multidisciplinary team consisting of a psychologist, a psychotherapist, an addiction medicine physician, a psychiatrist, a dietician, and two coaches at the end of part 1 of the intervention. Furthermore, all patient files were regularly checked by the coaches of the Web-based CBT intervention and these coaches were also present daily for consultation and advice. The multidisciplinary team was available remotely for consultation throughout the intervention and gave also expert advice to the therapists at the end of part 1 of the intervention of each participant. For all participants, the intervention was covered by Dutch health insurance, although the costs were set off against participants' deductible. Therefore, several of them needed to pay a contribution of up to €350 for their participation. This is a standard procedure in the Dutch health care system.

Figure 3. Participant's personal online file.



Onderwerp	Datum
Stof tot nadenken: de fasen van verandering	25-11-2011
Stof tot nadenken: op eigen kracht verder	25-11-2011
Beslissingen	25-11-2011
Kwaliteiten en sterke punten	22-11-2011
Stof tot nadenken: de gewone eetbui	19-11-2011

Beste Carola,

Mooi dat je de opdracht zo leerzaam vond. Het is heel belangrijk om te weten dat je ook sterke punten en goede kwaliteiten hebt en daarbij stil te staan. Belangrijk voor je zelfvertrouwen en een positief zelfbeeld!

Je hebt al veel bereikt; je hebt goed inzicht in je eetpatroon gekregen. Dat is een goed teken en maakt dat je goede kans hebt op succesvol veranderen.
Het is heel normaal dat het soms nog lastig is en misschien niet altijd lukt om je doel te bereiken. Je bent nog maar net op weg met veranderen. Het aanleren van een eetgewoonte kost tijd, dit gebeurt niet van de één op de

Waiting List Control Group

Participants in the WL received seven supportive email messages sent by the researcher during the waiting period. The first message included a brief explanation about the design of the Web-based CBT intervention as well as information about when participants could start the intervention, the aim of the supporting email messages, and what participants should do in case they needed urgent help. The second message provided information about the different topics discussed on the website [19], such as information about factors that could effect participants' eating behavior and information about the physical effects of eating disorders. The third email contained information about the online forum, which was part of the website [19] and provided contact with fellow sufferers. The fourth email included information about the scientific research project of the Web-based CBT intervention and the results of the pilot study. The fifth message contained a summary of comments from former participants of the Web-based CBT intervention (verbatim and anonymous) about their experiences with the treatment. The sixth email focused on common misconceptions about food, losing weight, and eating disorders. This message also referred to relevant information on the website [19] and possible consequences of these misconceptions. The last email included a description of the basic principles of the intervention and therapeutic support, and a description of what was expected of the participants during the intervention. Finally, participants were thanked for their patience during the waiting period and were wished success with their participation in the intervention. All email messages also concluded with information about what participants had to do when they needed urgent help. Specifically, they were referred to their GP in case of urgent medical problems. This information was also presented in the prior email that participants received the moment they were allocated to the control group.

Measures

The primary outcome measure was eating disorder psychopathology measured with the global mean scale of the Eating Disorder Examination-Questionnaire (EDE-Q) [27], a widely used self-report questionnaire to measure eating disorder severity. Items were scored on a 7-point Likert scale (range 0-6), with a higher score reflecting more psychopathology. Additionally, the scores on the EDE-Q subscales Restraint, Eating Concern, Shape Concern, and Weight Concern were calculated. Secondary outcome measures included the Body Attitude Test (BAT) [28,29] to assess body dissatisfaction, the Maudsley Addiction Profile-Health Symptom Scale (MAP-HSS) [30] and 15 eating disorder-specific physical complaints to measure physical health, the Depression Anxiety Stress Scale (DASS) [31] to measure mental health, the Rosenberg Self-Esteem Scale (RSES) [32] to examine self-esteem, the EuroQol visual analogue scale (EQ-5D VAS) [33] to assess quality of life, the Measurements in the Addictions for Triage and Evaluation - International Classification of Functioning, Disability and Health (MATE-ICN) [34] to examine social functioning, and Body Mass Index (BMI). Participants completed these measures at baseline and posttest (approximately 15 weeks after baseline). Additionally, demographic data, DSM-IV diagnosis, prior care for eating disorders and other psychological problems, duration of illness, and suicide risk were measured at baseline. DSM-IV eating disorder diagnosis was assessed using the online self-report questionnaire Eating Disorder Questionnaire-Online (EDQ-O) and related baseline data, optionally with additional questions by email. The EDQ-O was developed as a diagnostic instrument for establishing all DSM-IV-TR eating disorder diagnoses without using face-to-face contact, as in-person clinical interviews are not suitable for Web-based interventions. A recent study into the validity of the EDQ-O showed that this self-report questionnaire performs acceptably as a diagnostic instrument for all eating disorder classifications [35]. However, as the results of the validation study were not available at the start of the current RCT, the EDQ-O was not used as the only tool to set a DSM-IV eating disorder classification in this study, but also other baseline data of participants were taken into consideration such as their BMI and completed EDE-Q. If there were doubts about the diagnosis established with the EDQ-O based on the combination of all data, participants were asked additional questions by email. Based on the responses of the participants, a final diagnosis was established by the psychologist and researcher. Suicide risk was measured using Part C of the MINI-Plus [36,37], consisting of 6 self-report items examining suicidal tendencies. Participants for whom the current risk of suicide was classified as “high” were excluded from the study. The exclusion criterion of severe underweight was assessed in case the body weight of participants was less than 85% of ideal body weight, determined using the table of height/weight limits of the MINI-Plus [36,37]. At posttest, participants were also asked if they had other support for their eating disorder during the intervention or waiting period. Furthermore, participants' evaluation of the intervention and their personal therapist was measured as well as reasons for non-completion (if applicable).

Statistical Analyses

Our sample size was calculated based on an expected mean difference score of 1.0 (SD 1.2) on the EDE-Q global score (primary outcome measure) between the Web-based CBT and WL at posttest. This expected difference was based on the results of our before-after study, adjusted for an estimated improvement in the WL. Power analysis (G*Power) revealed a sample size of 25 participants in each condition based on a significance level of 5%, a power of 80%, the same number of participants per condition, 2 measurements, and a correlation among repeated measures of 0.95. However, we expected 40% of the participants not to complete the Web-based CBT, therefore, 42 participants in each condition (Web-based CBT and WL) were needed. To determine the efficacy of the Web-based CBT for the specific eating disorder subtypes, the total sample size was determined at 84 participants with BN, 84 participants with BED, and 84 participants with EDNOS (total of 252 participants).

All analyses were conducted using SPSS version 21. Data are presented as numbers (percentage) for categorical data and as the means (SD) for continuous data. Baseline differences between the Web-based CBT and WL are expressed as differences in proportion for categorical data and as the mean differences for continuous data. Chi square or Fisher's exact tests (as appropriate) were used to compare categorical measures between the groups, and *t* tests or Mann-Whitney tests to compare continuous measures. To measure baseline differences between the three subgroups, Chi square or Fisher's exact tests were used to compare categorical measures, and analysis of variance (ANOVA) with Tukey's post-hoc tests or Kruskall-Wallis tests were used to compare continuous measures. Post-hoc tests for categorical variables were conducted by pairwise comparisons, with a Holm-Bonferroni post-hoc correction.

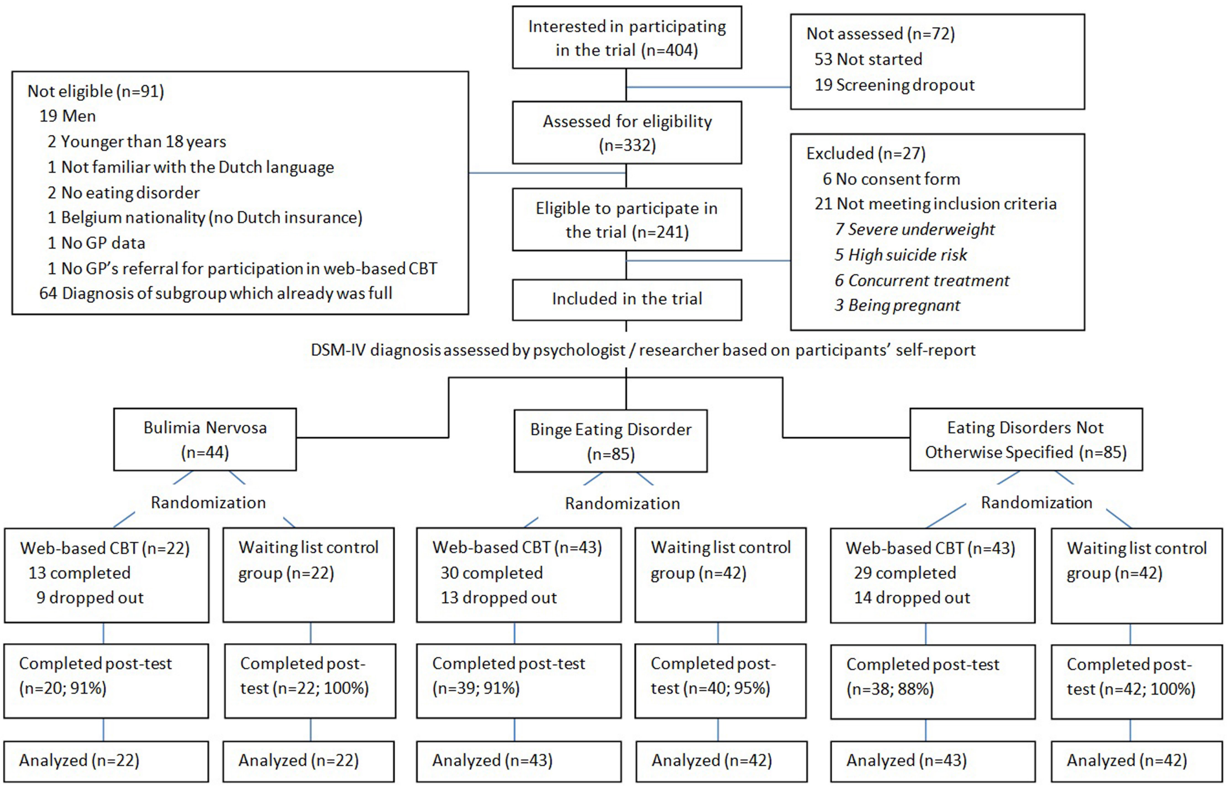
To measure the efficacy of the Web-based CBT in terms of primary and secondary outcome measures, Mixed Models for repeated measures were used, allowing for the inclusion of all participants, regardless of missing data. The intervention*time interaction effect was used to measure whether the change over time was different for the Web-based CBT compared to the WL. Between-group effect sizes were calculated according to Cohen's *d* by subtracting the average difference score between pretest and posttest of the control group from the corresponding difference score of the Web-based CBT group, and dividing the result by the pooled standard deviation of the pretest. Additionally, the effects over time within the Web-based CBT and WL group were measured. Within-group effect sizes were calculated by subtracting the average score at posttest from the average score at pretest and dividing the result by the pooled standard deviation of the pretest. Effect sizes of 0.8 were considered large, effect sizes of 0.5 moderate, and effect sizes of 0.2 small [38].

Results

Participant Flow

From the 404 subjects initially interested in participating in the trial, a total of 214 participants were randomized to one of the two conditions (Web-based CBT or WL), stratified by type of eating disorder (subgroups BN, BED, EDNOS). As shown in Figure 4, the predetermined sample size of 84 participants per eating disorder subtype had not been reached for the subtype BN (n=44). Within the Web-based CBT group, a total of 72 participants (66.7%, 72/108) completed the intervention and 36 participants (33.3%, 36/108) were considered treatment non-completers. Posttest assignments were completed for 201 participants (93.9%, 201/214) with a higher study dropout in the Web-based CBT (10.2%, 11/108) than WL (1.9%, 2/106) (χ^2_1=6.46, *P*=.01). Participants who withdrew from the study more often lived alone (χ^2_1=5.74, *P*=.04) and had less self-esteem (t_{212}=2.53, *P*=.01) at baseline than participants who completed the posttest. Within the Web-based CBT group, 99% of the treatment completers (71/72) and 56% of the treatment non-completers (20/36) completed the questions regarding treatment acceptability. Reasons for treatment non-completion were given by 67% of the non-completers (24/36).

Figure 4. Flow chart of study participants.



Participant Characteristics

The baseline characteristics of the participants are presented in Table 1. Participants were 214 females with an average age of 39 years, mostly living with others (74.3%, 159/214), employed (79.0%, 169/214), higher educated (50.9%, 109/214) and being overweight (85.0% BMI >25, 182/214). For most participants, the history of the eating disorder was long (68.2% >10 years, 146/214), whereas only 44.4% (95/214) received prior care for their disorder.

The three subgroups BN, BED, and EDNOS differed significantly in terms of age, living situation, and daily routine (data presented in Table 1). BN participants were younger, lived more often alone, and less frequently had a regular daily routine. Also the illness-related variables differed significantly between the three groups, except for the quality of life. Eating disorder psychopathology and mental health were most severe within the BN subgroup and least severe within the EDNOS subgroup. BN participants also had a significantly lower BMI and more problems in social functioning than the other two groups, and significantly more physical complaints and less self-esteem than the EDNOS participants. Participants of the BED subgroup were significantly less satisfied with their body than participants of the EDNOS subgroup.

Baseline characteristics showed no significant differences between the Web-based CBT (n=108) and WL (n=106), nor between these two groups within the three subgroups (data not shown), with the exception of a higher EDE-Q subscale “Restraint” in the EDNOS Web-based CBT subgroup (t_{83}=2.05, P=.04).

Table 1. Baseline characteristics of participants.

Variable	Overall (n=214)	BN (n=44)	BED (n=85)	EDNOS (n=85)	Test statistic	df	*P* value
Age (years), mean (SD)	39.4 (11.6)	33.2 (10.4)	40.2 (11.4)	41.9 (11.3)	9.00	2	<.001[hi]
Living situation, n (%)					7.93	2	.02
Alone	55 (25.7)	17 (39)	24 (28)	14 (16)			
With others	159 (74.3)	27 (61)	61 (72)	71 (84)			
Level of education, n (%)					3.59	4	.47
Low	25 (11.7)	2 (4)	13 (15)	10 (12)			
Intermediate	80 (37.4)	19 (43)	31 (36)	30 (35)			
High	109 (50.9)	23 (52)	41 (48)	45 (53)			
Employed, n (%)	169 (79.0)	33 (75)	66 (78)	70 (82)	1.09	2	.58
Regular daily routine, n (%)	156 (72.9)	24 (55)	60 (71)	72 (85)	13.73	2	.001
Prior eating disorder treatment, n (%)	95 (44.4)	17 (39)	39 (46)	39 (46)	0.74	2	.69
Prior psychological treatment, n (%)	143 (66.8)	35 (80)	58 (68)	50 (59)	5.74	2	.06
Body Mass Index, n (%)					67.01	4	<.001
<18.5	6 (2.8)	3 (7)	-	3 (4)			
18.5-25	26 (12.1)	20 (46)	-	6 (7)			
>25	182 (85.0)	21 (48)	85 (100)	76 (89)			
Duration of eating disorder, n (%)					11.99	6	.06
1-5 years	33 (15.4)	10 (23)	7 (8)	16 (19)			
6-10 years	35 (16.4)	8 (18)	17 (20)	10 (12)			
11-20 years	64 (29.9)	12 (27)	32 (38)	20 (24)			
>20 years	82 (38.3)	14 (32)	29 (34)	39 (46)			
Eating disorder psychopathology [a] , mean (SD)	3.5 (0.9)	4.1 (0.9)	3.6 (0.8)	3.0 (0.9)	23.54	2	<.001[hij]
Restraint	2.6 (1.3)	3.5 (1.1)	2.4 (1.3)	2.3 (1.3)	15.00	2	<.001[hi]
Eating concern	2.9 (1.3)	3.5 (1.1)	3.3 (1.2)	2.2 (1.3)	23.39	2	<.001[ij]
Shape concern	4.4 (1.1)	4.8 (1.1)	4.6 (0.9)	3.9 (1.1)	13.53	2	<.001[ij]
Weight concern	4.1 (1.1)	4.5 (1.2)	4.2 (0.8)	3.7 (1.1)	7.95	2	.001[ij]
Body dissatisfaction[b], mean (SD)	59.3 (14.8)	59.9 (17.4)	62.6 (14.1)	55.7 (13.5)	4.71	2	.01[j]
Physical health[c], mean (SD)	24.2 (12.3)	28.9 (13.4)	24.3 (11.3)	21.7 (12.1)	5.16	2	.006[i]
MAP-HSS, mean (SD)	12.3 (6.2)	13.1 (6.7)	12.7 (5.8)	11.5 (6.3)	1.31	2	.27
Mental health [d] , mean (SD)	33.1 (19.4)	43.0 (20.7)	33.5 (18.1)	27.6 (18.0)	19.83	3	<.001[hij]
Depression	11.4 (8.4)	14.9 (8.7)	11.6 (7.9)	9.5 (8.2)	13.82	3	.001[hi]
Anxiety	6.0 (6.4)	8.5 (7.4)	6.3 (6.5)	4.5 (5.3)	10.85	3	.004[i]
Stress	15.7 (8.6)	19.7 (9.0)	15.7 (8.3)	13.6 (8.0)	7.78	2	.001[hi]
Self-esteem[e], mean (SD)	16.7 (7.2)	14.1 (7.4)	16.4 (6.6)	18.4 (7.2)	5.57	2	.004[i]
Quality of life[f], mean (SD)	62.0 (17.1)	61.0 (17.2)	60.1 (15.6)	64.5 (18.4)	1.50	2	.23
Social functioning[g], mean (SD)	6.7 (4.8)	8.8 (5.5)	6.6 (4.6)	5.7 (4.4)	6.21	2	.002[hi]

[a] Eating Disorder Examination-Questionnaire.

[b] Body Attitude Test.

[c] Total score of Maudsley Addiction Profile-Health Symptom Scale (MAP-HSS) and 15 additional eating disorder-specific physical complaints.

[d] Depression Anxiety Stress Scale.

[e] Rosenberg Self-Esteem Scale.

[f] EuroQol visual analogue scale.

[g] Measurements in the Addictions for Triage and Evaluation – International Classification of Functioning, Disability and Health.

[h] Significant difference between BN and BED.

[i] Significant difference between BN and EDNOS.

[j] Significant difference between BED and EDNOS.

Efficacy of Web-Based Treatment for Eating Disorders in General

Table 2 reports the primary and secondary treatment outcomes of this study for all participants at posttest. Participants of the Web-based CBT improved significantly more on eating disorder psychopathology than participants of the WL, although both groups showed improvements over time. Web-based CBT participants also had significantly fewer concerns about their eating, shape, and weight after participating in the intervention than participants of the WL.

Body dissatisfaction, physical health, mental health, self-esteem, and social functioning all improved significantly more in participants of the Web-based CBT than in participants of the WL. Therefore, participating in the intervention resulted in an overall improvement on health indicators related to eating disorders, although effect sizes were generally small. For BMI, no significant effects were found in either group.

During the intervention or waiting period, several participants in both groups had other support for their eating disorder, for example, from family or friends, or through a self-help program (Web-based CBT 24%, 23/96; WL 30.8%, 32/104). Some of these participants had face-to-face contact with a professional such as a therapist, dietician, or GP (Web-based CBT 11%, 11/96 and WL 18.3%, 19/104). For the Web-based CBT group, the additional support had no added value as there were no significant differences between participants who had other (professional face-to-face) support and those who did not have this support during the intervention period. For the WL, additional support had only a significant effect on body dissatisfaction in case of face-to-face contact with a professional. Participants who had face-to-face contact with a professional during the waiting time improved significantly more on body dissatisfaction than participants who did not have this kind of support (F_{102}=5.16; P=.025). For all other outcome measures, no significant effects were found between those two WL groups. Overall effects of the Web-based CBT intervention did also not change significantly by correcting for additional face-to-face support (data not shown). As we used an intention-to-treat (ITT) analysis, we therefore did not correct for this.

Table 2. Treatment outcomes for all participants and for the individual subgroups BN, BED, and EDNOS.

	Web-based CBT (n=108)				WL (n=106)				Interaction effect of group x time[a]			
	Baseline	Posttest	*P*	*d*	Baseline	Posttest	*P*	*d*	F	df	*P*	*d*
All participants (n=214)												
Eating disorder psychopathology [b]	3.5 (0.9)	2.6 (1.2)	<.001	.82	3.5 (1.0)	3.0 (1.1)	<.001	.43	9.42	201	.002	.44
Restraint	2.7 (1.3)	2.2 (1.4)	<.001	.39	2.5 (1.4)	2.1 (1.4)	.046	.23	0.90	206	.34	.15
Eating concern	2.8 (1.2)	1.7 (1.3)	<.001	.83	3.0 (1.4)	2.4 (1.5)	<.001	.40	6.65	205	.01	.35
Shape concern	4.4 (1.1)	3.4 (1.5)	<.001	.72	4.4 (1.1)	3.9 (1.3)	<.001	.38	7.87	200	.006	.43
Weight concern	4.0 (1.1)	3.1 (1.4)	<.001	.73	4.1 (1.0)	3.7 (1.2)	<.001	.36	11.13	200	.001	.48
Body dissatisfaction[c]	58.4 (14.9)	50.5 (17.6)	<.001	.49	60.3 (14.8)	58.6 (15.3)	.05	.11	13.16	201	<.001	.42
Body Mass Index												
<18.5	17.2 (0.3)	17.2 (0.6)	.93	.03	17.9 (-)	17.9 (-)	-	-	<0.01	4	.98	.04
18.5 – 25	22.4 (1.7)	23.2 (1.2)	.09	.58	22.0 (1.7)	22.3 (2.5)	.56	.10	1.43	23	.24	.38
>25	33.5 (5.7)	33.3 (5.6)	.23	.04	34.0 (5.4)	34.0 (5.3)	.93	.00	0.94	169	.33	.04
Physical health [d]	22.8 (12.7)	18.1 (12.7)	<.001	.37	25.7 (11.8)	24.3 (12.2)	.045	.11	12.55	200	<.001	.28
MAP-HSS	11.7 (6.5)	9.5 (6.4)	<.001	.34	13.0 (5.9)	12.2 (5.9)	.046	.13	6.94	202	.009	.23
Mental health [e]	31.7 (18.6)	23.6 (18.5)	<.001	.43	34.6 (20.2)	31.2 (20.7)	.03	.17	4.88	203	.03	.24
Depression	11.2 (8.4)	7.5 (7.5)	<.001	.46	11.7 (8.5)	10.2 (8.6)	.06	.17	4.37	204	.04	.26
Anxiety	5.5 (5.8)	4.7 (5.4)	.06	.16	6.5 (7.0)	6.4 (6.4)	.84	.02	1.09	206	.30	.12
Stress	15.0 (8.5)	11.3 (8.4)	<.001	.43	16.4 (8.6)	14.6 (9.0)	.009	.20	3.55	203	.06	.22
Self-esteem[f]	16.2 (7.1)	18.6 (6.8)	<.001	.34	17.3 (7.2)	18.3 (6.9)	.01	.14	5.06	202	.03	.20
Quality of life[g]	62.8 (17.2)	69.2 (15.8)	<.001	.39	61.2 (17.1)	65.4 (15.0)	.03	.27	0.85	206	.36	.13
Social functioning[h]	6.8 (5.0)	5.2 (4.3)	<.001	.35	6.5 (4.7)	6.3 (4.3)	.50	.05	7.93	205	.005	.29
BN subgroup (n=44)												
Eating disorder psychopathology [b]	3.9 (1.0)	2.9. (1.1)	.003	.94	4.2 (0.8)	3.7 (1.1)	.02	.55	1.92	41	.17	.55
Restraint	3.3 (1.3)	2.6 (1.2)	.01	.64	3.7 (1.0)	3.4 (1.2)	.43	.27	1.11	42	.30	.44
Eating concern	3.3 (1.1)	1.9 (1.1)	<.001	1.30	3.7 (1.1)	3.0 (1.5)	.04	.53	2.59	41	.12	.65
Shape concern	4.7 (1.2)	3.8 (1.5)	.02	.67	4.9 (1.0)	4.4 (1.3)	.02	.49	0.74	41	.39	.31
Weight concern	4.4 (1.3)	3.5 (1.6)	.03	.62	4.7 (1.2)	4.2 (1.2)	.02	.43	0.86	41	.36	.32
BED subgroup (n=85)												
Eating disorder psychopathology [b]	3.5 (0.8)	2.6 (1.3)	<.001	.87	3.7 (0.7)	3.2 (0.9)	.002	.60	4.25	78	.04	.61
Restraint	2.5 (1.4)	1.9 (1.5)	.005	.42	2.3 (1.3)	2.0 (1.3)	.31	.23	0.60	80	.44	.22
Eating concern	3.1 (1.2)	2.0 (1.4)	<.001	.86	3.4 (1.1)	2.7 (1.4)	.001	.61	1.36	80	.25	.31
Shape concern	4.5 (1.0)	3.5 (1.6)	<.001	.76	4.7 (0.8)	4.2 (1.1)	.004	.52	4.13	77	.046	.60
Weight concern	4.1 (0.9)	3.1 (1.4)	<.001	.83	4.3 (0.8)	3.9 (0.9)	.02	.39	6.67	78	.01	.77
EDNOS subgroup (n=85)												
Eating disorder psychopathology [b]	3.2 (0.8)	2.5 (1.1)	<.001	.76	2.9 (0.9)	2.5 (1.0)	.002	.39	3.31	78	.07	.38
Restraint	2.6 (1.2)	2.3 (1.3)	.16	.26	2.0 (1.3)	1.7 (1.2)	.11	.28	0.02	80	.90	.03
Eating concern	2.2 (1.1)	1.4 (1.2)	<.001	.70	2.2 (1.5)	1.8 (1.4)	.04	.25	3.41	79	.07	.35

	Web-based CBT (n=108)				WL (n=106)				Interaction effect of group x time[a]			
	Baseline	Posttest	*P*	*d*	Baseline	Posttest	*P*	*d*	F	df	*P*	*d*
Shape concern	4.1 (1.0)	3.2 (1.4)	<.001	.72	3.8 (1.1)	3.4 (1.4)	.02	.32	3.27	78	.07	.44
Weight concern	3.8 (1.1)	3.0 (1.3)	<.001	.73	3.6 (1.0)	3.2 (1.2)	.005	.37	5.01	78	.03	.44

[a] Treatment outcomes were measured with repeated measures and mixed model analysis. Effect sizes were measured with Cohen's *d*.

[b] Eating Disorder Examination-Questionnaire.

[c] Body Attitude Test.

[d] Total score of Maudsley Addiction Profile-Health Symptom Scale (MAP-HSS) and 15 additional eating disorder-specific physical complaints.

[e] Depression Anxiety Stress Scale.

[f] Rosenberg Self-Esteem Scale.

[g] EuroQol visual analogue scale.

[h] Measurements in the Addictions for Triage and Evaluation – International Classification of Functioning, Disability and Health.

Efficacy of Web-Based Treatment for Specific Eating Disorder Subgroups

Among participants of the Web-based CBT, eating disorder psychopathology significantly improved over time for all subgroups (data presented in Table 2). However, participants in the control group also improved, and significant differences in effects between the Web-based CBT and WL on primary outcome were found only in the BED subgroup.

On secondary outcome measures, Web-based CBT participants in all three subgroups improved significantly over time with regard to physical health, mental health, self-esteem, quality of life, and social functioning, with small to moderate effect sizes (data presented in Multimedia Appendix 3). Furthermore, BED and EDNOS participants of the Web-based CBT group also improved significantly regarding body dissatisfaction. In the WL group, no significant time effects were found on the secondary outcome measures with the exception of self-esteem for participants with BN and body dissatisfaction and mental health for participants with EDNOS. However, significant between-group differences (Web-based CBT and WL) within the three subgroups were found only for body dissatisfaction and physical health in BED participants and for body dissatisfaction and mental health in EDNOS participants.

Treatment Acceptability

In general, participants were satisfied with the Web-based CBT intervention and their therapist. Most participants evaluated the intervention as rather (46%, 42/91) or very (35%, 32/91) useful, and according to the participants the intervention was especially effective for their eating behavior. Four out of five participants (79%, 72/91) indicated that the treatment had resulted in a healthier diet in their daily lives. Furthermore, for several participants the treatment also led to improvement of mental health (56%, 51/91), self-esteem (49%, 45/91), physical health (47%, 43/91), body image (46%, 42/91), and exercise habits (45%, 41/91). On a scale from 0-10, participants evaluated the intervention with a 7.6 (SD 1.3) and their therapist with an 8.1 (SD 1.0). The majority of participants considered the online contact to be (very) pleasant (77%, 70/91), personal (60%, 55/91), and safe (92%, 84/91). Almost all participants evaluated the support of the therapist to be of added value. For participants who completed the intervention, the therapeutic support was one of the most valuable and important components of the treatment. However, some participants had missed other forms of contact (eg, face-to-face or via telephone) a little (33%, 30/91), quite a lot (5%, 5/91), or very much (8%, 7/91). The participants who did not complete the intervention often mentioned several reasons for dropping out or stopping the intervention prematurely, which can be divided into three main categories: (1) personal reasons or problems (33%, 8/24; eg, lack of time, psychological problems, lack of motivation), treatment content/protocol (29%, 7/24; eg, eating diary annoying/too time consuming, assignments not supportive, not enough attention for weight loss), and the online method (21%, 5/24; eg, lack of personal contact, too open-ended). Furthermore, two participants were discharged by the therapist due to the seriousness of their problems and referred to a more appropriate treatment, and one participant stopped because her GP and psychologist considered the intervention not suitable and had reported her to an outpatient mental health facility for face-to-face treatment.

Discussion

Principal Results and Comparison With Prior Work

Our study shows that Web-based CBT is effective in reducing eating disorder psychopathology in participants with eating disorders in comparison with a waiting list control group. This finding is consistent with the results of a recent review on Internet-based treatments of eating disorders [10]. Participants of the Web-based CBT reported significant reductions in eating disorder psychopathology and were also less concerned with their eating, shape, and weight after participating in the intervention. Participation in the Web-based CBT also resulted in a significant reduction in body dissatisfaction and an improvement of physical and mental health, self-esteem, and social functioning. However, participants of the control group also improved on almost all eating disorder and health-related outcomes resulting in small to moderate effect sizes for the Web-based CBT on interaction effects. The reasons for the improvements in the control group are not totally clear, but several participants received other professional face-to-face support during the waiting period and this resulted in an improvement of participants' body dissatisfaction. Furthermore, it is possible that the no-reply email messages that participants received during the waiting period activated them to start behavioral change. Possibly the waiting list condition thus better resembled an unguided self-help condition than a no-intervention condition. Additionally, the process of seeking help or knowing that treatment would start shortly could also have contributed to the improvement, as some other studies showed similar results [39,40].

For BMI, no effects were found. Most participants were overweight at baseline and therefore improvement of BMI would be desirable. It would be interesting to use the follow-up data to investigate whether the BMI of participants will improve in the long term.

In comparison to the results of our pilot study [18], similar significant effects were found, but with somewhat lower effect sizes for the Web-based CBT. For eating disorder psychopathology, the effect size was d=1.14 in the pilot study and d=.82 in the current trial. This difference is most likely because the pilot results included only posttest data of participants who had completed the intervention, whereas the data of this study also included posttest data of treatment non-completers (33% of Web-based CBT group). As the content and protocol of the intervention, and the online method were important reasons for dropping out or stopping the intervention prematurely, it is likely that the results of the intervention are less positive for treatment non-completers than for treatment completers. For future research, it would be interesting to compare the results of treatment completers and treatment non-completers in terms of efficacy and acceptability of the intervention.

The treatment adherence of 67% in this study was remarkably higher than in our pilot study (54%) [18]. A possible explanation could be that the higher threshold to participate in the current study (because of the randomized design, GP referral, and informed consent) has resulted in selection bias. It is rather difficult to compare the treatment adherence of our Web-based CBT intervention to the adherence of other Web-based interventions, as the definition of adherence is quite diverse in the different studies [10]. However, based on the results of a systematic review on other Web-based interventions for patients with eating disorders [10], the compliance rate of 67% with participants completing all treatment modules of the intervention, can be considered as moderate to high. Compared to the compliance rate of 50% found in a systematic review on adherence in Web-based interventions focused on broader health issues [41], the treatment adherence in our study was significantly higher.

In addition to other studies, we compared the effects of the Web-based CBT between patients with different eating disorder diagnoses (BN, BED, and EDNOS). Web-based CBT was primarily effective for participants with BED. For this subgroup, significant interaction effects were found for eating disorder psychopathology, body dissatisfaction, and mental health, with medium effect sizes. According to the results of a recent review [10], the improvement in eating disorder psychopathology among BED participants in our study is similar to the results found for participants with BED participating in a 6-month self-help treatment with weekly therapist support [42]. For the EDNOS subgroup, participating in the Web-based CBT did not result in a significant interaction effect on eating disorder psychopathology, although the within-group effect size was rather high (P<.001, d=.71). As participants' body dissatisfaction and mental health did improve significantly, and also high within-group effects were found for the other outcome measures, the intervention was partly effective for this subgroup as well. Because EDNOS is not a homogeneous group, further research should elucidate whether the intervention may be more or less effective for specific subgroups of patients with an EDNOS diagnosis. For example, it would be interesting to investigate the results of the intervention among the participants of the EDNOS category who did meet a specified eating disorder diagnosis using the DSM-5 criteria. This was applicable for 29% of all EDNOS participants since 11 participants of the Web-based CBT group (26%) and 8 participants of the WL group (19%) met the DSM-5 criteria of BED, 2 participants of the Web-based CBT group (5%) met the DSM-5 criteria of AN, and 3 participants of the WL group (7%) met the DSM-5 criteria of BN. Though participants with BN improved during treatment with a high within-group effect size for eating disorder psychopathology and small to moderate within-group effect sizes for most secondary outcome measures, no significant interaction effects were found for any outcome measure. This could be explained by the smaller sample size of this subgroup (n=44). However, other explanations are possible, for example, that participants with BN need a more intensive (face-to-face) treatment as results showed that these subgroups had more severe eating disorder psychopathology and related health problems at baseline than the other subgroups. Therefore, it would be interesting to further evaluate the results of the intervention for this specific subgroup.

Participants were generally satisfied with the Web-based CBT intervention, and the support of the therapist was considered as very valuable and important. Several methods were used to ensure quality and consistency in the treatment of participants (eg, formats for each intervention module, intensive training, and supervision of therapists, and retrospective control of therapists' messages). Therefore we expect only minor differences in the support provided by the therapists, equivalent to differences between therapists in clinical face-to-face treatment. However, differences in therapeutic support possibly resulted in differences between therapists in participant outcome. As the 17 therapists were not stratified by type of eating disorder, and the number of participants assigned to each therapist differed significantly with a minimum of 2 and a maximum of 31 participants, this might have affected the results of the intervention. Therefore, it would be interesting to investigate the consistency in the support of the therapists, their adherence to the protocol, and their competencies, for example, by conducting a study including directive and summative content analysis of the treatment of several participants of the current study. Furthermore, additional research into the differences between therapists in participant outcome would be very valuable.

Because the effectiveness of the Web-based CBT was investigated within a naturalistic setting, results of this study are likely to approximate those of the Web-based intervention in everyday practice. Another strength of our study is the low study dropout, as posttest data were available for 94% of the participants.

Limitations

Although the study shows promising results, these should be considered in the context of several limitations. The first and most important limitation is that, despite extending the sampling period, we were unable to recruit the sample size of 84 participants with BN and therefore also did not reach the planned total sample size of 252 participants. Although the total sample size of 252 participants was not achieved, significant time and interaction effects for the overall group were found for almost all outcome measures. For participants with BN on the other hand, only significant time effects were found; there were no significant interaction effects.

Second, the number of participants per diagnostic category was low. Sample sizes were calculated based on the results of our pilot study. However, we did not sufficiently take into account the differences in effectiveness between the subgroups and the improvements among participants of the control group, resulting in a low number of participants per diagnostic category.

Third, eating disorder diagnoses were not assessed using an in-person clinical interview as required in formal diagnoses, but by using online self-report questions including the EDQ-O. As an in-person interview would probably increase the threshold to participate in the Web-based CBT intervention, we found this incompatible with the main objective of this treatment. Therefore, we decided to use only self-report assessments to measure all primary and secondary outcomes. A recent study [35] showed acceptable performance of the EDQ-O as a diagnostic instrument for all DSM-IV eating disorder classifications, although improvement was desirable. However, this study did not use the EDQ-O as the only tool to assess eating disorder diagnoses. Also, other baseline data were taken into consideration, and if necessary, additional questions were asked by email. Nevertheless, the lack of any personal interviews is an important limitation of this study. For future research, this topic should be considered carefully, weighing pros (validity) and cons (excluding patients).

Conclusions

Eating disorders have a considerable impact on the quality of life and psychological and physical health of patients. The participants in this study suffered from BN, BED, and EDNOS for several years, and more than half of the participants had never had treatment for their eating disorder. The results of this study provide support for the use of a Web-based CBT intervention to improve eating disorder psychopathology, body dissatisfaction, and related health among patients with eating disorders. Furthermore, our findings confirm that new technologies can be used to treat patients who otherwise refrain from seeking or receiving help.

Acknowledgments

This study received no specific grant from any funding agency, or commercial or not-for-profit sectors. We would like to thank all participants and therapists for their contribution to this study.

Conflicts of Interest

EH and HH participated in the development of the Web-based CBT intervention, but they did not derive financial benefit from the intervention.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [42].

[PDF File (Adobe PDF File), 146KB - jmir_v17i6e152_app1.pdf]

Multimedia Appendix 2

Content of Web-based CBT intervention.

[PDF File (Adobe PDF File), 48KB - jmir_v17i6e152_app2.pdf]

Multimedia Appendix 3

Treatment results on secondary outcome measures for participants of the subgroups BN, BED, and EDNOS.

[PDF File (Adobe PDF File), 79KB - jmir_v17i6e152_app3.pdf]

References

1. Preti A, Girolamo GD, Vilagut G, Alonso J, Graaf RD, Bruffaerts R, et al. The epidemiology of eating disorders in six European countries: results of the ESEMeD-WMH project. J Psychiatr Res 2009 Sep;43(14):1125-1132. [doi: 10.1016/j.jpsychires.2009.04.003] [Medline: 19427647]

2. Banasiak SJ, Paxton SJ, Hay PJ. Evaluating Accessible Treatments for Bulimic Eating Disorders in Primary Care. Aust. J. Prim. Health 1998;4(3):147. [doi: 10.1071/PY98042]
3. Becker AE, Hadley AA, Perloe A, Fay K, Striegel-Moore RH. A qualitative study of perceived social barriers to care for eating disorders: perspectives from ethnically diverse health care consumers. Int J Eat Disord 2010 Nov 1;43(7):633-647 [FREE Full text] [doi: 10.1002/eat.20755] [Medline: 19806607]
4. Cachelin FM, Striegel-Moore RH. Help seeking and barriers to treatment in a community sample of Mexican American and European American women with eating disorders. Int J Eat Disord 2006 Mar;39(2):154-161. [doi: 10.1002/eat.20213] [Medline: 16252278]
5. Evans EJ, Hay PJ, Mond J, Paxton SJ, Quirk F, Rodgers B, et al. Barriers to help-seeking in young women with eating disorders: a qualitative exploration in a longitudinal community survey. Eat Disord 2011;19(3):270-285. [doi: 10.1080/10640266.2011.566152] [Medline: 21516551]
6. Schmidt U. Getting technical. Eur Eat Disorders Rev 2003 May;11(3):147-154. [doi: 10.1002/erv.522]
7. 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]
8. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
9. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):178-186. [doi: 10.1016/j.cpr.2010.10.002] [Medline: 21095051]
10. Aardoom JJ, Dingemans AE, Spinhoven P, Van Furth EF. Treating eating disorders over the internet: a systematic review and future research directions. Int J Eat Disord 2013 Sep;46(6):539-552. [doi: 10.1002/eat.22135] [Medline: 23674367]
11. Smink FR, van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr Psychiatry Rep 2012 Aug;14(4):406-414 [FREE Full text] [doi: 10.1007/s11920-012-0282-y] [Medline: 22644309]
12. Jacobi C, Völker U, Trockel MT, Taylor CB. Effects of an Internet-based intervention for subthreshold eating disorders: a randomized controlled trial. Behav Res Ther 2012 Feb;50(2):93-99. [doi: 10.1016/j.brat.2011.09.013] [Medline: 22137366]
13. Beintner I, Jacobi C, Taylor CB. Effects of an Internet-based prevention programme for eating disorders in the USA and Germany--a meta-analytic review. Eur Eat Disord Rev 2012 Jan;20(1):1-8. [doi: 10.1002/erv.1130] [Medline: 21796737]
14. Moessner M, Bauer S. Online counselling for eating disorders: reaching an underserved population? J Ment Health 2012 Aug;21(4):336-345. [doi: 10.3109/09638237.2011.643512] [Medline: 22251091]
15. ter Huurne ED, Postel MG, DeJong CAJ. Web-based intensive therapeutic contact for eating disorders. Psychiatr Serv 2013 Jul 1;64(7):711. [doi: 10.1176/appi.ps.640109] [Medline: 23821172]
16. Postel MG, de Haan HA, ter Huurne ED, Becker ES, DeJong CAJ. 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]
17. Postel MG, DeJong CAJ, de Haan HA. Does e-therapy for problem drinking reach hidden populations? Am J Psychiatry 2005 Dec;162(12):2393. [doi: 10.1176/appi.ajp.162.12.2393] [Medline: 16330613]
18. ter Huurne ED, Postel MG, de Haan HA, Drossaert CHC, DeJong CAJ. Web-based treatment program using intensive therapeutic contact for patients with eating disorders: before-after study. J Med Internet Res 2013;15(2):e12 [FREE Full text] [doi: 10.2196/jmir.2211] [Medline: 23380291]
19. Tactus Addiction Treatment. Web-based CBT intervention Etendebaas. URL: http://www.etendebaas.nl [accessed 2015-06-15] [WebCite Cache ID 6ZIh0pHmx]
20. ter Huurne ED, Postel MG, de Haan HA, DeJong CAJ. Effectiveness of a web-based treatment program using intensive therapeutic support for female patients with bulimia nervosa, binge eating disorder and eating disorders not otherwise specified: study protocol of a randomized controlled trial. BMC Psychiatry 2013;13:310 [FREE Full text] [doi: 10.1186/1471-244X-13-310] [Medline: 24238630]
21. Eysenbach G, Consort- E. 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]
22. Brownley KA, Berkman ND, Sedway JA, Lohr KN, Bulik CM. Binge eating disorder treatment: a systematic review of randomized controlled trials. Int J Eat Disord 2007 May;40(4):337-348. [doi: 10.1002/eat.20370] [Medline: 17370289]
23. Shapiro JR, Berkman ND, Brownley KA, Sedway JA, Lohr KN, Bulik CM. Bulimia nervosa treatment: a systematic review of randomized controlled trials. Int J Eat Disord 2007 May;40(4):321-336. [doi: 10.1002/eat.20372] [Medline: 17370288]
24. Wilson GT, Grilo CM, Vitousek KM. Psychological treatment of eating disorders. Am Psychol 2007 Apr;62(3):199-216. [doi: 10.1037/0003-066X.62.3.199] [Medline: 17469898]
25. Britt E, Hudson SM, Blampied NM. Motivational interviewing in health settings: a review. Patient Educ Couns 2004 May;53(2):147-155. [doi: 10.1016/S0738-3991(03)00141-1] [Medline: 15140454]
26. Miller JH, Moyers T. Motivational interviewing in substance abuse: applications for occupational medicine. Occup Med 2002;17(1):51-65, iv. [Medline: 11726336]

27. Fairburn CG, Beglin SJ. Assessment of eating disorders: interview or self-report questionnaire? Int J Eat Disord 1994 Dec;16(4):363-370. [Medline: 7866415]
28. Probst M, Van Coppenolle H, Vandereycken W. Further experience with the Body Attitude Test. Eat Weight Disord 1997 Jun;2(2):100-104. [Medline: 14655849]
29. Probst M, Vandereycken W, Van Coppenolle H, Vanderlinden J. The Body Attitude Test for Patients with an Eating Disorder: Psychometric Characteristics of a New Questionnaire. Eating Disorders 1995 Jan;3(2):133-144. [doi: 10.1080/10640269508249156]
30. Marsden J, Gossop M, Stewart D, Best D, Farrell M, Lehmann P, et al. The Maudsley Addiction Profile (MAP): a brief instrument for assessing treatment outcome. Addiction 1998 Dec;93(12):1857-1867. [Medline: 9926574]
31. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [Medline: 7726811]
32. Rosenberg M. Society and the adolescent self-image. Princeton, NJ: Princeton University Press; 1965.
33. 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]
34. Schippers GM, Broekman TG, Buchholz A, Koeter MW, van den Brink W. Measurements in the Addictions for Triage and Evaluation (MATE): an instrument based on the World Health Organization family of international classifications. Addiction 2010 May;105(5):862-871. [doi: 10.1111/j.1360-0443.2009.02889.x] [Medline: 20331557]
35. ter Huurne ED, de Haan HA, ten Napel-Schutz MC, Postel MG, Menting J, van der Palen J, et al. Is the Eating Disorder Questionnaire-Online (EDQ-O) a valid diagnostic instrument for the DSM-IV-TR classification of eating disorders? Compr Psychiatry 2015 Feb;57:167-176. [doi: 10.1016/j.comppsych.2014.10.019] [Medline: 25464837]
36. Sheehan D, Lecrubier Y, Sheehan K, 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 34-57. [Medline: 9881538]
37. 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]
38. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
39. Ruwaard J, Lange A, Broeksteeg J, Renteria-Agirre A, Schrieken B, Dolan CV, et al. Online cognitive-behavioural treatment of bulimic symptoms: a randomized controlled trial. Clin Psychol Psychother 2013;20(4):308-318. [doi: 10.1002/cpp.1767] [Medline: 22298417]
40. Sánchez-Ortiz VC, Munro C, Stahl D, House J, Startup H, Treasure J, et al. A randomized controlled trial of internet-based cognitive-behavioural therapy for bulimia nervosa or related disorders in a student population. Psychol Med 2011 Feb;41(2):407-417. [doi: 10.1017/S0033291710000711] [Medline: 20406523]
41. 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]
42. Carrard I, Crépin C, Rouget P, Lam T, Golay A, Van der Linden M. Randomised controlled trial of a guided self-help treatment on the Internet for binge eating disorder. Behav Res Ther 2011 Aug;49(8):482-491. [doi: 10.1016/j.brat.2011.05.004] [Medline: 21641580]

Abbreviations

BAT: Body Attitude Test
BED: binge-eating disorder
BMI: body mass index
BN: bulimia nervosa
CBT: cognitive behavioral therapy
DASS: Depression Anxiety Stress Scale
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th revision
EDE-Q: Eating Disorder Examination-Questionnaire
EDNOS: eating disorder not otherwise specified
EDQ-O: Eating Disorder Questionnaire-Online
EQ-5D VAS: EuroQol-5D visual analogue scale
GP: general practitioner
MAP-HSS: Maudsley Addiction Profile-Health Symptom Scale
MATE-ICN: Measurements in the Addictions for Triage and Evaluation – International Classification of Functioning, Disability and Health
MINI-Plus: Mini International Neuropsychiatric Interview-Plus
NTR: Nederlands Trial register

RCT: randomized controlled trial
RSES: Rosenberg Self-Esteem Scale

Edited by G Eysenbach; submitted 14.10.14; peer-reviewed by M Jones, J Aardoom; comments to author 30.10.14; revised version received 27.01.15; accepted 28.04.15; published 18.06.15.

Please cite as:
ter Huurne ED, de Haan HA, Postel MG, van der Palen J, VanDerNagel JEL, DeJong CAJ
Web-Based Cognitive Behavioral Therapy for Female Patients With Eating Disorders: Randomized Controlled Trial
J Med Internet Res 2015;17(6):e152
URL: http://www.jmir.org/2015/6/e152/
doi:10.2196/jmir.3946
PMID:26088580

Original Paper

A Mobile Health Intervention Supporting Heart Failure Patients and Their Informal Caregivers: A Randomized Comparative Effectiveness Trial

John D Piette[1], PhD; Dana Striplin[1], MHSA; Nicolle Marinec[1], MPH; Jenny Chen[1], MPH; Ranak B Trivedi[2], PhD; David C Aron[3], MD, MS; Lawrence Fisher[4], PhD; James E Aikens[5], PhD

[1]Center for Clinical Management Research and Center for Managing Chronic Disease, VA Ann Arbor Healthcare System and University of Michigan School of Public Health, Ann Arbor, MI, United States

[2]Center for Innovation to Implementation and Department of Psychiatry and Behavioral Sciences, VA Palo Alto Health Care System and Stanford University, Palo Alto, CA, United States

[3]Louis Stokes VA Cleveland Medical Center and Department of Medicine, Case Western Reserve University, Cleveland, OH, United States

[4]Department of Family and Community Medicine, University of California-San Francisco, San Francisco, CA, United States

[5]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
John D Piette, PhD
Center for Clinical Management Research and Center for Managing Chronic Disease
VA Ann Arbor Healthcare System and University of Michigan School of Public Health
1415 Washington Heights, SPH-1
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 1457
Fax: 1 734 763 9115
Email: jpiette@umich.edu

Abstract

Background: Mobile health (mHealth) interventions may improve heart failure (HF) self-care, but standard models do not address informal caregivers' needs for information about the patient's status or how the caregiver can help.

Objective: We evaluated mHealth support for caregivers of HF patients over and above the impact of a standard mHealth approach.

Methods: We identified 331 HF patients from Department of Veterans Affairs outpatient clinics. All patients identified a "CarePartner" outside their household. Patients randomized to "standard mHealth" (n=165) received 12 months of weekly interactive voice response (IVR) calls including questions about their health and self-management. Based on patients' responses, they received tailored self-management advice, and their clinical team received structured fax alerts regarding serious health concerns. Patients randomized to "mHealth+CP" (n=166) received an identical intervention, but with automated emails sent to their CarePartner after each IVR call, including feedback about the patient's status and suggestions for how the CarePartner could support disease care. Self-care and symptoms were measured via 6- and 12-month telephone surveys with a research associate. Self-care and symptom data also were collected through the weekly IVR assessments.

Results: Participants were on average 67.8 years of age, 99% were male (329/331), 77% where white (255/331), and 59% were married (195/331). During 15,709 call-weeks of attempted IVR assessments, patients completed 90% of their calls with no difference in completion rates between arms. At both endpoints, composite quality of life scores were similar across arms. However, more mHealth+CP patients reported taking medications as prescribed at 6 months (8.8% more, 95% CI 1.2-16.5, *P*=.02) and 12 months (13.8% more, CI 3.7-23.8, *P*<.01), and 10.2% more mHealth+CP patients reported talking with their CarePartner at least twice per week at the 6-month follow-up (*P*=.048). mHealth+CP patients were less likely to report negative emotions during those interactions at both endpoints (both *P*<.05), were consistently more likely to report taking medications as prescribed during weekly IVR assessments, and also were less likely to report breathing problems or weight gains (all *P*<.05). Among patients with more depressive symptoms at enrollment, those randomized to mHealth+CP were more likely than standard mHealth patients to report excellent or very good general health during weekly IVR calls.

Conclusions: Compared to a relatively intensive model of IVR monitoring, self-management assistance, and clinician alerts, a model including automated feedback to an informal caregiver outside the household improved HF patients' medication adherence and caregiver communication. mHealth+CP may also decrease patients' risk of HF exacerbations related to shortness of breath and sudden weight gains. mHealth+CP may improve quality of life among patients with greater depressive symptoms. Weekly health and self-care monitoring via mHealth tools may identify intervention effects in mHealth trials that go undetected using typical, infrequent retrospective surveys.

Trial Registration: ClinicalTrials.gov NCT00555360; https://clinicaltrials.gov/ct2/show/NCT00555360 (Archived by WebCite at http://www.webcitation.org/6Z4Tsk78B).

(J Med Internet Res 2015;17(6):e142) doi:10.2196/jmir.4550

KEYWORDS

telehealth; mobile health; heart failure; disease management; self-management

Introduction

Chronic heart failure (HF) is associated with reduced quality of life, preventable hospitalizations, and early mortality [1,2]. For effective disease management, patients must systematically monitor symptoms, including shortness of breath, weight gain, and edema, and follow strict self-care practices including limiting salt and fluid intake, and taking medications as prescribed [3,4]. Because HF management is challenging, patients frequently experience life-threatening exacerbations that are responsible for $40 billion in US health care costs each year [5]. Telephone care management can improve HF patients' prognosis [6-10]. However, telephone follow-up is inadequately reimbursed and competes with in-person care for clinicians' time [11].

A number of recently completed clinical trials and evidence syntheses have shown that mobile health (mHealth) interventions can improve self-care behaviors and physiologic risk factors for poor outcomes of cardiovascular disease, including heart failure [12-16]. For example, risk factor management using interactive voice response (IVR) calls can improve dietary behaviors and blood pressure control among hypertension patients in the United States and Latin America [17-19], and remote monitoring coupled with self-management assistance has been shown to improve outcomes of cardiovascular disease in a number of countries [20-25]. Despite these encouraging findings, not all trials of HF self-management support via mHealth tools have shown positive outcomes [26,27]. Without substantial restructuring of financial incentives for health care organizations and systems to follow up on identified problems, increased monitoring may be insufficient to fill the gap between what HF patients need and what health systems can provide [28,29].

One potential solution to bridging the gap between the promise and the practice of mHealth self-management assistance may be to expand the reach of interventions so that they support not only patients but also their informal caregivers. Informal caregivers often help chronically ill patients follow self-management recommendations by providing support that is unavailable through professional care management [30-33]. However, in-home caregivers are often elderly, ill, and overwhelmed [34,35]. Most in-home caregivers lack the training and resources needed to systematically monitor HF patients and provide self-management assistance. Moreover, chronically ill patients increasingly have caregivers outside of the household, making health and self-care monitoring much more difficult [36,37].

The CarePartner program was developed through a series of Veterans Affairs (VA) and non-VA pilot and feasibility studies to address these challenges by enabling structured support by informal caregivers (CarePartners) who reside outside the patient's home. Through this program, patients receive regular monitoring and tailored self-management education via IVR calls with feedback to their clinician. While evidence suggests that between-visit mHealth assistance could be effective in improving HF self-care and outcomes, it remains unclear whether feedback to CarePartners is helpful over and above the support provided directly to patients and clinicians.

This study reports the results of a randomized comparative effectiveness trial testing the impact of systematic feedback to patients' CarePartners, compared to patients receiving standard mHealth monitoring and self-management education. Analyses focused on changes in patients' HF-related quality of life, self-care, and patient-CarePartner communication reported via 6- and 12-month surveys, as well as on patients' medication adherence and symptoms reported via weekly IVR calls.

CarePartners also completed surveys at 6- and 12-months post enrollment. The primary results of those assessments are presented elsewhere [38]. In brief, CarePartners who experienced significant caregiving strain and depression at baseline experienced significant decreases in those symptoms if randomized to receive systematic feedback about their patient-partner's health and self-care, and also reported increased engagement in self-management support. In order to provide additional information about the intervention experience from the CarePartners' perspective, here we briefly describe qualitative feedback from CarePartners at follow-up as well as their unsolicited replies to email reports sent automatically based on patients' IVR assessment calls.

Methods

Recruitment

Patients were recruited from VA Cleveland Medical Center outpatient clinics between June 2009 and January 2012 and were followed for 12 months. To be eligible, patients had to have an HF diagnosis, New York Heart Association classification of II or III, and a documented ejection fraction <40% (see Multimedia Appendix 1). Patients also had to have attended at least one VA outpatient visit within the previous 12 months, have a VA primary care provider, and be able to participate in automated telephone calls in English. Patients needed to nominate an eligible CarePartner, that is, a relative or friend living outside their home. Patients were excluded if they lived in a skilled nursing facility; were prescribed oxygen supplementation; were receiving palliative care; had a life-threating condition such as lung cancer; or had ICD-9 coded diagnoses indicating dementia, bipolar disorder, or schizophrenia.

Potentially eligible patients identified from electronic medical records were sent an invitation letter, followed by a screening and recruitment call. Eligible and interested patients were mailed informed consent forms and were assisted in identifying potential CarePartners using the Norbeck Social Support Questionnaire (NSSQ) [39]. To be eligible, CarePartners had to live outside the patient’s home, speak English, have access to a telephone and email, and report at least monthly contact with the patient. CarePartners provided verbal consent to participate.

Randomization

After completing baseline surveys, patient-CarePartner dyads were randomized by a research associate to a patient-focused mHealth service (standard mHealth) or a service that included feedback to patients’ CarePartners (mHealth+CP). Pairs were randomized within strata defined by whether the patient had an in-home caregiver. Sealed randomization envelopes were created by the study coordinator in blocks using an online random number generator. It was impossible to blind patients to their random assignment because patients and CarePartners were aware whether the CarePartner received email feedback.

Standard mHealth Intervention

Patients, CarePartners, and in-home caregivers (when present) randomized to standard mHealth were mailed information about HF self-care [37]. Patients received weekly IVR monitoring and self-management support calls for 12 months. Up to nine call attempts per week were made at times the patient indicated were convenient. IVR calls included recorded information and questions that patients answered using their touchtone keypad. The IVR calls were developed by a panel including primary care physicians, cardiologists, nurses, and experts in health behavior change and mHealth. Calls lasted roughly 10 minutes and followed a tree-structured algorithm to ask about overall health, HF symptoms, and self-management behaviors. Patients received pre-recorded information tailored to their reported symptoms and self-care practices. See Figures 1 and 2 for screenshots of the website used for enrollment and call scheduling.

When patients reported an urgent issue via IVR (ie, worsening shortness of breath or a significant weight increase), the system automatically issued a fax notification to their clinician. A significant weight increase was defined as a 5-lb increase over 1 or 2 weeks, a 7-lb increase over 3 weeks, or an average gain of 2 lbs per week since the last automated call if more than 3 weeks had elapsed. Actions taken by clinicians based on the faxes were not tracked.

Figure 1. Patient enrollment page.

Select an option from the drop-down menu to either add a new Patient or edit an existing Patient. Use the search box to skip directly to a selection -- search by first name, last name, or study id. Fields denoted with an asterisk(*) are required.

<< New Patient >> Search:
*Study ID: *Category:
*First Name: *Last Name:
*Gender: ● Male ○ Female *Birthyear (yyyy):
*Phone Number (primary):
Phone Number (secondary): Email Address:

Care Partner: Select Partner Search:
Care Manager: Select Manager Receive Fax: ● Yes ○ No
Doctor: Select Doctor Receive Fax: ● Yes ○ No
Send Shortness of Breath Fax: ● Yes ○ No
Send Weight Gain Fax: ● Yes ○ No
*Time Zone: Select Time Zone *Daylight Savings: ● Yes ○ No
*Withdrawn: ○ Yes ● No

Comment:

Figure 2. Call scheduling page.

Select an option from the drop-down menu to adjust the schedule of Patient or Care Partner. Please provide the information in the required fields.

When selecting call days, the priority will be determined by the Call Begin Date. If the Call Begin Date is a Wednesday, a Friday call schedule will take priority over Tuesday. Select an appropriate Call Begin Date to ensure priority scheduling.

*Phone Number: Not Used

Number of Weeks for Calls: 52

To suspend calls to this Patient enter the dates to begin and end the suspension period.

mHealth+CP Intervention

The mHealth+CP intervention was based on self-regulation theory, which emphasizes communication of expectations of behavior (“standards”), promotion of motivation to meet standards, and monitoring with feedback regarding the gap between behavior and standards [40,41]. Patients and CarePartners randomized to mHealth+CP received identical intervention elements described above.

mHealth+CP CarePartners were automatically emailed a structured report after each completed IVR call. CarePartner reports were sent to their personal, individual email addresses, which were stored in the system’s secure database at the University of Michigan. Reports described in lay language what patients’ responses meant in terms of risk for HF exacerbations and included suggestions for how CarePartners could support self-management. Email reports referred to the patient using gender-specific pronouns, for example, “Your partner did not weigh himself last week”, but were otherwise de-identified. Reports included feedback about the patient’s most recent issues as reported during their IVR call, including shortness of breath, medication adherence, salt, and fluid intake, and increases in weight. CarePartners were asked to call their patient-partner weekly to review the reports and address identified problems.

CarePartners received guidelines about how to communicate in a positive motivating way, avoid conflict by respecting boundaries, include in-home caregivers, and respect confidentiality. Patients received a notebook including reminders and tips for their weekly patient-CarePartner calls. CarePartners received logbooks for tracking IVR reports, upcoming patient contacts, clinical encounters, and medication refills.

Measurement

Baseline, 6-Month, and 12-Month Surveys

Patients' HF-specific quality of life, self-care, and patient-CarePartner communication were measured via quantitative telephone surveys. Baseline sociodemographic variables included patients' age, race, marital status, employment status, educational attainment, and income. Patients' baseline depressive symptoms were measured using the 10-item version of the CES-D [42]. CarePartners completed online surveys at each time point; the current analyses include baseline CarePartner characteristics relevant to the comparability of groups at the time of randomization, and qualitative feedback from CarePartners' 12-month surveys.

The primary outcome was HF-specific quality of life at 12 months, as measured by the Minnesota Living with Heart Failure Questionnaire (MLHFQ) [43]. HF self-care behaviors were measured using the Revised Heart Failure Self-Care Behavior Scale (HFSCB) [3]. A measure of HF medication adherence was created using the HFSCB adherence items with which patients reported how often they "took [their] pills every day", "took [their] pills as the doctor prescribed, ie, took all of the doses of [their] pills", "always refilled prescriptions for [their] pills on time", and "had a system to help tell [them] when to take [their] pills". The adherence measure based on these items was designed to identify patients reporting perfect adherence (ie, a binary measure identifying patients reporting "always" engaging in all four behaviors). Binary indicators for perfect adherence tend to correct for inflated adherence reporting [44,45].

To identify changes in patient-CarePartner communication, three relationship dimensions were measured at each time point. First, as an objective measure of communication intensity, patients were asked how often over the prior 6 months they communicated with their CarePartner by phone. Analyses examined patients' likelihood of reporting that they spoke with their CarePartner at least twice per week. Second, the affective dimension of CarePartner support was measured using items based on prior studies of caregiving relationships [46,47]. Patients were asked how often they experienced each of six negative emotions when talking with their CarePartner (sadness, loneliness, anger, tension, guilt, or frustration), and analyses examined patients' likelihood of reporting that they regularly experience one or more of these emotions. Third, to understand patients' perspective of the difficulty involved in CarePartner communication, analyses examined participants' likelihood of agreeing or strongly agreeing that it was "difficult to talk with [their] CarePartner about [their] illness".

Weekly Interactive Voice Response Adherence and Symptom Reports

Patients' IVR medication adherence and symptom reports were examined as potential indicators of differences across arms in intervention effectiveness, because short-term reporting intervals often provide information that is more reliable and less prone to bias than retrospective recall surveys [48-50]. Patients were considered adherent if they reported "always" taking their HF medication exactly as prescribed in the past week. Patients were classified as experiencing shortness of breath if they reported being bothered by shortness of breath "daily" or "several days" in the prior week. Patients were coded as having a significant weight gain if their reported weight met criteria described above. Finally, patients were coded as having positive self-reported health if they reported that their overall health was "excellent" or "very good".

CarePartner Feedback

Although replies were not solicited to email reports sent to CarePartners based on the patient's IVR feedback, if CarePartners did reply, that message was sent to the study coordinator. Also, in 12-month follow-up interviews, CarePartners were asked an open-ended question regarding what they felt were the strengths of the program. Here we briefly summarize both types of CarePartner feedback and include exhaustive lists of CarePartner comments in Multimedia Appendices 2 and 3.

Statistical Analysis

The sample included all patients with 12-month surveys plus 22 patients for whom 6-month survey data were carried forward. Initial analyses compared the baseline characteristics of patients who did versus did not have 12-month data in the imputed sample. Subsequent analyses compared patients and CarePartners across arms in the sample with outcome data. IVR call completion rates were calculated using one record per week of attempted IVR calls, that is, 52 call-weeks per patient minus weeks in which the patient was on vacation or hospitalized. Logistic models were used to predict patients' likelihood of completing each weekly call as a function of arm, baseline characteristics, and the number of weeks since enrollment. Statistical tests for the analyses of call completion rates were adjusted for clustering of call-weeks within patients.

The primary outcome was change in HF-specific quality of life between baseline and 12 months. The study was powered to detect a medium/small effect (d=.351) assuming a 20% loss to follow-up, similar to that observed in the prior HF trial by Sisk et al [51]. All outcomes were analyzed on an intent-to-treat basis. The xtmixed and logistic regression commands in Stata version 13.1 [52] were used to identify intervention effects on patients' HF-related quality of life, self-care, and patient-CarePartner communication. Predictors for each analysis included an indicator for arm, time (baseline, 6-month, and 12-month), and an arm-by-time interaction. Effect estimates represent differences across arms adjusted for baseline values. To examine differences across arms in IVR reports of medication adherence and symptoms, graphical displays were created illustrating the proportion of patients reporting a given outcome each week, separately by arm. Logistic regression models were fitted to predict patients' weekly IVR-reported outcomes, with weekly reports clustered within patient. Models included the following predictors: arm, time, and an arm-by-time interaction term. Variances for the estimated intervention effects were adjusted for the within-patient correlation of IVR reports across weeks [53-55]. To illustrate the magnitude of intervention-control differences in IVR-reported outcomes, the probability for each outcome at week 26 and 52 was predicted based on the logistic model separately for mHealth+CP and standard mHealth groups.

Among patients with chronic medical problems, depressive symptoms may influence their perceived health status even more than objective symptoms and impairments resulting from their medical condition [56]. Depressed patients often attribute their difficulties to insufficient social support [57,58], and CarePartners' support may have counteracted their tendency to over-generalize health problems [59-61]. To test this hypothesis, we examined a potential interaction between patients' baseline level of depressive symptoms (CES-D) and arm, with respect to IVR reports of excellent/very good health. Specifically, patients' unadjusted frequency of reporting excellent/very good health was examined graphically as described above, within subgroups defined by baseline CES-D scores. Because graphical displays suggested an inflection point with two very different slopes, we fit logistic models separately for patients with CES-D scores that were low (0-4) versus high (5-10). Each model included terms for arm, baseline CES-D score, an arm-by-CES-D interaction, time, and an arm-by-time interaction.

Human Subjects Approval

The study protocol was approved by the Ann Arbor VA Human Subjects Committee, and all patients provided written informed consent. Patients and CarePartners received US $20 for completed surveys; patients did not have financial incentives for completing IVR calls. None of the authors had any financial conflict of interest.

See Multimedia Appendix 4 for the CONSORT-EHEALTH checklist [62].

Results

Recruitment and Baseline Characteristics

A total of 4140 potentially eligible patients were identified from electronic medical records. Of these, 372 were randomized, and 331 (89%) had outcome data at 12 months (see Figure 3). Patients lost to follow-up were less likely to report at baseline that they spoke with their CarePartner at least twice per week (43.9% versus 65.9%, *P*=.006) and had better baseline HF self-care scores as measured by the HFSCB (*P*=.002) but were not significantly different from patients with follow-up data on any other characteristic shown in Table 1 (see also Multimedia Appendix 5).

Patients in both arms had similar baseline characteristics, except that mHealth+CP patients were more likely to have a high school education or less (Table 1). Education was included in outcome analyses as a control variable, although analyses not including education as a covariate produced essentially the same results. There were no significant baseline differences across arms in measures of patient-CarePartner communication or in CarePartner characteristics. As expected in a VA population, most participants were male. Patients were on average 67.8 years of age (SD 10.2), 77.0% (255/331) were white, 48.0% (159/331) had a high school education or less, 32.6% (108/331) lived alone, and 87.6% (290/331) were retired or unemployed. While most patients (65.9%, 218/331) reported at baseline that they talked with their CarePartner by phone at least twice per week over the prior 6 months, 44.8% (147/328) reported regularly experiencing one or more negative emotions during those conversations, and 21.5% (71/331) agreed that it was difficult to talk with their CarePartner about their illness. Compared to patients, CarePartners were younger, more likely to be female, more likely to be employed, and had more years of education. A total of 41.4% (137/331) of CarePartners were the patients' daughters/daughter-in-laws, 20.2% (67/331) were sons/son-in-laws, 11.2% (37/331) were friends, 9.1% (30/331) were sisters/sisters-in-laws, and the remaining 18.1% (60/331) were other family and social network members.

Table 1. Baseline characteristics of the sample.

	Overall (n=331)	Standard mHealth (n=165)	mHealth+CP (n=166)
Patient characteristics			
Age in years, mean (SD)	67.8 (10.2)	68.1 (10.1)	67.6 (10.3)
Male, % (n)	99.4 (329)	98.8 (163)	100.0 (166)
White race, % (n)	77.0 (255)	77.0 (127)	77.1 (128)
Married/Partnered, % (n)	58.9 (195)	61.2 (101)	56.6 (94)
High school or less, % (n)	48.0 (159)	41.8 (69)	54.2 (90)
Live alone, % (n)	32.6 (108)	32.7 (54)	32.5 (54)
Unemployed/retired, % (n)	87.6 (290)	86.1 (142)	89.2 (148)
Income <$15,000, % (n)	31.4 (104)	30.3 (50)	32.5 (54)
CES-D Depression, mean (SD)	3.0 (2.5)	3.0 (2.5)	3.0 (2.5)
MLHFQ[a], mean (SD)	43.3 (25.3)	43.0 (26.4)	48.8 (24.3)
HFSCB[b], mean (SD)	82.8 (17.9)	82.6 (19.2)	83.0 (16.5)
Adherent to HF Rx[c],% (n)	52.3 (173)	50.3 (83)	54.2 (90)
Relationship quality [d] , % (n)			
Talk 2+ times/ week	65.9 (218)	66.1 (109)	65.7 (109)
Negative emotions[e]	44.8 (147)	45.4 (74)	44.2 (73)
Perceived difficulty[f]	21.5 (71)	18.8 (31)	24.1 (40)
CarePartner characteristics			
Age in years, mean (SD)	46.7 (13.2)	47.2 (14.5)	46.2 (11.9)
Male, % (n)	35.0 (116)	32.7 (54)	37.3 (62)
Married/Partnered, % (n)	68.6 (227)	67.3 (111)	69.9 (116)
High school or less, % (n)	27.8 (92)	23.6 (39)	31.9 (53)
Unemployed/retired, % (n)	36.9 (122)	38.8 (64)	34.9 (58)

[a] Minnesota Living with Heart Failure Questionnaire Scores. Lower scores indicate better functioning.

[b] Revised Heart Failure Self-Care Behavior Scale. Higher scores indicate better HF self-care.

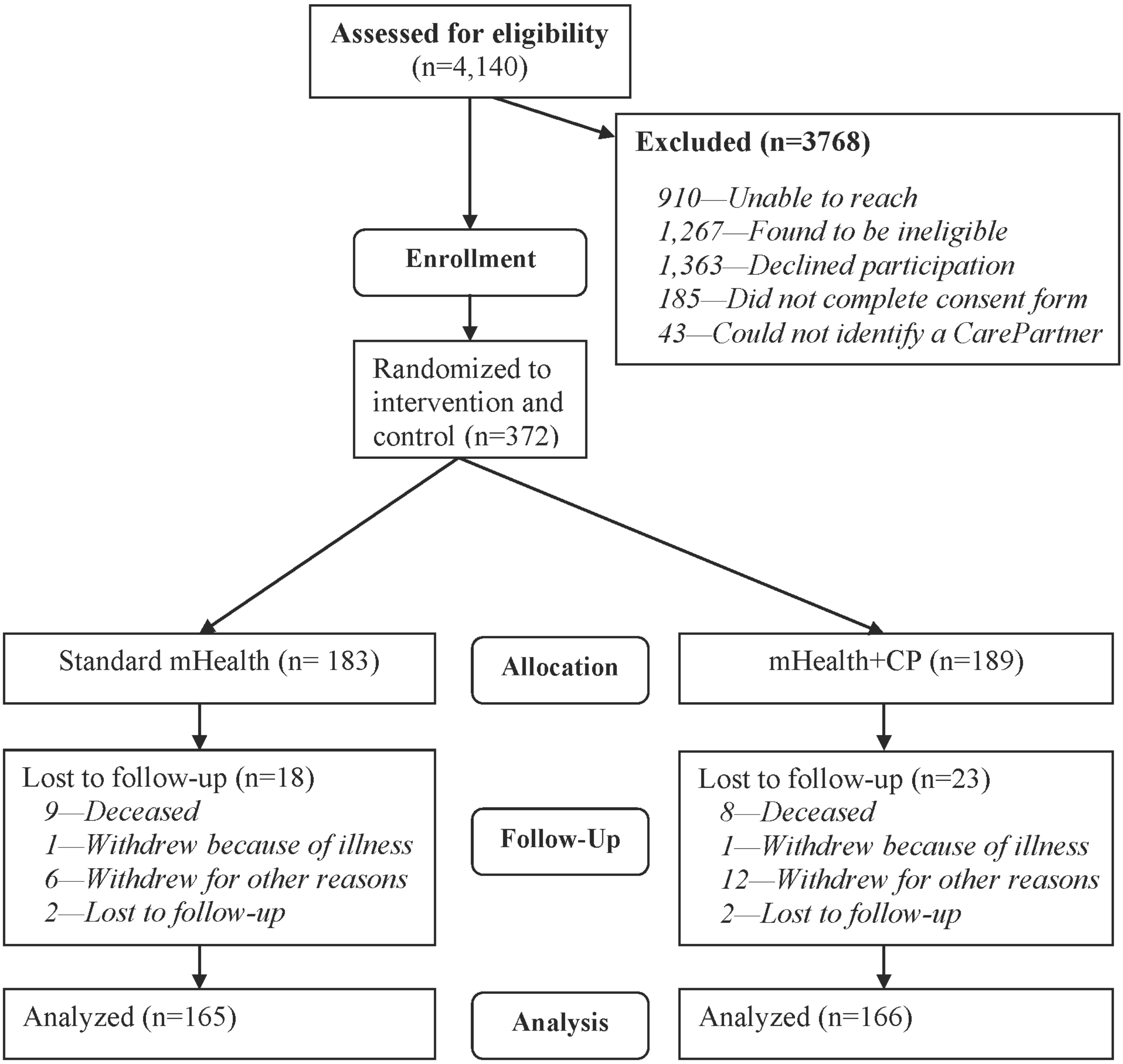
[c] Percent of patients with perfect HF medication adherence over the prior month as measured by the four HFSCB items focused on adherence (see Methods).

[d] Patients' reports regarding their relationship with their CarePartner.

[e] Percent of patients who report regularly experiencing any of six negative emotions when talking with their CarePartner (sadness, loneliness, anger, tension, guilt, or frustration).

[f] Percent of patients who agree that it is "difficult to talk to [their] CarePartner about [their] illness".

Figure 3. CONSORT Diagram for participants in the trial.



Interactive Voice Response Call Completion

Patients participated for a total of 15,709 call-weeks, during which they completed 14,175 calls, for a completion rate of 90.2%. IVR completion rates were essentially the same between mHealth+CP and standard mHealth arms (90.8% versus 89.7%), and there was no change in patients' likelihood of completing IVR calls throughout follow-up (P=.19). The likelihood of call completion was unrelated to patients' baseline HF-specific quality of life (MLHFQ) scores, HF self-management scores, CES-D scores, or measures of patient-CarePartner relationship quality (all P values ≥.15). IVR calls generated fax notifications to clinicians 1606 times (11.3% of completed calls), including 743 for weight gain, 774 for shortness of breath, and 89 for both problems. There were no differences in the number of fax alerts to clinicians between arms (P=.52).

Intervention Effects

Effects on Quality of Life, Self-Care, and CarePartner Communication Measured via Surveys at 6 and 12 Months

There were no differences by arm at either 6 or 12 months in HF quality of life (MLHFQ) scores (Table 2; both P>.21). Overall, there were no differences by arm in HF self-care behaviors measured by the HFSCB composite score. However, based on the four HFSCB items addressing HF medication adherence, mHealth+CP patients were 8.8% more likely than standard mHealth patients to report taking medication exactly as prescribed at 6 months (62.8% versus 54.0%, P=.02) and 13.8% more likely at 12 months (66.4% versus 52.6%, P=.01).

Table 2. Intervention effects measured via 6- and 12-month surveys.

	Baseline to 6 months		Baseline to 12 months	
	mHealth+CP effect (95% CI)	*P* value	mHealth+CP effect (95% CI)	*P* value
Quality of life and self-care				
MLHFQ[a]	+2.66 (-1.51 to 6.82)	.21	0.74 (-4.62 to 4.77)	.98
HFSCB[b]	-2.33 (-6.00 to 1.35)	.21	-1.08 (-4.74 to 2.58)	.56
Adherent to HF Rx[c]	+8.8% (1.2-16.5)	.02	+13.8% (3.7-23.8)	.01
Relationship quality [d]				
Talk 2+ times/ week	+10.2% (0.0-20.5)	.048	0.02% (-8.8%, 12.1%)	.76
Negative emotions[e]	-9.9% (-19.8 to -0.1)	.049	-13.8% (-23.4 to -4.2)	.01
Perceived difficulty[f]	-2.3% (-10.1 to 5.5)	.56	-8.3% (-16.6 to 0.0)	.049

[a] Minnesota Living with Heart Failure Questionnaire Scores. Lower scores indicate better functioning.

[b] Revised Heart Failure Self-Care Behavior Scale. Higher scores indicate better HF self-care.

[c] Patients' likelihood of reporting perfect HF medication adherence over the prior 30 days as measured by the four HFSCB items focused on heart failure medication use (see Methods).

[d] Patients' reports regarding their relationship with their CarePartner.

[e] Patients' likelihood of reporting regularly experiencing any of six negative emotions when talking with their CarePartner (sadness, loneliness, anger, tension, guilt, or frustration).

[f] Patients' likelihood of agreeing that it is "difficult to talk to [their] CarePartner about [their] illness".

Patients' survey responses indicated that dyadic communication with their CarePartner was more active and positive in the mHealth+CP arm. For example, in the 6-month survey, mHealth+CP patients had an absolute 10.2% greater likelihood than standard mHealth patients of reporting talking with their CarePartner at least twice per week over the prior 6 months (70.2% versus 60.0%; *P*=.048). mHealth+CP patients were significantly less likely than standard mHealth patients to report regularly experiencing negative emotions when talking with their CarePartner at the 6-month (31.9% versus 41.8%, *P*=.049) and 12-month follow-up (26.6% versus 40.4%, *P*=.01). Also, at the 12-month follow-up, mHealth+CP patients were 8.3% less likely than standard mHealth patients to agree that it was difficult for them to talk with their CarePartner about their illness (16.2% versus 24.5%; *P*=.049).

Effects on Adherence and Symptoms Reported Weekly via Interactive Voice Response

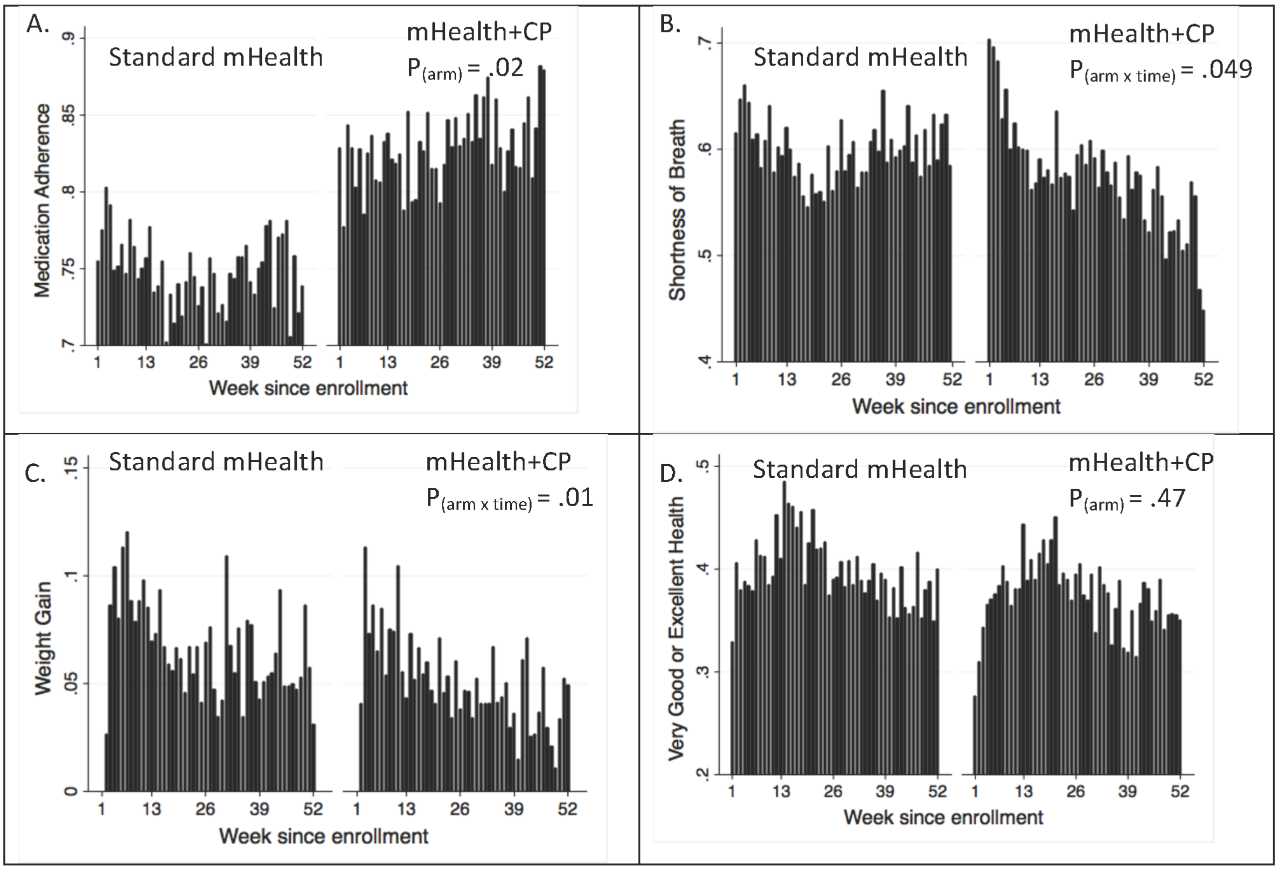
Displays of the unadjusted proportion of patients reporting perfect medication adherence, shortness of breath, and concerning weight changes via IVR suggested differences that favored mHealth+CP (Figure 4). These findings were substantiated by logistic regression analyses. Throughout the 1-year intervention, mHealth+CP patients were consistently more likely than standard mHealth patients to report perfect HF medication adherence over the prior week (main effect for arm, ie, ß=.5092; 95% CI 0.0857-0.9329; *P*=.02). There were no differences in time-trends in adherence reports across arms (*P*=.41), and the arm-by-time interaction term was excluded from the final model. Based on the logistic model, mHealth+CP patients had an 8.3% absolute greater likelihood of reporting perfect HF medication adherence in the prior week at 6 months (83.7% versus 75.4% for standard mHealth) and a 10.0% greater likelihood at 12 months (84.9% versus 74.9%).

Over the course of follow-up, mHealth+CP patients became increasingly less likely than standard mHealth patients to report shortness of breath during the prior week (arm-by-time interaction ß=-.0114; 95% CI -0.0206 to -0.0022; *P*=.049). The main effect of arm was not statistically significant (ß=.0894; 95% CI -0.2857 to 0.4644; *P*=.64). mHealth+CP patients had a 4% absolute reduction compared to standard mHealth patients in the likelihood of reporting shortness of breath at 6 months (57% versus 61%) and an 11.1% reduction at 12 months (50.1% versus 61.2%).

A significant arm-by-time interaction indicated that mHealth+CP patients were significantly less likely than standard mHealth patients to experience clinically significant weight increases (ß=-.0148; 95% CI -0.0232 to -0.0064; *P*=.01). The main effect of arm was not statistically significant (ß=.0454; 95% CI -0.2147 to 0.3055; *P*=.73). At 12 months, mHealth+CP patients had an absolute 2.4% decrease in the likelihood of generating a clinician notification for weight gain, relative to standard mHealth patients. Given the expected 12-month rate of significant weight increase in the standard mHealth group (5.4%), the reduction in the mHealth+CP arm represents a 44.4% relative improvement.

With respect to patients' reports of excellent/very good health, arm had neither a main effect (ß=-.1469; 95% CI -0.5366 to 0.2427; *P*=.39) nor an interaction with time (*P*=.70).

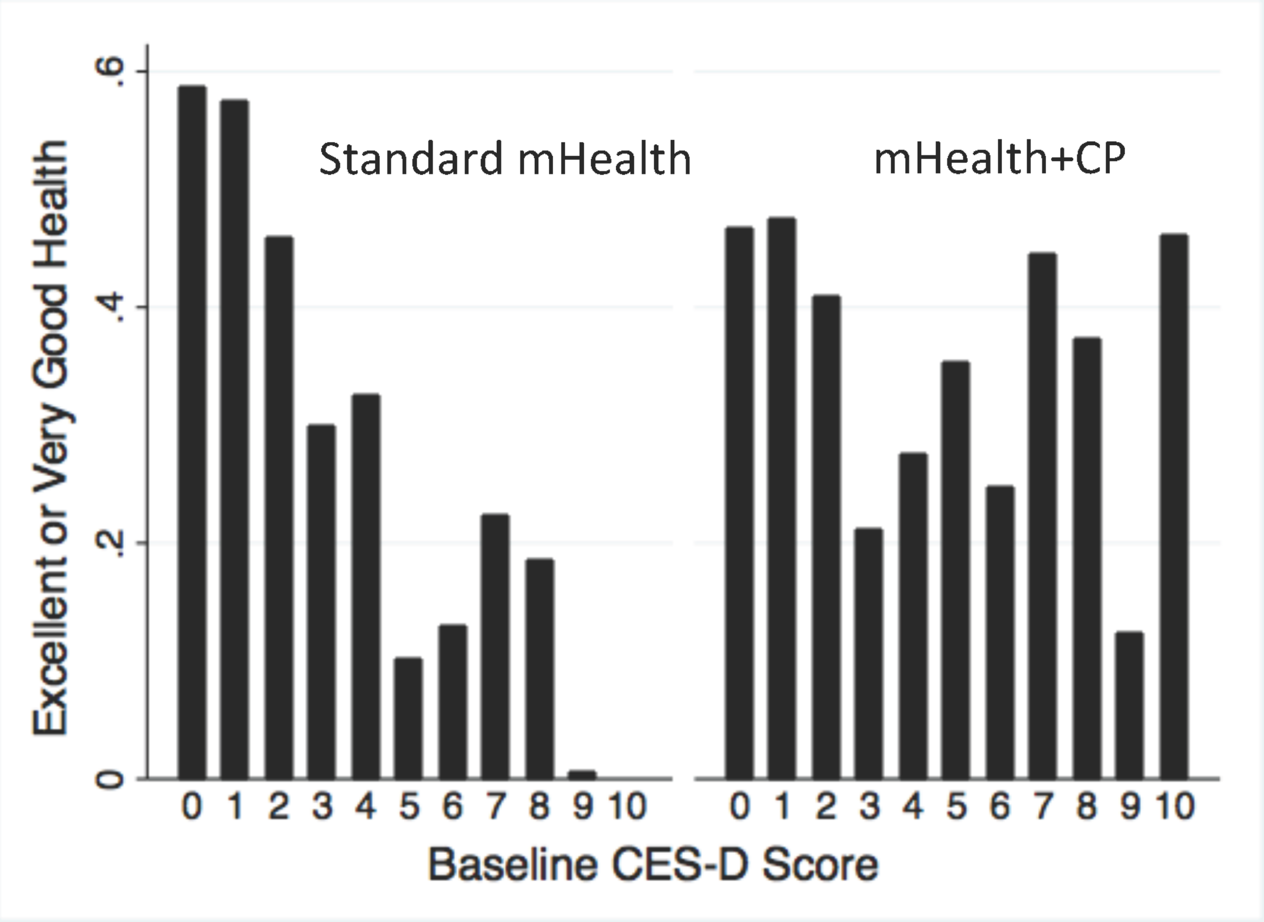
Figure 4. Unadjusted self-care and health status reports for patients in each randomization group by week since enrollment: Standard mHealth=patients randomized to IVR monitoring and self-care support with clinician alerts; mHealth+CP=patients randomized to the same intervention + weekly feedback to patients' CarePartners. The Y-axis for each panel differs in scale; bars represent the proportion of patients responding with that report. P values are from logistic regression models testing differences across arms. P values <.05 represent significant effects favoring mHealth+CP. A: Reports of always taking heart failure medication exactly as prescribed in the prior week. B: Reports of being bothered by shortness of breath every day or several days in the prior week. C: Clinically significant weight gain generating a notification to patients' healthcare team. D: Reports of very good or excellent health (versus good, fair, or poor health) in the prior week.



Auxiliary Analysis of the Interaction Between Randomization Arm and Baseline Depression Scores With Respect to Perceived Health Status Reported via Interactive Voice Response

Among patients randomized to standard mHealth, there was a strong negative association between higher (ie, worse) baseline CES-D depression scores and patients' likelihood of reporting excellent health via IVR (see Figure 5). In contrast, IVR reports of excellent health status were roughly constant in the mHealth+CP arm, regardless of the patient's baseline level of depressive symptoms. The leveling of mHealth+CP patients' perceived health reports across baseline CESD-levels reflected a somewhat lower proportion of mHealth+CP patients reporting excellent/very good health relative to standard mHealth patients when baseline CES-D scores were low, as well as a substantially higher proportion reporting excellent/very good health among those with greater baseline depressive symptoms. In multivariate analyses examining the effect of arm on patients' likelihood of reporting excellent health status separately in groups with low CES-D (scores 0-4) and high baseline CES-D (5+) scores, the effect of mHealth+CP was significant in both groups. mHealth+CP had a positive effect among patients with higher baseline CES-D scores (ß=1.27; 95% CI 0.42-2.12; *P*<.01), and a smaller negative effect among patients with lower baseline CES-D scores (ß=-.46; CI -0.90 to -0.028; *P*=.04). According to these models, patients with a baseline CES-D score of 1 were 11% less likely to report excellent/very good health if randomized to mHealth+CP, while patients with a baseline CES-D score of 8 were 22% more likely to report excellent/very good health if randomized to mHealth+CP relative to the control group.

Figure 5. Unadjusted reports of excellent/very good health for patients in each randomization group by baseline CES-D depression score. Higher scores indicated greater depressive symptoms.



CarePartner Feedback

Although mHealth+CP CarePartners' responses to IVR reports were not systematically tracked, many CarePartners did reply to those email reports, and their messages suggested that the structured alerts were read and acted upon (see Multimedia Appendix 2). Examples of text from those CarePartner replies include:

> *Hi. Thanks, there is nothing to report. He is doing quite well, thank you for your continuing caring and support.*
>
> *Hi. [Patient-partner's name] is coming alone fine, he was hospitalized for a few days due to an infection from his dialysis treatment, he is doing better today, he just returned from dialysis treatment. Thank you.*
>
> *Yes he has had a little shortness of breath and has sought council [sic] from his doctor. Thank you.*

Qualitative feedback from mHealth+CP CarePartners in their 12-month online follow-up survey also suggested that they felt that the feedback about their patient-partner was useful and that they were using that information as the basis for a stronger, more active relationship related to their partner's HF (see Table 3 for example quotes and Multimedia Appendix 3 for an exhaustive list of CarePartner comments). Comments suggest that CarePartners found the intervention useful not only for increasing the information base of their self-care assistance, but that it also served as a vehicle for strengthening their relationship with their patient-partner more generally.

Table 3. Example of responses to open-ended questions to mHealth+CP CarePartners in their 12-month follow-up survey regarding the perceived strengths of the program.

Category	Responses
Informational support and general knowledge about heart failure	I learned a lot about heart failure by being in the program. My father learned a lot too! [The program] gave me better insight into my dad’s health. It kept my relative in a reporting mode where he had to think about what he needed to do because someone would be checking in with him. I appreciated the weekly update regarding medications. [I liked] the CarePartner calls. The monitoring program is awesome. [I liked that] even if I hadn’t spoke with him yet, I knew from the email, he was ok.
Improved communication, reassurance, and relationship quality	[The program] helped my brother and I to get closer and communicate better. Communication about heart failure was more open. [The program] helped me understand my dad better. I liked that my dad told me a lot more about his health. I felt more comfortable talking to my brother about his heart failure. [The program] helps me to keep in touch with my cousin on a regular basis.
Ease of use and general positive comments	[mHealth+CP was] friendly, easy to understand, the questionnaire was easy to navigate. It was not very intrusive. As far as what I liked about the program, the fact that it even exists! It a wonderful idea and hopefully will yield results that are helpful to your patients. I think it made my Dad a little more responsible because he was more accountable to an outside party.

Discussion

Principal Findings

In this randomized comparative effectiveness trial, no group differences were identified at 6 or 12 months in the primary outcome of HF-specific quality of life or the composite measure of HF self-care. However, a number of potentially important differences in the process and outcomes of care were identified that favored mHealth+CP compared to standard mHealth. For example, in both follow-up surveys, a greater proportion of mHealth+CP patients reported perfect medication adherence, and mHealth+CP patients were consistently more likely throughout the 1-year follow-up to report via IVR that they took their HF medications as prescribed during the prior week. mHealth+CP patients also had a significantly greater decrease in their likelihood of reporting shortness of breath via IVR and were less likely to report clinically significant weight gains.

The strongly negative association between patients’ baseline depressive symptoms and IVR reports of perceived general health that we observed in the standard mHealth group was not apparent among patients who were randomized to mHealth+CP. In particular, patients with more severe depressive symptoms at baseline were relatively likely to make positive self-assessments about their health via IVR if they were in the mHealth+CP arm. This finding (as well as the feedback from CarePartners presented here) is consistent with studies suggesting that social support can have powerful impacts on patients’ well-being over and above the concrete benefits in terms of specific self-management behaviors [58].

These intervention effects represent positive impacts in some of the most fundamental areas of HF-related self-care and morbidity. Medication adherence is vital for HF patients, and poor adherence is a major predictor of acute events [63]. Shortness of breath and rapid weight gain are correlates of patients’ functional decline and used as sentinel events to identify patients at high risk for acute episodes. If these risk factors can be effectively addressed via mHealth services such as this one that focus on increasing caregiver support instead of the use of costly medical services, it would represent a major advance.

It is important to emphasize that these intervention effects were observed in a comparative effectiveness trial, over and above potential changes in health and self-care among patients receiving an active control intervention. All participants identified a CarePartner prior to randomization, and control patients and CarePartners received considerable information about HF self-care and self-management support. Control patients also received weekly IVR monitoring and self-management support calls with feedback to their clinician.

A soon-to-be-published companion paper using survey data from CarePartners in this same trial provides additional positive information consistent with the patient information reported here [38]. Compared to CarePartners in the standard mHealth arm, those randomized to mHealth+CP reported greater involvement in the patient’s medication adherence at both endpoints (both $P<.05$). mHealth+CP CarePartners also were more likely to report attending the patient’s medical visits at the 6-month follow-up. Importantly, CarePartners reporting the most symptoms of depression and strain at baseline had those symptoms significantly reduced if the CarePartner was in the mHealth+CP versus standard mHealth arm. These CarePartner reports as well as the qualitative feedback from CarePartners reported here suggest that involvement in mHealth information exchange may significantly improve relationship quality and self-management assistance for patients with chronic health problems. The qualitative feedback presented in Table 3 and Multimedia Appendix 3 is particularly interesting—since many mHealth+CP CarePartners volunteered that the intervention served to strengthen their relationship with their patient-partner.

The 1-week reporting interval used for the IVR-based outcome measures may have been more sensitive to intervention effects than the 6- and 12-month surveys. Differences across arms in IVR-reported adherence were consistent over the 12-month follow-up, and improvements in shortness of breath and weight became evident only after several months of program participation. This suggests that the pattern of effects is not the result of biased reporting, which tends to be immediate and short-lived [64]. Also, improvements in health and self-care measured via IVR were consistent with patients' improved medication adherence reported in both follow-up surveys and with reports of more frequent supportive communication with CarePartners. More generally, reports of health behaviors are more reliable when reporting intervals are brief, avoiding the biases associated with longer periods of retrospective recall [44,45,49,50,65].

Limitations

This trial had several limitations. It is possible that patients were biased about their medication adherence reporting in order to avoid burden for their CarePartner or conflict in the relationship. However, prior studies have shown that patients' medication self-reports are highly correlated with objective measures of medication use, especially when the recall interval is short and the measure is designed to identify even mild forms of non-adherence [44,45]. Also, other positive reports from both patients and CarePartners in this trial corroborate patients' IVR reports of medication adherence when randomized to mHealth+CP. Nevertheless, it would be important to confirm these findings with medication refill data. Similarly, it would be useful to verify patients' self-reported weights using data-storing electronic scales. Another limitation is that the trial was conducted among VA patients, nearly all of whom were men. Caregiving dynamics differ by patients' demographic and clinical characteristics, and future studies should determine whether results can be replicated in other populations, including non-VA patients and women. Some important clinical information about participants was not collected during the trial. For example, we do not know whether patients underwent cardiac surgery, resynchronization therapy, or revascularization. While we have no indication that randomization was unsuccessful, and patients in both groups were well matched on a wide range of baseline characteristics, it remains possible that unobserved differences in patients' clinical status at the time of enrollment may have contributed to the intervention effects observed. Finally, our study had several outcomes measured at two time points, and multiple comparisons may have contributed to the findings. However, results were consistent with the study's theoretical framework, and significant results were consistently in the same direction, that is, favoring mHealth+CP over standard mHealth.

Conclusions

This comparative effectiveness trial suggests that, although not all outcomes were different across arms at follow-up (notably HF-specific quality of life and a composite measure of HF self-care), providing caregivers with automated updates and guidance on self-care support may enhance the beneficial effects of mHealth for HF patients' health and self-management. Given increasing numbers of patients with chronic illness and the growing strains on clinical resources, health systems using mHealth approaches should consider creative ways to engage patients' social supporters to play a more active role. Finally, trials such as this one that include frequent mHealth monitoring may uncover intervention effects that are missed through more intermittent surveys and lengthy retrospective recall intervals.

Acknowledgments

Kimberly Schaub, PhD, and Julie Gee, NP, were key collaborators and clinicians at the Louis Stokes VA in Cleveland Ohio. Ananda Sen provided invaluable statistical consultation.

The study was funded by VA Health Services Research and Development Program (HSR&D) grant #IIR 07-185. JP is a VA Senior Research Career Scientist. RT is a VA HSR&D Career Development Awardee. Additional financial support came from grant #P30DK092926 from the National Institute of Diabetes and Digestive and Kidney Diseases.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Operationalization of inclusion and exclusion criteria.

[PDF File (Adobe PDF File), 44KB - jmir_v17i6e142_app1.pdf]

Multimedia Appendix 2

Unsolicited email replies from CarePartners when receiving feedback about the status of their patient-partner.

[PDF File (Adobe PDF File), 53KB - jmir_v17i6e142_app2.pdf]

Multimedia Appendix 3

Responses to open-ended questions to mHealth+CP CarePartners in their 12-month online follow-up survey regarding the perceived strengths of the program.

[PDF File (Adobe PDF File), 98KB - jmir_v17i6e142_app3.pdf]

Multimedia Appendix 4

CONSORT-EHEALTH checklist V1.6.2 [65].

[PDF File (Adobe PDF File), 147KB - jmir_v17i6e142_app4.pdf]

Multimedia Appendix 5

Comparison of baseline characteristics for patients with and without follow-up data.

[PDF File (Adobe PDF File), 68KB - jmir_v17i6e142_app5.pdf]

References

1. Harrison MB, Browne GB, Roberts J, Tugwell P, Gafni A, Graham ID. Quality of life of individuals with heart failure: a randomized trial of the effectiveness of two models of hospital-to-home transition. Med Care 2002 Apr;40(4):271-282. [Medline: 12021683]
2. Benatar D, Bondmass M, Ghitelman J, Avitall B. Outcomes of chronic heart failure. Arch Intern Med 2003 Feb 10;163(3):347-352. [Medline: 12578516]
3. Artinian NT, Magnan M, Sloan M, Lange MP. Self-care behaviors among patients with heart failure. Heart Lung 2002;31(3):161-172. [Medline: 12011807]
4. Ni H, Nauman D, Burgess D, Wise K, Crispell K, Hershberger RE. Factors influencing knowledge of and adherence to self-care among patients with heart failure. Arch Intern Med 1999 Jul 26;159(14):1613-1619. [Medline: 10421285]
5. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, American Heart Association Advocacy Coordinating Committee, Stroke Council, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation, Council on Cardiovascular Nursing, Council on the Kidney in Cardiovascular Disease, Council on Cardiovascular Surgery and Anesthesia, Interdisciplinary Council on Quality of Care and Outcomes Research. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation 2011 Mar 1;123(8):933-944 [FREE Full text] [doi: 10.1161/CIR.0b013e31820a55f5] [Medline: 21262990]
6. Whellan DJ, Hasselblad V, Peterson E, O'Connor CM, Schulman KA. Metaanalysis and review of heart failure disease management randomized controlled clinical trials. Am Heart J 2005 Apr;149(4):722-729. [doi: 10.1016/j.ahj.2004.09.023] [Medline: 15990759]
7. Windham BG, Bennett RG, Gottlieb S. Care management interventions for older patients with congestive heart failure. Am J Manag Care 2003 Jun;9(6):447-59; quiz 460 [FREE Full text] [Medline: 12816174]
8. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
9. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
10. Konstam MA. Home monitoring should be the central element in an effective program of heart failure disease management. Circulation 2012 Feb 14;125(6):820-827 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.031161] [Medline: 22331919]
11. Piette JD. Interactive behavior change technology to support diabetes self-management: where do we stand? Diabetes Care 2007 Oct;30(10):2425-2432. [doi: 10.2337/dc07-1046] [Medline: 17586735]
12. Piette JD, List J, Rana G, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015:0000-0000 (forthcoming).

13. Wakefield BJ, Boren SA, Groves PS, Conn VS. Heart failure care management programs: a review of study interventions and meta-analysis of outcomes. J Cardiovasc Nurs 2013;28(1):8-19. [doi: 10.1097/JCN.0b013e318239f9e1] [Medline: 22343208]
14. Pandor A, Gomersall T, Stevens JW, Wang J, Al-Mohammad A, Bakhai A, et al. Remote monitoring after recent hospital discharge in patients with heart failure: a systematic review and network meta-analysis. Heart 2013 Dec;99(23):1717-1726. [doi: 10.1136/heartjnl-2013-303811] [Medline: 23680885]
15. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
16. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
17. Migneault JP, Dedier JJ, Wright JA, Heeren T, Campbell MK, Morisky DE, et al. A culturally adapted telecommunication system to improve physical activity, diet quality, and medication adherence among hypertensive African-Americans: a randomized controlled trial. Ann Behav Med 2012 Feb;43(1):62-73. [doi: 10.1007/s12160-011-9319-4] [Medline: 22246660]
18. Magid DJ, Ho PM, Olson KL, Brand DW, Welch LK, Snow KE, et al. A multimodal blood pressure control intervention in 3 healthcare systems. Am J Manag Care 2011 Apr;17(4):e96-103. [Medline: 21774100]
19. Piette JD, Datwani H, Gaudioso S, Foster SM, Westphal J, Perry W, et al. Hypertension management using mobile technology and home blood pressure monitoring: results of a randomized trial in two low/middle-income countries. Telemed J E Health 2012 Oct;18(8):613-620 [FREE Full text] [doi: 10.1089/tmj.2011.0271] [Medline: 23061642]
20. Boriani G, Da CA, Ricci RP, Quesada A, Favale S, Iacopino S, et al. The MOnitoring Resynchronization dEvices and CARdiac patiEnts (MORE-CARE) randomized controlled trial: phase 1 results on dynamics of early intervention with remote monitoring. J Med Internet Res 2013;15(8):e167 [FREE Full text] [doi: 10.2196/jmir.2608] [Medline: 23965236]
21. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, et al. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
22. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
23. Agboola S, Jethwani K, Khateeb K, Moore S, Kvedar J. Heart failure remote monitoring: evidence from the retrospective evaluation of a real-world remote monitoring program. J Med Internet Res 2015;17(4):e101 [FREE Full text] [doi: 10.2196/jmir.4417] [Medline: 25903278]
24. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
25. Klersy C, De SA, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
26. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
27. Vuorinen A, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van GM, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
28. Fihn SD, McDonell MB, Diehr P, Anderson SM, Bradley KA, Au DH, et al. Effects of sustained audit/feedback on self-reported health status of primary care patients. Am J Med 2004 Feb 15;116(4):241-248. [doi: 10.1016/j.amjmed.2003.10.026] [Medline: 14969652]
29. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JGF. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
30. Martire LM, Lustig AP, Schulz R, Miller GE, Helgeson VS. Is it beneficial to involve a family member? A meta-analysis of psychosocial interventions for chronic illness. Health Psychol 2004 Nov;23(6):599-611. [doi: 10.1037/0278-6133.23.6.599] [Medline: 15546228]
31. Lett HS, Blumenthal JA, Babyak MA, Strauman TJ, Robins C, Sherwood A. Social support and coronary heart disease: epidemiologic evidence and implications for treatment. Psychosom Med 2005;67(6):869-878. [doi: 10.1097/01.psy.0000188393.73571.0a] [Medline: 16314591]
32. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
33. Pinquart M, Sörensen S. Spouses, adult children, and children-in-law as caregivers of older adults: a meta-analytic comparison. Psychol Aging 2011 Mar;26(1):1-14. [doi: 10.1037/a0021863] [Medline: 21417538]

34. Rosland AM, Piette JD, Choi H, Heisler M. Family and friend participation in primary care visits of patients with diabetes or heart failure: patient and physician determinants and experiences. Med Care 2011 Jan;49(1):37-45 [FREE Full text] [doi: 10.1097/MLR.0b013e3181f37d28] [Medline: 21102357]
35. Mårtensson J, Dracup K, Canary C, Fridlund B. Living with heart failure: depression and quality of life in patients and spouses. J Heart Lung Transplant 2003 Apr;22(4):460-467. [Medline: 12681424]
36. Vespa J, Lewis JM, Kreider RM. Current Population Reports. Washington, DC: US Census Bureau; 2013. America's families and living arrangements: 2012 URL: http://www.mwcog.org/uploads/committee-documents/kV1YX19a20140512145731.pdf [accessed 2015-06-03] [WebCite Cache ID 6Z16gjWYV]
37. National Institute on Aging. NIH Pub. So Far Away: Twenty Questions and Answers about Long Distance Caregiving URL: http://www.nia.nih.gov/sites/default/files/so_far_away_twenty_questions_about_long-distance_caregiving.pdf [accessed 2015-06-03] [WebCite Cache ID 6Z14AhccT]
38. Piette JD, Striplin D, Marinec N, Chen J, Aikens JE. A randomized trial of mobile health support for heart failure patients and their informal caregivers: impacts on caregiver-reported outcomes. Medical Care 2015:0000-0000 (forthcoming).
39. Norbeck JS, Lindsey AM, Carrieri VL. Further development of the Norbeck Social Support Questionnaire: normative data and validity testing. Nurs Res 1983;32(1):4-9. [Medline: 6549842]
40. Baumeister RF, Vohs KD. Self-regulation, ego depletion, and motivation. Social Pers Psych Compass 2007 Nov;1(1):115-128. [doi: 10.1111/j.1751-9004.2007.00001.x]
41. Zimmerman BJ. Attaining Self-Regulation: A Social Cognitive Perspective. In: Boekaerts M, Pintrich PR, Zeidner M, editors. Handbook of Self-Regulation. San Diego: Academic Press; 2000:13-35.
42. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). Arch Intern Med 1999;159(15):1701-1704. [Medline: 10448771]
43. Rector T, Kubo S, Cohn JN. Validity of the Minnesota Living with Heart Failure questionnaire as a measure of therapeutic response to enalapril or placebo. Am J Cardiol 1993 May 1;71(12):1106-1107. [Medline: 8475878]
44. Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: a summary of the literature. Med Care 2004 Jul;42(7):649-652. [Medline: 15213489]
45. Stirratt M, Crane H, Simoni J, Czajkowski S, Hillard M, Kahana S, et al. Self-report measures of medication adherence behavior: Recommendations on optimal use. Translational Behavioral Medicine 2015:0000-0000 (forthcoming).
46. Rolland Y, Aquino JP, Andrieu S, Beard J, Benetos A, Berrut G, et al. Identification of the main domains for quality of care and clinical research in nursing homes. J Nutr Health Aging 2011 May;15(5):410-424. [Medline: 21528170]
47. Lieberman MA, Fisher L. The effects of family conflict resolution and decision making on the provision of help for an elder with alzheimer's disease. The Gerontologist 1999 Apr 01;39(2):159-166. [doi: 10.1093/geront/39.2.159]
48. Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. J Pers Soc Psychol 2002 Jul;83(1):198-215. [Medline: 12088126]
49. Baranowski T. Validity and reliability of self report measures of physical activity: an information-processing perspective. Research Quarterly for Exercise and Sport 1988 Dec;59(4):314-327. [doi: 10.1080/02701367.1988.10609379]
50. Walsh JC, Mandalia S, Gazzard BG. Responses to a 1 month self-report on adherence to antiretroviral therapy are consistent with electronic data and virological treatment outcome. AIDS 2002 Jan 25;16(2):269-277. [Medline: 11807312]
51. Sisk JE, Hebert PL, Horowitz CR, McLaughlin MA, Wang JJ, Chassin MR. Effects of nurse management on the quality of heart failure care in minority communities: a randomized trial. Ann Intern Med 2006 Aug 15;145(4):273-283 [FREE Full text] [Medline: 16908918]
52. Stata Statistical Software: Release 13.1. College Station, TX: StataCorp LP; 2014. URL: http://www.stata.com/ [accessed 2015-06-03] [WebCite Cache ID 6Z171uwIa]
53. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980 May;48(4):817-830. [doi: 10.2307/1912934]
54. Rogers WH. Stata Technical Bulletin. 1993. Regression standard errors in clustered samples URL: http://wwwstatacom/support/faqs/statistics/stb13_rogerspdf [accessed 2015-06-03] [WebCite Cache ID 6Z1767371]
55. Froot KA. Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. The Journal of Financial and Quantitative Analysis 1989 Sep;24(3):333-355. [doi: 10.2307/2330815]
56. Ormel J, Kempen GI, Deeg DJ, Brilman EI, van Sonderen E, Relyveld J. Functioning, well-being, and health perception in late middle-aged and older people: comparing the effects of depressive symptoms and chronic medical conditions. J Am Geriatr Soc 1998 Jan;46(1):39-48. [Medline: 9434664]
57. Strine TW, Kroenke K, Dhingra S, Balluz LS, Gonzalez O, Berry JT, et al. The associations between depression, health-related quality of life, social support, life satisfaction, and disability in community-dwelling US adults. J Nerv Ment Dis 2009 Jan;197(1):61-64. [doi: 10.1097/NMD.0b013e3181924ad8] [Medline: 19155812]
58. Gladstone GL, Parker GB, Malhi GS, Wilhelm KA. Feeling unsupported? An investigation of depressed patients' perceptions. J Affect Disord 2007 Nov;103(1-3):147-154. [doi: 10.1016/j.jad.2007.01.019] [Medline: 17328958]
59. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life: the Heart and Soul Study. JAMA 2003 Jul 9;290(2):215-221 [FREE Full text] [doi: 10.1001/jama.290.2.215] [Medline: 12851276]

60. Carels RA. The association between disease severity, functional status, depression and daily quality of life in congestive heart failure patients. Qual Life Res 2004 Feb;13(1):63-72. [Medline: 15058788]
61. Bekelman DB, Havranek EP, Becker DM, Kutner JS, Peterson PN, Wittstein IS, et al. Symptoms, depression, and quality of life in patients with heart failure. J Card Fail 2007 Oct;13(8):643-648. [doi: 10.1016/j.cardfail.2007.05.005] [Medline: 17923356]
62. 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]
63. Ghali JK, Kadakia S, Cooper R, Ferlinz J. Precipitating factors leading to decompensation of heart failure. Traits among urban blacks. Arch Intern Med 1988 Sep;148(9):2013-2016. [Medline: 3046541]
64. Orne MT. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist 1962;17(11):776-783. [doi: 10.1037/h0043424]
65. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]

Abbreviations

HF: chronic heart failure
HFSCB: Revised Heart Failure Self-Care Behavior Scale
IVR: interactive voice response
HSR&D: Health Services Research & Development
mHealth: mobile health
mHealth+CP: mobile health plus CarePartner
MLHFQ: Minnesota Living with Heart Failure Questionnaire
NSSQ: Norbeck Social Support Questionnaire
VA: Veterans Health Administration

Edited by G Eysenbach; submitted 17.04.15; peer-reviewed by S Woods, A Patel; comments to author 07.05.15; revised version received 22.05.15; accepted 24.05.15; published 10.06.15.

Please cite as:
Piette JD, Striplin D, Marinec N, Chen J, Trivedi RB, Aron DC, Fisher L, Aikens JE
A Mobile Health Intervention Supporting Heart Failure Patients and Their Informal Caregivers: A Randomized Comparative Effectiveness Trial
J Med Internet Res 2015;17(6):e142
URL: http://www.jmir.org/2015/6/e142/
doi:10.2196/jmir.4550
PMID:26063161

Original Paper

Pharmaceutical Companies and Their Drugs on Social Media: A Content Analysis of Drug Information on Popular Social Media Sites

Jennifer Tyrawski[1*], MA; David C DeAndrea[1*], PhD

The Ohio State University, Columbus, OH, United States

[*] all authors contributed equally

Corresponding Author:
David C DeAndrea, PhD
The Ohio State University
154 N. Oval Mall, 3066 Derby Hall
Columbus, OH, 43210
United States
Phone: 1 614 292 3400
Fax: 1 614 292 2055
Email: deandrea.1@osu.edu

Abstract

Background: Many concerns have been raised about pharmaceutical companies marketing their drugs directly to consumers on social media. This form of direct-to-consumer advertising (DTCA) can be interactive and, because it is largely unmonitored, the benefits of pharmaceutical treatment could easily be overemphasized compared to the risks. Additionally, nonexpert consumers can share their own drug product testimonials on social media and illegal online pharmacies can market their services on popular social media sites. There is great potential for the public to be exposed to misleading or dangerous information about pharmaceutical drugs on social media.

Objective: Our central aim was to examine how pharmaceutical companies use social media to interact with the general public and market their drugs. We also sought to analyze the nature of information that appears in search results for widely used pharmaceutical drugs in the United States on Facebook, Twitter, and YouTube with a particular emphasis on the presence of illegal pharmacies.

Methods: Content analyses were performed on (1) social media content on the Facebook, Twitter, and YouTube accounts of the top 15 pharmaceutical companies in the world and (2) the content that appears when searching on Facebook, Twitter, and YouTube for the top 20 pharmaceutical drugs purchased in the United States. Notably, for the company-specific analysis, we examined the presence of information similar to various forms of DTCA, the audience reach of company postings, and the quantity and quality of company-consumer interaction. For the drug-specific analysis, we documented the presence of illegal pharmacies, personal testimonials, and drug efficacy claims.

Results: From the company-specific analysis, we found information similar to help-seeking DTCA in 40.7% (301/740) of pharmaceutical companies' social media posts. Drug product claims were present in only 1.6% (12/740) of posts. Overall, there was a substantial amount of consumers who interacted with pharmaceutical companies through commenting (23.9%, 177/740). For the drug-specific analysis, we found that the majority of search results contained drug product claims (69.4%, 482/695); more claims mentioned only benefits (44.8%, 216/482) relative to only risks (27.2%, 131/482). Additionally, approximately 25% (150/603) of posts on Twitter and YouTube were presented as personal testimonials. A considerable percentage of content on Facebook contained advertisements for illegal online pharmacies (17%, 16/92).

Conclusions: Pharmaceutical companies avoid making drug product claims on their social media accounts but frequently post content that is consistent with FDA definitions for help-seeking DTCA. Thousands of people often view content posted by pharmaceutical companies on social media; users also share company postings making both direct and indirect influence possible. Finally, people are likely to be exposed to drug product claims and information about illegal pharmacies when searching for information about popular pharmaceutical drugs on social media.

(J Med Internet Res 2015;17(6):e130) doi:10.2196/jmir.4357

KEYWORDS

social media; eHealth; direct-to-consumer advertising; eDTCA; pharmaceutical drugs; online pharmaceutical services; illegal online pharmacies; Facebook; Twitter; YouTube

Introduction

Background

Direct-to-consumer advertising (DTCA) of pharmaceutical products is an increasingly used but widely debated practice [1,2]. Electronic DTCA (eDTCA), in particular, is a rapidly growing marketing strategy [3] that was recently declared a "global health challenge" [4]. In particular, the features and affordances of social media (ie, interactive Web platforms where users can connect, collaborate, and exchange user-generated content) add complexity to pharmaceutical drug marketing. For instance, pharmaceutical companies can quickly and cheaply reach a variety of consumers online with multimodal, interactive, promotional activities, and consumers can produce promotional content as well [3]. Despite growing concerns about harmful effects, there is a lack of academic research on eDTCA [5]. Given that approximately 75% of adults online in the United States use social media frequently [6], it is critical to examine how social media are being used for eDTCA [4,5]. This study seeks to further our understanding of eDTCA by examining how pharmaceutical companies use social media to interact with the general public and market their drugs.

In addition to pharmaceutical companies' official social media accounts, it is important to document what information consumers are exposed to when searching popular social media sites for drug information. Researchers have noted that other consumers' reviews and testimonials are often quite persuasive [3-5]. The extent to which nonexperts make drug efficacy claims and share personal testimonials on social media currently has not been well documented despite the potential for such information to highly influence viewers. Public health officials are also greatly concerned that social media sites are being used to promote or host illegal pharmacies that directly harm patients [3,7,8]. The presence of drug efficacy claims and illegal pharmacies on social media sites is important to examine because these media have the potential to convey a degree of credibility to content they host [9]. Put differently, people might trust the claims made by illegal pharmacies or nonexperts more when the claims are hosted on popular social media sites than on strange or unknown websites. To better understand the prevalence of these concerns and how severely the public might be affected by drug information on social media, we analyzed the nature of information resulting from searches for the 20 most highly sold drugs in the United States on Facebook, Twitter, and YouTube.

Pharmaceutical Drug Marketing Via Social Media

The practice of DTCA is controversial. Proponents suggest DTCA has positive effects, such as generating disease awareness and increasing patient involvement in health decisions, but opponents suggest DTCA promotes misinformation, overemphasizes the benefits of pharmaceutical treatment over the risks, increases inappropriate prescribing, and more [2,10,11]. Due to these concerns, the US Food and Drug Administration (FDA) regulates the content of DTCA, banning all untruthful or misleading advertisements [2]. Additionally, the FDA requires product claim advertisements, a specific type of DTCA that names the drug and the condition(s) it treats, to present a "fair balance" of the benefits and risks of product use. In print advertisements, pharmaceutical companies must provide a brief summary of all risks associated with product use to meet fair-balance requirements. For broadcast advertisements, a statement of the major risks and information on where to locate complete risk information is required. The 2 other types of DTCA, reminder advertisements and help-seeking advertisements, do not indicate which condition(s) a product treats and thus are not subject to fair-balance rules. Reminder advertisements name the drug and often include information on dosage form or price. Help-seeking advertisements describe a health condition and encourage consumers to discuss the condition and potential treatment options with their doctor.

Online promotional activities, or eDTCA, now occupy an increased share of pharmaceutical companies' marketing budgets and more companies are marketing through social media [2,3]. Public health researchers have documented the negative effects that can occur from frequent and widespread eDTCA [3,4]. However, it remains unclear how pharmaceutical companies are currently using social media to market their drugs. Prior to changes in Facebook's commenting policy, many companies had specific social media pages for their products [3,12]. Although most product-specific pages have since been discontinued, pharmaceutical companies still maintain official social media accounts. As such, the first step of this study was to assess the extent to which information akin to the 3 forms of DTCA is present on major pharmaceutical companies' official social media accounts. We also documented the audience reach of eDTCA and whether companies are adhering to the FDA's fair-balance guidelines on social media.

In addition to eDTCA shared directly by companies, the interactive nature of social media has raised concerns that consumers might provide inaccurate and dangerous information about drugs on the official social media platforms of pharmaceutical companies [5]. People might be more likely to trust information posted by an outside source, particularly if the source claims to have personal experience with the topic at hand [3-5]. Additionally, pharmaceutical companies can potentially delete or alter negative consumer reviews, leaving only the most flattering portrayals behind [4]. Accordingly, we examined whether pharmaceutical companies provide formal policies that regulate what users can post to their official social media accounts (hereafter user postings/contributions are referred to as "user-generated content") and the frequency and nature of the posted user-generated content. Specifically, we examine whether users posted personal testimonials about health-related issues, the tone of user-generated comments, and the degree to which companies interacted with consumers.

Information About Pharmaceutical Drugs on Social Media

Although people can get information directly from pharmaceutical companies' sites, they can also search for information about particular drugs within popular social media sites. In particular, Facebook, Twitter, and YouTube are 3 of the most common social media platforms [5,6,13,14] that provide search capabilities; in a recent survey, 40% of participants had searched for health information on general social media sites such as these before [13]. The pharmaceutical drug information shared on these sites could have a large impact on their users' treatment decisions. Specifically, personal testimonials and drug efficacy claims, particularly from people unaffiliated with the pharmaceutical company, can be highly influential [3-5]. What information are people exposed to when they search for pharmaceutical drugs on social media? To address this question, we analyzed the nature of information people are exposed to when searching for the 20 most highly sold drugs in the United States on Facebook, Twitter, and YouTube.

Of critical interest to public health researchers is the extent to which illegal pharmacies are allowed to persist online. Illegal pharmacies are sites where consumers can purchase prescription drugs without a prescription and can compromise public safety by providing drugs to people who have not consulted medical officials and/or by providing counterfeit drugs that are ineffective, lead to injury, or cause death [7,8,12,15]. Given these serious implications for public health safety, we assessed the extent to which people are exposed to illegal pharmacies when searching on popular social media sites for commonly purchased pharmaceutical drugs.

In analyzing the results that appear when people search for pharmaceutical drugs on Facebook, Twitter, and YouTube, we more broadly documented the audience reach of the resulting pages and classified who controls the social media accounts (ie, is the site proprietor the pharmaceutical company or a consumer). We also documented the format and tone of the information posted as well as the nature of the associated user-generated comments.

To summarize, we sought to answer the following research questions:

1. To what extent is eDTCA present on pharmaceutical companies' social media accounts?
2. What is the nature of the user-generated content present on pharmaceutical companies' social media accounts?
3. To what extent are (1) drug efficacy claims, (2) personal testimonials, and (3) illegal pharmacies present when searching on popular social media sites for pharmaceutical drugs?

Methods

Two content analyses (company-specific and drug-specific) were conducted. For the company-specific analysis, the social media content of the top 15 pharmaceutical companies in the global and US Fortune 500 rankings were analyzed [16,17]. The drug-specific content analysis examined information on Facebook, Twitter, and YouTube about the top 20 drugs in 2013 based on US spending [18]. Table 1 lists the pharmaceutical companies and drugs examined.

Table 1. Pharmaceutical companies and drugs examined.

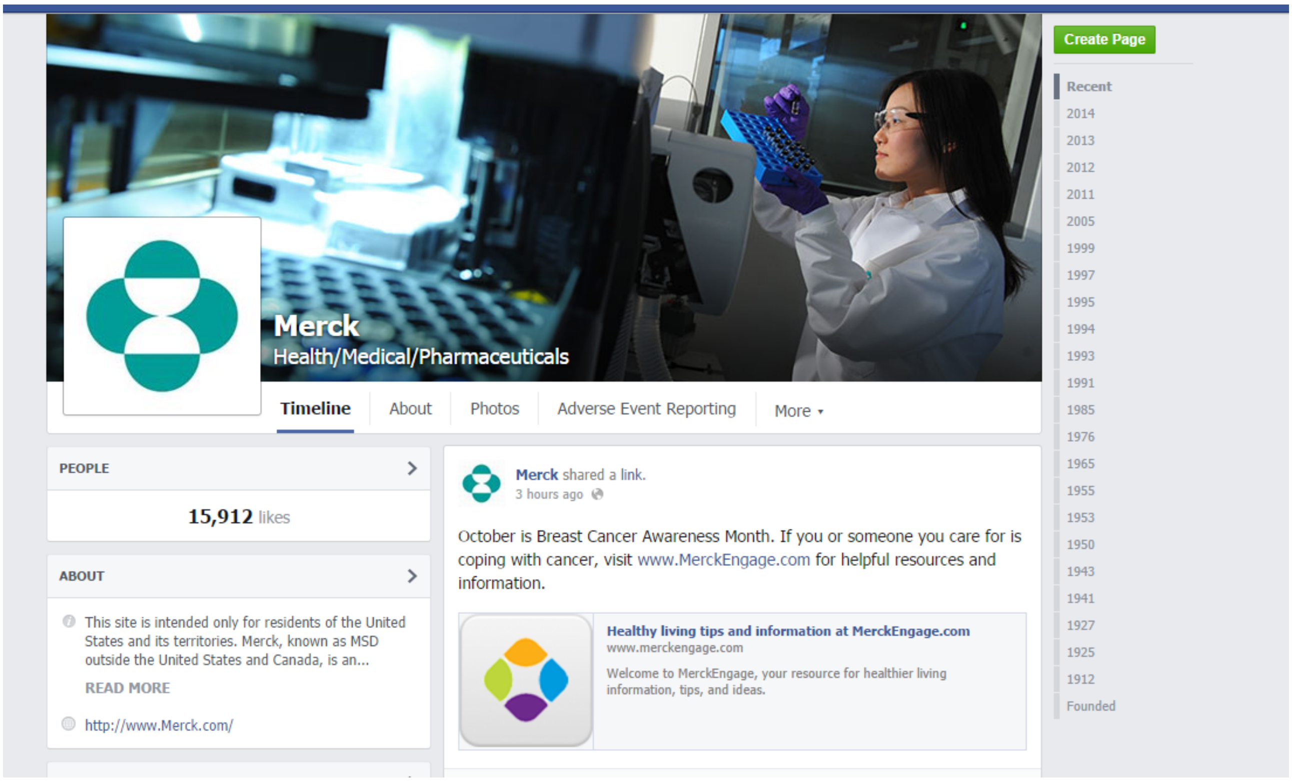
Companies	Drugs
Johnson & Johnson	Abilify
Novartis	Nexium
Pfizer	Humira
Roche Group	Crestor
Sanofi	Cymbalta
Merck	Advair Diskus
GlaxoSmithKline	Enbrel
Sinopharm	Remicade
AstraZeneca	Copaxone
Eli Lilly & Company	Neulasta
AbbVie Inc	Rituxan
Bristol-Myers Squibb Co	Lantus SoloSTAR/Lantus
Gilead Sciences, Inc	Spiriva Handihaler
Biogen Idec Inc	Atripla
Mylan Inc	Januvia, Avastin, OxyContin, Lyrica, Epogen, and Celebrex

Sample

Company-Specific Analysis

We analyzed (1) the social media information on the company's website; (2) each company's Facebook, Twitter, and YouTube page-level characteristics (eg, overall number of followers, commenting policies); (3) randomly selected posts appearing on those pages; and (4) user-generated comments on the randomly selected posts. For the individual posts, we randomly selected 20 posts from each site during a 1-year time frame (October 1, 2013-September 30, 2014). For pages with fewer than 20 posts in the time frame, the 20 most recent posts were selected. A total of 740 posts and 348 user-generated comments were analyzed. See Figures 1 and 2 for examples of content analyzed in the company-specific analysis.

Figure 1. Example Facebook page from company-specific analysis.



This site is intended only for residents of the United States and its territories. Merck, known as MSD outside the United States and Canada, is an...

http://www.Merck.com/

Merck shared a link.
3 hours ago

October is Breast Cancer Awareness Month. If you or someone you care for is coping with cancer, visit www.MerckEngage.com for helpful resources and information.

Healthy living tips and information at MerckEngage.com
www.merckengage.com
Welcome to MerckEngage, your resource for healthier living information, tips, and ideas.

Figure 2. Example tweet from company-specific analysis.

You'd ask your server a menu question. Why not ask your doctor about #diabetes? Tips on starting the conversation: ow.ly/zkl0K

8:15 AM - 12 Aug 2014

Drug-Specific Analysis

For Facebook and YouTube, each drug's name was entered into the site's search bar. Because most people do not venture past the first page of search results [19], the top 10 results were selected. Additionally, we collected the 10 most recent user-generated comments on the selected pages. For Twitter, we searched for each drug using a hashtag with the drug name (eg, #abilify) and randomly selected 20 tweets made within a 1-year time period (October 1, 2013-September 30, 2014). See Figure 3 for an example tweet from the drug-specific analysis. As in previous social media analyses, the sample was limited to content written in English [20,21]. A total of 800 pages/tweets/videos were analyzed.

Figure 3. Example tweet from drug-specific analysis.

A year of injections to help keep #MS at bay. It's my #Copaxone journey so far ... alifecopingwithms.blogspot.co.uk/2014/09/a-year…

2:18 PM - 14 Sep 2014

Company-Specific Variables

Social Media Sites

Any social media site links on the company's official webpage, including Facebook, Twitter, YouTube, Google+, LinkedIn, Flickr, Instagram, Pinterest, or blogs, were recorded.

Audience Reach

As in other content analyses of social media, the page likes (Facebook), followers (Twitter), and subscribers (YouTube) were coded to assess audience reach [20,22].

Page Commenting Policy

For each page, the presence or absence of a policy for user-generated comments was recorded. If a commenting policy existed, we assessed whether it prohibited discussions of (1) drug products, (2) drug benefits, and (3) drug risks, and whether the policy stated (4) the company would remove misinformation and (5) how users should report adverse events to the FDA.

Post/Comment Source

We assessed whether the content was originally authored by (1) the pharmaceutical company, (2) other for-profit company, (3) media outlet (news, television, radio, etc), (4) government agency, (5) nonprofit or academic organization, (6) consumer, or (7) other source. These categories were adapted from previous social media content analyses [23,24].

Post/Comment Content

We coded the presence or absence of the following content for each post/comment. Using the FDA's DTCA definitions, a post/comment could include (1) drug product claims or information about a specific drug and condition(s) it treats, (2) reminder information or information about a specific drug without uses, or (3) help-seeking information or information about a health condition without mentioning a treatment. For drug product claims, it was also noted whether the content included benefit and/or risk information. Additionally, content could include (4) nondrug treatment or information about nonpharmaceutical options to treat conditions and/or improve physical or mental health, (5) company information or news, or (6) job information/career opportunities.

Post/Comment Format

Based on previously used categories [19], we assessed the format of the information posted online. Information could be presented as either one or a combination of the following: (1) video, (2) image, (3) audio, and/or (4) text. Additionally, a post/comment could be an (5) interactive click-and-choose activity (poll, quiz, contest, or game) or (6) personalized/tailored content, where users receive a unique response based on provided information. We also coded whether a post/comment was presented as a testimonial (personal experience or story) or as didactic information (facts, reasons, or opinions without personal experience).

Post Interactivity

The interactivity of the post was assessed in multiple ways. First, following previous social media studies, we coded audience engagement as the number of "likes" (Facebook, YouTube), views (YouTube), shares (Facebook), and retweets and favorites (Twitter) [20,24]. Second, we assessed whether commenting was allowed and, if so, if the post solicited comments (ie, directly asked users to comment, retweet, or share the content) [19]. Third, the number of user-generated comments on each post and the number of company replies were recorded.

Comment Valence and Relevance

The valence of user-generated comments was coded as either (1) positive (ie, expressing support for the company, its products, or the content of the initial post), (2) negative (ie, expressing opposition to the company, its products, or the content of the initial post), or (3) mixed/neutral (ie, expressing both support and opposition). User-generated comments could also either be (1) relevant to the original post and on-topic or (2) irrelevant to the original post and clearly off-topic.

Drug-Specific Variables

Source/Site Proprietor

In addition to using the source options from the company-specific analysis, we also noted whether the site proprietor or account holder/creator was (1) an individual, (2) pharmaceutical company/representative, (3) another organization/group, or (4) other.

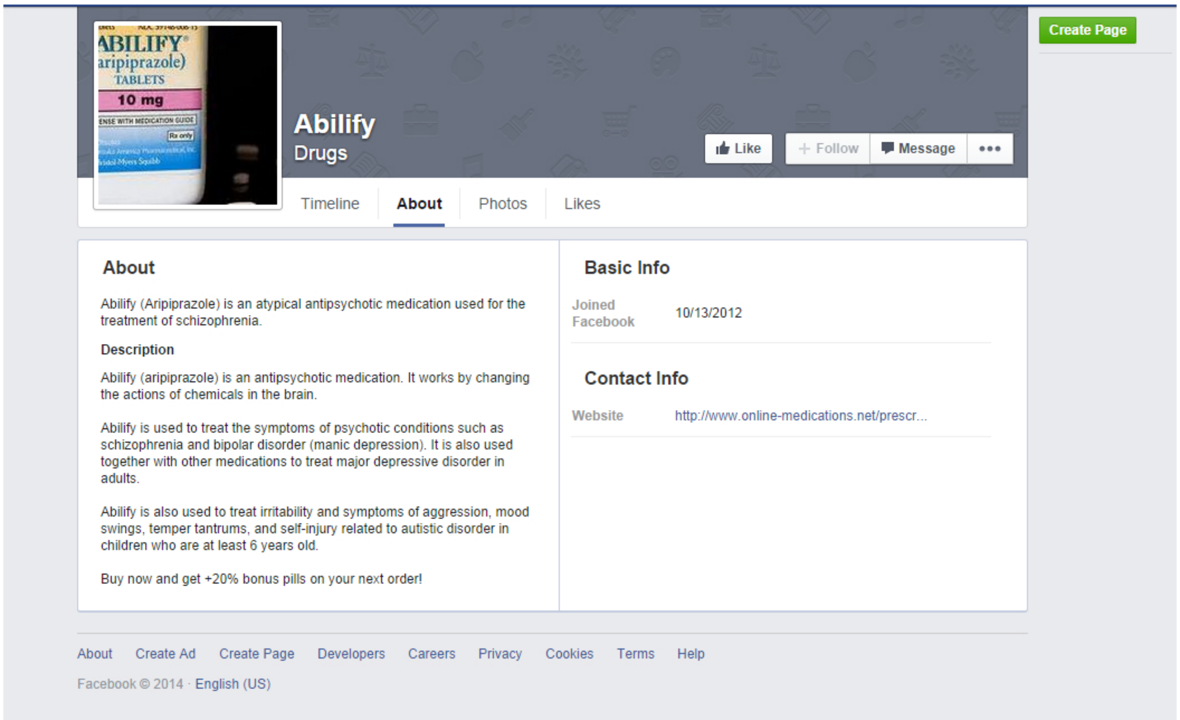
Content

The presence or absence of the following information was recorded for both the main posts and the user-generated comments. First, it was assessed whether the content was actually about the drug. Additionally, the content could make a claim about the drug's efficacy; if coded, we assessed whether the claim included benefit and/or risk information. Other content included (1) alternative treatment options, including other drugs or behaviors; (2) pharmaceutical company news; (3) emotional/informational support from other patients; (4) illegal pharmacies; and (5) lawsuits against the pharmaceutical company. See Figure 4 for an example of an illegal pharmacy on Facebook.

Format and Tone

The format codes from the company-specific analysis were used to classify the format of the content in the drug-specific analysis. We also coded whether the content was presented as humorous, such as joking about the side effects of the drug, or serious/nonhumorous.

Figure 4. Example Facebook page from drug-specific analysis.



About

Abilify (Aripiprazole) is an atypical antipsychotic medication used for the treatment of schizophrenia.

Description

Abilify (aripiprazole) is an antipsychotic medication. It works by changing the actions of chemicals in the brain.

Abilify is used to treat the symptoms of psychotic conditions such as schizophrenia and bipolar disorder (manic depression). It is also used together with other medications to treat major depressive disorder in adults.

Abilify is also used to treat irritability and symptoms of aggression, mood swings, temper tantrums, and self-injury related to autistic disorder in children who are at least 6 years old.

Buy now and get +20% bonus pills on your next order!

Basic Info

Joined Facebook 10/13/2012

Contact Info

Website http://www.online-medications.net/prescr...

Coder Training and Intercoder Reliability

Two separate teams of 3 coders each practiced extensively to clarify definitions and coding decisions. Each coder averaged a training time of approximately 30 hours. Each team coded 10% of their respective samples for reliability testing and intercoder reliability was established for all reported variables (Krippendorff's α>.70). For the company-specific analysis, Krippendorff's alpha scores ranged from .73 to 1.00. For the drug-specific analysis, scores ranged from .81 to 1.00.

Results

Company-Specific Analysis

Pharmaceutical Companies' Social Media Accounts

Overview

With the exception of Sinopharm, all pharmaceutical companies linked to at least one social media account on their website. Twitter was the most common social media site used (93%, 14/15), followed by Facebook (66%, 10/15), YouTube (66%, 10/15), and LinkedIn (60%, 9/15). Other less common social media sites included blogging platforms (26%, 4/15), Pinterest (26%, 4/15), Instagram (13%, 2/15), Flickr (13%, 2/15), and Google+ (6%, 1/15).

Company Facebook, Twitter, and YouTube Pages

The audience reach of 38 pages (10 Facebook, 17 Twitter, and 11 YouTube) was analyzed. The Facebook pages ranged in likes from 4716 to 642,816 (mean 105,806, SD 194,560; median 21,342.50, IQR 113,799). The Twitter pages ranged from 1521 to 98,589 followers (mean 36,723, SD 32,770). The YouTube accounts had a mean 2074 subscribers (SD 3169; median 924, IQR 1879), ranging from zero to 11,096 subscribers.

Across sites, the majority of pages did not have a formal commenting policy (63%, 24/38). Of the existing policies, most suggested misinformation would be removed (92%, 13/14), but did not explicitly prohibit consumers from making claims about their pharmaceutical products (85%, 12/14). The majority of policies did, however, provide information on how to report adverse events to the FDA (85%, 12/14).

Company Posts and User-Generated Comments

A total of 740 posts on pharmaceutical companies' social media accounts (200 Facebook, 340 Twitter, and 200 YouTube) and 348 user-generated comments (225 Facebook, 69 Twitter, and 54 YouTube) were analyzed.

Electronic Direct-to-Consumer Advertising

Overview

Table 2 displays the percentage of posts and user-generated comments on company-run Facebook, Twitter, and YouTube pages that included information that matched the FDA's definition of DTCA. Help-seeking information was the most common form of eDTCA; it was present in approximately 40% of all main posts (301/740), but was more commonly found on YouTube and Twitter than on Facebook (χ^2_2=14.6, P=.001). Drug product claims were present in only 1.6% of posts (12/740); of these, all posts mentioned the benefits of the drug (12/12) and only 33% (4/12) also mentioned its risks. Only 0.1% (1/740) of posts contained reminder information. Overall, most eDTCA found in pharmaceutical companies' social media posts could be classified as help-seeking advertisements; specific information about drug products was rare. However, when drug product claims were made, the majority did not follow fair-balance rules.

Table 2. Electronic direct-to-consumer advertising (eDTCA) on company-run social media pages.[a]

Content	Facebook	Twitter	YouTube	Total
Main posts, n	200	340	200	740
eDTCA, n (%)				
Help-seeking	59 (29.5)[b]	149 (43.8)	93 (46.5)[c]	301 (40.7)
Drug product claims	0 (0)	8 (2.4)	4 (2.0)	12 (1.6)
Benefits only[d]	0 (0)	5 (63)	3 (75)	8 (67)
Risks only[d]	0 (0)	0 (0)	0 (0.0)	0 (0)
Benefits and risks[d]	0 (0)	3 (38)	1 (25)	4 (33)
Reminder	1 (0.5)	0 (0)	0 (0)	1 (0.1)
User-generated comments, n	225	69	54	348
eDTCA, n (%)				
Help-seeking	15 (6.7)	12 (17)[c]	3 (6)	30 (8.6)
Drug product claims	4 (1.8)	0 (0)	0 (0)	4 (1.1)
Benefits only[e]	2 (50)	0 (0)	0 (0)	2 (50)
Risks only[e]	2 (50)	0 (0)	0 (0)	2 (50)
Benefits and risks[e]	0 (0)	0 (0)	0 (0)	0 (0)
Reminder	2 (0.9)	0 (0)	0 (0)	2 (0.6)

[a] Percentages in table based on column N, except where noted.

[b] Statistically underrepresented in sample.

[c] Statistically overrepresented in sample.

[d] Percentages based on drug product claim posts only (n=12).

[e] Percentages based on drug product claim comments only (n=4).

The user-generated comments on pharmaceutical companies' social media posts followed a similar pattern. Information that matched the FDA's definition of a help-seeking advertisement was the most common in comments (8.6%, 30/348) and was primarily found in user replies to company tweets (χ^2_2=8.5, *P*=.02). Drug product claim information was present in 1.1% (4/348) of comments, with half of these comments mentioning only benefits and half mentioning only risks. Similar to posts, reminder information was rare in comments (0.6%, 2/348). Overall, user-generated comments did not contain much DTCA-related information, but of those containing drug product claims, half did not provide any risk information.

Other Content

Table 3 displays the percentage of non-DTCA content in posts and user-generated comments on company-run Facebook, Twitter, and YouTube pages. The majority of pharmaceutical companies' posts shared company news (63.4%, 469/740), with this information most commonly shared on Twitter (χ^2_2=15.3, *P*<.001). A small portion of posts shared job information (5.0%, 37/740) and approximately 15% (112/740) of posts shared nondrug treatments for improving health. User-generated comments followed a similar pattern, as company news was the most common type of content (21.3%, 74/348) followed by nondrug treatments (2.9%, 10/348) and job information (1.4%, 5/348). Overall, both pharmaceutical companies' social media posts and user-generated comments primarily discussed company news.

Table 3. Non–electronic direct-to-consumer advertising (eDTCA) content on company-run social media pages.[a]

Content	Facebook	Twitter	YouTube	Total
Main posts, n	200	340	200	740
Non-eDTCA, n (%)				
Nondrug treatment	34 (17.0)	44 (12.9)	34 (17.0)	112 (15.1)
Company news	138 (69.0)	190 (55.9)[b]	141 (70.5)[c]	469 (63.4)
Job information	16 (8.0)	14 (4.1)	7 (3.5)	37 (5.0)
User-generated comments, n	225	69	54	348
Non-eDTCA, n (%)				
Nondrug treatment	4 (1.8)	3 (4)	3 (6)	10 (2.9)
Company news	44 (19.6)	17 (25)	13 (24)	74 (21.3)
Job information	5 (2.2)	0 (0)	0 (0)	5 (1.4)

[a] Percentages in table based on column n.

[b] Statistically underrepresented in sample.

[c] Statistically overrepresented in sample.

Source

Pharmaceutical companies authored the vast majority of content on their social media sites (91.9%, 680/740). However, pharmaceutical companies also shared information from media sources (3.8%, 28/740), advocacy groups (1.8%, 13/740), government agencies (1.2%, 9/740), and other companies and groups (1.3%, 10/740). Consumers posted the majority of user-generated comments (79.6%, 277/348), although pharmaceutical company employees or representatives posted 11.8% (41/348) of the comments. Other sources of comments included advocacy groups (4.0%, 12/348) and other companies or groups (4.5%, 16/348).

Format and Interactivity

The majority of pharmaceutical companies' social media posts were text-based (51.1%, 373/740) or video-based (26.3%, 199/740), and 20.0% (148/740) included both text and images. Interactive click-and-choose activities (0.1%, 1/740) and personalized/tailored content (0.1%, 1/740) were uncommon. Testimonials were used in 16.7% of posts (123/740).

Table 4 displays the degree of interaction found on the pharmaceutical companies' social media posts. In terms of audience engagement, Facebook posts averaged 65.53 likes (SD 75.98) and 8.5 shares (SD 15.18). Tweets averaged 2.11 favorites (SD 2.94) and 3.94 retweets (SD 4.98). YouTube videos averaged 1597.38 views (SD 31,886.88) and 211.76 likes (SD 2361.65). Close to 25% of posts had comments present (177/740), with an average of 0.50 comments per post (SD 1.32). Replies from the company were less common (mean 0.03, SD 0.20). Most interaction occurred on Facebook; of the posts with comments, half were on Facebook (χ^2_2=74.0, $P<.001$). Additionally, Facebook posts were more likely to solicit user-generated comments (χ^2_2=26.5, $P<.001$) and have replies from the company (χ^2_2=13.8, $P=.001$). Only YouTube allowed companies to disable comments on their posts and almost half of the YouTube videos sampled (96/200) had disabled the commenting function. Overall, audience engagement with pharmaceutical companies' social media posts was high, as users often interacted through liking and sharing the content. Additionally, a quarter of the posts included interaction through comments and pharmaceutical companies used Facebook to both solicit comments and have discussions with consumers.

Table 4. Interactivity on company-run social media pages.[a]

Interactivity	Facebook, n (%) n=200	Twitter, n (%) n=340	YouTube, n (%) n=200	Total, n (%) n=740
Comments allowed	200 (100.0)	200 (100.0)	104 (52.0)	644 (87.0)
Comments present	92 (46.0)[b]	57 (16.8)[c]	28 (14.0)[c]	177 (23.9)
Comments solicited	24 (12.0)[b]	9 (2.6)[c]	5 (2.5)[c]	38 (5.0)
Company replied	13 (6.5)[b]	3 (0.9)[c]	6 (3.0)	22 (3.0)

[a] Percentages in table based on column n.

[b] Statistically overrepresented in sample.

[c] Statistically underrepresented in sample.

User-Generated Comment Format, Valence, and Relevance

Approximately 14% (47/348) of user-generated comments on pharmaceutical companies' social media posts were testimonials. More than half of the user-generated comments were positive (186/348), 37.4% (130/348) were classified as mixed/neutral, and 9.2% (32/348) as negative. Positive comments were overrepresented on YouTube and negative comments were overrepresented on Twitter (χ^2_2=17.0, *P*=.002). The majority of comments were also relevant to the initial post (83.0%, 289/740). Relevant comments were overrepresented on YouTube (χ^2_2=8.0, *P*=.02). Positive comments were more likely to be relevant to the initial post (177/289), whereas mixed/neutral comments were more likely to be irrelevant (42/59; χ^2_2=42.6, *P*<.001). The majority of positive comments were on pages with a commenting policy (153/186), whereas most negative comments were on pages without a commenting policy (19/32; χ^2_2=8.7, *P*=.01). There was no relationship between comment relevance and presence of a commenting policy (χ^2_1=0.1, *P*=.75). Overall, it appeared that user-generated comments were mostly supportive of the pharmaceutical company and its products, particularly when the company had a commenting policy in place.

Drug-Specific Analysis

Of the 800 Facebook pages, tweets, and YouTube videos sampled from social media searches for pharmaceutical drugs, 86.9% (695/800) were actually about the searched-for drug. The following analyses included this portion of the sample.

Source/Site Proprietor

Of the 695 main posts about the searched-for drug, the majority of site proprietors were individuals (51.1%, 355/695) or nonpharmaceutical organizations (48.3%, 336/695). Pharmaceutical companies ran 0.6% of accounts (4/695). On Twitter and YouTube, consumers created most of the content (41.1%, 248/603), closely followed by media sources (37.0%, 233/603). Other sources included advocacy groups (6.0%, 36/603), pharmaceutical companies (3.3%, 20/603), other for-profit companies (9.0%, 54/603), and government agencies (1.0%, 6/603). Overall, most information from searches for drugs on Facebook, Twitter, or YouTube was attributed to members of the public rather than pharmaceutical companies.

Drug Product Claims

Table 5 displays the percentage of drug product claims in the search results on Facebook, Twitter, and YouTube. The majority included drug product claims (69.4%, 482/695), most of which were on YouTube (χ^2_2=13.7, *P*=.001). Of the drug product claims, posts mentioning only the benefits (44.8%, 216/482) were significantly more common than both risk-only posts (27.2%, 131/482) and posts that discussed both benefits and risks (28.0%, 135/482; χ^2_2=28.6, *P*<.001). The majority of user-generated comments on Facebook and YouTube videos also contained drug product claims (85.4%, 140/164). In contrast to the main posts, risk-only information (39.2%, 55/140) was significantly more common in comments than benefit-only information (22.9%, 32/140; χ^2_2=7.0, *P*=.03). Overall, results indicate that when the public searches for drugs on Facebook, Twitter, or YouTube, they are likely to come into contact with claims about those drugs' effectiveness. Although the main posts often highlight the benefits of the drug, the user-generated comments often present a contrasting view.

Table 5. Drug product claims in the drug-specific analysis.[a]

Content	Facebook	Twitter	YouTube	Total
Main posts, n	92	409	194	695
Drug product claims, n (%)	68 (73.9)	262 (64.1)	152 (78.4)[b]	482 (69.4)
Benefits only[c]	24 (35.5)	148 (56.5)[b]	44 (28.9)[d]	216 (44.8)
Risks only [c]	14 (20.6)[b]	85 (32.4)[b]	32 (21.1)[d]	131 (27.2)
Benefits and risks[c]	30 (44.1)[b]	29 (11.1)[d]	76 (50.0)[b]	135 (28.0)
User-generated comments, n	51		113	164
Drug product claims, n (%)	46 (90)		94 (83.1)	140 (85.4)
Benefits only[e]	15 (33)		17 (18.1)	32 (22.9)
Risks only[e]	14 (30)		41 (43.6)	55 (39.3)
Benefits and risks[e]	17 (37)		36 (38.3)	53 (37.8)

[a] Percentages in table based on column n, except where noted.

[b] Statistically overrepresented in sample.

[c] Percentages in row based on drug product claim posts only (n=482).

[d] Statistically underrepresented in sample.

[e] Percentages in row based on drug product claim comments only (n=140).

Illegal Pharmacies

Table 6 presents the other content found in the social media search results. Illegal pharmacies were present in 17.4% (16/92) of Facebook pages. Illegal pharmacies were less common on YouTube (χ^2_2=29.6, P<.001). Links to illegal pharmacies were also present in 9.1% (15/164) of user-generated comments on Facebook and YouTube; these comments were also more common on Facebook (χ^2_1=13.7, P<.001). When searching for drug information on social media, consumers were likely to come into contact with at least one link to an illegal pharmacy, particularly if consumers conducted the search on Facebook.

Table 6. Other content in drug-specific analysis.[a]

Content	Facebook	Twitter	YouTube	Total
Main posts, n	92	409	194	695
Illegal pharmacies, n (%)	16 (17)[b]	21 (5.1)	3 (1.5)[c]	40 (5.8)
Lawsuits, n (%)	11 (12)[b]	26 (6.4)	2 (1.0)[c]	39 (5.6)
Patient support, n (%)	8 (9)	6 (1.5)[c]	32 (16.5)[b]	46 (6.6)
Alternative treatments, n (%)	6 (7)	16 (3.9)	8 (4.1)	30 (4.3)
Company news, n (%)	20 (22)	100 (24.4)[b]	11 (5.7)[c]	131 (18.8)
User-generated comments, n	51		113	164
Illegal pharmacies, n (%)	11 (22)[b]		4 (3.5)[c]	15 (9.1)
Lawsuits, n (%)	10 (20)[b]		1 (0.9)[c]	11 (6.7)
Patient support, n (%)	12 (24)[c]		47 (41.6)[b]	59 (36.0)
Alternative treatments, n (%)	7 (14)		27 (23.9)	34 (20.7)
Company news, n (%)	13 (26)[b]		3 (2.7)[c]	16 (9.8)

[a] Percentages in table based on column n.

[b] Statistically overrepresented in sample.

[c] Statistically underrepresented in sample.

Other Content

Lawsuit information was present in 5.6% (39/695) of all drug-specific social media posts and was more common on Facebook and Twitter than YouTube (χ^2_2=15.1, P=.001). Patient support (6.6%, 46/695) and alternative treatment information (4.3%, 30/695) were present in fewer posts than company news (18.8%, 131/695). The majority of patient support was on YouTube (χ^2_2=48.8, P<.001). Most of the company news was on Twitter (χ^2_2=30.9, P<.001). There was no difference in alternative treatment information based on social media site (χ^2_2=1.3, P=.53). Thus, when searching for drug information on social media sites, consumers were likely to find information about the pharmaceutical company on Twitter, but support from other patients on YouTube. Information regarding lawsuits was found slightly less often than illegal pharmacies on Facebook. Alternative treatment options were relatively uncommon on all social media sites.

In contrast to the posts, patient support (36.0%, 59/164) and alternative treatment information (20.7%, 34/164) were more common than company news (9.8%, 16/164) and lawsuit information (6.7%, 11/164) in user-generated comments on Facebook and YouTube. Lawsuits (χ^2_1=19.7, P<.001) and company news (χ^2_1=20.8, P<.001) were more common on Facebook, whereas support was more common on YouTube (χ^2_1=5.0, P=.04). Overall, these results indicate that other consumers commented to provide alternative treatment options and support, even though this content was largely absent in the main posts.

Reach, Format, and Tone

Approximately 25% of posts on Twitter and YouTube were testimonials (150/603). Additionally, a large majority of tweets and YouTube videos had a serious, nonhumorous tone (96.7%, 583/603). The Facebook pages ranged from zero to 62,427 likes (median 69.0, IQR 328.3). Approximately 80% of tweets had zero favorites and zero retweets, with a mean of 0.42 favorites (SD 1.53) and 1.25 retweets (SD 16.32). YouTube video views ranged from 2 to 1,077,399 (median 4707.0, IQR 14,009). YouTube video likes ranged from zero to 1671 (median 51.83, IQR 30.0). In general, the majority of posts that arose from searches about specific drugs on social media provided didactic, nonhumorous information. The degree of audience engagement with drug information on social media sites varied widely.

Discussion

Principal Findings

The results of this study directly address critical concerns raised by researchers and public health officials about the marketing of pharmaceutical drugs via social media. Importantly, novel evaluations are provided about (1) how pharmaceutical companies use social media for DTCA, (2) how greatly companies reach and interact with consumers through social media, and (3) how likely people are to be exposed to drug efficacy claims and information about illegal pharmacies when searching for information about pharmaceutical drugs via social media. Respectively, the results suggest that (1) pharmaceutical companies avoid making drug product claims but frequently post help-seeking content, (2) thousands of people often view and share content posted by pharmaceutical companies, and (3) people are likely to be exposed to drug product claims and information about illegal pharmacies when searching for information about popular pharmaceutical drugs on social media.

More specifically, approximately 40% of all pharmaceutical companies' Facebook, Twitter, and YouTube posts in our sample met the FDA's definition of a help-seeking advertisement. This content focuses on generating awareness of a health condition or disease and often suggests that the audience should learn about potential treatment options from their doctor or other source. Despite concerns that specific drugs would be heavily advertised through pharmaceutical companies' social media accounts, product claim advertisements were uncommon. Only approximately 1% of posts contained a product claim. However, one-third of the product claim posts did not include any information on drug risks, thus failing to adhere to FDA regulations for traditional DTCA. Although this occurred in a relatively small number of posts overall, the problems surrounding the absence of risk information in product claims is well documented [2,4,5,25]. The FDA has developed draft guidelines for eDTCA regulations [26], and the inclusion of risk information is required for all company postings about specific products. To increase compliance with fair-balance rules on social media, the FDA should finalize the eDTCA regulations and formally detail how regulatory oversight will be enacted. Although monitoring every single post is likely unfeasible, the FDA could follow the procedures of this study to regularly monitor a random selection of posts and require pharmaceutical companies to notify the FDA whenever they use any media to share information with the public that is consistent with traditional forms of DTCA.

It is particularly important for the FDA to monitor pharmaceutical companies' social media accounts because they can have rather large audiences. Pharmaceutical companies' social media pages averaged approximately 45,000 followers or subscribers. Additionally, our results indicate that audience members are actively interacting with companies and sharing the content that the companies' post with people in their own social networks. For instance, posts are often liked and shared on Facebook and favorited and retweeted on Twitter. The public approval of this information on users' social network pages increases the potential for these posts to influence a large portion of the public. For example, research suggests that health behaviors and attitudes often spread through social networks and a number of social media-based interventions have shown that exposure to health information on social networking sites leads to health behavior change [27,28]. Thus, the messages that pharmaceutical companies share through social media channels have the potential to reach and influence millions of people worldwide; estimates of direct exposure grossly underestimate the cumulative influence of eDTCA on social media.

One common concern regarding eDTCA is that positive (and potentially misleading) product claim testimonials would populate the user-generated comments on pharmaceutical companies' pages [3,4]. Approximately 25% of posts had at least one comment and most were supportive of the company. However, very few comments contained information that would be classified as DTCA if the pharmaceutical company had produced the comments. Most commonly, user-generated comments contained information that resembled content that would appear in a help-seeking ad. When users did make drug product claims, however, they tended to focus either exclusively on benefits or risks. Additionally, most companies did not explicitly prohibit users from making product claims in their commenting policies. According to current FDA draft guidance documents, pharmaceutical companies are not responsible for the content of user-generated comments unless the comments were created by, paid for, or edited by the company [29,30]. Most commenting policies did, however, suggest that inaccurate information would be removed. Under current FDA draft guidance documents, pharmaceutical companies can, but are not required to, correct misinformation about their products in user-generated comments [30]. Interestingly, companies with commenting guidelines had significantly more positive comments than those without a commenting policy. Although it cannot be determined through this analysis, companies who are more aware of user-generated comments (and thus have a commenting policy) might be deleting negative user-generated comments [4]. If companies selectively delete user-generated comments, the information in the remaining user-generated comments would be applicable to FDA regulations [29,30].

Although uncommon on pharmaceutical companies' sites, product claims and testimonials were largely present in posts resulting from general searches for drug information on Facebook, Twitter, and YouTube. The majority of the top search results contained drug efficacy claims. Troublingly, most claims were made by nonexpert sources and mentioned only benefits of the drug rather than presenting a balanced view of both the benefits and risks of product use. Given that approximately 20% of Internet users check online reviews of particular drugs [31], our results suggest that consumers are likely getting incomplete drug information through social media. Additionally, around 25% of these posts were testimonials, a format that often enhances the credibility of the claims made [3-5]. Furthermore, it is likely that well-known and trusted media such as Facebook lend credibility to the health information posted within their pages relative to other online channels [9]. For example, young adults, the most prolific users of social networking sites [6], are the most likely age group to trust health information on social media [32] and to search for health advice and others' health experiences on social media [13].

The potential credibility afforded to information on social media is also problematic when we consider the continued presence of illegal pharmacies in the top search results. Illegal pharmacies were most prominent on Facebook; approximately 20% of the Facebook pages in the drug-specific analysis advertised illegal pharmacies. Although Facebook's terms of use prohibit illegal activity [33] and the organization is partnered with the Center for Safe Internet Pharmacies [8], our results suggest Facebook is not adequately policing its site for illegal pharmacies. We echo the calls of other scholars for social media companies to actively monitor their sites and make meaningful policy changes to eliminate this type of content [7,8,15]. Past policy changes demonstrate that social media organizations can have a measurable impact on the presence of pharmaceutical drug information on their sites. For example, Facebook eliminated companies' ability to block the commenting feature on their pages in 2011, so many of the drug product pages that existed in previous analyses [12] were discontinued [3]. Thus, these sites need to take an active role in protecting their users from harmful illegal pharmacies.

Limitations

There are some limitations to the present study. First, we focused our analysis on the top pharmaceutical companies and best-selling drugs. As a cost-effective marketing strategy, smaller companies might rely on social media advertising more so than larger companies and pharmaceutical companies might use social media channels to introduce newer, less established drugs to the marketplace [3]. Additionally, although we analyzed a large number of postings, our review only focused on 1 year of social media activity. As marketing trends constantly change [12], future research should investigate if the presence of eDTCA changes over time. Last, we focused on the presence of product claims, benefits, and risks and did not examine the accuracy of the claims or whether benefits or risks were emphasized within in a single post. To get a more complete picture of the pharmaceutical drug information that appears on social media sites, future research should explore these areas.

Conclusions

Social media sites are an accessible channel for pharmaceutical companies and others to easily deliver drug information to millions of people across the globe. Although pharmaceutical companies are not directly marketing specific products through their social media accounts often, they are posting content similar to help-seeking DTCA, which describes a health condition without providing a specific solution. If people search for drug solutions to these ailments via social media sites, they will likely be exposed to testimonials that highlight pharmaceutical drug benefits over risks as well as links to pharmacies where they can illegally purchase these drugs. Thus, pharmaceutical drug information on social media sites is potentially quite dangerous to public health and should be monitored accordingly.

Acknowledgments

The authors wish to thank Jennifer Brewer, Abigail Cohen, Juho Ma, Jenna McGlade, and Taylor Walton for their assistance with the coding of the data.

Conflicts of Interest

None declared.

References

1. Gilbody S, Wilson P, Watt I. Benefits and harms of direct to consumer advertising: A systematic review. Qual Saf Health Care 2005;14:246-250 doi: 10.1136/qshc.2004.012781. [doi: 10.1136/qshc.2004.012781]
2. Ventola CL. Direct-to-consumer pharmaceutical advertising: therapeutic or toxic? P T 2011 Oct;36(10):669-684 [FREE Full text] [Medline: 22346300]
3. Gibson S. Regulating direct-to-consumer advertising of prescription drugs in the digital age. Laws 2014 09;3(3):410-438. [doi: 10.3390/laws3030410]
4. Liang BA, Mackey TK. Direct-to-consumer advertising with interactive internet media: Global regulations and public health issues. JAMA 2011;305(8):824. [doi: 10.1001/jama.2011.203]
5. Greene JA, Kesselheim AS. Pharmaceutical marketing and the new social media. N Engl J Med 2010 Nov 25;363(22):2087-2089. [doi: 10.1056/NEJMp1004986] [Medline: 21105789]
6. Duggan M, Ellison NB, Lampe C, Lenhart A, Madden M. Social Media Update 2014. Washington, DC: Pew Internet & American Life Project; 2015 Jan 09. URL: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf [accessed 2015-02-08] [WebCite Cache ID 6WCS7cMcQ]
7. Mackey TK, Liang BA. Global reach of direct-to-consumer advertising using social media for illicit online drug sales. J Med Internet Res 2013 May;15(5):e105 [FREE Full text] [doi: 10.2196/jmir.2610] [Medline: 23718965]
8. Mackey TK, Liang BA, Strathdee SA. Digital social media, youth, and nonmedical use of prescription drugs: The need for reform. J Med Internet Res 2013;15(7):e143. [doi: 10.2196/jmir.2464]
9. Flanagin AJ, Metzger MJ. Digital media and youth: unparalleled opportunity and unprecedented responsibility. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; 2008:5-28.
10. Deshpande A, Menon A, Perri IM, Zinkhan G. Direct-to-consumer advertising and its utility in health care decision making: a consumer perspective. J Health Commun 2004;9(6):499-513. [doi: 10.1080/10810730490523197] [Medline: 15764449]

11. Krezmien E, Wanzer MB, Servoss T, LaBelle S. The role of direct-to-consumer pharmaceutical advertisements and individual differences in getting people to talk to physicians. J Health Commun 2011 Sep;16(8):831-848. [doi: 10.1080/10810730.2011.561909] [Medline: 21512934]
12. Liang BA, Mackey TK. Prevalence and global health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011 Aug;13(3):e64 [FREE Full text] [doi: 10.2196/jmir.1775] [Medline: 21880574]
13. Health Research Institute. Social media “likes” healthcare: from marketing to social business. 2012. URL: http://pwchealth.com/cgi-local/hregister.cgi/reg/health-care-social-media-report.pdf [accessed 2015-02-08] [WebCite Cache ID 6WCSCxM6I]
14. Liang Y, Zheng X, Zeng DD, Zhou X, Leischow SJ, Chung W. Exploring how the tobacco industry presents and promotes itself in social media. J Med Internet Res 2015 Jan;17(1):e24 [FREE Full text] [doi: 10.2196/jmir.3665] [Medline: 25608524]
15. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 internet pharmacies: Illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013;15(9):e199. [doi: 10.2196/jmir.2606]
16. Fortune. 2014. Fortune 500 2014 URL: http://fortune.com/fortune500/johnson-johnson-39/ [accessed 2015-02-09] [WebCite Cache ID 6WDdU7T7u]
17. Fortune. 2014. Fortune Global 500 2014 URL: http://fortune.com/global500/johnson-johnson-121/ [accessed 2015-02-09] [WebCite Cache ID 6WDdY7Fmd]
18. IMS Institute for Healthcare Informatics. Medicine use and shifting costs of healthcare: A review of the use of medicines in the United States in 2013. 2014. URL: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/Reports/Secure/IIHI_US_Use_of_Meds_for_2013.pdf [accessed 2015-02-14] [WebCite Cache ID 6WKEcVxWd]
19. Freeman B, Chapman S. Measuring interactivity on tobacco control websites. J Health Commun 2012 Aug;17(7):857-865. [doi: 10.1080/10810730.2011.650827] [Medline: 22574732]
20. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014 Aug;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
21. Tian Y. Organ donation on Web 2.0: content and audience analysis of organ donation videos on YouTube. Health Commun 2010;25(3):238-246. [doi: 10.1080/10410231003698911]
22. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
23. Paek HJ, Hove T, Jeon J. Social media for message testing: A multilevel approach to linking favorable viewer responses with message, producer, and viewer influence on YouTube. Health Commun 2013;28(3):226-236. [doi: 10.1080/10410236.2012.672912]
24. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012;27(5):478-485. [doi: 10.1080/10410236.2011.610258]
25. Davis JJ. Riskier than we think? The relationship between risk statement completeness and perceptions of direct to consumer advertised prescription drugs. J Health Commun 2000 Oct;5(4):349-369. [doi: 10.1080/10810730050199141] [Medline: 11191018]
26. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, Center for Devices and Radiological Health. Guidance for Industry: Internet/Social Media Platforms with Character Space Limitations—Presenting Risk and Benefit Information for Prescription Drugs and Medical Devices. 2014 Jun. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM401087.pdf [accessed 2015-03-19] [WebCite Cache ID 6X8QFAUAl]
27. 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]
28. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: A systematic review and meta-analysis. J Am Med Inform Assoc 2015;22:243-256. [doi: 10.1136/amiajnl-2014-002841]
29. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine. Guidance for Industry: Fulfilling Regulatory Requirements for Postmarketing Submissions of Interactive Promotional Media for Prescription Human and Animal Drugs and Biologics. 2014 Jan. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM381352.pdf [accessed 2015-03-19] [WebCite Cache ID 6X8QGad56]
30. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, Center for Devices and Radiological Health. Guidance for Industry Internet/Social Media Platforms: Correcting Independent Third-Party Misinformation About Prescription Drugs and Medical Devices. 2014 Jun. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM401079.pdf [accessed 2015-03-19] [WebCite Cache ID 6X8QCChgD]

31. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2015-02-14] [WebCite Cache ID 6WKEp8JDd]
32. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. “Friending” teens: systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015;17(1):e4. [doi: 10.2196/jmir.3692]
33. Facebook. 2015. Terms of service URL: https://www.facebook.com/unsupportedbrowser [accessed 2015-02-12] [WebCite Cache ID 6WHCA5xE9]

Abbreviations

DTCA: direct-to-consumer advertising
eDTCA: electronic direct-to-consumer advertising
FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 18.02.15; peer-reviewed by S Gibson, R Briones; comments to author 11.03.15; revised version received 19.03.15; accepted 22.03.15; published 01.06.15.

Please cite as:
Tyrawski J, DeAndrea DC
Pharmaceutical Companies and Their Drugs on Social Media: A Content Analysis of Drug Information on Popular Social Media Sites
J Med Internet Res 2015;17(6):e130
URL: http://www.jmir.org/2015/6/e130/
doi:10.2196/jmir.4357
PMID:26032738

Original Paper

The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness

Panos Balatsoukas[1,2], PhD; Catriona M Kennedy[3,4], PhD; Iain Buchan[1,2], MD, FFPH, FACMI; John Powell[5], PhD, FFPH; John Ainsworth[1,2], MSc

[1]Institute of Population Health, Centre for Health Informatics, University of Manchester, Manchester, United Kingdom
[2]Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, Manchester, United Kingdom
[3]School of Computer Science, University of Birmingham, Birmingham, United Kingdom
[4]School of Computer Science, University of Manchester, Manchester, United Kingdom
[5]Nuffield Department of Primary Care, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Panos Balatsoukas, PhD
Institute of Population Health
Centre for Health Informatics
University of Manchester
G304 J.McFarlane Bldg
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1613060673
Fax: 44 1613060673
Email: panagiotis.balatsoukas@manchester.ac.uk

Abstract

Background: Social network technologies have become part of health education and wider health promotion—either by design or happenstance. Social support, peer pressure, and information sharing in online communities may affect health behaviors. If there are positive and sustained effects, then social network technologies could increase the effectiveness and efficiency of many public health campaigns. Social media alone, however, may be insufficient to promote health. Furthermore, there may be unintended and potentially harmful consequences of inaccurate or misleading health information. Given these uncertainties, there is a need to understand and synthesize the evidence base for the use of online social networking as part of health promoting interventions to inform future research and practice.

Objective: Our aim was to review the research on the integration of expert-led health promotion interventions with online social networking in order to determine the extent to which the complementary benefits of each are understood and used. We asked, in particular, (1) How is effectiveness being measured and what are the specific problems in effecting health behavior change?, and (2) To what extent is the designated role of social networking grounded in theory?

Methods: The narrative synthesis approach to literature review was used to analyze the existing evidence. We searched the indexed scientific literature using keywords associated with health promotion and social networking. The papers included were only those making substantial study of both social networking and health promotion—either reporting the results of the intervention or detailing evidence-based plans. General papers about social networking and health were not included.

Results: The search identified 162 potentially relevant documents after review of titles and abstracts. Of these, 42 satisfied the inclusion criteria after full-text review. Six studies described randomized controlled trials (RCTs) evaluating the effectiveness of online social networking within health promotion interventions. Most of the trials investigated the value of a "social networking condition" in general and did not identify specific features that might play a role in effectiveness. Issues about the usability and level of uptake of interventions were more common among pilot studies, while observational studies showed positive evidence about the role of social support. A total of 20 papers showed the use of theory in the design of interventions, but authors evaluated effectiveness in only 10 papers.

Conclusions: More research is needed in this area to understand the actual effect of social network technologies on health promotion. More RCTs of greater length need to be conducted taking into account contextual factors such as patient characteristics and types of a social network technology. Also, more evidence is needed regarding the actual usability of online social networking and how different interface design elements may help or hinder behavior change and engagement. Moreover, it is crucial to investigate further the effect of theory on the effectiveness of this type of technology for health promotion. Research is needed linking theoretical grounding with observation and analysis of health promotion in online networks.

(J Med Internet Res 2015;17(6):e141) doi:10.2196/jmir.3662

KEYWORDS

health behaviors; health promotion; health behavior change; health education; social media; social technology; social networking; content analysis; theoretical grounding

Introduction

Background

Social networking sites (SNS)—such as YouTube, Facebook, and Twitter—have been used extensively in public health and prevention interventions to change behavior and improve health outcomes [1,2]. Several aspects of SNS—including social support, empowerment, peer pressure, and interactive information-emotion sharing—have the potential to influence patients' health behaviors and increase adherence to and engagement with such interventions [3-5]. Yet little is known about the actual effect of SNS on behavior change and on the factors that may influence user interaction and experience, such as usability, user satisfaction, and level of technology acceptance or engagement. Therefore, there is a need to understand the effectiveness of SNS in the context of wider health promotion methods and evidence—not simply assuming that interventions can be ported from one medium to another.

Previous reviews of the literature have provided mixed results about the effectiveness of SNS for health promotion with many authors characterizing the effect of online social networking on behavior change as positive, but not statistically significant [1,6]. For example, Korda and Itani [7] identified both positive and less successful examples of the application of social media (including blogs, forums, video-sharing, and wikis) for health promotion. However, the authors also concluded that there is a need for precise evaluation metrics and for behavior change interventions to be grounded in theory in order to successfully measure and assess their effectiveness. The previous work in this area suggests that the lack of clear evidence can be attributed to the following factors.

First, there are a small number of randomized controlled trials (RCTs) with considerable heterogeneity used to evaluate the actual effect of online social networking on behavior change. This was evident in two recent systematic reviews, with a meta-analysis, by Maher and Lewis [1] and Lavanjo et al [2], which showed mixed results. Maher and Lewis showed a modest effect for the examined interventions on behavior change when magnitudes of the effect sizes were calculated, while Lavanjo et al reported a slight positive effect of SNS interventions on health behavior-related outcomes. However, the findings of these two studies should be interpreted with caution since, in the case of both reviews, the authors analyzed a small number of RCTs (six studies in [1] and eight in [2]), the majority of which were short-term trials, with a study duration not exceeding 6 months, while there was considerable heterogeneity of study designs, evaluation metrics, health topics, and types of SNS.

Further, there is a lack of ecological validity due to the difficulty in assessing the true effect of SNS in the context of multi-component interventions. There is a lack of clarity over whether a positive effect could be attributed to the SNS or the non-SNS component of an intervention [1,2,8]. A typical example of this phenomenon was highlighted by Chang et al [9] who reviewed the evidence about the effect of SNS on weight management behaviors. From the 20 studies that met the eligibility criteria for this review, only one study measured the "isolated effect" of social media. The authors cautioned that in the case of the remaining studies it was difficult to assess whether a reported effect was related to a social media component alone or was a synergistic effect. This problem was also reported in other reviews of social media use in behavior change and health promotion, such as Schein et al [10] who reviewed the effectiveness of social media in public health communication, or the review by Hamm et al [6] who were focused on the behaviors of patients and caregivers.

There is also a lack of knowledge about the role of theory in the effectiveness of SNS-enabled interventions. Although studies have shown a positive effect of theory-driven Internet-based interventions on behavior change [11], there is little evidence in the context of SNS [1,2]. Understanding this phenomenon is important for the design of interventions. Yet, more research is needed to review existing evidence in this context and identify the type of theories and models currently used in the delivery of interventions through SNS, but also for the design of the social networking application itself.

Finally, previous literature reviews in the area of SNS for health promotion have focused on summative and outcome evaluations rather than formative and process assessments. For example, most reviewers in this field have attempted to examine the effect of SNS on objectively measured behavior change usually though the use of RCTs or some form of experimental study, like pre-test and post-test evaluations [1,2]. However, other factors that may have an important influence on the effectiveness of SNS, such as usability, user satisfaction, and level of technology acceptance or engagement, have rarely been synthesized. While these types of evaluation cannot provide direct evidence on effectiveness, they may provide very useful insights to guide future intervention development and implementation. For example, usability factors may influence which features of the delivered intervention are actually used, thus limiting its actual effectiveness. This type of information is usually included in research and technical papers reporting work in progress or complete research documenting the results of an iterative evaluation process. To date, a significant number of this type of studies has not met the eligibility criteria for inclusion in traditional RCT-focused systematic reviews.

Therefore, the aim and originality of this current review is to extend our knowledge about the effectiveness of SNS for health promotion by addressing some of these gaps in the existing literature, in particular, (1) extending the focus on effectiveness by reviewing studies reporting findings relevant to the usability, user satisfaction, acceptance, and level of engagement with SNS, as well as studies using different research methods and techniques, beyond traditional RCTs, to evaluate effectiveness, such as observational, qualitative, and pilot studies; (2) focusing on studies and findings that apply directly to the isolated effect of SNS (wherever this is possible); and (3) to investigate the extent to which theory has contributed to the design of SNS-driven interventions.

This paper is structured as follows. First, we present definitions of concepts that are central to this review. The next section presents the methods used to review the literature as well as the decisions made to select studies for review. In the following section, we present the findings of this review, while the final section includes a discussion and some conclusions.

Definitions

Our use of the term “social networking sites” (SNS) or “social networking” includes the broader concepts of Health 2.0 and Medicine 2.0. The definitions of these concepts have been previously reviewed [12]. They identify the two most important features as (1) patient/consumer participation and (2) Web 2.0 technology (user-generated content). There are several examples of different types of SNS that have been used for health promotion. For example, YouTube has been frequently used for the promotion of information about cancer screening, as well as obesity and dietary problems [13,14], Facebook has been used in interventions related to sexual health issues [15], and Twitter has been incorporated in the design of interventions about prenatal health promotion and education [16]. In addition to publicly available popular SNS (like Facebook), there is also a considerable number of standalone health-focused social networking applications used for conditions like obesity [17], healthy living [18], as well as various chronic diseases, like diabetes [19].

In the context of this review, the term “health promotion” is used in a broad sense to include health education initiatives (eg, in schools), social marketing campaigns (eg, using advertising), community development, and behavior change interventions (eg, smoking cessation websites). It can also take the form of educators in social networks to direct non-experts towards relevant and accurate health information. Agents with this role (which may be people or tools) have been called “apomediaries” [20]. Examples include knowledgeable collaborative filtering and recommendation agents. Despite the fact that health promotion is not synonymous with health prevention strategies, like social marketing and health education, in the context of our study, health promotion is used as an umbrella term to include also interventions grounded in social marketing and health education approaches. This decision was made because to date there are several successful examples of integrative health promotion interventions using social marketing methods and approaches, like audience segmentation [21,22], or health promotion interventions applying health education strategies to promote behavior change [23].

In this paper, we consider studies of “effectiveness” to encompass evaluation of measured behavior change (eg, RCTs and controlled studies), as well as aspects of the user experience and interaction with the SNS application that might help or hinder behavior change, such as usability, user satisfaction, technology acceptance, and level of engagement. “Usability” refers to the ease of use of the SNS application and is normally measured using behavioral metrics, like effectiveness, efficiency, learnability, and errors [24]. “User satisfaction” reports on the subjective satisfaction with the interface components of a given application [25]. “User engagement/adoption” includes the reporting of statistical figures about the level of adherence with a given intervention. This information may be reported both in terms of participation rate in the online intervention, but also in terms of Google analytics indicators, like number of hits or posts, and time spent. Finally, the term “technology acceptance” is used in a broad manner to include both the level of uptake of a given technology, but also more formal studies focused on modeling factors influencing user acceptance of technology, such as the Technology Acceptance Model [26].

Expectations about social networking, such as motivational support and peer-pressure, may be grounded in social or behavioral theories. For example, the Theory of Planned Behavior [27] predicts that norms of significant people in an individual’s social circles (subjective norms) have a strong influence on the individual’s behavioral intentions. Similarly, Social Cognitive Theory [28] predicts social learning by observation, which can take place in social networks. In the context of this review, the term “theory” is used broadly to include any theory used as the basis for the design of an intervention delivered through online social networking. In the absence of specific theory, we examined for the presence of a specific model or technological approach used to inform the design and delivery of interventions through SNS.

Methods

Overview

The narrative synthesis approach to literature review was used to analyze the existing evidence. This decision was made because the aim of this review was to synthesize evidence from a heterogeneous body of literature with studies representing different health promotion initiatives with a range of effectiveness evaluation measures and mixed-method research designs [29].

As guidance to this review, we followed the method of narrative synthesis prescribed by Rodgers et al [29]. Key elements of this method were (1) the development of a preliminary synthesis, and (2) the exploration of relationships (differences and similarities) within and between homogeneous groups of studies.

For the development of a preliminary synthesis, we used two techniques: (1) tabulation, as a means of extracting and organizing data from the primary studies in tables, and (2) grouping/clustering, which involved an interpretivist analysis of the contents of the primary studies in order to identify dominant groups of studies that shared a common set of characteristics. More details about the preliminary synthesis are presented in the following subsections. After the preliminary synthesis, the data collected were used to explore relationships between primary studies both at the individual and group level.

Scoping Search and Searching Process

We undertook an initial scoping search of the literature using Google Scholar. The purpose of this initial search was to gain a feel about the important aspects of the topic of this review, and more specifically to identify the different types of SNS available and to explore different areas of health promotion where SNS can play an important role. The results of the initial scoping review informed the design of our search strategy.

We searched Google Scholar and PubMed using a search strategy conceptualized as the following: Health AND "behavior change" AND <health promotion keywords> AND <social networking technology keywords>. The full search terms were health AND "behavior change" AND ("health promotion" OR "health education" OR "social marketing" OR "intervention" OR "persuasive" OR "therapy") AND ("social networking" OR "social media" OR "peer-to-peer" OR "online forum" OR "online community" OR "virtual community OR "online discussion" OR "electronic support groups" OR "participatory" OR "citizen-led" OR "web 2.0" OR "medicine 2.0" OR "user-generated content" OR "social software" OR "collaborative software").

The identification of a broad range of studies was one of the main challenges of this review. For this purpose, we decided to search using the Google Scholar (in addition to the PubMed database). Empirical studies [30,31] have shown that Google Scholar provides sufficient coverage to be used reliably in literature reviews of this kind. The date range was January 2005 to December 2013. Only articles written in English were included. Keyword searches were conducted in January 2014.

Inclusion/Exclusion Criteria

We included articles on health promotion (HP) interventions, where online SNS was a major theme in the study. In particular, these included the following: (1) Evaluation of interventions combining HP with SNS, including studies of effectiveness in terms of behavior change, usability, user satisfaction, level of engagement, and technology acceptance; (2) Observational studies of a social network within an existing HP intervention, including those involving content analysis, social network analysis or other usage patterns, but excluding studies of general social networks where health was one topic, unless the discussions were connected to an HP initiative; and (3) Designs and planned interventions were included if they addressed the relationship between HP and the anticipated emergent features of SNS. We also included papers reporting planned methodologies for the evaluation of interventions, as well as papers reporting work in progress, such as evaluation of early prototype designs. Information extracted from these papers contributed to our understanding of the different methods available for the evaluation of the effectiveness of interventions, and the presence of theories as evidence for guiding the design of interventions with an HP and an SNS component.

The following were excluded: mention of social networking in a generic, non-specific way; use of a discussion board as an "added extra" in an intervention without any significant role in the study; use of the term "social networking" to indicate "top-down" dissemination only (eg, using mobile phones or text messaging) without mention of peer-to-peer communication or other emergent SNS effects; study of health discussions on general social networks in which there is no HP initiative; and discussion/position papers, including definitions and research roadmaps (but some are cited as background).

Data Extraction and Synthesis Process

Two of the authors (PB and CK) performed the review working independently. They extracted data on effectiveness (broadly defined) and theoretical grounding. The items extracted are shown in Multimedia Appendix 1. Disagreements during the study selection and data extraction process were solved after consultation with the other authors (IB, JA, and JP).

We did not use a specific quality assessment tool due to the heterogeneity of study designs and the varying level of completeness of the studies included in this review. However, we did make individual assessments of the internal validity of the studies. In the results, we present the research design used by each selected study and the nature of the findings reported in the individual studies, including objectively and subjectively reported measures; long-term and short-term designs; strong and weak associations, or no associations (for observational studies); positive, negative, or mixed results (in the case of pilot and qualitative studies); and significant/not significant findings (for RCTs and controlled studies) (a detailed description is provided in Multimedia Appendix 1). This information was assessed during the tabulation process. Finally, we performed an interpretivist analysis to categorize primary studies into groups and examine the relationship between them.

Results

Overview

The search identified 162 potentially relevant documents after review of titles and abstracts. Of these, 42 satisfied the inclusion criteria after full-text review (Figure 1). Results on effectiveness, with details about the type of study design and main findings are shown in Table 1. The use of theory in interventions, as well as the extent of top-down, theory-based approaches, and bottom-up participation (observation) is shown in Table 2.

Table 1. Effectiveness evaluation (summary of study types and findings).

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
An et al, 2008 [32] (Quitplan)	Smoking cessation/adults	Active and passive online community participation	Observational study: Bi, multivariate, and path analysis to determine association between online activities and abstinence	Weak association between active community engagement and abstinence	SNS; Abstinence: +
Baghaei et al, 2009 [33] (SOFA)	Obesity/families	Motivational support; involve families	Pilot trial: will users engage with educational content? What kind of profile increases engagement?	Educational content attracted positive attention; individual profiles better than whole family	SNS+HP; Acceptance: +
Burke & Oomen-Early, 2008 [34]	General/ High School students	Blogging; community debates; advocacy campaigns	Education idea	N/A (concept only)	N/A
Cobb et al, 2010 [35] (QuitNet)	Smoking cessation/ QuitNet users	Online social support	Social network analysis: determine SNS effects (persistence, peer-to-peer communication, heterogeneity); compare with other SNS; characterize participants and subgroups	SNS effects are present; most integrated are female and older	N/A
Cunningham et al, 2008[b] [36] (Alcohol_HelpCenter)	Problem drinkers	Online social support	Usage patterns and message content analysis: determine quality of interactions	Qualitative: content appears valuable and supportive	SNS+HP; Acceptance: +
Falan et al, 2011 [37] (SCEDES)	Diabetics	Community support and education	Concept: minimize hospitalizations	N/A	N/A
Foster et al, 2010 [38] (StepMatron)	PA/ office workers	Social influence: competitive step-counting (FaceBook app)	Pilot trial:10 nurses, 9F, 1M	9/10 walked more in social condition than in non-social (Stat. significance tested)	SNS; Objectively measured behavior change (walking): +
Fukuoka et al, 2011 [39]	Diabetes prevention/ overweight, sedentary adults	Mobile peer to peer support	Qualitative focus-group analysis to determine desired features of planned mobile intervention	Real-time peer support emerged as desirable (also, tailored advice, self-monitoring)	N/A
Gasca et al, 2009 [17] (pHealthNet)	Obesity/ adults with weight-related health problems	Persuasive and SNS technology for existing support-groups (pedometer, Web portal, mobile app)	Field study of support groups: low sustainability of behavior changes; technology evaluation: 12 patients: compare behavior during and after technology-assisted group sessions (2 subgroups of 6)	Semi-quantitative: sustained PA changes 2 wks after technology-enabled session (3 wks). Positive acceptance of technology	SNS+HP; Observational study weak association (low sustainability of behavior change): +; Acceptance: +
Gay et al, 2011 [40] (AURORA)	Emotional awareness/ adults	Mobile sharing of emotions (Web and mobile app)	Pilot study, 65 adults, 7 days. Random (EMA) assessments and post-study survey	EMA and post-study results positive for emotion awareness, sharing and social support (also among strangers), but danger of negative contagion	SNS; Emotional health: + contagion danger: -
Kamal et al, 2010 [41]	Nutrition/ general	Theory-based social networking software	Prototype development	N/A	N/A

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
Kharrazi et al, 2011 [42]	Obesity/ general	Online sharing of progress and peer-pressure (Facebook app)	Technology design	N/A	N/A
Krukowski et al, 2008 [43] (VTrim)	Obesity/ adults	Weight loss websites with online social support as a feature.	Observational study: Determine what elements of a website (VTrim) are associated with actual weight loss. Exploratory factor analysis; 123 overweight adults; 1 yr: treatment: months 0-6; maintenance months 7-12	In maintenance phase, “social support” was best predictor for additional weight loss. “Feedback” was best predictor during initial phase	SNS; Weight loss maintenance: +++
Lindsay et al, 2009 [44]	Exercise, smoking, diet/ coronary heart patients in deprived urban area	Online support community	RCT: determine effects of removing moderator support from online community: 108 participants, 12 months, non-moderated phase after 6 months; randomly assign half to Web-portal access and half to non-Web portal group	Significant reduction in self-reported health behaviors 3 months after moderator withdrawal (for both groups); during moderated phase, Web portal access led to positive behavior changes	HP; Self-reported health behavior: +++
Linehan et al, 2010 [45] (Tagliatelle)	Obesity/ adults	Social photo tagging of meals for nutritional content	Pilot usage and acceptability study: 14 participants	9/14 participants regularly used system over 7-day trial	SNS; Acceptance: +
Liu & Chan, 2010 [46]	General health	Seeking help in virtual communities	Research design: determine relation between social identity, beliefs, and help-seeking behavior (planned survey)	N/A	N/A
Maibach et al, 2007 [47]	General health	Social networks as ecological fields of influence	Conceptual framework for social marketing to mobilize health-promoting dynamics in social networks	N/A	N/A
Munson et al, 2010 (3GT) [48]	Positive psychology/ adults	Facebook app (3GT) for sharing positive experiences (“good things”)	Survey of 3GT users (190 participants) to record usage patterns and attitudes	Positive acceptance of app, but concern about privacy; indifference about reminders	SNS+HP; Acceptance: +/-
Nahm et al, 2009 [49] (TSW)	Hip fracture prevention/ older adults	Educational discussion board	Exploratory qualitative analysis (316 forum posts; 245 participants)	Emergent themes included sharing of health behaviors, problems, and opportunities; also social support	N/A
Nordfelt et al, 2010 [19] (Diabit)	Diabetes/ children and parents	Peer-to-peer chat and blogging on a Web 2.0 portal	Qualitative content analysis of essays written by portal users (19 parents, 5 young people 11-18 years)	Message boards and chats found to provide valuable information that could not be provided by clinicians (attitudes to website itself were mixed)	N/A
O’Grady et al, 2008 [50]	General health	SNS for collaborative learning	Proposal of Experiential Health Information Processing Model	N/A	N/A

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
Olsen & Kraft, 2009 [51]	General health	SNS role in providing social support and adherence	Pilot study to determine which aspects of SNS are important in social support and adherence (semi-structured interviews, 5 participants, qualitative analysis)	Social support provided mostly by close friends or family; adherence may be improved with dynamic and interactive features (eg, games, contests)	N/A
Potente et al, 2011 [52]	Sun protection/ Australian youth	Social Media Marketing (SMM)	Online survey and thematic analysis of comments to determine effects of an SMM music video on attitudes and risk-awareness	Positive stat. significant difference in attitudes between video-exposed respondents and non-video-exposed	SNS+HP; Self-reported risk-awareness: ++
Rhodes et al, 2010 (CyBER/ M4M) [53]	Human immunodeficiency virus (HIV) prevention/ men who have sex with men (MSM)	Educators in Internet chat rooms	Quantitative analysis of participant survey (n=210); qualitative analysis of chat content (n=1851): private and public messages	Inconsistent condom use: 27% (77% of HIV positive chatters): Qualitative: need for prevention information; privacy, and trust important; educators had to respect culture	N/A
Richardson et al, 2010 [54] (Stepping Up to Health - SUH)	PA/ adults	Online community in Stepping Up to Health website	RCT: effect of online community in website. n=324; (5:1 randomization, larger number in community condition); Objective measures: pedometer data, community usage (activity) and intervention completion rates	Online community more engaged and more likely to complete intervention than non-community; otherwise no great difference in walking. However, within online community, active participants (with more posts and page views) walked more than less active participants	SNS; Adherence: ++++
Roblin, 2011 [55]	Diabetes/ patients and families	Mobile peer support for glucose management	Pilot study: experience of patients and their peer supporters using mobile technology for encouraging and reminding	Self-reported improved self-monitoring and encouragement through mobile communication with peer-supporter	SNS+HP; Acceptance: +
Stoddard et al, 2008 [56] (Smokefree.gov)	Smoking cessation/ adults	Bulletin board in website	RCT: effect of bulletin board (BB) in website. n=1375 (50:50 allocation BB vs usual)	In BB condition, only 11% posted or viewed messages; no significant difference in cessation; more time on website for BB condition; no difference in satisfaction	SNS; Abstinence: 0
Toscos et al, 2010[b] [57]	Barriers to Physical activity/female forum users of *GetFit!*	Online forum on PA	Qualitative Analysis of *GetFit!* Forum content; compare with literature survey on barriers.	Differences between PA barriers emerging in forums and those from surveys; *GetFit!* intervention not aware of them	N/A
Waters et al, 2011 [58]	Student health	Facebook profiles of University Health Centers	Content analysis to determine the extent of “dialogic principles” (eg, usability, conversation of visitors, feedback options)	Least applied dialogic principles were feedback options (contact details) and promoting return visits. Significant relation between social networking extent (friends, fans) and use of dialogic principles	N/A

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
West et al, 2011 [59]	Breastfeeding	Blogging	Determine extent of blogging to support breastfeeding behavior: qualitative and quantitative analysis of posts and comments; 32 active blogs, 354 posts, 881 comments	Reports on one's own behavior and personal experience sharing were more likely to elicit behavioral intention than advice or information. Attitude (like/dislike) most common theme in blog posts (28%); praise (support) for breastfeeding most frequent comment (43%)	SNS; Behavioral intention: ++
Woodruff et al, 2007 [60]	Smoking/ adolescents	Virtual chat room	RCT: determine effect of intervention with MI and virtual chat room (n=136)	Short-term: self reported smoking reduction for intervention group; long-term: not significant	SNS+HP; Self-reported behavior: ++
Young et al, 2010 [61]	PA/ teenage girls	Micro-blogging	Pilot study: 4 students; determine if peer-pressure and SNS technology can influence girls to exercise	Positive behavior change, gradual increase in number of steps over 4 weeks	SNS; Behavior change: +
Kamal et al, 2013 [18]	Healthy living / Adults	VivoSpace	Pilot study: interviews, questionnaires, and prototyping. Aim was evaluation in terms of usability of a novel theoretical framework (Appeal, Belonging, Commitment) for design of a social networking tool for healthy living	Findings showed ABC framework in combination with iterative usability evaluation to be promising for user engagement; but, since the study was focused on prototypes and not fully working systems, no tangible data on actual nature of engagement and its effect on health behavior change	SNS +HP; Engagement: +
Baelden et al, 2012 [62]	Acquired Immune Deficiency Syndrome (AIDS) and HIV/ Adults	Online discussion group	Pilot study: examining suitability of an anonymized discussion forum for increasing interpersonal communication and engagement in the area of HIV / evaluation through usage statistics & focus group interviews	Mixed on suitability of online discussion forums for interpersonal communication about AIDS. Use of discussion forum was successful when integrated into the curriculum. Usage was lower when participants had to use the forum on a voluntary basis	SNS; Adherence and technology engagement: +/-
Ploderer et al, 2013 [63]	Smoking cessation / Adults	Facebook support group	Pilot study: Examining the relationship between stage of health identity change and seek for social support / thematic analysis of messages posted in a public Facebook support group	Findings showed that supportive responses and leadership came from users who just started their behavior change process rather than people who had successfully completed it	SNS + HP; Self-reported behavior change: ++
Gold et al, 2012 [64]	Sexual health / Young people	Facebook + YouTube	Pilot study: Review of challenges related to promotion of sexual health behavior through Web 2.0 / usage statistics, satisfaction questionnaires, and focus groups	Mixed results in terms of adherence and engagement with technology	SNS; Adherence or technology engagement: +/-

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
Nguyen et al, 2013 [65]	Sexual health / Young adults	Facebook + SNS	Pilot study: Review of challenges related to promotion of sexual health behavior through Web 2.0 / usage statistics and questionnaires	Mixed results on effectiveness. The project reached 900 fans across 5 Facebook pages. Key challenges included a lack of viral recruitment, evoking substantial interest, and maintaining user engagement	SNS; Adherence or technology engagement: -
Kolt et al, 2013 [66]	Physical activity	Walk 2.0 project (blogs, social networking, virtual walking groups, forums)	RCT: A methodology to compare the effectiveness between Web 1.0, Web 2.0 and control interventions) using larger sample size and repeated measures data collection	N/A (the paper presented the methodology of the evaluation, but no results were presented or discussed)	SNS; Self-reported behavior change: +; Objectively measured behavior change: + ; Engagement: N/A
Gabarron et al, 2012 [67]	Sexual health / Young adults	Virtual Clinic for Sexually Transmitted Diseases (VCSTD) / Avatars	Impact evaluation: Methodology to examine usefulness of service / user experience through online feedback forms—behavior change through online questionnaires—usage data / effect of the interventions on (1) number of abortions, (2) number of chlamydia tests, (3) amount of emergency contraception information sold	N/A (presented the methodology of the evaluation, but not the results)	SNS; Acceptability/ user engagement: N/A Self-reported behavior change: N/A
Kelty et al, 2012 [68]	Physical activity/ teenage girls	Facebook / “Girls’ recreational activity support program using information technology”	RCT: evaluating a baseline intervention (based on face-to-face support) and an intervention based on Facebook pages; data collected during a 3-month period. Study aimed to evaluate the effectiveness of social networking intervention for improving physical activity and behavior change, as well as the feel of support to the users of the service	Although intervention group increased physical activity, the difference between the 2 interventions was not significant. Engagement with the online component was low. Additional strategies are required to improve engagement and compliance with social networking interventions based on Facebook	SNS; Adherence-engagement: +; Objectively measured behavior change (based on physiological data, BMI): ++++
Laakso et al, 2012 [69]	Self-management of chronic disease	HOFA (Healthy Outcomes for Australians): Social media platform for information sharing, community building, and social networking for those with chronic disease	Lit review: No evaluation of effectiveness. Lit review informed the design of the intervention. Paper presents the results of the review and a general description of the HOFA website	N/A (paper included a review of the relevant literature)	N/A

Reference/ project or intervention name	Health topic/ Study population	Social networking topic/key words/ technology	Type of study/methods	Main findings	Effectiveness evaluations (if any)[a]
Hwang et al, 2012 [70]	Weight loss/ Adults	SparkPeople.com/ Discussion forum and blogs	Observational study: finding an association between frequency of use of social media & social support in the context of weight loss/ survey	Using social media tools of an online weight loss program at least 1x/wk is strongly associated with receiving encouragement, but not information or shared experiences	SNS; Self-reported behavior change: ++

[a] Abbreviations and symbols used in this column are explained in Multimedia Appendix 1.

[b] Conflict of interest declared.

Figure 1. Flow of studies through the review.

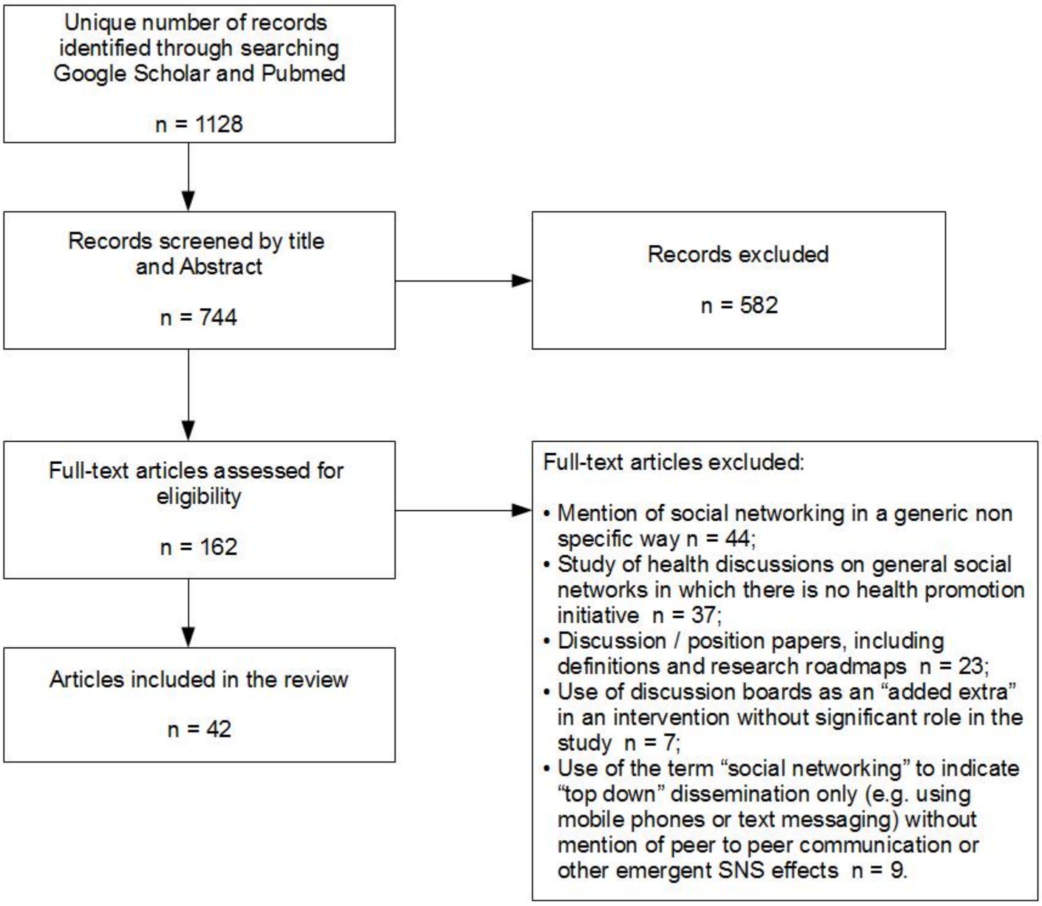


Table 2. Role of theory and relationship between top-down and bottom-up features in interventions.[a]

Reference/ intervention name	Theories or models used (if any)	Role of top-down design (HP) in intervention	Role of bottom-up or emergent SNS features	Citizen-led or participatory elements (if any)	Relation between HP and SNS in study
An et al, 2008 [32]	N/A	Quitplan website	Observed usage patterns	N/A	[b]HP ⟵⟶ SNS
Baghaei et al, 2009 (SOFA) [33]	N/A	Educational content	Usage patterns	N/A	HP ⟵⟶ SNS
Burke & Oomen-Early, 2008 [34]	Bloom's Taxonomy of Educational Objectives	High school teaching idea (guided use of SNS)	Learning from SNS expected	Students learn advocacy campaigning and citizen debates	HP ⟶ SNS
Cobb et al 2010 (QuitNet) [35]	Social Network Analysis	Design of smoking interventions	Social networking analysis results inform HP	N/A	HP ⟵ SNS
Cunningham et al, 2008[c] (Alcohol_HelpCenter) [36]	N/A	Expert forum moderator	Observed usage and content inform HP	N/A	HP ⟵ SNS
Falan et al, 2011 (SCEDES) [37]	N/A	Nurses, educators in community	Planned bottom-up flow of knowledge	Planned consumer empowerment	HP ⟵⟶ SNS
Foster et al, 2010 (StepMatron) [38]	N/A	Design of intervention	Peer pressure	N/A	HP ⟵⟶ SNS
Fukuoka et al, 2011 [39]	N/A	Planned anti-diabetes intervention	Planned social support in community	Focus group emergent themes help determine intervention	HP ⟵⟶ SNS
Gasca et al, 2009 (pHealthNet) [17]	N/A	Design of intervention based on existing hospital support groups	Peer-to-peer challenges, games, experience sharing, community attachment	Researchers consulted support groups to determine technology design	HP ⟵⟶SNS
Gay et al, 2011 (AURORA) [40]	N/A	Design of intervention based on effects of emotional health on physical health	Visual emotion sharing (selecting Flickr pictures)	N/A	HP ⟵⟶ SNS
Kamal et al, 2010 [41]	Social Science Theories (U&G; CICB; SI; OC; SNT; DI) and Behavior Change Theories (TTM; HBM; SCT; TRA)	Intervention design based on survey of models and theories	Planned SNS should promote social belonging, identity and comparison (grounded in theories)	N/A	HP ⟶ SNS
Kharrazi et al, 2011 [42]	TPB	Educational materials + pedometer linked to personal health record	Planned SNS should enable peer pressure, competition, and rewards	Interactive personal health record should empower consumer	HP ⟶ SNS
Krukowski et al, 2008 [43]	N/A	Website design with educational content	Bulletin board, Web chats, stories, biographies	Focus groups help to determine website features	HP ⟵⟶ SNS
Lindsay et al, 2009 [44]	N/A	Moderator support	Online closed community	N/A	HP ⟵⟶ SNS
Linehan et al, 2010 [45]	N/A	Planned intervention for general nutrition education	Participants upload photos of meals to be tagged anonymously for nutrition value	Nutrition tagging generated by participants	HP ⟶ SNS
Liu & Chan, 2010 [46]	Social Support Theory; Social Identity Theory (SI); HBM	Virtual community management based on theories and evidence	Observed social support patterns in SNS inform interventions	N/A	HP ⟵ SNS

Reference/ intervention name	Theories or models used (if any)	Role of top-down design (HP) in intervention	Role of bottom-up or emergent SNS features	Citizen-led or participatory elements (if any)	Relation between HP and SNS in study
Maibach et al, 2007 [47]	Ecological models: people-based and place-based fields of influence	Planned framework for Social Marketing to promote behavior change in SNS	Theory of SNS as people-based fields of influence	Participatory model considered	HP → SNS
Munson et al, 2010 (3GT) [48]	Positive Psychology	Encouraging sharing of positive events in SNS	Real attitudes of SNS users	N/A	HP ←→ SNS
Nahm et al, 2009 (TSW) [49]	Social Cognitive Theory	Theory-based website with moderated discussion	Emerging themes from discussion	N/A	HP ←→ SNS
Nordfelt et al, 2010 (Diabit) [19]	N/A	Educational materials on website	Attitudes from essays written by participants	Attitudes and suggestions provide input for further development of website	HP ← SNS
O’Grady et al, 2008 [50]	Kolb Model of Experiential Learning	Design of collaborative health education	Harnessing of SNS technology to support learning	Patients may be considered as authoritative due to their experience	HP → SNS
Olsen & Kraft, 2009 [51]	N/A	Future designs based on observed SNS features	Aspects of SNS perceived by users as promoting social support and adherence	Attitudes of SNS users provide input to technical design of SNS technology (positive and negative experiences/ concerns)	HP← SNS
Potente et al, 2011 [52]	N/A	Social Marketing use of social media	Sharing and debating video online (YouTube, Twitter, forums)	N/A	HP ←→ SNS
Rhodes et al, 2010 (CyBER/ M4M) [53]	Social Cognitive Theory (SCT); Grounded Theory used for data analysis	Chat room educators	Observed chat rooms interactions with educators inform intervention design	Methodology: Community-Based Participatory Research (CBPR)	HP ←→ SNS
Richardson et al, 2010 [54] (SUH)	SCT	SUH intervention	Observed community engagement and peer support	N/A	HP ←→ SNS
Roblin, 2011 [55]	Social support	Planned diabetes intervention	Peer-to-peer mobile messages	Participatory model for diabetes management	HP ←→ SNS
Stoddard et al, 2008 [56] (Smokefree.gov)	N/A	Smoking intervention	Observed bulletin board usage and effectiveness	N/A	HP ←→ SNS
Toscos et al, 2010 [57]	For qualitative analysis: Presentation of Self in Everyday Life & Cognitive Dissonance	Future designs based on SNS observations	Commonly mentioned barriers to PA in forum to inform HP design	N/A	HP ← SNS
Waters et al, 2011 [58]	Dialogic Theory	University Health Centers	Health Center SNSs’ use of Dialogic Principles	N/A	HP ←→SNS
West et al, 2011 [59]	Integrated Behavioral Model (IBM): to code constructs for behavioral support.	Health education on breastfeeding	Observed peer support via blogging to inform HP interventions	N/A	HP ← SNS

Reference/ intervention name	Theories or models used (if any)	Role of top-down design (HP) in intervention	Role of bottom-up or emergent SNS features	Citizen-led or participatory elements (if any)	Relation between HP and SNS in study
Woodruff et al, 2007 [60]	MI	MI used within virtual chat room	Peer pressure and social support	Participatory research involving schools and academics	HP ←→ SNS
Young et al, 2010 [61]	Persuasion Design Principles (PSD)	PA website with pedometer	Harness peer pressure using microblogging	Teenagers were consulted about design principles	HP → SNS
Kamal et al, 2013 [18]	ABC: A theoretical framework encompassing concepts from 13 individual theoretical models	Design & content components of a social networking tool were informed from the ABC theoretical framework	N/A (the study involved only a prototype)	Researchers involved users in the prototype design and evaluation phase	HP → SNS
Ploderer et al, 2013 [63]	N/A	Smoking cessation Facebook support group	Analysis of posts made to a Facebook support group by 180 users	Analysis of users’ posts	SNS → HP
Baelden et al, 2012 [62]	N/A	Design of the tool was based on participatory approaches	Observation of usage statistics following 3 implementation scenarios: (1) voluntary (with 15,000 users), (2) semi-voluntary (with 1431 users), & (3) curriculum integration (with 161 users). Each implementation phase lasted ~1 month	Researchers involved users in prototype design and evaluation phase (through focus group interviews)	HP← SNS
Gold et al, 2012 [64]	N/A	Design of intervention was based on collaboration between public health professionals, experts in user experience, and people from creative industries	Observation of usage statistics	N/A	HP ← SNS
Nguyen et al, 2013 [65]	Concept of edutainment	Design of tool was based on the concept of edutainment	Observation of usage statistics + online surveys	N/A	HP ← SNS
Kolt et al, 2013 [66]	N/A	N/A	Observation of participants self-reported behavior including data on physical activity levels, self-reported quality of life, user satisfaction, psychosocial correlates	N/A	SNS → HP
Gabarron et al, 2012 [67]	Gaming and eLearning approach	Design of tools involved an avatar, which was influenced by gaming and eLearning concepts	Feedback forms; online questionnaires and publicly available usage data	N/A	SNS → HP

Reference/ intervention name	Theories or models used (if any)	Role of top-down design (HP) in intervention	Role of bottom-up or emergent SNS features	Citizen-led or participatory elements (if any)	Relation between HP and SNS in study
Kelty et al, 2012 [68]	N/A	N/A	Objectively measured effect (eg, use of pedometers; digital scales, calculation of BMI and CRF scores)	N/A	SNS → HP
Laakso et al, 2012 [69]	N/A (based on lit review of the barriers to accessing and managing health information)	Interdisciplinary input from specialists in physiotherapy, exercise science, nutrition, education, human services, psychology	N/A	N/A	N/A
Hwang et al, 2012 [70]	N/A	N/A	Questionnaire survey, interviews, qualitative analysis of posts in discussion forums	N/A	SNS → HP

[a] PA: Physical activity; Social Science theories: U &G: Uses and Gratification theory [71], CICB: Common Identity and Common Bond theories [72], OT: Organizational Commitment theory [73], SI: Social Identity theory [74,75], SST: Social Support Theory [76,77], SNT: Social Network Threshold [78], DI: Diffusion of Innovation theory [79]; Behavior change theories: SCT: Social Cognitive Theory [28], TTM: Transtheoretical Model [80], TPB: Theory of Planned Behavior [27], TRA: Theory of Reasoned Action (see TPB), HBM: Health Belief Model [81], MI: Motivational Interviewing.

[b] The following notations have been used to denote the relationship between HP and SNS in the study: HP ⟷ SNS (emphasis on top-down design); HP ← SNS (emphasis on bottom-up flow of knowledge through observation and/or participation); HP → SNS (both aspects included in the study).

[c] Conflict of interest declared.

Effectiveness Studies

Overview

A total of 26 studies (Table 1) had an explicit focus on effectiveness. These were RCTs (n=6), fully powered and explicitly designed observational studies (n=5), and pilot studies (n=15). A total of 17 articles (Table 1) did not report results on the effectiveness of social networking for health promotion. The studies presented in these articles were either planned interventions, conceptual frameworks, and early prototypes—usually coupled with findings from a literature review [34,37,39,41,42,46,47,50,58,67,69] or showed results other than those related to the measurement of the effectiveness of social networking applications. For example, findings were focused on the information seeking and sharing behavior of users of social media, or the application of social network analysis to show the growth and characteristics of Web 2.0 applications [35,49,19,51,53,57]. The main findings of the 26 studies with a focus on effectiveness are summarized below.

Randomized Controlled Trials

Six studies were RCTs [44,54,56,60,66,68]. Of these, three studies [54,66,68] examined the effect of online social networking on objectively measured behavior, while the remaining studies attempted to examine this effect on self-reported behaviors. In the case of objectively measured behaviors, Kolt et al [66] presented the methodology, but not actual results from the study. Richardson et al [54] and Kelty et al [68] showed no significant effect on physical activity (in terms of walking behavior) between the baseline and online social networking interventions. However, the two studies showed mixed results in terms of the level of engagement and adherence with socially mediated interventions. Richardson et al [54] reported a positive effect of an online community on adherence (ie, engagement and completion of the intervention) while Kelty et al [68] showed a low level of engagement.

Researchers who examined self-reported behavior change using RCTs presented a mixed picture of online social networking versus behavior change in the context of smoking cessation, healthy eating, and physical activity. Stoddard et al [56] measured the effect of a bulletin board on smoking abstinence (n=1375, 50:50 allocation to bulletin board vs usual care)—only 11% in the intervention arm viewed or posted to the bulletin board, and no significant effect was found. Woodruff et al [60] found a short-term self-reported effect on smoking abstinence. However, the study evaluated the whole intervention (which included motivational interviewing) thus making it difficult to determine the effect of the social networking aspects. The effect of a specific HP component in a health care social network was evaluated by Lindsay et al [44], who studied the effect of removing a moderator from an online community. The 12-month study involved 108 coronary heart patients, half of whom were randomly assigned to Web portal access. For both groups, moderation was removed after 6 months. After 3 months of non-moderated usage, there was a significant reduction in self-reported healthy behaviors for both groups. During the moderated phase, there was a positive effect for the portal (intervention) group.

Observational Studies

Four studies determined effectiveness through controlled observational designs. An et al [32] found a weak association between community engagement and abstinence (smoking) using multivariate and path analyses. Krukowski et al [43] used exploratory factor analysis to determine which website features were associated with actual weight loss (n=123). “Social support” was the highest predictor. Similar findings were presented by Hwang et al [70]. The researchers found that using the social networking tools of an online weight loss website was strongly associated with receiving encouragement and support from the community. However, no strong associations were observed between the use of social networking tools and the amount of new information or shared experiences received. Ploderer et al [63] examined the relationship between stages of health identity change and seeking social support. They performed a quantitative analysis of messages posted in a public Facebook support group for smoking cessation. The findings showed that supportive responses and leadership came from users who had just started their behavior change process rather than people who successfully completed it. Finally, West et al [59] performed both qualitative and quantitative analyses of a large set of blog posts to determine whether blogging can promote breastfeeding. The findings showed that sharing personal experiences was more likely to elicit behavioral intention than generic advice or information.

Pilot Studies

A total of 14 articles examined the effectiveness of social networking interventions in studies that were pilots (with regard to the power to detect the effect of interest) or qualitative explorations. In the majority of cases, researchers recruited small sample sizes and employed mixed (qualitative and quantitative) methods. Typical data collection techniques were focus groups, online questionnaire surveys, interviews, and quantitative analysis of user-generated content (such as posts in blogs, discussion forums, and other social networking sites).

Nine studies [18,33,36,38,40,45,52,55,61] showed a positive effect of social networking interventions on engagement/acceptance of technology and behavior change. In particular, several studies [18,33,45] showed that social networking interventions enhanced user engagement and acceptance of technology in the contexts of obesity, healthy eating, and physical activity. Similar findings were reported in the case of interventions related to alcohol misuse and diabetes [36,38]. In addition to positive user engagement, two studies [38,61] demonstrated promotion of walking (gradual increase in the number of steps). Positive behavior changes were self-reported [40,52]. Gay et al [40] focused on the application of social networking in the context of emotional health. The results were positive for emotion awareness, sharing, and social support. Finally, Potente et al [52] showed a high level of self-reported risk awareness in the context of sun protection.

The remaining five studies [17,48,62,64,65] presented mixed results regarding the effectiveness of social networking interventions in health promotion. Several studies [62,64,65] were focused on sexual health promotion (including HIV protection). The findings of these studies showed that social networking can be a useful tool for initiating online discussions. However, several limitations were identified, such as low level of participation and engagement on a voluntary basis, lack of expected “viral” recruitment through online networks, and problems maintaining user engagement in the long term. In addition to sexual health, two studies [17,48] that were focused on obesity and emotional health reported similarly mixed effectiveness. In particular, Gasca et al [17] showed a high level of acceptance of technology, but the authors reported also that social networking did not support long-term behavior change (ie, low sustainability of behavior change). In Munson et al [48], the positive engagement with technology was counteracted by concerns about privacy and personal information management.

Theoretical Grounding

Twenty studies involved interventions that were grounded in social and psychological theories, or technological model and approaches. Most of these were early stage designs that we classed as top-down studies in Table 2. Many were based on the expected emergent properties of social networks. In particular, Kamal et al 2010 [41] grounded their intervention design on a survey of theories relating to social networking and behavior change. The social networking theories employed were Uses and Gratification (U&G) theory [71]: participants use media actively and search for specific resources (for usefulness or gratification); Common Identity and Common Bond (CICB) theories [72]: online communities need to be managed in a way that facilitates attachment to a group (Common Identity) and attachment to group members (Common Bond) in order to sustain voluntary participation; Organizational Commitment theory (OT) [73]: a model of different kinds of commitment (or attachment) to an organization, which can be relevant to an online community; Social Identity (SI) theory [74,75]: motivation for behavior change is influenced by the sense of belonging to a group; Social Support Theory (SST) [76,77]: in social networks, social support might take the form of messages showing empathy, encouragement and caring (among others), which may be beneficial for health and positive mental attitude, including motivation for behavior change; Social Network Threshold (SNT) [78]: this theory distinguishes critical/threshold numbers of individuals’ contacts influencing their adoption behavior from the effects of structural aspects regarding individuals’ positions in social networks; and Diffusion of Innovation (DI) theory [79]: populations comprise a theoretical distribution of people with different propensities for adopting innovations, from “innovators” and their “early adopters” to “laggards”.

The planned social network should promote a sense of belonging and social identity (based on SI and CICB theories) as well as social support (based on SST) among other features. Social support theory was also applied in other interventions [46,55]. In a follow-up paper, Kamal et al [18] summarized the individual theoretical models into the ABC framework. This informed the design of the VivoSpace, a social networking tool focused on healthy living.

Other theories used were as follows: People-based and Place-based fields of influence, where people are influenced by the places they are in, as well as other people (norms, etc) [47]; Positive psychology [82], used by Munson et al 2010 [48] (3GT), in which sharing of positive stories and experiences promotes emotional health (acceptance evaluation); Social

Cognitive Theory used for the whole intervention design in three studies with moderated discussion [49,53,54]; Theory of Planned Behavior, in which peer-pressure (norms) should emerge in planned social network for sharing step count data [42]; Kolb Model of Experiential Learning [83], in which learning happens through experience, and experience sharing [50]; Dialogic Theory [84] used in one study [58] to evaluate university health center use of Facebook; Motivational Interviewing (MI) used for chat room educators [60]; and Persuasion Design Principles (PSD) used for website design [61].

A few studies were not theoretically grounded but instead based on commonly held expectations about the effects of social networking. For example, AURORA [40] was focused on the expected positive effects on emotional health if positive experiences are shared. However, this can also be negative, due to contagion of negative emotions. Another was Tagliatelle [45], which is based on the expectation of constructive social tagging of meals. Nguyen et al [65] designed an intervention for sexual education using Facebook. The intervention followed the concept of edutainment to support adherence and engagement. Finally, the Virtual Clinic for Sexually Transmitted Diseases [67] was an Avatar-supported intervention, the design of which was based on concepts from gaming and eLearning to support adherence and promote behavior change among the users of the service.

Discussion

Principal Considerations

The aim of this study was to review the existing evidence about the effectiveness of SNS in health promotion. As opposed to existing systematic reviews, this study took a different approach by including a broader range of studies for review. The selected papers reflected different dimensions of effectiveness and types of a research design. This decision was made in order to address some of the gaps identified in previous reviews of the relevant literature, and in particular, the focus on RCTs (ignoring other types of research designs), as well as the narrow focus of effectiveness on behavior change (excluding other types of effectiveness that may have an impact on our understanding of behavior change, like usability, user satisfaction, level of adherence, and technology acceptance). By reviewing a larger pool of papers in this context, our objectives were to extend our existing knowledge about how effectiveness is being measured and identify the level of uptake of theories in the design of interventions based on online social networking.

Effectiveness of Social Networking Sites

In accordance with findings from previous reviews [1,2], the RCTs included in this review showed no clear effect of SNS on objectively measured behavior change (eg, no significant increase in walking behavior in the context of obesity-related interventions [54,68]). However, more positive effects on both self-reported and objectively measured behavior change were reported in the case of small pilot studies [38,61]. It is well recognized that small pilot studies often show a more promising positive effect of an intervention than later larger and more pragmatic evaluations [85].

The review of controlled observational studies showed some interesting aspects about the role of social support in behavior change. It appears that not all aspects of SNS (eg, social support, peer pressure, or information sharing) have an equal role. In particular, social support was the highest predictor of behavior change in the context of weight loss [43]. Also, the use of SNS in weight loss interventions was more strongly associated with receiving encouragement and support from the community rather than the amount of new information and experiences received [70]. Finally, there was evidence that social support is not manifested equally among members of an online community. The level of completion of behavior change appeared to be an important predictor of social support, with users who had just started their behavior change being more supportive than their peers who successfully completed it [63]. In previous reviews of the literature [1,2], social support was identified as a positive aspect of interventions delivered through SNS. However, this review goes a step further by highlighting its role in relation to other aspects of SNS, like peer pressure and information sharing, but also among different members of the online community. Future research should investigate in more depth the role of social support as a specific component of health promotion interventions and for interface design. For example, what is the effect of different contextual factors on online social support? Or how can the interface design of SNS applications be enhanced with features that could motivate social support among different members of the online community?

Broader influences on effectiveness, such as usability or level of engagement, were reported more frequently in pilot studies, rather than RCTs and observational research. The majority of pilot studies showed results about the level of engagement with an online social networking application over a short period of time (normally between 1-4 weeks). Despite the fact that all authors reported systematically a good level of engagement at the beginning of the trial period, in many cases the number of active users dropped considerably in the long term [17,48,62,64,65]. Only a few authors attempted to explain the reasons for this phenomenon. However, when this information was reported, the most common reasons included concerns about privacy, problems related to personal information management, and lack of motivation [48,53]. Only in one pilot study did the authors examine what actions should be taken to improve the level of adherence and engagement with SNS [51]. They found that dynamic and interactive elements (such as online games and contests) could improve adherence. The lack of active participation and long-term engagement with SNS technology was an issue also in the case of RCTs. For example, Stoddard et al [56] reported that only 11% of participants were active users (ie, posted or viewed comments/messages), while Woodruffe et al [60] found a significant self-reported behavior change only in the short term. A reduction in the level of engagement in RCTs has been reported by other authors as well [2,86,87]. Also, it is interesting that almost all RCTs in our review, except for one, did not exceed a 12-week trial period. This shows a lack of evidence about the level of user engagement and retention in the case of longer trial periods (such as 12 months or more). The lack of long-term RCTs (ie, more than a year) is a typical phenomenon in this context and similar concerns regarding long-term user engagement and retention have been expressed by other authors in the past [2].

Lack of clear evidence was evident in the case of the evaluation of the usability and technology acceptance of the SNS. Despite the fact that usability was frequently mentioned in several papers as a feature of a well-designed social networking application, there was no evidence of complete usability tests or heuristic evaluations. In the majority of cases, usability was reduced to the evaluation of the quality of the contents and information in an SNS [58]. In other cases, some authors reported the application of a participatory design approach to inform the development of usable interfaces for SNS. This was more common in interventions with a health-focused SNS component rather than the mainstream SNS channels, like Facebook. Evaluating the usability (ie, interface design) of SNS applications is important for both user engagement and behavior change [88]. Also, this type of evaluation will provide some of the evidence needed by informaticians to design ease-of-use SNS interfaces for health promotion interventions. Finally, the review showed a lack of studies examining technology acceptance (ie, studies focused on identifying and modeling factors of technology acceptance and intention to use the specific technology).

Use of Theory in the Design of Social Networking Sites

As opposed to previous reviews of the literature [2], the papers included in this review showed a wider range of social and behavioral theories and design approaches used to inform the design of interventions. This finding shows that more researchers are choosing a more theory-driven approach as a means of achieving powerful effects [11]. Although a wide range of theories were mentioned in the studies, the social networking concepts that they emphasized were often overlapping. The most common were peer pressure, social support, and sense of identify (ie, belonging to a community).

Of the 20 papers that showed evidence about the use of theory to inform the design of interventions, the authors evaluated effectiveness in only half. In the context of physical activity, smoking cessation, and diabetes, the findings showed a positive effect of interventions grounded on persuasion design [61], motivational interviewing [60], and social support theory [55] on behavior change (both self-reported and objectively measured) respectively. Also, interventions based on the ABC framework [18] and positive psychology [48] showed good level of engagement and a positive effect on behavioral intention to share personal experiences. Positive, but not statistically significant, effect on behavior change was reported by authors who applied social cognitive theory to the design of an SNS intervention for physical activity [54], while the results were mixed in terms of engagement in the case of an SNS intervention grounded on the concept of edutainment [65].

Theories were used a priori to inform the design and contents of the online intervention. However, in the majority of cases, authors were not clear as to which aspects of the theory were applied specifically for the delivery of the SNS. This was common for interventions encompassing a website, part of which was the social networking application. In a few cases, the researchers also used bottom-up approaches to enhance the design and the contents of SNS. Bottom-up approaches were based on the use of observation using information extraction tools and social network analysis [49,59,57,35]. Yet, no study showed clearly how both top-down and bottom-up approaches to the design of health promotion interventions can be integrated into an iterative design life-cycle or how top-down design of health promotion can be linked with bottom-up observation and user participation.

Limitations

This review has several limitations. Only articles indexed in Google Scholar or PubMed were included. However, most academic publications are found by Google Scholar. We did not include gray literature such as white papers and unpublished reports. In addition, our search terms may have missed some relevant articles, especially in the context of health prevention and preventive strategies. However, health prevention was not the focus of this review and a decision was made to include in the search for relevant papers only terms representing prevention strategies that are known examples of integrative (mixed) health promotion interventions, that is, health promotion interventions that incorporate methods from prevention strategies, like social marketing and health education. Finally, due to the exploratory nature of this review, we decided to include a range of study designs, at various stages of completeness. This made it difficult to assess the risk of bias or perform a meta-analysis of the papers included in the analysis. Therefore, the findings should be interpreted with caution.

Conclusions

Narrative approaches to evidence synthesis that incorporate diverse literature can be valuable in highlighting issues beyond simple summary measures of effect. Indeed, a simple meta-analysis of this evidence base would be misleading given the heterogeneity of the interventions. Instead, this review has identified theoretical and empirical issues related to the success of health promoting interventions that harness social media. We have shown that more, and longer, RCTs need to be conducted that take into account contextual factors such as patient characteristics and types of SNS. Also, more evidence is needed regarding the actual usability of SNS and how different interface design elements may help or hinder behavior change and engagement. It will be crucial to investigate further the effect of theory on the effectiveness of SNS for health promotion. The informatics research in this field needs better designed experiments. Public health practitioners need to prepare for more action research whereby theoretically founded interventions generate evidence that helps them to evolve—reflecting the emergent nature of social technologies.

Authors' Contributions

PB and CK conducted the review of the literature, including the tabulation and synthesis of the results. IB, JA, and JP methodologically guided the synthesis process and contributed to the discussions/conclusions of the results, as well as the editing.

Conflicts of Interest

JP was initially a reviewer for the paper and was added as a co-author after the initial editorial decision. He did not take part in the re-review of the manuscript.

Multimedia Appendix 1

List of extracted items.

[PDF File (Adobe PDF File), 42KB - jmir_v17i6e141_app1.pdf]

References

1. Maher C, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
2. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, Lau Annie Y S. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
3. Valente TW. Social networks and health: models, methods and applications. UK: Oxford University Press; 2010:296.
4. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
5. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]
6. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Klassen TP, et al. Social media use by health care professionals and trainees: a scoping review. Acad Med 2013 Sep;88(9):1376-1383. [doi: 10.1097/ACM.0b013e31829eb91c] [Medline: 23887004]
7. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
8. 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]
9. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
10. Schein R, Wilson K, Keelan JE. Peel Public Health Report. 2011. Literature Review on Effectiveness of the Use of Social Media: A Report for Peel Public Health URL: https://www.peelregion.ca/health/resources/pdf/socialmedia.pdf [accessed 2015-06-03] [WebCite Cache ID 6Z0co8fAT]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Van De Belt TH, Engelen LJLPG, Berben SAA, 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]
13. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
14. Chou W, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
15. Pedrana A, Hellard M, Gold J, Ata N, Chang S, Howard S, et al. Queer as F**k: reaching and engaging gay men in sexual health promotion through social networking sites. J Med Internet Res 2013;15(2):e25 [FREE Full text] [doi: 10.2196/jmir.2334] [Medline: 23391459]
16. Mackert M, Kim E, Guadagmo M, Donovan-Kicken E. Using Twitter for prenatal health promotion: encouraging a multivitamin habit among college-aged females. Stud Health Technol Inform 2012;182:93-103. [Medline: 23138084]
17. Gasca E, Favela J, Tentori M. Assisting Support Groups of Patients with Chronic Diseases through Persuasive Computing. Journal of Universal Computer Science 2009;15(16):3081-3100. [doi: 10.3217/jucs-015-16-3081]
18. Kamal N, Fels S, Blackstock M, Ho K. The ABCs of designing social networksfor health behaviour change: the Vivospace social network. In: Advances in network analysis. Berlin: Springer-Verlag; 2013:323-348.

19. Nordfeldt S, Hanberger L, Berterö C. Patient and parent views on a Web 2.0 Diabetes Portal--the management tool, the generator, and the gatekeeper: qualitative study. J Med Internet Res 2010;12(2):e17 [FREE Full text] [doi: 10.2196/jmir.1267] [Medline: 20511179]
20. 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]
21. Evans D. How social marketing works in healthcare. BMJ 2006;332(7551):1207-1210. [doi: 10.1136/bmj.332.7551.1207-a]
22. Griffiths J. Social marketing for health and specialised health promotion. 2008. URL: https://www.rsph.org.uk/download.cfm?docid=1391e86b-6137-4f44-8ffcf407420f7b29 [accessed 2015-06-03] [WebCite Cache ID 6Z0ixb6Zk]
23. World Health Organization. Health Education: theoretical concepts, effective strategies and care competences. URL: http://applications.emro.who.int/dsaf/EMRPUB_2012_EN_1362.pdf [accessed 2015-06-03] [WebCite Cache ID 6Z0iyuIrn]
24. Nielsen J. Usability 101: introduction to usability. 2012. URL: http://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2015-06-03] [WebCite Cache ID 6Z0j19bnw]
25. Nielsen J. User satisfaction vs performance metrics. 2012. URL: http://www.nngroup.com/articles/satisfaction-vs-performance-metrics/ [accessed 2015-06-03] [WebCite Cache ID 6Z0j2JlSn]
26. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
27. Ajzen I. The Theory of Planned Behaviour. Organizational Behavior and Human Decision Processes 1991;50(2):179-211 [FREE Full text] [doi: 10.1016/07495978(91)90020-T]
28. Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
29. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Effectiveness of Interventions to Promote Smoke Alarm Ownership and Function. Evaluation 2009 Jan 01;15(1):49-73. [doi: 10.1177/1356389008097871]
30. Howland J, Wright TC, Boughan RA, Roberts BC. How Scholarly Is Google Scholar? A Comparison to Library Databases. College & Research Libraries 2009 May 01;70(3):227-234. [doi: 10.5860/crl.70.3.227]
31. Walters WH. Google Scholar coverage of a multidisciplinary field. Information Processing & Management 2007 Jul;43(4):1121-1132. [doi: 10.1016/j.ipm.2006.08.006]
32. An LC, Schillo BA, Saul JE, Wendling AH, Klatt CM, Berg CJ, et al. Utilization of smoking cessation informational, interactive, and online community resources as predictors of abstinence: cohort study. J Med Internet Res 2008;10(5):e55 [FREE Full text] [doi: 10.2196/jmir.1018] [Medline: 19103587]
33. Baghaei N, Freyne J, Kimani S, Smith G, Berkovsky S, Bhandari D, et al. SOFA: an Online Social Network for Engaging and Motivating Families to Adopt a Healthy Lifestyle. In: OZCHI 09. 2009 Presented at: 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design; November 23-27, 2009; Melbourne, Australia URL: http://unitec.researchbank.ac.nz/handle/10652/2200 [doi: 10.1145/1738826.1738871]
34. Burke S, Oomen-Early J. That's Blog Worthy. American Journal of Health Education 2008 Nov;39(6):362-364. [doi: 10.1080/19325037.2008.10599064]
35. Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. Am J Public Health 2010 Jul;100(7):1282-1289 [FREE Full text] [doi: 10.2105/AJPH.2009.165449] [Medline: 20466971]
36. Cunningham JA, van MT, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
37. Falan S, Han B, Rea A. A Smart Consumer-empowered Diabetes Education System (SCEDES): Integrating Human Wellbeing and Health Care in the Community Environment. In: AMCIS 2011 Proceedings. 2011 Presented at: Americas Conference on Information Systems; 2011; Detroit, Michigan URL: http://aisel.aisnet.org/amcis2011_submissions/370/
38. Foster D, Linehan C, Kirman B, Lawson S, James G. MindTrek 2010. Motivating Physical Activity at Work: using Persuasive Social Media for Competitive Step Counting URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.369.7693 [accessed 2015-06-03] [WebCite Cache ID 6Z0kKgL5y]
39. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]
40. Gay G, Pollak J, Adams P, Leonard JP. Pilot study of Aurora, a social, mobile-phone-based emotion sharing and recording system. J Diabetes Sci Technol 2011 Mar;5(2):325-332 [FREE Full text] [Medline: 21527101]
41. Kamal N, Fels S, Ho K. Online Social Networks for Personal Informatics to Promote Positive Health Behavior. 2010 Presented at: Second ACM SIGMM Workshop on Social Media; October 25-29, 2010; Florence, Italy. [doi: 10.1145/1878151.1878167]
42. Kharrazi H, Vincz L, Stephanidis C. Increasing Physical Activity by Implementing a Behavioral Change Intervention Using Pervasive Personal Health Record System: An Exploratory Study. In: Stephanidis C, editor. Universal Access in Human-Computer Interaction: Applications and Services. Berlin Heidelberg: Springer-Verlag; 2011:366-375.

43. 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]
44. Lindsay S, Smith S, Bellaby P, Baker R. The health impact of an online heart disease support group: a comparison of moderated versus unmoderated support. Health Educ Res 2009 Aug;24(4):646-654 [FREE Full text] [doi: 10.1093/her/cyp001] [Medline: 19251770]
45. Linehan C, Doughty M, Lawson S, Kirman B, Olivier P, Moynihan P. Tagiatelle: Social Tagging to Encourage Healthier Eating. In: CHI EA. 2010 Presented at: 28th International Conference on Human Factors in Computing Systems; April 10-15, 2010; Atlanta, GA URL: http://eprints.lincoln.ac.uk/2429/ [doi: 10.1145/1753846.1753980]
46. Liu N, Chan H. Understanding the Influence of Social Identity on Social Support Seeking Behaviors in Virtual Healthcare Communities. In: ICIS 2010. 2010 Presented at: International Conference on Information Systems (ICIS); 2010; St Louis, MO URL: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1159&context=icis2010_submissions
47. Maibach EW, Abroms LC, Marosits M. Communication and marketing as tools to cultivate the public's health: a proposed "people and places" framework. BMC Public Health 2007;7:88 [FREE Full text] [doi: 10.1186/1471-2458-7-88] [Medline: 17519027]
48. Munson SA, Lauterbach D, Newman MW, Resnick P. Happier Together: Integrating a Wellness Application Into a Social Network Site. 2010 Presented at: Persuasive Technology, 5th International Conference - Persuasive; 2010; Copenhagen, Denmark p. 27-39. [doi: 10.1007/978-3-642-13226-1_5]
49. Nahm ES, Resnick B, DeGrezia M, Brotemarkle R. Use of discussion boards in a theory-based health web site for older adults. Nurs Res 2009;58(6):419-426. [doi: 10.1097/NNR.0b013e3181bee6c4] [Medline: 19918152]
50. O'Grady LA, Witteman H, Wathen CN. The experiential health information processing model: supporting collaborative web-based patient education. BMC Med Inform Decis Mak 2008;8:58 [FREE Full text] [doi: 10.1186/1472-6947-8-58] [Medline: 19087353]
51. Olsen E, Kraft P. ePsychology: a pilot study on how to enhance social support and adherence in digital interventions by characteristics from social networking sites. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive 2009; April 26-29, 2009; Claremont, CA URL: http://dl.acm.org/citation.cfm?doid=1541948.1541991
52. Potente S, McIver J, Anderson C, Coppa K. “It's a Beautiful Day … for Cancer”: An Innovative Communication Strategy to Engage Youth in Skin Cancer Prevention. Social Marketing Quarterly 2011 Sep;17(3):86-105. [doi: 10.1080/15245004.2011.595604]
53. Rhodes SD, Hergenrather KC, Duncan J, Vissman AT, Miller C, Wilkin AM, et al. A pilot intervention utilizing Internet chat rooms to prevent HIV risk behaviors among men who have sex with men. Public Health Rep 2010;125 Suppl 1:29-37 [FREE Full text] [Medline: 20408385]
54. 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]
55. Roblin DW. The potential of cellular technology to mediate social networks for support of chronic disease self-management. J Health Commun 2011;16 Suppl 1:59-76. [doi: 10.1080/10810730.2011.596610] [Medline: 21843096]
56. Stoddard JL, Augustson EM, Moser RP. Effect of adding a virtual community (bulletin board) to smokefree.gov: randomized controlled trial. J Med Internet Res 2008;10(5):e53 [FREE Full text] [doi: 10.2196/jmir.1124] [Medline: 19097974]
57. Toscos T, Consolvo S, McDonald DW. is it normal to be this sore?: using an online forum to investigate barriers to physical activity. In: Proceedings of the First ACM International Health Informatics Symposium. 2010 Presented at: First ACM International Health Informatics Symposium; 2010; Arlington, VA p. 346-355. [doi: 10.1145/1882992.1883041]
58. Waters RD, Canfield R, Foster JM, Hardy E. Applying the dialogic theory to social networking sites. Journal of Social Marketing 2011 Oct 11;1(3):211-227. [doi: 10.1108/20426761111170713]
59. West J, Hall PC, Hanson C, Thackeray R, Barnes M, Neiger B, et al. Breastfeeding and Blogging. American Journal of Health Education 2011 Mar;42(2):106-115. [doi: 10.1080/19325037.2011.10599178]
60. Woodruff SI, Conway TL, Edwards CC, Elliott SP, Crittenden J. Evaluation of an Internet virtual world chat room for adolescent smoking cessation. Addict Behav 2007 Sep;32(9):1769-1786. [doi: 10.1016/j.addbeh.2006.12.008] [Medline: 17250972]
61. Young M. Twitter Me: Using Micro-blogging to Motivate Teenagers to Exercise. In: Global Perspectives on Design Science Research -. 2010 Presented at: DESRIST 2010; 2010; St Gallen, Switzerland p. 448. [doi: 10.1007/978-3-642-13335-0_30]
62. Baelden D, Van Audenhove L, Vergnani T. Using new technologies for stimulating interpersonal communication on HIV and AIDS. Telematics and Informatics 2012 May;29(2):166-176. [doi: 10.1016/j.tele.2011.05.002]
63. Ploderer B, Smith W, Howard S, Pearce J, Borland R. Patterns of support in an online community for smoking cessation. In: Proceedings of the 6th international conference on communities and technologies. 2013 Presented at: 6th International Conference on Communities and Technologies; 2013; Munchen, Germany p. 26-35. [doi: 10.1145/2482991.2482992]

64. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
65. Nguyen P, Gold J, Pedrana A, Chang S, Howard S, Ilic O, et al. Sexual health promotion on social networking sites: a process evaluation of The FaceSpace Project. J Adolesc Health 2013 Jul;53(1):98-104. [doi: 10.1016/j.jadohealth.2013.02.007] [Medline: 23583509]
66. Kolt GS, Rosenkranz RR, Savage TN, Maeder AJ, Vandelanotte C, Duncan MJ, et al. WALK 2.0 - using Web 2.0 applications to promote health-related physical activity: a randomised controlled trial protocol. BMC Public Health 2013;13:436 [FREE Full text] [doi: 10.1186/1471-2458-13-436] [Medline: 23642010]
67. Gabarron E, Serrano JA, Wynn R, Armayones M. Avatars using computer/smartphone mediated communication and social networking in prevention of sexually transmitted diseases among North-Norwegian youngsters. Medical Informatics and Decision Making 2012;12:120.
68. Kelty T, Morgan P, Lubans D. Efficacy and feasibility of the "Girls Recreational Activity Support Program Using Information Technology": a pilot randomised controlled trial. Advances in Physical Education 2012;2(1):10-16.
69. Laakso E, Armstrong K, Usher W. Cyber-management of people with chronic disease: A potential solution to eHealth challenges. Health Education Journal 2011 May 12;71(4):483-490. [doi: 10.1177/0017896911408813]
70. Hwang KO, Etchegaray JM, Sciamanna CN, Bernstam EV, Thomas EJ. Structural social support predicts functional social support in an online weight loss programme. Health Expect 2014 Jun;17(3):345-352. [doi: 10.1111/j.1369-7625.2011.00759.x] [Medline: 22212418]
71. Ruggiero T. Uses and Gratifications Theory in the 21st Century. Mass Communication and Society 2000 Feb;3(1):3-37. [doi: 10.1207/S15327825MCS0301_02]
72. Yuqing R, Kraut R, Kiesler S. Applying Common Identity and Bond Theory to Design of Online Communities. Organization Studies 2007 Mar 01;28(3):377-408. [doi: 10.1177/0170840607076007]
73. Allen NJ, Meyer JP. The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization J Occup Psychol, (63). Journal of Occupational Psychology 1990;63(1):1.
74. Tajfel H. Social identity and intergroup behaviour. Social Science Information 1974 Apr 01;13(2):65-93. [doi: 10.1177/053901847401300204]
75. Tajfel H, Turner JC. The Social Identity Theory of Intergroup Behavior. In: The Social Psychology of Intergroup Relations. Chicago: Nelson-Hall; 1986:7.
76. Cobb S. Social support as a moderator of life stress. Psychosomatic Medicine 1976;38(5):300.
77. Minkler M. Applications of Social Support Theory to Health Education: Implications for Work with the Elderly. Health Education Behavior 1981;8:147.
78. Valente TW. Social network thresholds in the diffusion of innovations. Social Networks 1996 Jan;18(1):69-89. [doi: 10.1016/0378-8733(95)00256-1]
79. Rogers EM. Diffusion of innovation. New York: The Free Press, a division of Simon & Schuster Inc; 1983.
80. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
81. Rosenstock IM, Strecher VJ, Becker MH. Social Learning Theory and the Health Belief Model. Health Education & Behavior 1988 Jan 01;15(2):175-183. [doi: 10.1177/109019818801500203]
82. Seligman MEP, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
83. Kolb DA. Experiential learning: experience as the source of learning development. Englewood Cliffs, NJ: Prentice-Hall; 1984.
84. Kent ML, Taylor M. Building dialogic relationships through the world wide web. Public Relations Review 1998;24(3):321.
85. Ioannidis JPA. Why most published research findings are false? PloS Med 2005;2(8):e124.
86. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
87. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
88. Caperchione CM, Kolt GS, Savage TN, Rosenkranz RR, Maeder AJ, Vandelanotte C, et al. WALK 2.0: examining the effectiveness of Web 2.0 features to increase physical activity in a 'real world' setting: an ecological trial. BMJ Open 2014;4(10):e006374 [FREE Full text] [doi: 10.1136/bmjopen-2014-006374] [Medline: 25304191]

Abbreviations

CICB: Common Identity and Common Bond theories
DI: Diffusion of Innovation theory

HBM: Health Belief Model
HP: health promotion
MI: Motivational Interviewing
OT: Organizational Commitment theory
PA: physical activity
RCT: randomized controlled trials
SCT: Social Cognitive Theory
SI: Social Identity theory
SNS: social networking sites
SNT: Social Network Threshold
SST: Social Support Theory
TPB: Theory of Planned Behavior
TRA: Theory of Reasoned Action
TTM: Transtheoretical Model
U&G: Uses and Gratification theory

Edited by G Eysenbach; submitted 11.07.14; peer-reviewed by J Powell, J Loss; comments to author 19.09.14; revised version received 11.12.14; accepted 23.03.15; published 11.06.15.

Please cite as:
Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J
The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness
J Med Internet Res 2015;17(6):e141
URL: http://www.jmir.org/2015/6/e141/
doi:10.2196/jmir.3662
PMID:26068087

Original Paper

ClickDiary: Online Tracking of Health Behaviors and Mood

Ta-Chien Chan[1], PhD; Tso-Jung Yen[2], PhD; Yang-Chih Fu[3], PhD; Jing-Shiang Hwang[2], PhD

[1]Research Center for Humanities and Social Sciences, Academia Sinica, Taipei City, Taiwan
[2]Institute of Statistical Science, Academia Sinica, Taipei City, Taiwan
[3]Institute of Sociology, Academia Sinica, Taipei City, Taiwan

Corresponding Author:
Jing-Shiang Hwang, PhD
Institute of Statistical Science
Academia Sinica
128 Academia Road, Section 2
Taipei City, 115
Taiwan
Phone: 886 2 6614 5677
Fax: 886 2 2783 1523
Email: hwang@sinica.edu.tw

Abstract

Background: Traditional studies of health behaviors are typically conducted using one-shot, cross-sectional surveys. Thus, participants' recall bias may undermine the reliability and validity of the data. To capture mood changes and health behaviors in everyday life, we designed an online survey platform, ClickDiary, which helped collect more complete information for comprehensive data analyses.

Objective: We aim to understand whether daily mood changes are related to one's personal characteristics, demographic factors, and daily health behaviors.

Methods: The ClickDiary program uses a Web-based platform to collect data on participants' health behaviors and their social-contact networks. The name ClickDiary comes from the platform's interface, which is designed to allow the users to respond to most of the survey questions simply by clicking on the options provided. Participants were recruited from the general population and came from various backgrounds. To keep the participants motivated and interested, the ClickDiary program included a random drawing for rewards. We used descriptive statistics and the multilevel proportional-odds mixed model for our analysis.

Results: We selected 130 participants who had completed at least 30 days of ClickDiary entries from May 1 to October 31, 2014 as our sample for the study. According to the results of the multilevel proportional-odds mixed model, a person tended to be in a better mood on a given day if he or she ate more fruits and vegetables, took in more sugary drinks, ate more fried foods, showed no cold symptoms, slept better, exercised longer, and traveled farther away from home. In addition, participants were generally in a better mood during the weekend than on weekdays.

Conclusions: Sleeping well, eating more fruits and vegetables, and exercising longer each day all appear to put one in a better mood. With the online ClickDiary survey, which reduces the recall biases that are common in traditional one-shot surveys, we were able to collect and analyze the daily variations of each subject's health behaviors and mood status.

(J Med Internet Res 2015;17(6):e147) doi:10.2196/jmir.4315

KEYWORDS

health behaviors; mood; diet; physical exercise; quality of sleep; personality

Introduction

Happiness has been regarded as an important indicator correlated to an individual's mental and physical health [1,2]. Although it is well known that emotional state is an important piece of information to consider in health or psychological studies, such information has been either overlooked [3] or confounded by one's daily activities [4], social contacts [5], personal health behaviors [6,7], and personality [8]. In conventional surveys, for example, respondents are often asked how happy or unhappy they have been, in general, over a long period of time, such as the past month or the past year [9], a time frame that is often too vague or too long to recall specific details. The information collected from such surveys can be biased by more recent and memorable experiences, or confused with the participant's general mood, both undermining the extent of accuracy in data analysis. In some cases it may also be difficult for researchers to determine which factors play more critical roles in distinguishing one's mood. One way to overcome such limitations is to get each participant's dynamic daily mood changes properly documented.

In an effort to collect such longitudinal data about mood changes, some health studies use a journal-like design, on a daily basis, to tap participants' emotional status [10], headache symptoms (of children) [11], signs of depression and stress from working or learning [12], and instances of gastrointestinal illness and other physical illnesses [13]. Data collection in such studies usually lasts for only 1 week or 1 month, thus limiting their sample sizes despite having repeated measurements. Although some social network studies also use the contact diary format to collect data [14-16], they all rely on conventional paper-and-pencil instruments. To improve the process of diary taking, minimize participants' efforts, and enhance the accuracy of results, we designed an online-based diary platform for our study.

In previous studies, one's emotional stability has been linked to certain personal characteristics, such as personality—particularly agreeableness and neuroticism [8]—and demographic factors, such as age [17]. Those individual factors, however, are more or less fixed, either ascribed or predetermined since childhood. We believe that it is also important to examine how emotional stability or happiness varies by other variables that are more closely relevant to one's health behaviors or lifestyles, such as physical exercise [2,18], quality of sleep [19], consumption of vegetables and fruits [6,7], and so on.

The aim of this study is to use the online diary platform to collect health and contact data on a daily basis. We want to elucidate the extent to which a person's daily mood varies by his or her personal characteristics, demographic factors, and health behaviors, as well as the day of the week and the extent of social interactions in everyday life.

Methods

Ethics

This study was approved by the Institutional Review Board (IRB) on Humanities and Social Science Research, Academia Sinica (AS-IRB-HS 02-13022). Participants must be at least 20 years old and capable of making juridical acts in Taiwan. Before registering as a participant of the ClickDiary program, one must read the guidelines of the program and give informed consent. We have removed personal identifiers, such as names and email addresses, and assigned serial numbers to both the participants and everyone in their contact networks to ensure privacy.

The ClickDiary Program

The ClickDiary program uses a Web-based platform [20] (see Figure 1, section A) to collect data on participants' health behaviors and their social-contact networks. The interface of ClickDiary is designed so that users can input their responses by clicking options in the instrument, making it easier to record responses on a daily basis (see Figure 1, section B). At this stage, this program is specifically tailored to the Taiwanese population. The program is unlike traditional cross-sectional health behavior surveys [21] or one-shot paper-and-pencil contact surveys [15,16]. In addition to being user friendly, the ClickDiary program helps generate daily, real-time, longitudinal data. After participants successfully sign up for their accounts, they are asked to provide demographic information, including age, gender, place of residence, marital status, and current job. The program also collects the Big Five personality traits—openness, conscientiousness, extraversion, agreeableness, neuroticism (OCEAN) [14,22], whose exact wording was taken from the International Personality Item Pool (IPIP) [23]—height and weight for calculating body mass index (BMI, kg/m^2), perceived health status and happiness, the number (and characteristics) of people contacted during the day, and a baseline health survey, which borrows items from the Taiwan Social Change Survey [24]. After giving such basic information, participants can then proceed to fill out their health diary and contact diary. A reward system serves as incentive for the participants to keep both diaries at least three times a week. In the following sections we introduce our health diary, recruitment methods, quality-control process, reward system, and feedback design.

Figure 1. Screenshots of ClickDiary. (A) Home page; (B) The interface of the health diary.

隨時關心您的人脈，讓生活更精彩；
勤記每日生活行為，讓身心更健康。

歡迎大家一同來 點·日記！

著作權©中央研究院統計科學研究所. 版權所有.
Copyright© Institute of Statistical Science, Academia Sinica. All rights reserved.

我的日記

昨晚幾點睡覺?
時段 下午 時 11 分 00

今天幾點起床？
時段 上午 時 5 分 00

睡眠品質
○ 非常好 ● 很好 ○ 還好 ○ 不好 ○ 非常不好

1. 過去一天的心情如何？
● 非常好 ○ 很好 ○ 還好 ○ 不好

2. 過去一天飲食狀況？

蔬菜	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
水果	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
全穀根莖類	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
白飯麵粉類	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
豬/牛/羊 肉類	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
雞/鴨/鵝 肉類	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
有無吃肥肉	● 無	○ 有					
海鮮	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份
蛋	● 0	○ 0.5	○ 1	○ 2	○ 3	○ 3+	份

Health Diary

In this study, we used only the health diary and baseline profile data for our analysis. Thus, we will not introduce the contents of the contact diary here. In the health diary, we asked participants seven major questions regarding what happened in the past day, including their sleep behaviors (ie, what time they got up and went to bed, and how well they slept—very good, good, fair, poor, or very poor), their mood during the past day (very good, good, fair, or poor), their food intake during the past day, duration of physical exercise (no exercise, 1 to 30 minutes, 31 to 60 minutes, 61 to 120 minutes, or longer than 120 minutes), the number of people with whom they had contact in the past day, the number of suspected or confirmed influenza-inflicted people (and their symptoms) with whom they had contact, and the maximum distance they traveled from their residence (less than one kilometer, 1 to 9 kilometers, 10 to 49 kilometers, 50 to 300 kilometers, or farther than 300 kilometers).

In the section on food intake, we listed 16 categories of food, the quantities of which were measured in six degrees (ie, 0, 0.5, 1, 2, 3, and 3+) with different units. The 16 categories of food were vegetables, fruits, whole grains and rootstock, rice and flour, pork/beef/mutton, chicken/duck/goose, fatty meat, seafood, eggs, beans, milk and cheese, nuts, fried foods, processed foods, desserts, and sugary drinks. For this study, we selected and consolidated the food types into the following eight types: meat, seafood, milk and cheese, whole grains and rootstock, fried foods, sugary drinks, desserts, and vegetables and fruits. Mouse-over labels were available to inform users about the units for the different categories of food. One serving of fruit, for example, was equal to a fistful or 80% of a 240 milliliter bowl. These hints ensured that all participants had the same understanding of the units.

Time Definition

We designed the health diary to collect data from the past 24 hours. Because it was likely that participants would enter data several times a day, we reorganized the dates indicated on the diaries according to the logged time. If the logged time fell after noon of the current day, the entry would be included in the current day's health diary. If the logged time fell between noon of the previous day and noon of the current day, then it was considered the previous day's health diary entry.

Recruitment of Participants

Recruited from various channels, our participants included university students, school teachers and administrative employees, volunteers at health-promotion centers, hospital patients, and community college students, as well as other adults in the general population. Due to a limited budget and the longitudinal nature of the ClickDiary program, we were unable to recruit a representative sample as large as that of the one-shot cross-sectional surveys. To diversify the patterns of participation, however, we used two major recruitment strategies. First, we targeted unspecified individuals and groups to solicit volunteers using posters and other promotional campaigns, both online and offline, such as on Facebook and in classes held at different universities and community colleges scattered across northern, central, and southern Taiwan. Second, in several groups with delineated boundaries, we asked group leaders to monitor the participation rate by periodically encouraging group members to complete both health and contact diaries. To help such monitoring, our system issued a summary report of each group to its leader at the end of each week. If a group's diary entries met our standards for both quantity and quality, we sent out convenience store vouchers to the leader and each group member as a reward.

Quality Control

To ensure the quality of the data, we checked the data pattern of each participant every week. If we suspected that a specific participant had not been keeping his or her diaries properly, the participant's serial number was put on an alert list, and any data entered by the participant would be excluded from our analysis. In addition, anyone who provided incomplete and poor data was excluded from the random drawing and from receiving any reward. We did not publish details of our quality-control procedure, because quality control might become more difficult if the participants get acquainted with our rules of screening.

Incentives

We designed a random drawing that took place both weekly and monthly. Each entry into the drawing was assigned a weight based on each participant's corresponding weekly and monthly cumulative points. The more time and effort participants spent on completing their ClickDiary entries, the greater were their chances to win a reward. A participant automatically received 20 points after finishing the health diary each day. If someone submitted multiple health diary entries within a single day, the system still gave the participant only 20 points. Participants also received 3 points for each contacted person entered into the contact diary. Although one could record multiple interactions with the same contacted person on the same day, each contacted person would only yield 3 points. To prevent the participants from intentionally giving false data about daily contacts, we developed a screening program for checking the accuracy and quality of diary data every week.

Feedback

We did not give feedback to participants, based on the IRB's recommendations. We did, however, provide an interactive Web chart summarizing the records in each participant's contact and health diaries. Participants could then view the summarized reports that kept track of their health behaviors for up to one year. In addition, participants could gain insight from their overall contact patterns by checking the contact network tree we developed, which also allowed them to take a glance at how their mood changed when contacting different people over time.

Statistical Analysis

Participants' reports on their overall mood in the past 24 hours were coded on a 4-point scale—1 (Poor), 2 (Fair), 3 (Good), and 4 (Very good). Other health-related daily entries included diet, exercise, and the time and quality of sleep during the 184-day study period. For this study, we applied a multilevel proportional-odds mixed model to analyze the data [25,26]. With this model, we analyzed the relationship between mood swings and health behaviors, while adjusting for individual characteristics, such as age, gender, and personality.

Let Y_{ij} be the i^th individual's scalar response of mood recorded on day t_{ij}, where $j=1,2,...,T_i$, and T_i ranged from 26 to 184 for the n=130 participants. The random-effects ordinal regression model for analyzing such multilevel data is given as:

$$\text{logit}[\Pr(Y_{ij} \le k)] = \theta_k - a_i - b_1 \cdot I(Y_{i,j-1}=1) - b_2 \cdot I(Y_{i,j-1}=2) - b_3 \cdot I(Y_{i,j-1}=3) - b_4 \cdot I(Y_{i,j-1}=4) - \alpha' Z_i - \beta' X_{ij},\ k=1,2,3$$

Time-dependent covariates for the i^{th} participant are denoted by $X_{ij} = (X_{1ij},...,X_{qij})'$ and the other covariates are denoted by $Z_i = (Z_{1i},...,Z_{pi})'$. The parameters $\{\theta_k\}$ increasing in k are known as thresholds or cut-points. Random components of $a_i \sim N(0,\sigma^2_a)$, representing variations of these logits for each response level k among the n individuals, are added into the mean equation to adjust for the thresholds of each individual.

For the repeated recordings, each individual's scalar response may be associated with previously reported responses. To take this into account, we added another random component $b_l \sim N(0, \sigma^2_b)$, $l=1,...,4$ to the model for further adjustment of the thresholds. This model has the same fixed effects as determined by the parameters α and β of the covariates of interest for each response level k. For example, the ratio between the odds of making a response at each level k or below with covariates $X_{ij} = x^1_{ij}$ and x^0_{ij} is $\exp[-\beta'(x^1_{ij} - x^0_{ij})]$. In this application, our covariates are all categorical variables. For the h^{th} covariate of interest, X_{hij}, which is represented by an indicator variable corresponding to a level of a categorical variable, such as the quality of sleep, we report $\exp(-\beta_h)$ as odds ratio (OR) of cumulative probabilities between the presence of the level and the baseline level of this categorical variable.

We might expect that an individual with the level of this categorical variable had odds of worsening mood $\exp(-\beta_h)$ times compared to those with a baseline level of this variable if the odds ratio were larger than 1. On the contrary, an estimate of $\exp(-\beta_h)<1$ indicates better mood with the level of this categorical variable. We used the clmm function from the R package “ordinal” to estimate the model parameters [27].

Results

We selected 130 participants from 726 qualified participants (17.9%) who had completed at least 30 days of contact diaries from May 1 to October 31, 2014. The contact diary served better as our criterion because it required more time to complete—on average, it took 1 minute and 12 seconds to record all variables per contact, or about 12 minutes for 10 contacts per day—than the health diary, which averaged 1 minute and 39 seconds per day. Thus, a more complete contact diary normally indicates higher commitment by the participant, which in turn ensures that the health diary, as well as other variables, are of better quality.

The average participant kept the health diary for about 69 days, with a range from 26 to 184 days (see Table 1). The 130 participants had input 8824 complete health diary entries. Our sample was overrepresented by females (98/130, 75.4%), and the overall mean age was 33.1 (SD 13.4), ranging from 20 to 67 years. Participants averaged a BMI of 22.0 kg/m^2 (SD 3.3), and most were within the normal range—healthy BMI: 18.5 kg/m^2≤BMI<24 kg/m^2. Among the Big Five personality items, about 90% of participants said they were thorough in performing a task—a measure of being conscientious—or they sympathized with others’ feelings—a measure of being agreeable (see Table 2). We used the 10 personality measures as covariates for the multilevel proportional-odds mixed model.

Table 1. Summary of selected variables for the 32 male and 98 female participants (n=130).

Variable	Mean (SD)	Minimum	Maximum
Participating days	68.9 (38.1)	26	184
Age in years	33.1 (13.4)	20	67
Body mass index (kg/m^2)	22.0 (3.3)	16.2	36.0

Table 2. Summary of personality items for the 32 male and 98 female participants (n=130).

Big Five personality items	Responses			
	Not at all, n (%)	Not very, n (%)	Somewhat, n (%)	Very, n (%)
Extraversion				
Outgoing and sociable	13 (10.0)	47 (36.2)	58 (44.6)	12 (9.2)
Do not talk a lot	22 (16.9)	56 (43.1)	42 (32.3)	10 (7.7)
Agreeableness				
Sympathize with others’ feelings	5 (3.8)	8 (6.2)	81 (62.3)	36 (27.7)
Do not trust others	20 (15.4)	61 (46.9)	39 (30.0)	10 (7.7)
Conscientiousness				
Thorough	3 (2.3)	8 (6.2)	70 (53.8)	49 (37.7)
Careless	21 (16.2)	60 (46.2)	39 (29.9)	10 (7.7)
Neuroticism				
Relaxed most of the time	8 (6.1)	27 (20.8)	72 (55.4)	23 (17.7)
Get nervous easily	9 (6.9)	27 (20.8)	66 (50.8)	28 (21.5)
Openness to new experiences				
Have a vivid imagination	6 (4.6)	31 (23.8)	60 (46.2)	33 (25.4)
Conservative	8 (6.2)	29 (22.3)	71 (54.6)	22 (16.9)

Table 3 shows the results of parameter estimates for time-independent covariates. Participants’ gender, age, and BMI had no significant association with the odds of reporting a mood status during the study period. Those who were relaxed most of the time seemed to have lower odds of reporting a mood status at a level k or below (OR 0.78, 95% CI 0.58-1.04, *P*=.09), indicating a marginal association between better mood and emotional stability. Being quiet (ie, "do not talk a lot"), being sympathetic to others’ feelings, and not trusting others had marginally significant associations with higher odds of reporting a mood status at a level k or below, indicating the likelihood of reporting a negative mood up to 50%.

Table 3. Parameter estimates for the time-independent covariates in the multilevel proportional-odds model.

Covariate of interest			Estimate (SE)	OR[a] (95% CI)	*P*
Gender (female as reference)	Male		-0.114 (0.134)	1.12 (0.86-1.46)	.40
Age group (26≤age≤59 years as reference) in years					
	<26		0.084 (0.121)	0.92 (0.73-1.17)	.49
	>59		0.233 (0.242)	0.79 (0.49-1.27)	.34
BMI [b] (18.5≤BMI<24 kg/m^2 as reference), kg/m^2					
	<18.5		0.089 (0.159)	0.91 (0.67-1.25)	.58
	24≤BMI<27		-0.013 (0.160)	1.01 (0.74-1.38)	.94
	≥27		-0.105 (0.234)	1.11 (0.70-1.76)	.65
Big Five personality items					
	Extraversion				
		Outgoing and sociable	-0.109 (0.137)	1.12 (0.85-1.46)	.43
		Do not talk a lot	-0.214 (0.123)	1.24 (0.97-1.58)	.08
	Agreeableness				
		Sympathize with others' feelings	-0.409 (0.218)	1.50 (0.98-2.31)	.06
		Do not trust others	-0.215 (0.123)	1.24 (0.97-1.58)	.08
	Conscientiousness				
		Thorough	0.312 (0.225)	0.73 (0.47-1.14)	.17
		Careless	-0.093 (0.119)	1.10 (0.87-1.39)	.43
	Neuroticism				
		Relaxed most of the time	0.254 (0.150)	0.78 (0.58-1.04)	.09
		Get nervous easily	0.016 (0.140)	0.98 (0.75-1.30)	.91
	Openness to new experiences				
		Have a vivid imagination	0.092 (0.134)	0.91 (0.70-1.19)	.49
		Conservative	-0.042 (0.129)	1.04 (0.81-1.34)	.74

[a] Odds ratio (OR).

[b] Body mass index (BMI).

Compared to time-independent covariates, several time-dependent covariates were more closely associated with the odds of reporting a mood status at a level k or below (see Table 4). The odds ratio estimate of 0.36 (*P*<.001) for those who slept very well, for example, indicates that they had a 64% reduction in odds compared to those who slept just fairly, after taking the length of sleep, diet, and other lifestyle factors into account. In contrast, participants who slept very poorly reported odds of moods being at or below a level that was 147% greater than the odds of those who slept fairly (*P*<.001). This finding supports our expectation of a strong relationship between the quality of sleep and one's mood in daily life. Participants' moods were also closely associated with longer duration of physical exercise. The odds of reporting mood status at a level k or below was 0.86 (95% CI 0.76-0.98, *P*=.021) for people who exercised more than 60 minutes a day, compared to those who did not exercise at all. To a lesser degree, exercising for 1 to 30 minutes or 31 to 60 minutes per day also helped—OR 0.93 (95% CI 0.86-1.00) and 0.91 (95% CI 0.83-1.00), respectively.

Table 4. Parameter estimates for the time-dependent covariates in the multilevel proportional-odds model.

Covariate of interest			Estimate (SE)	OR[a] (95% CI)	*P*
Sleeping quality (Fair as reference)					
	Very poor		-0.903 (0.121)	2.47 (1.94-3.13)	<.001
	Poor		-0.364 (0.052)	1.44 (1.30-1.59)	<.001
	Good		0.529 (0.040)	0.59 (0.54-0.64)	<.001
	Very good		1.036 (0.077)	0.36 (0.31-0.41)	<.001
Length of sleep (6-8 hours as reference), hours					
	<6		0.489 (0.305)	0.61 (0.34-1.12)	.11
	>8		-0.110 (0.153)	1.12 (0.83-1.51)	.47
Duration of physical exercise (None as reference), minutes					
	1-30		0.075 (0.041)	0.93 (0.86-1.00)	.07
	31-60		0.093 (0.049)	0.91 (0.83-1.00)	.06
	>60		0.149 (0.064)	0.86 (0.76-0.98)	.02
Diet (None as reference within each category), servings					
	Meat				
		0-2	0.032 (0.049)	0.97 (0.88-1.07)	.51
		≥2	0.065 (0.057)	0.94 (0.84-1.05)	.25
	Seafood				
		0-2	0.047 (0.036)	0.95 (0.89-1.02)	.19
		≥2	0.061 (0.067)	0.94 (0.83-1.07)	.36
	Milk and cheese				
		0-2	-0.029 (0.035)	1.03 (0.96-1.10)	.41
		≥2	0.037 (0.066)	0.96 (0.85-1.10)	.57
	Whole grains and rootstock				
		0-2	0.057 (0.036)	0.94 (0.88-1.01)	.12
		≥2	-0.023 (0.063)	1.02 (0.90-1.16)	.72
	Fried food				
		0-1	0.000 (0.050)	1.00 (0.91-1.10)	.997
		≥1	0.081 (0.040)	0.92 (0.85-1.00)	.04
	Sugary drinks				
		0-2	0.013 (0.036)	0.99 (0.92-1.06)	.72
		≥2	0.133 (0.063)	0.88 (0.77-0.99)	.04
	Dessert				
		0-1	0.081 (0.047)	0.92 (0.84-1.01)	.08
		≥1	0.025 (0.038)	0.98 (0.90-1.05)	.51
	Vegetables and fruits				
		0-4	0.175 (0.085)	0.84 (0.71-0.99)	.04
		≥4	0.240 (0.093)	0.79 (0.66-0.94)	.01
Day of the week (Tuesday-Friday as reference)					
	Monday		-0.010 (0.040)	1.01 (0.93-1.09)	.79
	Saturday, Sunday		0.177 (0.034)	0.84 (0.78-0.90)	<.001
ILI[b] symptoms (Yes as reference)	Without ILI		0.109 (0.048)	0.90 (0.82-0.98)	.02

Covariate of interest		Estimate (SE)	OR[a] (95% CI)	*P*
Distance away from residence (<1 km as reference), km				
	1-9	0.036 (0.051)	0.96 (0.87-1.07)	.48
	10-49	0.150 (0.055)	0.86 (0.77-0.96)	.006
	≥50	0.336 (0.074)	0.71 (0.62-0.83)	<.001
Number of people contacted (<5 as reference), n				
	5-9	-0.072 (0.066)	1.07 (0.94-1.22)	.27
	≥10	-0.048 (0.068)	1.05 (0.92-1.20)	.48

[a] Odds ratio (OR).

[b] Influenza-like illness (ILI).

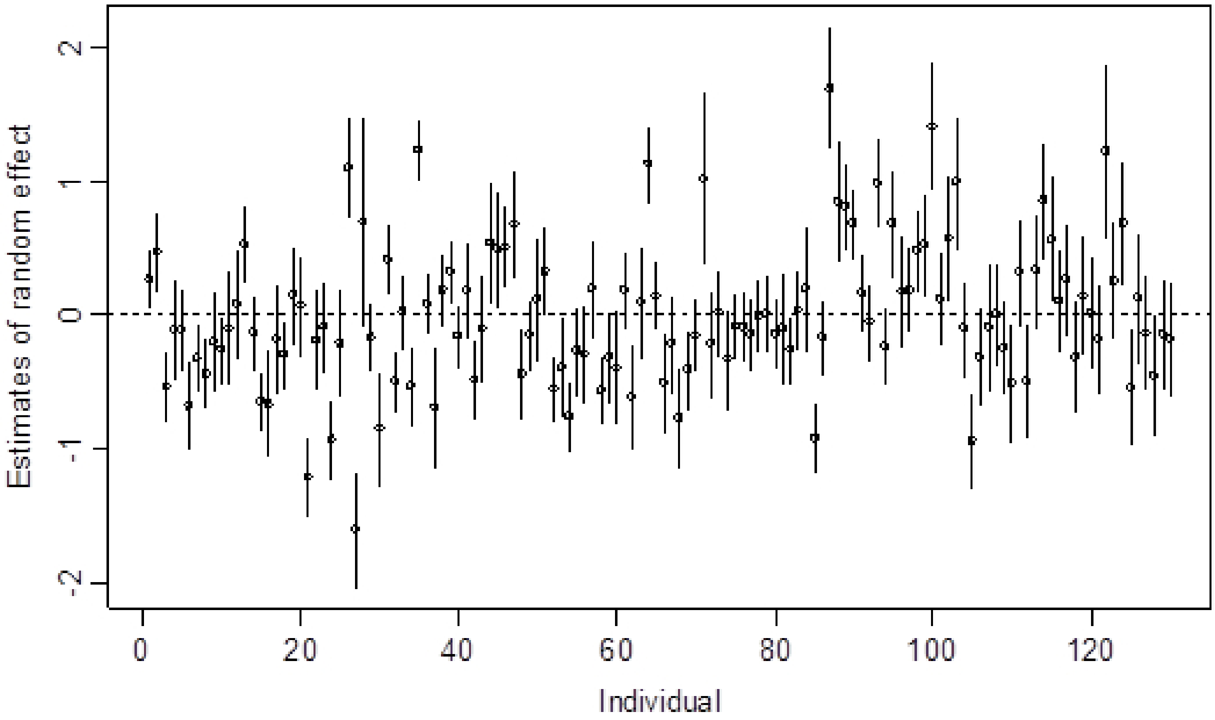
Some types of diets were positively related to the mood, too. Participants who ate four servings or more of vegetables or fruits had lower odds of reporting a mood status at a level of k or below (OR 0.79, 95% CI 0.66-0.94, *P*=.01) than those who did not eat any vegetables or fruits. Eating fewer servings of vegetables or fruits also resulted in lower odds (OR 0.84, 95% CI 0.71-0.99, *P*=.04). Those who had two bottles or more of sugary drinks had lower odds compared with those who did not have such drinks (OR 0.88, 95% CI 0.77-0.99, *P*=.04). Moreover, eating less than one serving of dessert was also marginally associated with daily mood (OR 0.92, 95% CI 0.84-1.01, *P*=.08). One surprising finding was that taking at least one serving of fried food was also linked to a somewhat better mood (OR 0.92, 95% CI 0.85-1.00, *P*=.04). Participants tended to have lower odds of reporting their mood status at a level of k or below on weekends (OR 0.84, 95% CI 0.78-0.90, *P*<.001). Traveling away from home by at least 50 kilometers was associated with significantly lower odds (OR 0.71, 95% CI 0.62-0.83, *P*<.001), as was traveling 10 to 49 kilometers away (OR 0.86, 95% CI 0.77-0.96, *P*=.006). As expected, people who did not suffer from any symptoms of influenza-like illness (ILI) tended to experience a better mood (OR 0.90, 95% CI 0.82-0.98, *P*=.02).The variance estimate, σ^2_a=0.327, for the random components representing variation of these thresholds among the n individuals, was not negligible. As shown in Figure 2, we saw a lot of random-effect estimates of this variance component deviated away from the zero mean. This indicates that there is still some uncertainty that remains unexplained by the covariates considered in the model. The estimate of variance, σ^2_b, for the other four variance components was 1.587. This result was expected because participants' reports on mood status were affected by their previous reports to various degrees, as shown in Table 5.

Table 5. Number of mood changes reported in two consecutive diary entries by the 130 participants during the study period of 184 days.

Response for current report, n (%)	Response for previous report, n (%)			
	Poor	Fair	Good	Very good
Poor	142 (40.3)	138 (3.99)	57 (1.45)	11 (1.02)
Fair	153 (43.5)	2834 (81.91)	442 (11.24)	36 (3.32)
Good	43 (12.2)	458 (13.23)	3262 (82.96)	164 (15.14)
Very good	14 (4.0)	30 (0.87)	171 (4.35)	872 (80.52)
Total	352 (100)	3460 (100)	3932 (100)	1080 (100)

Among the participants who reported having a poor mood, about 40% reported the same level during the following day, 43% moved to fair, and 12% moved to good. On the other end, among those who reported a very good mood, about 81% felt the same in the following report, 15% changed to good, and 3% changed to fair. For participants who reported fair or good moods, about 82% to 83% retained the same feeling the following day, while about 11% to 13% tended to swing between these two levels. This phenomenon had been adjusted by the four random components with effects estimates b_1 = -1.43, b_2 = -0.8, b_3 = 0.33, and b_4 = 1.88 in the proportional-odds model.

Figure 2. Estimates of random effects with 95% confidence intervals for the random components of the 130 individuals representing the participants' threshold deviations in the multilevel proportional-odds mixed model.



Discussion

Principal Findings

This study provides clear evidence that an individual’s mood can be associated with health behaviors and activities in everyday life. Our findings are based on longitudinal diaries collected through the user-friendly, online ClickDiary platform, which allows participants to select their mood status and health behaviors by simply clicking appropriate response items in the respective diaries. Using this platform, participants can record their daily activities during the past 24 hours at their convenience. The design should have substantially reduced recall bias.

On average, it takes only about one minute to complete one health diary entry. Such a low-demanding task helps keep users participating in the study for a longer period of time. In addition, the visualization of their own diary entries and the random drawings help new participants sign up and existing participants to remain committed. Long-term data on variations in participants’ moods and health behaviors are essential for understanding the dynamic interactions between the two. In contrast, a cross-sectional or short-term follow-up study design would not allow researchers to collect the wealth of information on the daily variations provided by each participant. The advantages of such longitudinal data can be further examined by comparing our findings to those of previous studies that focus on emotions and health behaviors.

Comparison With Previous Research

Earlier studies showed that a higher BMI or being overweight was closely linked to negative affect among young adults and adverse psychosocial outcomes among grade-school children [7,28]. Among adolescents with excess weight, BMI was also a good predictor of emotion-driven impulsivity and cognitive inflexibility [29]. Our results, however, do not show that BMI has a significantly negative correlation with an individual's mood. Note that BMIs for most of the individuals in this study were within the normal range. This may be a reason for not finding a significant association in our analysis.

Most of our participants perceived themselves as being agreeable and conscientious. From the parameter estimates in the model, we found that agreeableness tends to be negatively correlated with a person’s mood, while relaxation is positively correlated. The finding differs from that of a study conducted in Finland [8], which showed that agreeableness is associated with higher positive and lower negative effects on mood, and that neuroticism predicts higher negative and lower positive effects on mood. The exact underlying reasons for the differences need to be further examined. The Finnish study, however, employed a very different data collection approach and sample groups—it recruited 106 university students aged 19 to 35 and used portable devices, such as mobile phones, to collect information 10 times per day for 1 week. By contrast, the ClickDiary participants came from different backgrounds, including students, full-time workers, housewives, and retired persons, between the ages of 20 and 67. Our measure of daily mood was recorded once a day for an average of 69 days. While the Finnish study collected the mood changes detailed within each day, our study collected the mood variations on different days over a longer period of time.

With regard to daily diets, we found that eating vegetables and fruits is related to better mood, which is consistent with findings from previous studies [6,7,30]. The possible biological mechanism is from the polyphenols found in fruits and vegetables. They can battle oxidative stress and help stimulate the activation of the neural molecules that aid in synaptic plasticity, which is important for cognitive function [31]. Currently, there are over 8000 polyphenolic compounds that have been identified in the world. For example, cocoa polyphenols have been shown to cause positive mood in one randomized study [32]. We also found, however, that eating certain unhealthful foods, such as fried food and sweetened beverages, is correlated to a slightly positive mood. Because it is difficult to differentiate the temporal order between eating behaviors and mood changes, we cannot conclude that such unhealthful foods actually trigger more positive emotions. Although such foods might play a role in promoting a good mood, we need further experimentation and validation.

Our study confirms that the participants are clearly in a better mood on the weekends, as well as when they travel farther away from their homes. Those without ILI symptoms are also happier than their counterparts with ILI symptoms. Those who exercise more also tend to be in a better mood compared to those who do not exercise, a finding consistent with those of earlier studies [18,33,34]. In addition, the longer one exercises, the better his or her mood becomes.

Most important, we found the quality of sleep to be a strong factor in distinguishing how one's mood changes from day to day. Having slept well or not during the previous night has a clear effect on a person's mood the following day. As also found in other studies [19,35], better-quality sleep clearly leads to a more positive mood the following day.

Limitations

Despite various interesting findings from this study, some limitations remain to be addressed. First, our sample of participants was not representative of Taiwan's population. Due to the longitudinal nature of diary studies, we required long-term and demanding commitment from our participants. After further screening for complex data analyses, only 130 participants met our strict threshold. Even though we tried to recruit participants through various channels, the demographic distribution of our participants was still skewed to females and young adults. Such a skewed sample tends to be common in many online surveys as well [7,36,37], which might be due to health issues and patterns of computer use. As a result, we cannot infer our findings to the general population, but the internal validity of the study is retained.

Second, we cannot make definite causal inferences between moods and health behaviors. Our ClickDiary platform requires participants to record their moods and health behaviors during the past 24 hours, but the temporal order of emotions and health behaviors remains unclear. Therefore, we can identify only the overall mood of the participant on a given day and his or her corresponding health behaviors on the same day. In addition, there are still many uncollected factors such as working stress and other life behaviors affecting people's moods within the day. In the current study design, we were not able to capture and adjust all these factors in the model.

Third, some participants' contact and health diaries may not have been complete. In this study, the average length of participation was about 69 days during the 184 days of follow-up. Unlike previous diary research that managed to collect complete information about one's daily contacts during the full periods of study [38,39], the diary records of some of our participants may have lagged at different intervals. The resulting sporadic records may have somewhat inhibited further analyses that required continuous time-series data. The study is still ongoing, however, and mobile apps for the iOS and Android systems have been released in January 2015. With participants from more diverse sources, a longer observation period, and more complete diary entries on a continuous basis, we expect to reexamine and further validate our current findings in the near future.

Conclusions

Sleeping well, eating more fruits and vegetables, and exercising longer all contribute significantly to improving a person's mood in everyday life. Using our online ClickDiary program, which helps reduce the recall bias associated with traditional one-shot surveys, we collected data on a daily basis to carefully identify the links between health behaviors and mood.

Acknowledgments

This research was supported by a grant from the Academia Sinica, Taiwan, Republic of China (ROC) (AS-103-TP-C03), and partially supported by a research grant from the Ministry of Science and Technology (grant No. NSC 102-2118-M-001-013-MY2). We would like to thank Ms Tsuey-Hwa Hu and Ms Jie-Yu Sung for their data management, analysis, and administrative support.

Conflicts of Interest

None declared.

References

1. Siahpush M, Spittal M, Singh GK. Happiness and life satisfaction prospectively predict self-rated health, physical health, and the presence of limiting, long-term health conditions. Am J Health Promot 2008;23(1):18-26. [doi: 10.4278/ajhp.061023137] [Medline: 18785370]

2. Wang F, Orpana HM, Morrison H, de Groh M, Dai S, Luo W. Long-term association between leisure-time physical activity and changes in happiness: analysis of the Prospective National Population Health Survey. Am J Epidemiol 2012 Dec 15;176(12):1095-1100 [FREE Full text] [doi: 10.1093/aje/kws199] [Medline: 23171884]
3. Levine LJ, Safer MA. Sources of bias in memory for emotions. Curr Dir Psychol Sci 2002 Oct;11(5):169-173. [doi: 10.1111/1467-8721.00193]
4. Tadic M, Oerlemans WM, Bakker A, Veenhoven R. Daily activities and happiness in later life: the role of work status. J Happiness Stud 2012 Sep 28;14(5):1507-1527. [doi: 10.1007/s10902-012-9392-9]
5. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
6. Piqueras JA, Kuhne W, Vera-Villarroel P, van Straten A, Cuijpers P. Happiness and health behaviours in Chilean college students: a cross-sectional survey. BMC Public Health 2011;11:443 [FREE Full text] [doi: 10.1186/1471-2458-11-443] [Medline: 21649907]
7. White BA, Horwath CC, Conner TS. Many apples a day keep the blues away--daily experiences of negative and positive affect and food consumption in young adults. Br J Health Psychol 2013 Nov;18(4):782-798. [doi: 10.1111/bjhp.12021] [Medline: 23347122]
8. Komulainen E, Meskanen K, Lipsanen J, Lahti JM, Jylhä P, Melartin T, et al. The effect of personality on daily life emotional processes. PLoS One 2014;9(10):e110907 [FREE Full text] [doi: 10.1371/journal.pone.0110907] [Medline: 25343494]
9. Kim-Prieto C, Diener E, Tamir M, Scollon C, Diener M. Integrating the diverse definitions of happiness: a time-sequential framework of subjective well-being. In: Delle Fave A, editor. The Exploration of Happiness: Present and Future Perspectives. Dordrecht, Netherlands: Springer Netherlands; 2013:47-75.
10. Yip T, Cross WE. A daily diary study of mental health and community involvement outcomes for three Chinese American social identities. Cultur Divers Ethnic Minor Psychol 2004 Nov;10(4):394-408. [doi: 10.1037/1099-9809.10.4.394] [Medline: 15554801]
11. Connelly M, Bickel J. An electronic daily diary process study of stress and health behavior triggers of primary headaches in children. J Pediatr Psychol 2011 Sep;36(8):852-862 [FREE Full text] [doi: 10.1093/jpepsy/jsr017] [Medline: 21441572]
12. Schneider S, Stone AA. Distinguishing between frequency and intensity of health-related symptoms from diary assessments. J Psychosom Res 2014 Sep;77(3):205-212. [doi: 10.1016/j.jpsychores.2014.07.006] [Medline: 25149030]
13. Tokuda Y. Gastrointestinal symptoms in a Japanese population: A health diary study. WJG 2007 Jan 28;13(4):572-578 [FREE Full text] [doi: 10.3748/wjg.v13.i4.572] [Medline: 17278223]
14. Chan TC, Fu YC, Hwang JS. Changing social contact patterns under tropical weather conditions relevant for the spread of infectious diseases. Epidemiol Infect 2015 Jan;143(2):440-451. [doi: 10.1017/S0950268814000843] [Medline: 24725605]
15. Fu YC, Wang DW, Chuang JH. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. PLoS One 2012;7(10):e45113 [FREE Full text] [doi: 10.1371/journal.pone.0045113] [Medline: 23056193]
16. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social mixing patterns in rural and urban areas of southern China. Proc Biol Sci 2014 Jun 22;281(1785):20140268 [FREE Full text] [doi: 10.1098/rspb.2014.0268] [Medline: 24789897]
17. Brose A, Scheibe S, Schmiedek F. Life contexts make a difference: emotional stability in younger and older adults. Psychol Aging 2013 Mar;28(1):148-159. [doi: 10.1037/a0030047] [Medline: 23066802]
18. Khazaee-Pool M, Sadeghi R, Majlessi F, Rahimi Foroushani A. Effects of physical exercise programme on happiness among older people. J Psychiatr Ment Health Nurs 2015 Feb;22(1):47-57. [doi: 10.1111/jpm.12168] [Medline: 25492721]
19. Kalmbach DA, Pillai V, Roth T, Drake CL. The interplay between daily affect and sleep: a 2-week study of young women. J Sleep Res 2014 Dec;23(6):636-645. [doi: 10.1111/jsr.12190] [Medline: 25082413]
20. ClickDiary Research Team. ClickDiary. 2015. URL: http://clickdiary.stat.sinica.edu.tw/ [accessed 2015-01-23] [WebCite Cache ID 6VmpJEzjQ]
21. Yuan C, Lv J, VanderWeele TJ. An assessment of health behavior peer effects in Peking University dormitories: a randomized cluster-assignment design for interference. PLoS One 2013;8(9):e75009 [FREE Full text] [doi: 10.1371/journal.pone.0075009] [Medline: 24040377]
22. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
23. Oregon Research Institute. International Personality Item Pool. URL: http://ipip.ori.org/ [accessed 2015-01-25] [WebCite Cache ID 6VqUpeEAE]
24. Chang YH, Fu YC. The evolution of the Taiwan Social Change Survey. In: Proceedings of the JGSS Symposium 2003: Birth of JGSS and its Fruit-Social Surveys in Different Countries and Areas and JGSS. Osaka City, Japan: Institute of Regional Studies, Osaka University of Commerce; 2004 Presented at: JGSS Symposium 2003: Birth of JGSS and its Fruit; June 21-22, 2003; East Osaka City, Japan p. 149-160.
25. Tutz G, Hennevogl W. Random effects in ordinal regression models. Comput Stat Data Anal 1996 Sep;22(5):537-557. [doi: 10.1016/0167-9473(96)00004-7]
26. de Leeuw J, Meijer E, editors. Handbook of Multilevel Analysis. New York, NY: Springer-Verlag; 2008.

27. Christensen RHB. The Comprehensive R Archive Network. Ordinal: regression models for ordinal data URL: http://www.cran.r-project.org/web/packages/ordinal/index.html [accessed 2015-01-23] [WebCite Cache ID 6VmpZ3mwh]
28. van Grieken A, Renders CM, Wijtzes AI, Hirasing RA, Raat H. Overweight, obesity and underweight is associated with adverse psychosocial and physical health outcomes among 7-year-old children: the 'Be active, eat right' study. PLoS One 2013;8(6):e67383 [FREE Full text] [doi: 10.1371/journal.pone.0067383] [Medline: 23825655]
29. Delgado-Rico E, Río-Valle JS, González-Jiménez E, Campoy C, Verdejo-García A. BMI predicts emotion-driven impulsivity and cognitive inflexibility in adolescents with excess weight. Obesity (Silver Spring) 2012 Aug;20(8):1604-1610. [doi: 10.1038/oby.2012.47] [Medline: 22421897]
30. Conner TS, Brookie KL, Richardson AC, Polak MA. On carrots and curiosity: eating fruit and vegetables is associated with greater flourishing in daily life. Br J Health Psychol 2015 May;20(2):413-427. [doi: 10.1111/bjhp.12113] [Medline: 25080035]
31. Gomez-Pinilla F, Nguyen TT. Natural mood foods: the actions of polyphenols against psychiatric and cognitive disorders. Nutr Neurosci 2012 May;15(3):127-133 [FREE Full text] [doi: 10.1179/1476830511Y.0000000035] [Medline: 22334236]
32. Pase MP, Scholey AB, Pipingas A, Kras M, Nolidin K, Gibbs A, et al. Cocoa polyphenols enhance positive mood states but not cognitive performance: a randomized, placebo-controlled trial. J Psychopharmacol 2013 May;27(5):451-458. [doi: 10.1177/0269881112473791] [Medline: 23364814]
33. Kye SY, Park K. Health-related determinants of happiness in Korean adults. Int J Public Health 2014 Oct;59(5):731-738. [doi: 10.1007/s00038-014-0588-0] [Medline: 25033934]
34. Timperio AF, van Stralen MM, Brug J, Bere E, Chinapaw MJ, De Bourdeaudhuij I, ENERGY consortium. Direct and indirect associations between the family physical activity environment and sports participation among 10-12 year-old European children: testing the EnRG framework in the ENERGY project. Int J Behav Nutr Phys Act 2013;10:15 [FREE Full text] [doi: 10.1186/1479-5868-10-15] [Medline: 23374374]
35. Moturu ST, Khayal I, Aharony N, Pan W, Pentland AS. Sleep, mood and sociability in a healthy population. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5267-5270. [doi: 10.1109/IEMBS.2011.6091303] [Medline: 22255526]
36. Nelson EJ, Hughes J, Oakes JM, Pankow JS, Kulasingam SL. Estimation of geographic variation in human papillomavirus vaccine uptake in men and women: an online survey using facebook recruitment. J Med Internet Res 2014;16(9):e198 [FREE Full text] [doi: 10.2196/jmir.3506] [Medline: 25231937]
37. Forster H, Fallaize R, Gallagher C, O'Donovan CB, Woolhead C, Walsh MC, et al. Online dietary intake estimation: the Food4Me food frequency questionnaire. J Med Internet Res 2014;16(6):e150 [FREE Full text] [doi: 10.2196/jmir.3105] [Medline: 24911957]
38. Fu YC. Contact diaries: building archives of actual and comprehensive personal networks. Field methods 2007 May 01;19(2):194-217. [doi: 10.1177/1525822X06298590]
39. Fu YC, Ho HC, Chen HM. Weak ties and contact initiation in everyday life: Exploring contextual variations from contact diaries. Soc Networks 2013 Jul;35(3):279-287. [doi: 10.1016/j.socnet.2013.02.004]

Abbreviations

BMI: body mass index
ILI: influenza-like illness
IPIP: International Personality Item Pool
IRB: Institutional Review Board
OCEAN: openness, conscientiousness, extraversion, agreeableness, neuroticism
OR: odds ratio
ROC: Republic of China

Edited by G Eysenbach; submitted 04.02.15; peer-reviewed by F Wang, M Khazaeepool; comments to author 15.04.15; revised version received 23.04.15; accepted 23.04.15; published 15.06.15.

Please cite as:
Chan TC, Yen TJ, Fu YC, Hwang JS
ClickDiary: Online Tracking of Health Behaviors and Mood
J Med Internet Res 2015;17(6):e147
URL: http://www.jmir.org/2015/6/e147/
doi:10.2196/jmir.4315
PMID:26076583

Original Paper

A Scalable Framework to Detect Personal Health Mentions on Twitter

Zhijun Yin[1]; Daniel Fabbri[1,2], PhD; S Trent Rosenbloom[2,3,4,5], MD, MPH; Bradley Malin[1,2], PhD

[1]Dept. of Electrical Engineering & Computer Science, Vanderbilt University, Nashville, TN, United States
[2]Dept. of Biomedical Informatics, Vanderbilt University, Nashville, TN, United States
[3]Dept. of Medicine, Vanderbilt Univerisity, Nashville, TN, United States
[4]School of Nursing, Vanderbilt University, Nashville, TN, United States
[5]Dept. of Pediatrics, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Bradley Malin, PhD
Dept. of Electrical Engineering & Computer Science
Vanderbilt University
Department of Biomedical Informatics, Vanderbilt University
2525 West End Avenue, Suite 1030
Nashville, TN, 37203
United States
Phone: 1 615 343 9096
Fax: 1 615 936 8545
Email: b.malin@vanderbilt.edu

Abstract

Background: Biomedical research has traditionally been conducted via surveys and the analysis of medical records. However, these resources are limited in their content, such that non-traditional domains (eg, online forums and social media) have an opportunity to supplement the view of an individual's health.

Objective: The objective of this study was to develop a scalable framework to detect personal health status mentions on Twitter and assess the extent to which such information is disclosed.

Methods: We collected more than 250 million tweets via the Twitter streaming API over a 2-month period in 2014. The corpus was filtered down to approximately 250,000 tweets, stratified across 34 high-impact health issues, based on guidance from the Medical Expenditure Panel Survey. We created a labeled corpus of several thousand tweets via a survey, administered over Amazon Mechanical Turk, that documents when terms correspond to mentions of personal health issues or an alternative (eg, a metaphor). We engineered a scalable classifier for personal health mentions via feature selection and assessed its potential over the health issues. We further investigated the utility of the tweets by determining the extent to which Twitter users disclose personal health status.

Results: Our investigation yielded several notable findings. First, we find that tweets from a small subset of the health issues can train a scalable classifier to detect health mentions. Specifically, training on 2000 tweets from four health issues (cancer, depression, hypertension, and leukemia) yielded a classifier with precision of 0.77 on all 34 health issues. Second, Twitter users disclosed personal health status for all health issues. Notably, personal health status was disclosed over 50% of the time for 11 out of 34 (33%) investigated health issues. Third, the disclosure rate was dependent on the health issue in a statistically significant manner ($P<.001$). For instance, more than 80% of the tweets about migraines (83/100) and allergies (85/100) communicated personal health status, while only around 10% of the tweets about obesity (13/100) and heart attack (12/100) did so. Fourth, the likelihood that people disclose their own versus other people's health status was dependent on health issue in a statistically significant manner as well ($P<.001$). For example, 69% (69/100) of the insomnia tweets disclosed the author's status, while only 1% (1/100) disclosed another person's status. By contrast, 1% (1/100) of the Down syndrome tweets disclosed the author's status, while 21% (21/100) disclosed another person's status.

Conclusions: It is possible to automatically detect personal health status mentions on Twitter in a scalable manner. These mentions correspond to the health issues of the Twitter users themselves, but also other individuals. Though this study did not investigate the veracity of such statements, we anticipate such information may be useful in supplementing traditional health-related sources for research purposes.

(J Med Internet Res 2015;17(6):e138) doi:10.2196/jmir.4305

KEYWORDS

consumer health; information retrieval; machine learning; social media; twitter; infodemiology

Introduction

Background

Traditional methods for collecting data in support of clinical research include prospectively collected surveys (eg, [1]), retrospective analyses of existing medical records (eg, [2,3]), and a combination of the two (eg, [4]). Over the past decade, computerized methods for data collection have emerged, with traditional surveys for health research moving onto the Internet [5] and increasingly widespread electronic medical records (EMRs) able to be mined to investigate a wide range of acute and longitudinal phenotypes [6-8]. At the same time, these approaches tend to focus only on a medically centric worldview, and may provide only a partial view of a patient's life. Recognizing this limitation, investigators have suggested that the data contributed through non-traditional domains, such as mobile apps [9-11] and online forums where patients self-report on their status [12,13], will provide a more complete view of an individual's health and population-based health trends.

An increasing number of studies demonstrate that the data disseminated via social media platforms, such as Twitter, can inform health-related investigations. We review such studies in the following section, but we highlight that studies have shown, for instance, that such data can be mined to model aggregate trends about health (eg, detection of statistically significant adverse effects of pharmaceuticals [14,15]). Recent investigations have also demonstrated that an individual's health status can be corroborated by the statements they publish over social media platforms (eg, confirmation of flu diagnoses [16]). Despite the power of such investigations, they are limited in that the associated approaches do not filter data from social media streams for any arbitrary health-related concept.

Objective and Contribution

The objective of our work is to develop a scalable framework for detecting mentions about personal health on a specific social media platform, namely Twitter. The system introduced in this paper is composed of several core processes. First, the system filters the Twitter stream for tweets that are likely to contain health-related information. Next, a subset of the tweets are labeled with respect to the type of information that is communicated (eg, health status of the author versus a metaphorical statement) and applied to train a classifier. While it is possible to label a large number of tweets given a substantial budget, it is unlikely that a classifier could be specialized for each specific health issue. For instance, imagine a researcher is interested in studying 10,000 distinct health issues, each of which will require at least 500 tweets to train a robust classifier. If the cost to label each tweet is $0.10, it would cost $500,000 to build the necessary corpora! Our framework demonstrates that a scalable classifier, which discovers health mentions across a broad range of health issues, can be composed by leveraging a mixture of tweets from various health issues, which could make large-scale investigations much more cost-effective. In doing so, however, our system is oriented toward a high precision while maintaining a reasonable recall.

There are three primary contributions of this paper:

- Labeled Health Mention Corpus. We leverage Amazon Mechanical Turk to create a labeled corpus of tweets with health mentions for 34 health issues. These include certain high impact health issues investigated in the Medical Expenditure Panel Survey [17], such as arthritis, asthma, bronchitis, cancer, diabetes, hypertension, and stroke.
- Health Mention Detection. We introduce a system to automatically detect personal health mentions in tweet streams. We show that this system is trainable with a relatively small number of labeled tweets from several health issues. Moreover, it can effectively detect personal health mentions across a range of health issues on Twitter. For instance, training on 2000 tweets associated with four health issues (cancer, depression, hypertension, and leukemia) can yield a classifier that achieves a precision of 0.77 on the aforementioned corpus of tweets of 34 health issues.
- Health Mention Attribution. To demonstrate the potential for the data filtered from Twitter, we investigated how people reveal information about themselves and others. In doing so, we show that the likelihood an individual self-discloses is dependent on the health issues communicated. For example, personal health status is revealed more than 50% for 11 of the 34 health issues. For certain health issues (eg, allergies, bronchitis, insomnia, migraines, and ulcers), people are more likely to disclose their own health status, while for other health issues (eg, Alzheimer's, Down syndrome, leukemia, miscarriage, and Parkinson's), people are more likely to disclose another person's status.

Prior Work

Social Media and Health Research

As alluded to, various investigations have demonstrated that social media can be successfully leveraged to (1) enable individuals to discuss their health status, (2) influence an individual's health behavior, and (3) support the analysis of aggregate trends around health activities.

First, a certain portion of studies have focused on the extent to which, as well as how, social media enables self-reports of health information. Hale et al [18] showed that users discuss their health conditions on public Facebook pages, but recognized that such pages tend to be overly general to attract users to contribute to a discussion. However, Bodnar and colleagues [16] found that individuals who use social media discuss certain ailments with high accuracy on Twitter. Specifically, it was demonstrated that college students tend to talk about their influenza diagnosis and associated symptoms. More generally, Paul et al [19] performed latent topic model discovery over self-reported health status in Twitter to detect complex and potentially novel phenotypes. It has further been shown, that some Twitter users reveal genome sequencing results (in relation to ancestry information according to 23andme.com services) over Twitter [20].

Second, the previous investigations show that individuals publish information about themselves, but there is also a growing body of evidence to suggest that social media can influence an individual’s health behavior. In certain cases, exploitation of social media can bring about negative health behaviors. For instance, based on discussions about prescription abuse over Twitter, it was observed that social media may aggravate such problems [21,22]. In a similar vein, a content analysis of tweets, in association with the demographics of the followers of marijuana Twitter handles, showed that social media may allure young people to establish substance use patterns. Wilson et al also argued that social media enables more individuals to be involved in an anti-vaccination movement [23]. However, it was also shown that social media can encourage more positive changes in health behavior. Notably, it was shown that increasing communications with smokers on social media can promote free cessation services [24]. Moreover, Cobb and colleagues [25] developed a Facebook application that was able to track the significant elements of an intervention on smoke cessation. It was also found that the design and realization of a community opinion leader model may mitigate the spread of HIV [26].

Third, social media can be mined to learn and characterize aggregate trends with respect to health activities. For instance, it was shown that flu trends can be effectively extracted from Twitter using standard machine learning strategies [27]. More specifically, the analysis of daily tweets across a major metropolitan region (eg, New York) can enable the prediction of which health issues are currently influencing the health of the public [28]. Meanwhile, Nagel et al [29] showed that both the keywords chosen to filter and create subgroups of tweets affected prediction accuracy. Beyond health status, it has been illustrated that the rare or unknown side-effects of drugs can be discovered through sentiment analysis over Twitter [15].

Though social media can support a wide array of health-related investigations, there are a number of hurdles to making the associated methodologies scalable. As Curtis and colleagues [30] point out, for instance, insufficient procedures for protecting participants’ privacy was one of the challenges to recruiting members from social media to conduct HIV research. In addition, it was recently revealed that the unreliability of big data and continuous changes of search algorithms contributed to failures in the Google Flu Trends program [31].

Our work differs from the aforementioned studies in that we focus on personal health status disclosure on Twitter. We note that Mao et al [32] discussed a similar topic, but their work is limited in that (1) it relied on regular expressions for classification, (2) focused on a limited number of health issues, and (3) examined whether personal health status is disclosed on status or conversation, but did not differentiate when heath status was disclosed for authors versus others. Lamb et al [33] showed that a combination of tweets about infection with respect to both authors and others performed better than tweets about the authors alone when predicting flu trends, which lends credibility to our work. However, it should be noted that their classification only focused on a diagnosis of the flu instead of a broad range of health issues, as is addressed in our work.

Classification on Social Media

To mine health-related information from social media, it is critical to develop a classifier. However, tweets are constrained in size and, thus, are composed of limited content. Consequentially, it is essential to define and select discriminative features to support automated health status detection. In certain studies, tweets were enriched with features by referencing external sources, such as Wikipedia [34,35], to improve topic modeling, but their generality hampers them in the support of personal health mention detection.

As an alternative, it has been shown that punctuation, emoji characters, hashtags, and the @username designation, as well as text (including n-grams of words or characters [36]) from the webpage referenced by the URL in a tweet, can form meaningful features for classification purposes [34,37,38]. Features generated using natural language processing tools, such as part of speech tags and dependencies between terms were also successfully incorporated as features in social media classifiers [33,39]. Building on previous studies, our work illustrates that nouns, verbs, pronouns, punctuation, emoji, hashtags, as well as dependencies, can serve as effective features for personal health mention.

Social Media Corpus Construction

If we rely on a classifier to filter and analyze social media, then it is essential to obtain (or create) a labeled corpus to train the classifier. Crowdsourcing over Web-based platforms, such as Amazon Mechanical Turk (MT), has been employed to generate labeled gold standard corpora [37]. Notably, MT was leveraged to label when tweets were related to the health status of the author of a tweet in the latent topic modeling analysis discussed above [19]. However, it should be recognized that the survey utilized by [19] is limited in that it only related tweet content to the author and not another person’s health status.

The Personal Health Status Mention Problem

To formalize the problem, we define the notions of personal health status and mention: Definition 1 (Personal Health Status) is the health condition of a specific person regarding a health issue or symptom, and Definition 2 (Personal Health Mention) is a statement of personal health status in social media.

These definitions focus on the health information of the individuals who are potentially identifiable. For instance, tweets such as “my father is cancer free for ten years”, “I have to do chemo tomorrow”, and “my little cousin has leukemia” are representatives of personal health mentions. By contrast, “Local charity doing great work to help cancer patients” is not a personal health mention because the subject is a group of people as opposed to a specific person.

We treat the problem of personal health mention detection as binary classification. We say a tweet is positive if it reveals personal health status and negative otherwise. For example, two MT masters assigned positive labels to each of the first three tweets in Table 1 (details in Method Section). Yet a term associated with a health issue can be uttered on Twitter for many other reasons, such as in a metaphorical sense, to express a viewpoint about a health issue in general, or to communicate a worry. The next three tweets in Table 1 provide examples of these reasons respectively.

Given their brevity (140 characters at most), tweets often have limited context. Consequentially, assigning a class label to a tweet is substantially more challenging than detecting if a given tweet communicates status of the author. The last three tweets in Table 1 illustrates this observation, where MT masters assigned different option labels to the same tweet.

In this paper, we study how people disclose personal health statuses on Twitter and present a scalable personal health mentions detection system for the Twitter stream. Specifically, we decompose this investigation into the following four hypotheses: H1: People discuss personal health status on Twitter; H2: Personal health status disclosure rate is health issue dependent; H3: The likelihood that people disclose their own versus other people's personal health status is health issue dependent; and H4: Personal health status mention classifiers based on tweets of multiple health issues are more scalable than those based on a single health issue.

Table 1. Examples of tweets related to health issues and the labels obtained through the Mechanical Turk (MT) survey.

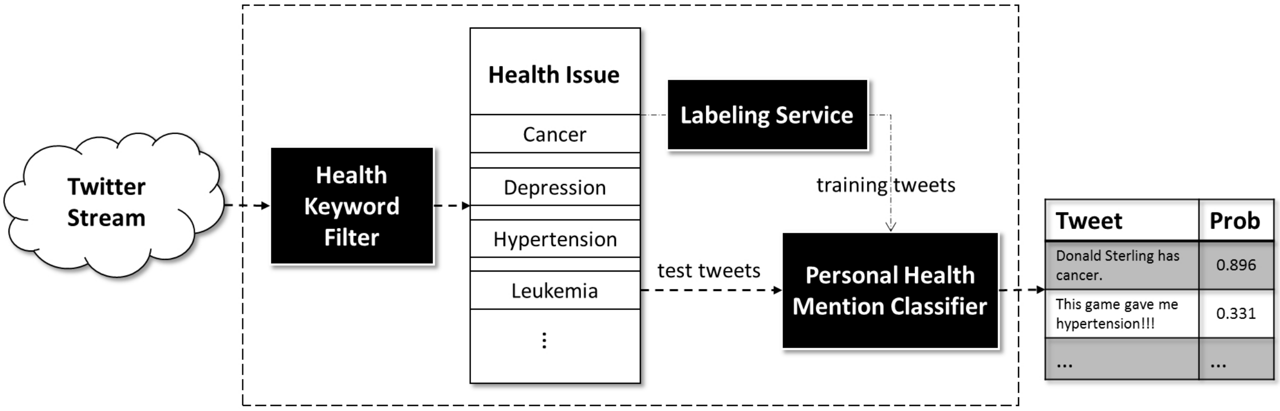
Tweet	Label via MT	
	Master 1	Master 2
Positive		
I'm suffering from schizophrenia and a little bit of insomnia.	author	author
Prayers for my dad would be appreciated. He has lymphoma. Thanks for the support everyone.	relative	relative
didn't she have a miscarriage like 3 days ago?	someone else	someone else
Negative		
you're gonna give Viv a heart attack	metaphor	metaphor
Even after Bill Gates relentless support and millions of dollars poured into Malaria research, we are not successful.	viewpoint	viewpoint
Praying I don't have pneumonia	worry	worry
Ambiguous		
Cheerios say she'll never have to worry about dieting. Too bad with 2:1 sodium to cal, she'll have to worry about high blood pressure.	metaphor	someone else
Yooo soo i walk out my apt and here this girl screaming for help. Apparently, she kneed her testicular cancer bf in the nuts repeatedly.	metaphor	someone else
memorial find. 10% of your bills went to leukemia and lymphoma research. when amber was around she brightened everyone's day in one way.	viewpoint	someone else

Methods

System Pipeline

Figure 1 provides a high-level summary of the system engineered to detect personal health mentions on Twitter. The system is composed of three primary components: (1) a filtering service (eg, a keyword filter based on health issues), (2) a labeling service, and (3) a health mention classification service. First, tweets collected via the Twitter streaming API are passed into a filter and stored in a bin indicative of a specific health issue. Next, a sample of the tweets associated with these health issues are sent to a labeling service (eg, MT). Once labeling is complete, a personal health mention classifier is trained and applied to report the probability that new incoming tweets correspond to such mentions.

Figure 1. Framework for personal health mention detection over Twitter. First, tweets are filtered into bins according to health issue topic. A portion of the tweets are supplied to a labeling service. The labeled data is then applied to train a classifier to detect personal health mentions.



Construction of a Health Mention Corpus

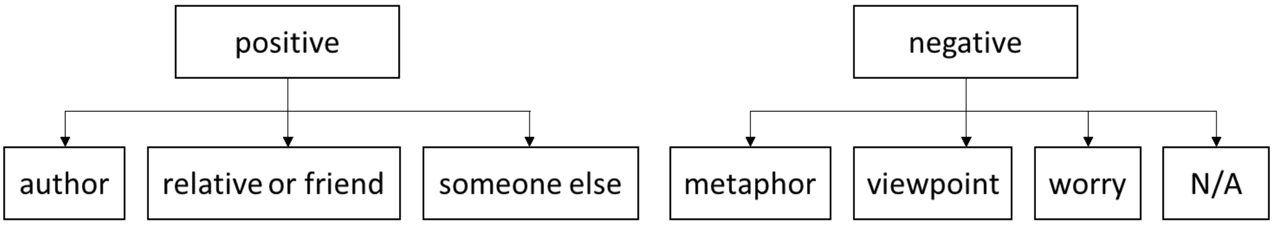
To create a labeled corpus of health status mentions, we solicited annotators through MT. Specifically, we set up a survey for labeling a corpus on MT, the details of which are in Multimedia Appendix 1. For each tweet, we directed two MT masters to select the best of seven options that describe how the tweet uses the health issue. These options represent the common usage of most health issues. We validated the reliability of the MT masters by illustrating that they exhibit high concordance in their labels (details in Tables A-2, A-3 in Multimedia Appendix 1, and in Multimedia Appendix 2). Figure 2 depicts how the options relate to the positive and negative labels.

The positive class includes the labels of author, relative or friend, and someone else. The negative class consists of labels for metaphor, viewpoint, and worry. Table 1 provides examples of tweets and the labels supplied by the MT masters. The last option label, N/A, which means none of the above, is also treated as a negative label in this investigation because it was observed (by the authors) that such labels were generally negative. For instance, these include tweets with job related information, which is spam that has nothing to do with a personal health mention.

For the purposes of this study, we created four types of datasets. The formalization of the design of these datasets is available in Table B-1 in Multimedia Appendix 3. We refer to the first as the gold standard dataset. It consists of all tweets with labels agreeing at the positive (negative) level. This dataset represents an ideal case where readers can determine when a tweet communicates personal health status. For example, this dataset treats tweets as positive when labeled as author by one MT master and someone else by a second MT master. By contrast, this dataset discards tweets labeled as relative or friend and worry.

Given the difficulty in labeling tweets in practice, we generated three additional datasets to resolve label conflicts. The first is the conflict as positive (CAP) dataset, which treats tweets with conflicting labels as positive. The second is the conflict as negative (CAN) dataset, which treats tweets with conflicting labels as negative. The third is the TieBreak dataset, which uses a third MT master to break the tie. These datasets represent the best case, the worst case, and the general case in the real world and we rely upon them to assess the system’s scalability.

Figure 2. Label hierarchy.



System Classifier Evaluation Roadmap

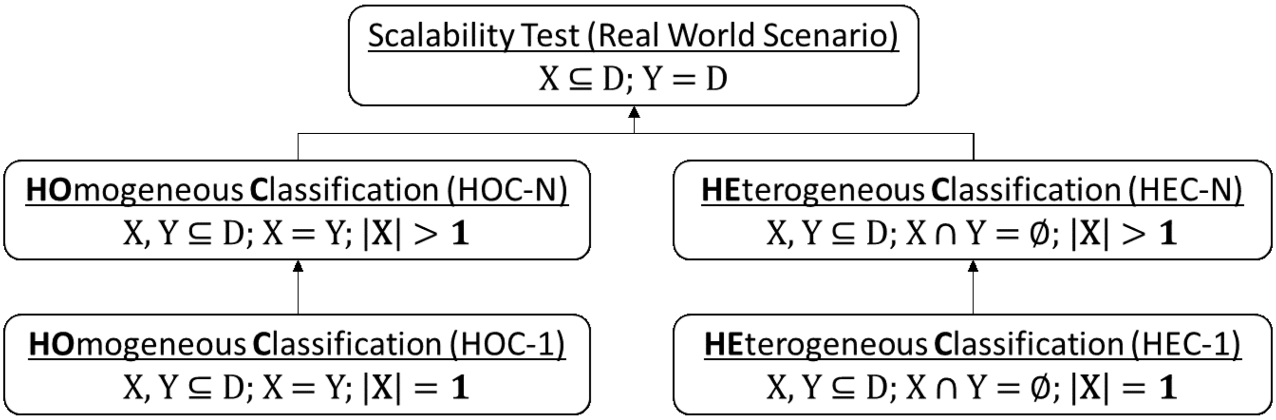
System scalability emphasizes the ability to detect mentions for many, potentially unknown, health issues communicated via social media, using the labeled tweets from a limited number of health issues.

To formalize the scenario, let D be the set of health issues and X and Y be the set of health issues selected to train and test the classifier, respectively. By default, $X, Y \subseteq D$.

As depicted in Figure 3, we assess two variations on classification. The first, which we refer to as homogeneous classification, corresponds to the traditional machine learning setting where a classifier is trained and tested on tweets from the same health issue. The second, which we refer to as heterogeneous classification, corresponds to when we train and test the classifier on tweets from disparate health issues. This type of scenario arises when a researcher attempts to reuse a classifier developed for one health issue on a different problem. Figure 3 further illustrates two training strategies to scale the system in a real-world scenario: train the classifier on tweets from (1) one health issue, which results in homogeneous classification with |*X*| = 1 (HOC-1) and heterogeneous classification with |*X*| = 1 (HEC-1), and (2) many health issues, which results in homogeneous classification with |*X*| > 1 (HOC-N) and heterogeneous classification with |*X*| > 1 (HEC-N).

The ideal scalability test is to train an HOC-1 classifier for every health issue in *D* with a sufficient quantity of labeled tweets. However, it is difficult to realize this scenario in practice because of limited budgets for gathering and annotating such corpora. As such, we performed a series of experiments to compare the performance of the various models (ie, HOC-1, HOC-N, HEC-1, and HEC-N) and leverage the best model to conduct scalability tests in a real-world scenario.

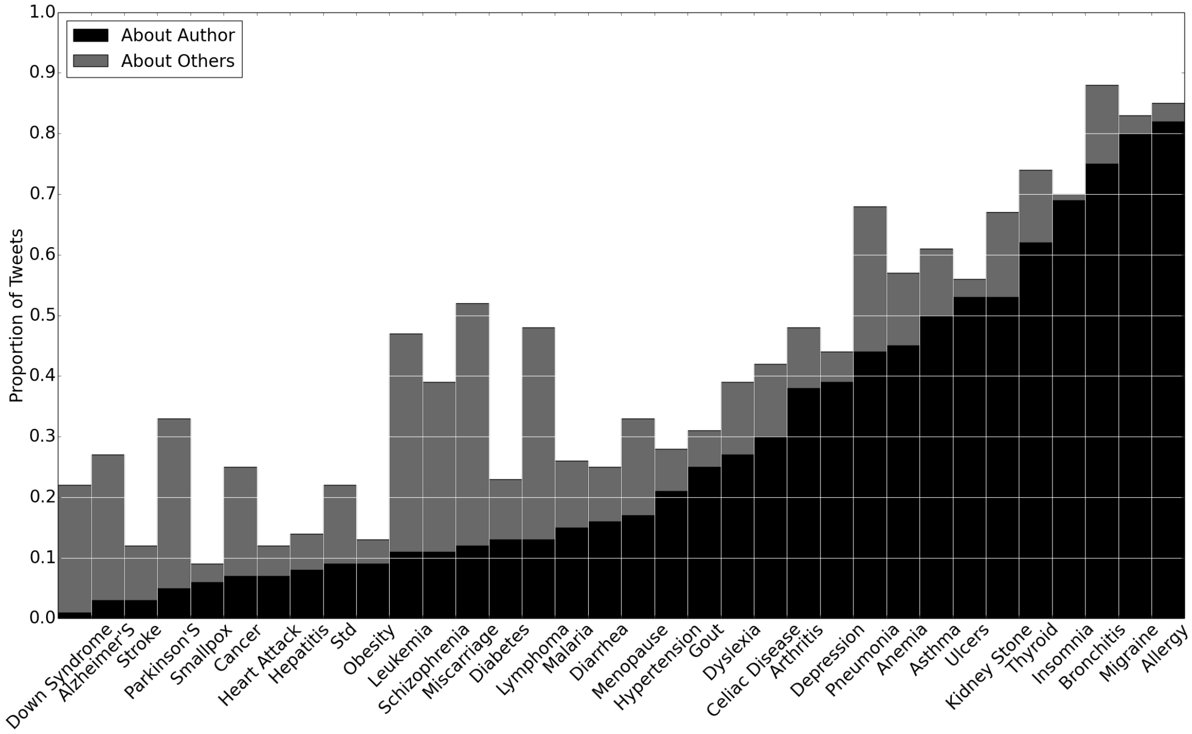
Figure 3. Overview of evaluation strategies for the personal health status mention classifier. Note, D={d1, d2, …, dn} is set of health issues, X is set of health issues selected to train classifier, and Y is set of health issues used to test classifier.



Performance Measures

To assess the performance of the system, we rely upon the standard measures of precision and recall. In our setting, precision (P) corresponds to the proportion of tweets classified as positive that are in fact positive. Recall (R)corresponds to the fraction of real positive tweets that are classified as positive. Given the large volume of tweets and the often unbalanced positive/negative class ratio per health issue (see Table 2 and Figure 4), we emphasize P while setting R to a reasonable level. Henceforth, we report the area under the PR curve (AUPRC) to evaluate how a classifier performs in general. We consider the PR curve, which can be more indicative of a classifier's performance when the class ratio is highly imbalanced [40]. To characterize general performance, we report on AUPRC when testing the scalability of the system.

Figure 4. The extent to which people tweet about themselves versus others when disclosing personal health status. Note that this is a stacked bar chart, such that the sum of the author and others proportions corresponds to the overall proportion of positive instances.



Health Status Classifier

One of the aims in this research is to examine whether we can use classifiers trained with tweets from multiple health issues to detect personal health mentions about other health issues. Hence, it should be noted that the goal of our research is to examine the effectiveness of classifiers when supplied with a set of known (or off-the-shelf) features. We use a Multinomial Naïve Bayes (MNB) binary classifier based on four types of features associated with tweets. Alternatively, we can plug other learning algorithms, such as logistic regression or a support vector machine, into the framework as the base classifier. Previous investigations verified the effectiveness of such features [33,34,37-39].

- Nouns, verbs, and pronouns. We transformed each word into its lemma form. Though pronouns are often defined as stop terms (which are discarded in traditional natural language processing), they are retained because they can disclose the personal health status of a friend or family member (eg, “My mom makes having cancer look good”).
- Dependencies. These are grammatical relations [41] between words in a tweet, such that one of the words is a health issue. We replaced terms for health issues with the keyword diagnosis to compact the feature space. For example, the dependency (“dobj”, “have”, “cancer”) is converted into a feature that can be supplied to MNB, dobj_have_diagnosis.
- Punctuation and Emoji. These can indicate an author’s emotion and may improve classification (e.g., “my uncle is cancer free !!!!!! lol”).
- HTTP LINK, #hashtags, and @username. These features represent the existence of link, hashtag, and @username in a tweet, respectively.

Experiment Design

Overview

In our experiments, we highlight the evaluation of two important factors that can affect the scalability of a classifier: (1) the diversity of health issues in the training data, and (2) the quantity of training tweets. When we compare different classifiers, we focus on the former. When we test system scalability, beside the system scalability, we also evaluate the performance of the classifiers with different size of training dataset. The following provides details of the experiment design.

Dataset

We use the 34 health issues depicted in Figure 4 to represent D and define a synthetic health issue, or SYND, as the union of cancer, depression, hypertension, and leukemia. We select cancer and leukemia, for which tweets are skewed toward communicating about other people’s health status, and depression and hypertension, for which tweets are skewed toward communicating about the author’s health status. We first applied the keywords (shown in Table D-1 in Multimedia Appendix 4), which were selected based on these health issues under the guidance of a clinical expert, to filter for tweets associated with the keywords. Then, we chose 1000 tweets, at random, for each of the four health issues to obtain the gold standard datasets. We also choose 100 tweets, at random, for each of health issue in D to generate gold standard, CAN, CAP and TieBreak datasets.

Comparison Between HOC-1 and HOC-N

We use the cancer, depression, hypertension, and leukemia gold standard datasets to train each homogeneous classifier. There are two situations where we can evaluate how the diversity of health issues in the training data influence the homogeneous classifiers. First, suppose that we aim to detect multiple health issues. Given a fixed number of training tweets, how does an HOC-N classifier (eg, trained with SYND) differ from a group of HOC-1 classifiers (eg, four HOC-1 classifiers)? Second, now imagine we wish to perform detection for only one single health issue (eg, cancer). Given a fixed number of training tweets, how does a HOC-N classifier (eg, trained with SYND and test on cancer) differ from the associated HOC-1 classifier (eg, cancer HOC-1 classifier)?

Comparison Between HEC-1 and HEC-N

To evaluate the diversity of health issues in training dataset, we compare HEC-1 with HEC-N ($2 \leq |X| \leq 4$). In particular, we use the cancer, depression, hypertension and leukemia gold standard datasets for training and the gold standard dataset of D SYND to test all of the heterogeneous classifiers.

System Scalability Test

When assessing system scalability, we test the classifier on the CAN, CAP, and TieBreak datasets of D. This enables the evaluation of the performance of the system in a real-world scenario. We also test the classifier trained with different number of tweets.

Experimental Methodology

For each experiment, we stratify the tweets and generate 30 train-test sets. In doing so, (1) each set preserves the proportion of samples for each positive (negative) class, and (2) the data is partitioned, such that we train on 80% of the tweets while we test on the remaining 20%. To control the comparison, the size of the training set for each compared classifier is equivalent.

Results

Dataset

We used the Twitter streaming API to filter for tweets between May 7, 2014 and July 23, 2014 that were (1) published in the contiguous United States according to their geolocation, and (2) written in the English language only. A total of 261,468,446 tweets were subject to a filter composed of keywords for 34 health issues, resulting in 281,357 tweets (0.11%) for further investigation.

How People Disclose Personal Health Status on Twitter

To demonstrate the opportunities for a personal health mention detection system, we conducted an investigation to test H1, H2, and H3. We chose 100 tweets, at random, for each of the 34 health issues as shown along the x-axis of Figure 4, to generate the TieBreak dataset. These health issues are based on common and high impact health issues as defined by the Medical Expenditure Panel Survey [17]. This figure illustrates how often people disclose their own health status as opposed to other individuals' status. The black bar, "About Author", represents the proportion of positive tweets with the author label. The gray bar, "About Others", represents the proportion of positive tweets with the label relative or friends and someone else. For a specific health issue, the sum of the two values is equal to the proportion of positive tweets for this health issue. For example, 40% of the tweets about miscarriages (40/100) disclosed other people's status, while only 12% (12/100) disclosed the author's status (such that 52%, 52/100, of the tweets were positive instances).

To test hypothesis H2 (personal health status disclosure rate) and H3 (who the disclosure is about), we define the following null hypotheses: $H2_o$: The rate of positive and negative tweets is independent of the health issues, and $H3_o$: The rate of tweets disclosing the author's health status and others' health status is independent of the health issues.

To test these hypotheses, we used the TieBreak dataset, which (due to randomness) represents 100 samples from each of the 34 distributions regarding how people disclose health status. To test H2, we applied a chi-square test on these two variables: the number of positive tweets and the number of negative tweets in each health issue samples. To test hypothesis H3, we applied a Spearman correlation test on these two variables: the rate of tweets disclosing the author's health status and the rate of tweets disclosing the others' health status. We set the alpha level of significance to .05.

The results reveal several notable pieces of evidence, which are related to the first three hypotheses posed above.

- People disclose personal health status on Twitter for a range of health issues (H1). The disclosure rate for each of the 34 health issues is greater than 9%. There are 29 health issues with disclosure rates greater than 20% and 11 health issues with disclosure rates greater than 50%. The latter group includes: allergies (85/100), anemia (57/100), arthritis (48/100), asthma (61/100), bronchitis (88/100), insomnia (70/100), kidney stones (67/100), migraines (83/100), miscarriages (52/100), pneumonia (68/100), thyroid (74/100) problems, and ulcers (56/100).
- Health status disclosure rate is dependent on the health issue, $\chi^2_{33}=697$, $P<.001$. For instance, more than 80% of the tweets about migraines (83/100) and allergies (85/100) communicate personal health status. By contrast, only ~10% of tweets about obesity (13/100) and heart attacks (12/100) communicate personal health status. Bronchitis (88/100) exhibits the largest proportion of tweets that disclose personal health status, while smallpox (9/100) exhibits the smallest proportion.
- The likelihood that people disclose their own versus other people's health status is dependent on the health issue, $Z=-5.745$, $P<.001$. For instance, 69% (69/100) of tweets about insomnia disclose the author's personal health statuses compared, while only 1% (1/100) disclose another person's status. By contrast, 1% (1/100) of the tweets for Down syndrome disclose the author's status, while 21% (21/100) disclose another person's status.

Classification Evaluation

Classification Data Set

We extracted the gold standard datasets for each of the four health issues mentioned in the Methods section. Table 2 summarizes the number of tweets in each class. Except leukemia, which has a balanced positive and negative instance space, there were substantially more negative than positive tweets. Due to the definition of SYND, the number of positive and negative tweets of the synthetic health issue is the sum of the four health issues.

Table 2. The number of positive and negative tweets in the gold standard datasets.

Tweet	Cancer	Depression	Hypertension	Leukemia	SYND[a]
Positive	166	261	211	436	1074
Negative	697	461	551	423	2132

[a] SYND: synthetic health issue (D).

Most Informative Features

Before conducting an in-depth empirical investigation, we inspected the classifiers and their corresponding features to determine if they are intuitive. Here, we report on the top 10 informative features by training in a homogeneous classification setting with tweets of each of the five health issues (cancer, depression, hypertension, leukemia, and SYND). Table 3 reports these features for each classifier.

The results show the effectiveness of feature selection in several ways. First, more than five features are pronouns, such as I, my, and she (which was also confirmed in [32]). These are stop words that are typically removed in the context of general text classification. However, in our scenario, they appear to signify users who disclose health information about themselves and others (eg, "my mom makes having cancer look easy"). Second, certain words, such as get, have, and battle, when applied in conjunction with a health issue, can disclose personal health status (eg, "my friend lost his battle to leukemia"). Third, dependencies, such as "obj_have_diagnosis", are strong positive indicators (eg, "I have seasonal allergy").

This table also provides several notable results about other behaviors when people disclose personal health status. For instance, people often include @someone in health mentions. They use links to provide additional information such as pictures, locations, or texts, or use exclamation mark to express strong feelings about personal health status.

The hypertension classifier was notable because it had specific health-related terminology ranked highly. Specifically, the term blood is highly informative for this classifier. We suspect this is because hypertension is commonly referred as high blood pressure.

Table 3. The most informative features for homogeneous health mention classification.

Rank	Cancer	Depression	Hypertension	Leukemia	SYND[a]
1	I	I	I	I	I
2	my	my	my	My	My
3	!		have		
4		you		HTTP LINK	!
5	you	it	dobj_have_diagnosis	!	Have
6	have	go	!	She	HTTP LINK
7	she	poss_diagnosis_my	get	Have	She
8	He	!	she	He	You
9	HTTP LINK	get	it	Battle	obj_have_diagnosis
10	obj_have_diagnosis	have	blood	Help	He

[a] SYND: synthetic health issue (D).

Homogeneous and Heterogeneous Classification

In this experiment, we compared the effectiveness of homogeneous and heterogeneous classifiers and then testing on tweets from each of the five health issues. Table 4 provides the AUPRCs for each homogeneous (along the diagonal) and heterogeneous (off diagonal cells) health mention classifier. Each row corresponds to the health issue relied upon for training the classifier, while each column corresponds to the health issue the classifier was applied to. To test the significance, we ran a *t* test when the results followed a normal distribution and a Kolmogorov-Smirnov (KS) test otherwise.

First, it should be noted that each homogeneous classifier outperforms the heterogeneous classifiers when testing the corresponding health issue tweets, but such classifiers do not generalize. It can be seen that the leukemia HOC-1 classifier achieved the highest AUPRC. This may be due to the balance in the positive and negative classes for this health issue. However, it was observed that the homogeneous classifiers exhibited much higher variance compared to the heterogeneous classifiers. This suggests that heterogeneous classifiers may yield stable results.

Second, the HEC-1 classifier may tend to obtain a better AUPRC when testing on health issues with a similar author-to-others disclosure rate. For instance, cancer achieved the best AUPRC when testing on leukemia tweets. Meanwhile, leukemia achieved the best AUPRC when testing on cancer tweets. Depression and hypertension also achieved the best AUPRC when testing on each other.

Third, it also shows that SYND heterogeneous classifier (HEC-N) was the second best heterogeneous classifier when testing on cancer, depression, and leukemia tweets, and the best heterogeneous classifier when testing on hypertension. Considering that the HEC-1 classifier is specialized to a certain health issue, the HEC-N classifier may provide a more scalable alternative when filtering for personal health mentions on other health issues.

Table 4. AUPRC for homogeneous and heterogeneous classifiers.[a]

	Cancer	Depression	Hypertension	Leukemia	SYND
	mean (SD)				
Cancer	0.732 (0.058)	0.528 (0.018) [b]	0.552 (0.014)[b]	0.869 (0.009)[b]	0.728 (0.009)[b]
Depression	0.441 (0.007)[b]	0.663 (0.054)	0.611 (0.014)[b]	0.821 (0.006)[b]	0.666 (0.006)[b]
Hypertension	0.451 (0.009)[b]	0.646 (0.011)	0.664 (0.062)	0.726 (0.008)[b]	0.616 (0.006)[b]
Leukemia	0.638 (0.011)[b]	0.603 (0.011)[b]	0.559 (0.019)[e]	0.936 (0.019)	0.579 (0.007)[b]
SYND[f]	0.625 (0.022)[e]	0.618 (0.026)[d]	0.626 (0.019)[c]	0.831 (0.023)[b]	0.820 (0.0180

[a] AUPRC: area under the precision recall curve. Classifiers were trained with row health issue tweets and tested on column health issue tweets. Within each column, a hypothesis test was conducted between HOC-1 and each model that is not HOC-1 (eg, HOC-1 vs HEC-1).

[b] *P*<.001

[c] *P*=.002

[d] *P*=.003

[e] *P*=.004

[f] SYND: synthetic health issue (D).

Table 5. AUPRC of homogeneous health mention classifiers, given the same number of training tweets.[a]

Classifier	Cancer mean (SD)	Depression	Hypertension	Leukemia
HOC-1[b]	0.732 (0.058)	0.663 (0.054)	0.664 (0.063)	0.936 (0.019)
HOC-N[c]	0.723 (0.061)	0.645 (0.053)	0.672 (0.070)	0.927 (0.022)
HOC-N‡	0.756 (0.050)	0.681 (0.050)	0.702 (0.059)[d]	0.940 (0.021)

[a] AUPRC: area under the precision recall curve. Within each column, the hypothesis test was conducted between HOC-1 and each model that is not HOC-1 (eg, HOC-1 vs HOC-N).

[b] HOC-1: homogeneous classification with |X| = 1

[c] HOC-N: homogeneous classification with |X| > 1

[d] *P*=.015

Comparison of Homogeneous Classifiers

In this experiment, we evaluated how homogeneous classifiers are influenced by (1) the number of health issues in the training set, and (2) the number of tweets used for training classifiers. Table 5 shows the results for the HOC-1 and HOC-N classifiers when testing on the tweets of each health issue. For each column, we trained homogeneous classifiers HOC-1 and HOC-N with the same number of training tweets. The number of training tweets for HOC-N‡ classifier equaled to the number of all the tweets training for each HOC-1 classifier. HOC-N‡ is introduced to compare classifiers in a scenario often encountered in practice. For instance, imagine there is a fixed budget (eg, monetary quantity) through which we can only label 2000 tweets. If we have four HOC-1 classifiers, then we can only allocate 500 tweets to each. However, we can allocate all 2000 tweets to the HOC-N classifier. Again, we ran a *t* test when the results failed to followed a normal distribution and a KS-test otherwise.

The hypothesis tests showed that only the HOC-1 and HOC-N‡ classifiers are statistically significant when testing on hypertension tweets (*P*=.015). This suggests that HOC-N classifiers are expected to have similar performance with HOC-1 classifiers when each classifier is trained with the same number of training tweets. However, if the total number of training tweets is fixed, the HOC-N classifier will outperform the combination of HOC-1 classifiers.

This indicates that the HOC-N classifier can serve as a substitute for HOC-1 classifiers.

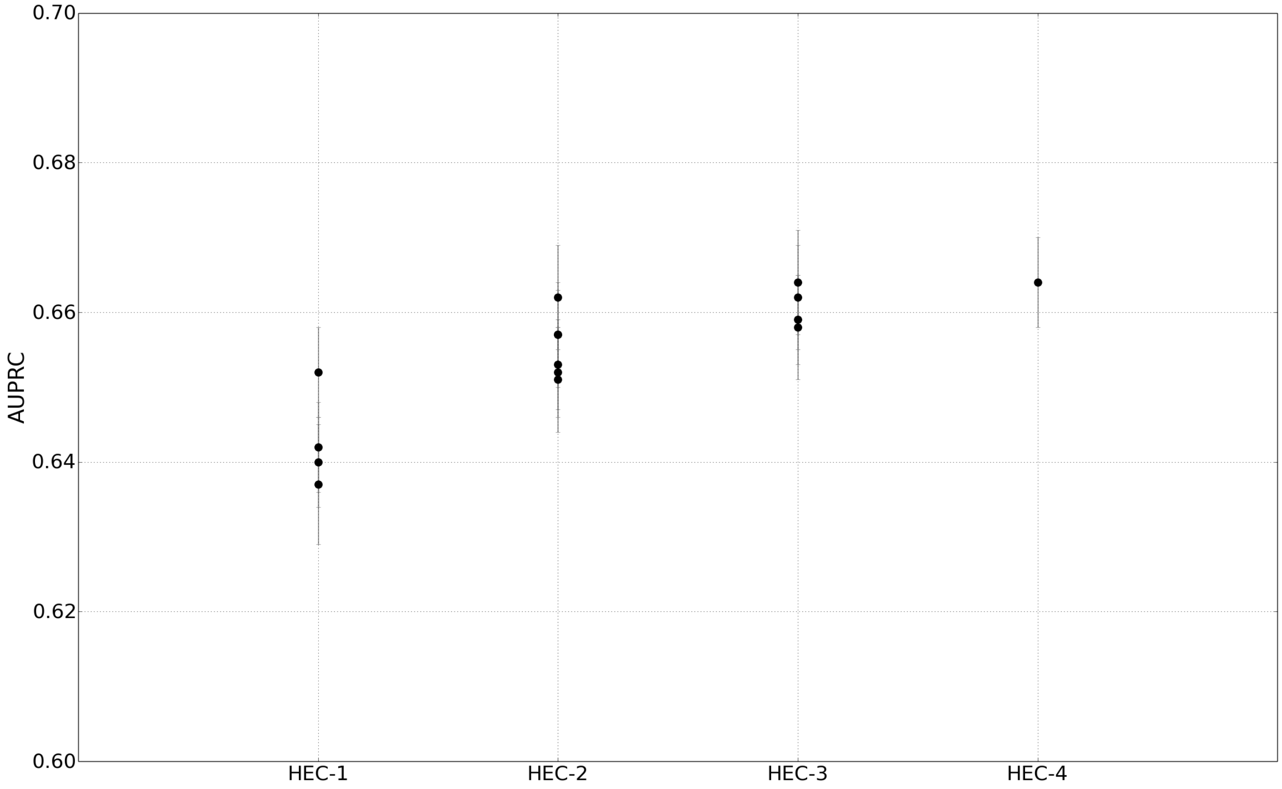
Comparison Between Heterogeneous Classifiers

In this experiment, we evaluated how the number of health issues in the training set influence the heterogeneous classifiers. Figure 5 shows the results of HEC-1 and HEC-N (N ∊ {2, 3, 4}) when testing on the other 30 health issues. For HEC-1, it should be noted that the cancer HEC-1 achieved the best AUPRC. This may stem from the fact that cancer can be invoked to communicate a wide variety of concepts beyond an individual’s health status, such as the Zodiac, the name of a physical building, or a metaphor. The results also indicate that HEC-N tends to outperform HEC-1.

This suggests hypothesis H4 may be true, provided the classifier is based on an appropriate mixture of health issues. However, determining an optimized group of health issues to achieve an HEC-N classifier with performance comparable to HEC-1 classifier is left to future investigation.

Based on these findings, we use HOC-N and HEC-N to conduct the system scalability test.

Figure 5. Comparison Between heterogeneous classifiers HEC-1 and HEC-N trained on cancer, depression, hypertension, and leukemia, and tested on the remaining 30 health issues. The tweets of each test health issue stratified with respect to their rate of observation.



System Scalability

After breaking ties, 43.7% of the TieBreak dataset are positive instances. Based on this proportion, there are approximately 120,260 positive instances out of 281,357 tweets in the health issue bins (or 0.046% of all the collected tweets). Table 6 reports the distribution of positive and negative tweets in each dataset.

We trained the SYND classifier with the gold standard datasets for cancer, depression, hypertension, and leukemia, and tested it on the other three types of datasets. Figure 6 depicts the PR curves for each dataset and shows the average and standard deviation of AUPRC. The upper line corresponds to testing on the CAP dataset (AUPRC 0.753, SD 0.005), the middle line corresponds to testing on the TieBreak dataset (AUPRC 0.685, SD 0.005) and the lower line corresponds to testing on the CAP dataset (AUPRC 0.594, SD 0.007). When fixing the recall to 0.4, it was observed that the CAP, TieBreak, and CAN scenarios yield a precision of 0.8, 0.77, and 0.61, respectively. These results demonstrate the scalability of the system classifiers to obtain a high precision with a reasonable recall when testing many other health issues in the Twitter environment.

Figure 7 shows how the size of the training set influences the AUPRC of the classifiers. For each training set, the mean AUPRC and a 95% confidence interval is illustrated in the gray area. For each dataset, the results suggest that AUPRC achieves stability when the training set consists of approximately 2000 tweets.

Table 6. Class distribution of tweets in the datasets.

Tweets	Gold	CAN[a]	CAP[b]	TieBreak
Positives	1082	1082	1718	1366
Negatives	1539	2175	1539	1891

[a] CAN: conflict as negative

[b] CAP: conflict as positive

Figure 6. PR (precision recall) curves for testing on the gold, CAN (conflict as negative), and CAP (conflict as positive) datasets.

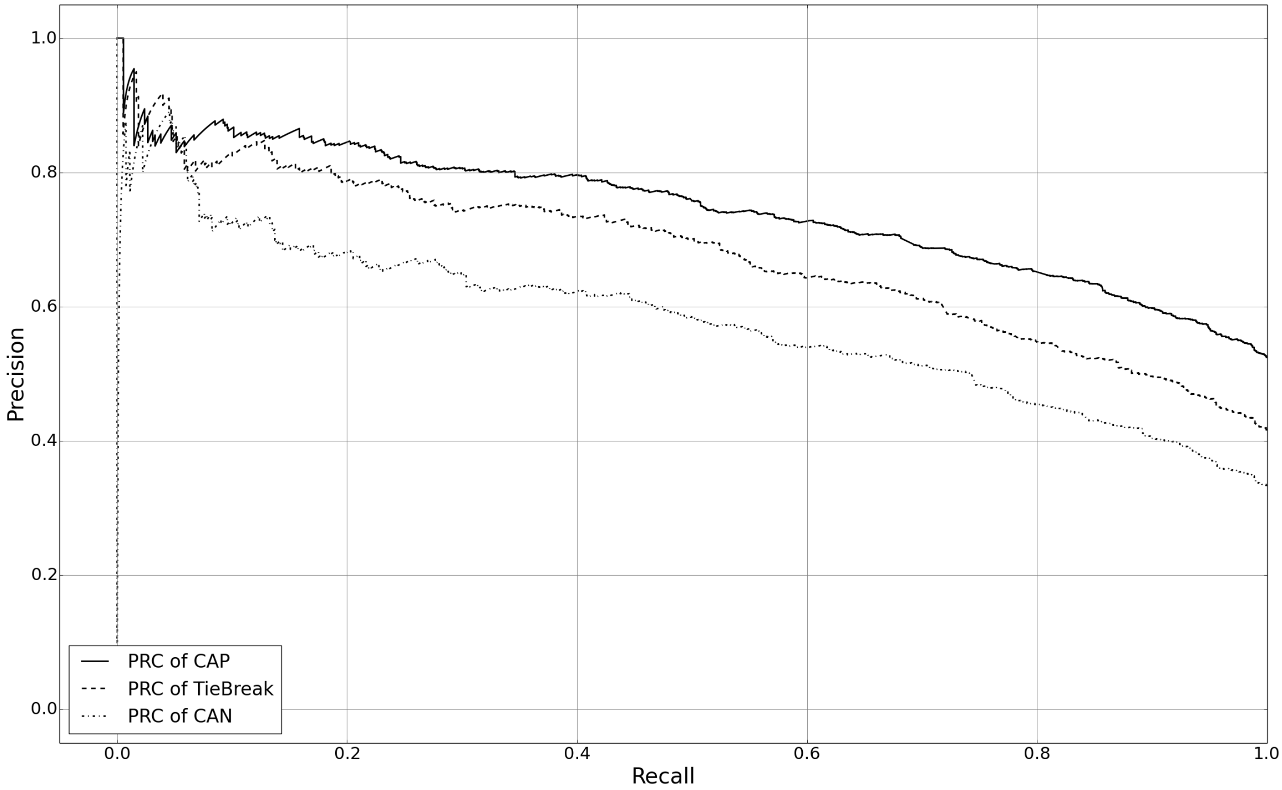
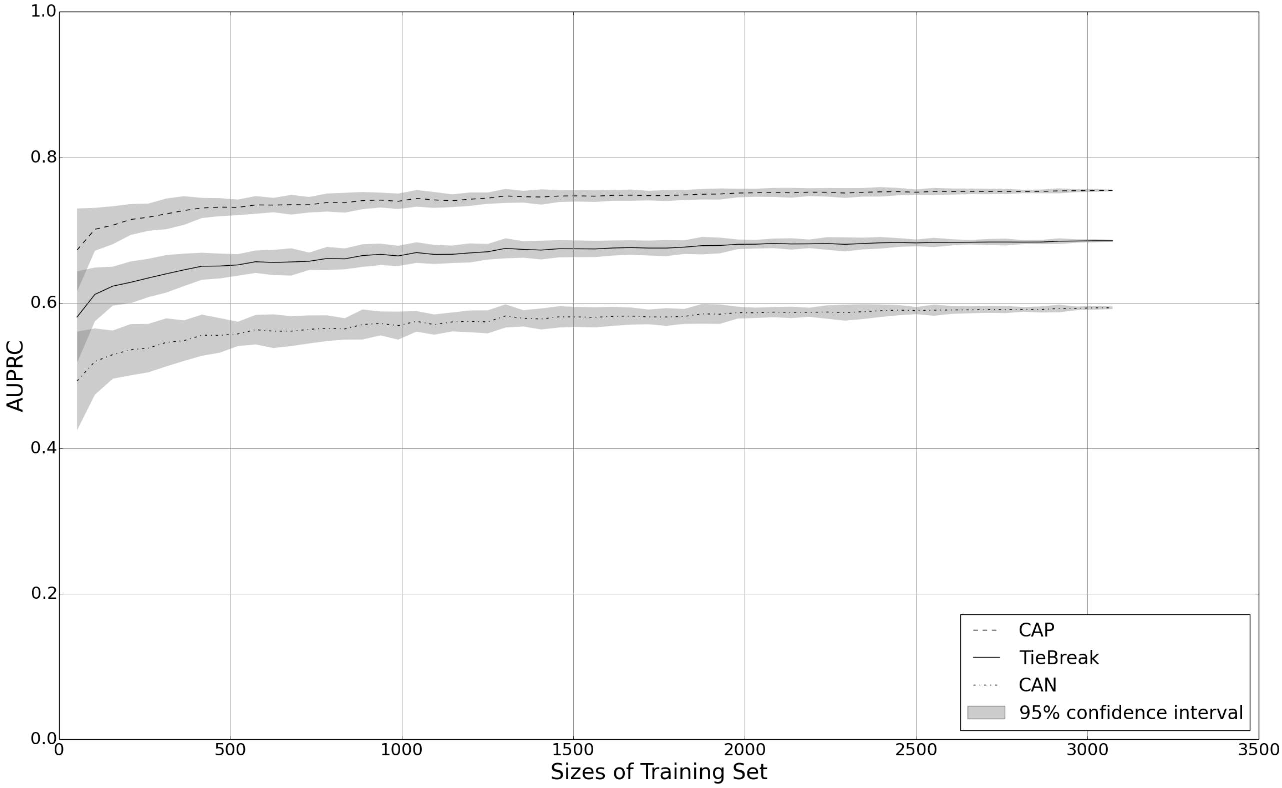


Figure 7. Performance of the SYND (synthetic health issue) classifier with a varying amount of training data.



Discussion

Principal Findings

There are several notable findings from this investigation. First, Twitter users disclose the health status of themselves and others. Second, the health status disclosure rate may depend on the health issue. Third, how people disclose their own and other people’s health status may also be health issue dependent. Fourth, tweets related with a small group of health issues can train a scalable classifier to detect health mentions on Twitter streams.

Another interesting phenomenon illustrated from the PR curves (Figure 6) is that the system classifier, trained with the tweets for which MT masters exhibited high concordance in their labels, is more likely than MT masters to classify tweets with conflict labels as positive. One possible explanation is that the classifier makes its decision based on thousands of examples, while most MT masters made decisions only with the description of the survey, which indicates that the classifier may be more familiar with the labeling task. This suggests there may be a difference between using an expert and crowdsourcing to generate the labeled corpus. However, determining how to best leverage the crowd to mimic an expert is beyond the scope of this investigation.

Impact on Health Related Research

According to our investigation, roughly 44% of the tweets containing health issue keywords disclose personal health status. We believe there is a potential for information to assist health care professionals in learning about their patients or their patients' family medical history, information often missing in the EMRs. This indicates that social media platforms, such as Twitter contains huge amount of personal health care related information that may complement traditional EMRs in research and practice. We recognized that we must still verify the veracity of such data, but an opportunity exists nonetheless.

Limitations

We wish to highlight several limitations of this investigation. First, two parameters to extract tweets from Twitter streams require configuration: (1) the set of keywords invoked in the filter, and (2) the geolocation applied to discover tweets. Compared to keywords, geolocation can filter tweets disseminated by authoritative organizations (due to the absence of "coordinates" and "place" information in these tweets), such as the American Cancer Society, and thus greatly reduce noise. However, it should be noted that invoking such a filter can also exclude the tweets of individuals who choose not to disclose their location. A second limitation exists in the survey provided to the MT masters for labeling the corpus. Specifically, we assumed the N/A option was a member of the negative class, but this could be an incorrect assumption in certain instances. Third, this investigation was restricted to only 34 health-related phenomena, which is clearly only a sample of all possible health issues. The keywords filter service can be enhanced by integrating a laymen health vocabulary [42]. Given that this study shows there is (1) high variability in the rate at which people tweet about a certain health issue, and (2) to whom the statement of health issue corresponds, it will be critical to investigate how these methods fare in the context of other health issues.

Conclusions

Recent studies demonstrate the information communicated through social media platforms, such as Twitter and Facebook, could supplement traditional medical and epidemiological research. In this paper, we showed that a health mention detection system can be designed and deployed for microblogging systems, such as Twitter. At the same time, we illustrated that the information communicated through such mentions can disclose the health status of the authors and other individuals at a wide range of rates. Our experimental investigation further showed that the combination of tweets from several health issues can yield a classifier that dominates a classifier based on the tweets of a single health issue. This may enable the system to use a small amount of training data to build a classifier that detects health status mentions across a range of health issues. We envision several opportunities for extending this work. First, we believe the scalability of the classifier may be improved by determining the minimal set of health issues and features (eg, more complicated grammar features). Second, we anticipate that the performance of the classifier could be improved be accounting for context, such as dialogue, relationships in the network, and profile information as new supplemental features. Finally, while the rate that health status is disclosed for the author versus other individuals is dependent upon the considered health issue, further investigation is required to determine what drives this disparity. We suspect, for instance, that it may be dependent on the sensitivity and severity of health issues, but this is only a conjecture.

Acknowledgments

This research was sponsored in part by grants from the National Science Foundation (CCF-0424422) and the Patient Centered Outcomes Research Institute (CDRN-1306-04869). The authors would like to thank the members of the Mid-South Clinical Data Research Network for useful discussions during the development of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example question posed to Mechanical Turk masters.

[PDF File (Adobe PDF File), 239KB - jmir_v17i6e138_app1.pdf]

Multimedia Appendix 2

Concordance between the system classifier and Mechanical Turk masters.

[PDF File (Adobe PDF File), 464KB - jmir_v17i6e138_app2.pdf]

Multimedia Appendix 3

Summary of four datasets.

[PDF File (Adobe PDF File), 221KB - jmir_v17i6e138_app3.pdf]

Multimedia Appendix 4

Keywords used to filter tweets.

[PDF File (Adobe PDF File), 216KB - jmir_v17i6e138_app4.pdf]

References

1. Garratt A, Ruta D, Abdalla M, Buckingham J, Russell I. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? BMJ 1993 May 29;306(6890):1440-1444 [FREE Full text] [Medline: 8518640]
2. Samsa GP, Matchar DB, Goldstein LB, Bonito AJ, Lux LJ, Witter DM, et al. Quality of anticoagulation management among patients with atrial fibrillation: results of a review of medical records from 2 communities. Arch Intern Med 2000 Apr 10;160(7):967-973. [Medline: 10761962]
3. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. Stroke 2000 Apr;31(4):858-862 [FREE Full text] [Medline: 10753988]
4. Quam L, Ellis LB, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research. The concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. Med Care 1993 Jun;31(6):498-507. [Medline: 8501997]
5. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
6. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. J Intern Med 2013 Dec;274(6):547-560. [doi: 10.1111/joim.12119] [Medline: 23952476]
7. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 Jun;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
8. Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. J Biomed Inform 2012 Aug;45(4):763-771 [FREE Full text] [doi: 10.1016/j.jbi.2012.01.009] [Medline: 22326800]
9. Estrin D. Small data, where n=me. Commun ACM 2014 Apr 01;57(4):32-34. [doi: 10.1145/2580944]
10. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
11. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai A. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
12. Riedl J, Riedl E. Crowdsourcing medical research. Computer 2013 Jan;46(1):89-92. [doi: 10.1109/MC.2013.15]
13. Wicks P, Vaughan T, Heywood J. Subjects no more: what happens when trial participants realize they hold the power? BMJ 2014;348:g368 [FREE Full text] [Medline: 24472779]
14. Bian J, Topaloglu U, Yu F. Towards large-scale Twitter mining for drug- related adverse events. In: Proceedings of the 2012 International Workshop on Smart Health and Wellbeing. 2012 Presented at: International Workshop on Smart Health and Wellbeing; 2012; Maui, Hawaii, USA p. 25-32.
15. Mukherjee S, Weikum G, Danescu-Niculescu-Mizil C. People on drugs: credibility of user statements in health communities. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014 Presented at: 20th ACM SIGKDD; 2014; New York, USA p. 65-74.
16. Bodnar T, Barclay VC, Ram N, Tucker CS, Salath´e M. On the ground validation of online diagnosis with Twitter and medical records. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion. 2014 Presented at: 23rd International Conference on World Wide Web; 2014; Seoul, Korea p. 651-656.
17. Medical Expenditure Panel Survey. URL: http://meps.ahrq.gov/mepsweb/ [accessed 2015-05-13] [WebCite Cache ID 6YUMTLNXr]
18. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
19. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
20. Olejnik L, Kutrowska A, Castelluccia C. I'M 2.8% Neanderthal - The beginning of genetic exhibitionism? In: Workshop on Genome Privacy. 2014 Presented at: Workshop on Genome Privacy; 2014; Amsterdam, Netherlands.

21. 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]
22. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
23. Wilson K, Keelan J. Social media and the empowering of opponents of medical technologies: the case of anti-vaccinationism. J Med Internet Res 2013;15(5):e103 [FREE Full text] [doi: 10.2196/jmir.2409] [Medline: 23715762]
24. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
25. Cobb NK, Jacobs MA, Saul J, Wileyto EP, Graham AL. Diffusion of an evidence-based smoking cessation intervention through Facebook: a randomised controlled trial study protocol. BMJ Open 2014;4(1):e004089 [FREE Full text] [doi: 10.1136/bmjopen-2013-004089] [Medline: 24448847]
26. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. AIDS Care 2012;24(5):593-600 [FREE Full text] [doi: 10.1080/09540121.2011.630355] [Medline: 22149081]
27. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; 2011; Edinburgh, United Kingdom p. 1568-1576.
28. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014;16(10):e236 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]
29. Nagel AC, Tsou MH, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. 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 [FREE Full text] [doi: 10.2196/jmir.2705] [Medline: 24158773]
30. Curtis BL. Social networking and online recruiting for HIV research: ethical challenges. J Empir Res Hum Res Ethics 2014 Feb;9(1):58-70 [FREE Full text] [doi: 10.1525/jer.2014.9.1.58] [Medline: 24572084]
31. Lazer DM, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
32. Mao H, Shuai X, Kapadia A. Loose tweets: an analysis of privacy leaks on Twitter. In: Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society. 2011 Presented at: ACM Workshop on Privacy in the Electronic Society; 2011; Chicago, Illinois p. 1-12.
33. Lamb A, Paul MJ, Dredze M. Separating fact from fear: tracking flu infections on Twitter. In: Proceedings of HLT-NAACL. 2013 Presented at: Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics; 2013; Atlanta, Georgia p. 789-795.
34. Gattani A, Doan A, Lamba D, Garera N, Tiwari M, Chai X, et al. Entity extraction, linking, classification, and tagging for social media. Proc VLDB Endow 2013 Aug 27;6(11):1126-1137. [doi: 10.14778/2536222.2536237]
35. Yang SH, Kolcz A, Schlaikjer A, Gupta P. Large-scale high-precision topic modeling on Twitter. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014 Presented at: ACM SIGKDD; 2014; New York, USA p. 1907-1916.
36. Banerjee N, Chakraborty D, Dasgupta K, Mittal S, Joshi A, Nagar S. User interests in social media sites: an exploration with micro- blogs. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. 2009 Presented at: ACM Conference on Information and Knowledge Management; 2009; Hong Kong, China p. 1823-1826.
37. Davidov D, Tsur O, Rappoport A. Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning. 2010 Presented at: Conference on Computational Natural Language Learning; 2010; Uppsala, Sweden p. 107-116.
38. Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M. Short text classification in Twitter to improve information filtering. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2010 Presented at: ACM SIGIR; 2010; Geneva, Switzerland p. 841-842.
39. Banerjee N, Chakraborty D, Joshi A, Mittal S, Rai A, Ravindran B. Towards analyzing micro-blogs for detection and classification of real-time intentions. In: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media. 2012 Presented at: International AAAI Conference on Weblogs and Social Media; 2012; Dublin, Ireland.
40. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning. 2006 Presented at: International Conference on Machine Learning; 2006; Pittsburgh, Pennsylvania p. 233-240.
41. Cer D, de Marneffe MC, Jurafsky D, Manning CD. Parsing to Stanford dependencies: trade-offs between speed and accuracy. In: Proceedings of the International Conference on Language Resources and Evaluation. 2010 Presented at: International Conference on Language Resources and Evaluation; 2010; Valletta, Malta.

42. Vydiswaran V, Mei Q, Hanauer DA, Zheng K. Mining consumer health vocabulary from community-generated text. In: Proceedings of the American Medical Informatics Association Annual Symposium (AMIA). 2014 Presented at: AMIA; 2014; Washington, DC.

Abbreviations

AUPRC: area under the precision recall curve
CAP: conflict as positive
CAN: conflict as negative
EMR: electronic medical record
HEC-1: heterogeneous classification with $|X| = 1$
HEC-N: heterogeneous classification with $|X| > 1$
HOC-1: homogeneous classification with $|X| = 1$
HOC-N: homogeneous classification with $|X| > 1$
KS: Kolmogorov-Smirnov Test
MNB: Multinomial Naïve Bayes
MT: Mechanical Turk
N/A: none of the above
SYND: synthetic health issue

Edited by G Eysenbach; submitted 31.01.15; peer-reviewed by J Bian, R Nagar; comments to author 19.02.15; revised version received 08.03.15; accepted 23.03.15; published 05.06.15.

Please cite as:
Yin Z, Fabbri D, Rosenbloom ST, Malin B
A Scalable Framework to Detect Personal Health Mentions on Twitter
J Med Internet Res 2015;17(6):e138
URL: http://www.jmir.org/2015/6/e138/
doi:10.2196/jmir.4305
PMID:26048075

Original Paper

Characterizing Sleep Issues Using Twitter

David J McIver[1*], PhD; Jared B Hawkins[1*], MMSc, PhD; Rumi Chunara[1,2], PhD; Arnaub K Chatterjee[3], MHA, MPA; Aman Bhandari[3], MPH, PhD; Timothy P Fitzgerald[4], PhD; Sachin H Jain[5], MD; John S Brownstein[1], PhD

[1]Boston Children's Hospital, Harvard Medical School, Boston, MA, United States
[2]New York University, New York, NY, United States
[3]Merck & Co, Inc, Boston, MA, United States
[4]Merck & Co, Inc, West Point, PA, United States
[5]CareMore Health System, Cerritos, CA, United States
*these authors contributed equally

Corresponding Author:
David J McIver, PhD
Boston Children's Hospital, Harvard Medical School
300 Longwood Ave.
Boston, MA, 02115
United States
Phone: 1 902 213 9005
Fax: 1 617 730 7547
Email: david.mciver@childrens.harvard.edu

Abstract

Background: Sleep issues such as insomnia affect over 50 million Americans and can lead to serious health problems, including depression and obesity, and can increase risk of injury. Social media platforms such as Twitter offer exciting potential for their use in studying and identifying both diseases and social phenomenon.

Objective: Our aim was to determine whether social media can be used as a method to conduct research focusing on sleep issues.

Methods: Twitter posts were collected and curated to determine whether a user exhibited signs of sleep issues based on the presence of several keywords in tweets such as insomnia, “can’t sleep”, Ambien, and others. Users whose tweets contain any of the keywords were designated as having self-identified sleep issues (sleep group). Users who did not have self-identified sleep issues (non-sleep group) were selected from tweets that did not contain pre-defined words or phrases used as a proxy for sleep issues.

Results: User data such as number of tweets, friends, followers, and location were collected, as well as the time and date of tweets. Additionally, the sentiment of each tweet and average sentiment of each user were determined to investigate differences between non-sleep and sleep groups. It was found that sleep group users were significantly less active on Twitter ($P=.04$), had fewer friends ($P<.001$), and fewer followers ($P<.001$) compared to others, after adjusting for the length of time each user's account has been active. Sleep group users were more active during typical sleeping hours than others, which may suggest they were having difficulty sleeping. Sleep group users also had significantly lower sentiment in their tweets ($P<.001$), indicating a possible relationship between sleep and pyschosocial issues.

Conclusions: We have demonstrated a novel method for studying sleep issues that allows for fast, cost-effective, and customizable data to be gathered.

(J Med Internet Res 2015;17(6):e140) doi:10.2196/jmir.4476

KEYWORDS

sleep issues; social media; insomnia; novel methods; sentiment; depression

Introduction

In 2006, between 50-70 million adults in the United States had perceived chronic sleep or wakefulness issues, which is an increasing trend, and more than 35% of adults report having insufficient sleep [1]. With the most common sleep issues reported by Americans as having <7 hours of sleep in a 24-hour period, restless leg syndrome, snoring, and insomnia, there are many areas where further exploration could be beneficial [1]. These forms of sleep-depriving conditions have been linked to decreased quality of life, excessive daytime sleepiness, depression, obesity, cardiovascular complications, diabetes, decreased productivity, increased chance of risky behaviors, increased risk of car accidents, and others [2-5]. Impaired sleeping can lead to serious impact on health; for example, the US Department of Transportation found that 2.2-2.6% of all fatal car crashes from 2005-2009 reportedly involved drowsy driving [6]. As well, depression has been an area of active research in attempting to determine its role in insomnia and sleep disorders, in either causal direction [7-11]. Due to the impact, both physical and psychosocial, of sleep-related issues on a large segment of the population, continued research in this area is needed.

For decades, interest in sleep issues has produced broad research and survey methods. In addition to studies and surveys being undertaken by private organizations such as the National Sleep Foundation, the Centers for Disease Control and Prevention (CDC), via the Behavioral Risk Factor Surveillance System (BRFSS), administers yearly questionnaires to the American, non-incarcerated population, regarding many types of health and risk factors. Beginning in 2009, the BRFSS has included a module dealing exclusively with sleep issues [1]. While the data gathered by the BRFSS have been instrumental in our understanding of sleep disorders, it does suffer from several limitations. The BRFSS is based on a random-digit-dialing system, and response rates can be low. Of all calls made, a response rate of between 40-67%, while respectable for epidemiological surveys, means much of the intended population is not being surveyed [1], though sample size and weighting calculations can correct for some of this bias. Of note, not all US states are included in the survey each year; therefore, the generalizability of the results to the entire US population is negatively impacted. Finally, because of the monumental amount of work involved in performing the surveys, gathering and combining data, analysis, and publication, the resulting BRFSS reports are expensive and are typically 7 months old by the time they are released. There have been many investigations of sleep disorders by independent researchers, but they too tend to suffer from some of these limitations, such as small sample size [12], high cost [13], long time frames [14], and lack of generalizability [15]. In light of these shortcomings, new supplemental methods of investigating the epidemiologic factors associated with sleep issues are needed to provide timely analyses that have greater external validity by incorporating a much larger sample size, and which are less costly, more quickly implemented and analyzed, and are malleable to allow for design restructuring based on new data.

We are interested in determining whether the way in which people with potential sleep issues interact with Twitter can be used as a method of identifying and characterizing those individuals. In recent years, there has been a great deal of interest in harnessing the massive amounts of data produced by social media websites, such as Facebook and Twitter, to try to glean insights into topics of interest to public health, and these platforms are increasingly being considered as valuable sources of patient information [16-19]. Recent examples include using social media to perform infectious and foodborne disease surveillance [20-22], chronic disease surveillance [23], prescription drug use [24], investigating hospital care quality [25], and many others [26,27]. With a greater focus on human behavior and characterization, researchers have used Twitter to investigate how people use social media in efforts of weight loss [28] and how suicide-related Twitter use compares to actual events [29]. Additionally, an increasing number of researchers have been experimenting with sentiment analysis on social media [30-35]. Sentiment can be determined in several ways, with the principle being to classify the underlying emotional information (within tweets, status updates, photos, etc) as either positive or negative; this can be done either purely by human input or by an algorithm trained to complete this process based on a human-classified set of objects. This process is useful for determining how people feel about products, events, other people, etc. Sentiment analysis has yet to be used on social media to help understand sleep disorders, but it does exhibit a diurnal characteristic [30] and offers interesting possibilities in investigating the links between sleep disorders and the overall sentiment or attitude of individuals displaying these characteristics. Demographics of Twitter users, while not entirely representative of the American population, have become more representative over time. Twitter is currently used by 23% of the adult Internet-using population and has seen increases in usage from hitherto underrepresented populations, such as men, whites, people aged 65 and older, and others. As of late 2014, 24% and 21%, respectively, of male and female adult Internet users used Twitter, and only 37% of that group were under 30 years of age [36].

We were interested in finding out if people who posted on Twitter about having sleep issues were more active on Twitter than people who did not, or if they had more friends or followers. As well, we wanted to know if people discussing sleep issues were posting more during traditional sleeping hours, suggesting that they may be having difficulty with sleeping. Furthermore, we were interested in the relationship between users who exhibited potential sleep issues and the sentiment of what it was they were tweeting, as a means of exploring the impact of sleep issues on emotions, feelings, and attitudes.

In this study, information posted on Twitter was used to identify people who may be exhibiting self-described signs or symptoms of sleep-related issues. By examining the content of tweets, users whose tweets contained specific sleep-related keywords were compared to a random population that did not contain these keywords. We then examined if there were observable differences between these groups in relation to their activity on Twitter.

Methods

Overview

Twitter is an online microblogging website where users "tweet", or post, statuses containing 140 characters or less. It boasts approximately 255 million monthly active users worldwide, including 33% residing in the United States [37]. Twitter allows conditional access to this wealth of information through their application programming interface (API), for data that users allow to be public. Using the Twitter API, one can collect tweets matching certain query criteria and access meta information including location (self-reported and geo-tagged), total number of tweets, number of "followers", number of friends, etc.

Twitter users who mentioned pre-defined keywords related to sleep or sleep issues in their tweets (sleep group) were compared to users whose tweets did not contain pre-defined keywords (non-sleep group). Sleep group tweets were identified on the basis of keywords being present in a curated tweet, on a prospective basis, starting on January 7, 2014, and ending on April 30, 2014, and were examined and curated on a "most recent tweet" basis. That is, during each curation session, tweets that were most recently posted to Twitter were analyzed first.

To build a corpus of both sleep group and non-sleep group users, code was written to access the Twitter API, which searched Twitter every 15 minutes for all new tweets containing any of the following keywords: "can't sleep", "insomnia", "melatonin", "Ambien", "Ambien-CR", "zolpidem", "Lunesta", "Intermezzo", "trazadone", "eszopiclone", "#teamnosleep", and "#cantsleep" (note that "#" is the symbol for a Twitter hashtag that denotes a user-identified topic within the tweet, and "teamnosleep" is a user-created hash tag often used by individuals who declare that they are unable to sleep). The list of Twitter search terms was identified through consultation with researchers with expertise in sleep-related fields of study and by experimentally querying the Twitter database to investigate which terms were most commonly used. By including keywords and hashtags that are related to specific medications (ie, zolpidem, Intermezzo, eszopiclone), we aimed to collect tweets that we were highly confident would be related to some type of sleep issue, even if the number collected was small. In contrast, by including keywords and hashtags that were broader (sleep, tired, insomnia, etc), we hoped to collect a large number of tweets, but not all of which would be strictly relevant. Since all tweets included in the study were manually curated, the low specificity of tweets collected under the more generic keywords was not an issue. This was not an exhaustive search across all possible search terms, but rather an exploratory approach to test the utility of this type of analysis.

To assess authenticity and ensure they met sleep group inclusion criteria, tweets that contained one or more of these keywords were manually curated, by a single individual (DM), looking for the following attributes. To be included as sleep group tweets, a tweet (and the Twitter account it is associated with) (1) must have been in the English language (as selected in user settings), (2) appeared to be from within the United States, (3) be owned by an "average" person (ie, not a company/corporation, celebrity, or spam account), and (4) was not a "re-tweet" (a re-post of a tweet originally posted by a different user). Re-tweets were removed because we were interested only in the experiences and expressed feelings of the individuals we were collecting information on, and not those of other people. Twitter accounts were qualitatively determined to be within the United States if the user-defined location was set to a US location or the account appeared to be located in the United States based on the nature of the user's profile information and previous posts. As well, tweets were examined to ensure that the keyword selected in the tweet was being used in the proper context. For example, a tweet that read "Just took my Ambien, hope I can sleep tonight" would be accepted as a sleep group user, but the tweet "A friend of mine just got prescribed Ambien" would not, because it did not pertain to the person who issued the tweet. Similarly, tweets that were ambiguous as to whether or not an action or outcome pertained to the individual who wrote the tweet were not treated as sleep group users. For example, the tweet "I took an Ambien, and now I'm sleepy" would be treated as a sleep group user, but the tweet "Ambien makes you sleepy" would not, because it did not indicate that this person took Ambien or was sleepy. They were simply making a statement.

A corpus of potential non-sleep group tweets was built by collecting tweets that did not contain any of the pre-defined keywords of interest. After initial manual curation to ensure tweets and users were in the English language, were from the United States, and were "normal" users, users were added to the non-sleep group if none of their tweets within the previous 10 days contained any of the pre-defined keywords of interest; text found in re-tweets was not considered. As an introductory and exploratory study, 10 days was chosen as a number of days that would allow for enough tweets to provide sufficient data for our purposes and was both computationally and financially achievable.

Tweets were automatically collected on an ongoing basis and selection of users into either the sleep group or non-sleep group was performed by the curator on a "most recent tweet" basis. That is, when the curator logged on to the curating tool, the most recent tweets to be collected were presented for curation. Therefore, if the curator were curating tweets at 9 am EST, the tweets they would be working on were the most recent tweets posted that matched the search criteria.

User Data

User-related data are data that are associated with a user's Twitter account as opposed to a particular tweet. For each user curated and included in the study, the metadata included in the analyses were total number of tweets, number of favorites (number of times that user favorited tweets from other users), total number of followers, total number of friends, user-submitted location, date of account creation, time zone of user, average number of tweets per day since account creation (calculated as total number of tweets divided by number of days that account has been active). For several of these collected variables, the count of the variable was also averaged over the lifetime of the user's account. This was done by dividing the variable count by the number of days the user had been active, which is equal to the number of days between account creation and the day the identified tweet was written. By creating data for the average number of counts/actions per day, the fact that some users have a higher number of friends, followers, or status updates, simply because they have had a Twitter account longer than some other users was accounted for. We also calculated the ratio of Twitter followers to friends for each user to create a way of measuring influence or impact on Twitter; a high follower:friend ratio indicates that a user has many people who follow their account but that they themselves follow relatively few people. This is often an indicator of high-impact Twitter users [38] and was included to ensure that both sleep- and non-sleep groups were equal in this respect.

To ensure that user data were collected at the same time for all study users, user metadata was collected after all tweets had been identified, rather than at the time of tweet approval. This was done primarily due to the increased time it took to identify sleep group tweets as compared to non-sleep group tweets. As a result, user metadata and tweet data presented in this study represent the state of a user's account as of May 1, 2014.

Tweet/Timeline Data

Tweet data are the data associated with a single tweet as opposed to the data associated with the user who issued the tweet. For each tweet that was included in the study, the analyzed tweet metadata included 140-character (maximum) tweet text, date and time of tweet creation (in Universal Time Code, UTC), and geo-tagged location of tweet (when available).

Similar to parsing non-sleep group users' previous 10 days of activity to search for keywords, additional information was gathered on all users to investigate the overall trend of non-sleep group users' behavior versus sleep group users' tweeting behavior. From the original tweet that was manually curated to classify a user, a minimum of 10 days' worth of previous tweets were collected from a user's timeline. The process proceeded such that the Twitter API was queried to return 200 tweets for a given user. If the returned 200 tweets represented less than 10 days' worth of tweets, the process was repeated until 10 days of tweets were collected, or until the Twitter API indicated that the user had no more data to retrieve.

For all study users, the number of tweets that were published during certain times of day (coded as 1: midnight-5:59 am, 2: 6 am-11:59 am, 3: 12 pm-5:59 pm, 4: 6 pm-11:59 am) and on which day of the week they were created was determined. All tweet times used in this analysis were converted from UTC to the user's local time (based on the user's time zone). While it is possible that a user has an incorrect time zone set, this is highly unlikely as it is based on the time zone of their computer or smart device.

Sentiment Analysis

To determine the difference in sentiment of tweets published by sleep group users and non-sleep group users, Amazon's Mechanical Turk (AMT) platform was used. Amazon's Mechanical Turk is an online tool that allows large, tedious jobs to be completed very quickly by harnessing the efforts of numerous personnel hired by Amazon [39]. For this study, we had AMT workers perform a sentiment analysis on select tweets. This is a popular AMT feature in which text (in this case tweets) is rated as having either a strongly positive, positive, neutral, negative, or strongly negative sentiment (recorded as 2, 1, 0, -1, -2, respectively). Ratings are of course based on each AMT worker's own subjective opinion. For each Twitter user included in the study, 20 of their tweets (the original, curated tweet plus the user's previous 19 tweets) were rated by AMT workers, in a randomized, de-identified, non-categorized format. Two AMT employees, who were classified by Amazon as being highly experienced in the field of sentiment analysis (Master Workers) [40], rated each tweet. The result was an average sentiment score for each tweet, across both sleep and non-sleep groups. Because only two users rated each tweet, the final average sentiment results were grouped into the following categories: Positive=0.5, 1.0, 1.5, 2.0; Neutral=0; Negative=-0.5, -1.0, -1.5, -2.0. By comparing the proportion of sleep group and non-sleep group tweets that were identified as positive, negative, or neutral, sentiment differences were assessed. While there are numerous software options for determining the sentiment of any string of text (such as tweets), we opted to use AMT as it involves human graders, which is the gold standard on which many automated methods are based [41,42]. Humans are better able to catch uses of language, such as irony or sarcasm, that are difficult for computers to identify. In addition, while machines may be better at identifying individual words attributed to positive and negative sentences, determining the sentiment of a complex sentence and taking word context into consideration is still quite difficult for a machine [43].

To ensure that AMT workers were rating tweet sentiment reliably, we calculated agreement and Cohen's kappa values between sets of workers. Because AMT can use hundreds of individual workers for a project, we focused our efforts on the AMT workers who were most prolific in rating tweet sentiment to capture at least 20% of rating jobs.

Statistical Methods

To investigate differences between sleep group users and non-sleep group users for variables with highly skewed distributions, permutation analyses with 10,000 iterations with re-sampling was used to investigate differences in median values. Variables based on proportions, such as the proportion of a user's tweets published on a certain day of the week, were compared between groups by performing two-tailed, two-group proportion tests, with statistical significance considered to be a *P* value of ≤.05. All analyses were performed in Stata 13.

Code and Database Structure

Custom code was written in PHP (hypertext preprocessor) to access the Twitter REST API (v1.1), which utilizes the open-source OAuth library tmhOAuth. Tweets are accessed via the Twitter API as "status objects", which are structured, JSON-formatted objects that contain all of the metadata about both the individual tweet and the user. Tweets were searched on the presence or absence of keywords using the GET search/tweets request. User timelines were collected using the GET statuses/user_timeline request. Returned tweets were stored in an Amazon Web Service (AWS) Relational Database Service (RDS) MySQL database as complete status objects in JSON format. Additionally, some tweet and user fields were stored in separate MySQL tables for faster access. Subsequent analysis and data cleaning were done using custom scripts written in PHP and Python.

Results

As of May 1, 2014, the total number of sleep group tweets that were collected over 115 days and stored in the database was 2,820,427. The number of tweets collected for each keyword are reported in Table 1. Due to the large number of tweets collected, only a small percentage could be analyzed. Of all collected tweets, 1000 of both sleep group and non-sleep group users (N=2000) were manually curated and approved for inclusion in the study. At the time of user account metadata collection on May 1, 2014, there were some accounts that had become inaccessible (eg, switched to a private setting, deleted, or banned from Twitter). After accounting for these changes, our final dataset included 896 sleep group users and 934 non-sleep group users. Summary statistics of the collected user metadata and tweet data, categorized by user group, are presented in Table 2 and Table 3, respectively.

Table 1. Number of tweets collected by various insomnia or sleep related keywords.[a]

Keyword	n	Proportion, %
#TeamNoSleep	119,378	4.23
Ambien	54,420	1.93
Can't Sleep	1,533,704	54.38
Eszopiclone	151	0.01
Insomnia	994,049	35.24
Intermezzo	10,145	0.36
Lunesta	3,734	0.13
Melatonin	103,674	3.68
Trazadone	1,149	0.04
Zaleplon	23	0.00
Total	2,820,427	100.00

[a] Number of tweets collected per keyword in this list represent different forms and combinations of each keyword (ie, Can’t Sleep includes “Can’t Sleep” as well as “#cantsleep”) as well as re-tweeted tweets. Some tweets may contain more than one keyword.

Table 2. Twitter user data.

Variable	Total Mean	Median	Per day[a] Mean	Median
Days active, n				
Non-sleep group	817	777		
Sleep group	1054	993		
P value		<.001		
Favorites, n				
Non-sleep group	1909	684	4.8	1.1
Sleep group	3257	1069	6.2	1.3
P value		<.001		.11
Followers, n				
Non-sleep group	817	319	5.5	0.5
Sleep group	792	295	1.2	0.3
P value		.08		<.001
Friends, n				
Non-sleep group	689	318	6.4	0.5
Sleep group	518	295	1.3	0.3
P value		.13		<.001
Follower:Friend ratio				
Non-sleep group	1.44	1.01		
Sleep group	1.45	0.99		
P value		0.901		
Statuses, n				
Non-sleep group	12609	5853	22	10
Sleep group	15253	7622	18	8
P value		<.001		.04

[a] Per day data refers to the total count of the variable divided by the total number of days a user's account has been active.

Table 3. Proportion of tweets posted at time of day by group.

	Proportion of tweets (%) by time 0:00-5:59	6:00-11:59	12:00-17:59	18:00-23:59
Non-sleep group	12.1	22.5	28.7	36.7
Sleep group	16.8	16.3	28.6	38.1
P value	<.001	<.001	.72	<.001

Sleep group users had Twitter accounts that were significantly older than other users (*P*<.001). The number of tweets overall were higher for users in the sleep group than for non-sleep group users (*P*<.001), but when calculated as the number of tweets per day since account creation, sleep group users had significantly fewer tweets (*P*=.04). The total number of tweets a user has favorited (other user's tweets) was significantly higher for sleep group users (*P*<.001), but this association was non-significant when considering the number of favorited tweets per day since the account was created. Sleep group users had both significantly fewer followers per day as well as friends per day (*P*<.001 for both).

For tweet-level data, the day-of-week and time-of-day data analyses were performed on a subset of data for which user-submitted time zone data were available. For all compiled timeline tweets (n=418,773), 73.5% had user-submitted time zones for which time zone specific date and time tweet data could be calculated. There was a significant difference between sleep group and non-sleep group users in the proportion that did or did not have user-submitted time zone information; 76.8%

of sleep group users disclosed their time zone while only 64.0% of others provided this data (*P*<.001).

A larger proportion of tweets between 12 am-5:59 am were from sleep group users (*P*<.001), as well as between 6 pm-11:59 pm (*P*<.001 for both). Conversely, more tweets from between 6 am-11:59 am were from non-sleep group users (*P*<.001). An hourly proportion of statuses posted by both groups is presented in Figure 1. In addition, a larger proportion of tweets that were submitted on Saturday, Sunday, Monday, and Tuesday, were from sleep group users (*P*<.001), whereas tweets on Wednesday, Thursday, and Friday, were more often from non-sleep group users (*P*<.001) (Figure 2).

Statistical analysis of sentiment scores calculated by AMT revealed that sleep group users (ie, those that were identified as expressing symptoms consistent with sleep issues) had significantly more negative tweet sentiment than non-sleep group users (*P*<.001), and conversely, that non-sleep group users had significantly more positive sentiment in their tweets (*P*<.001). There was no difference between groups in the neutral sentiment category (*P*=.45). To control for inherent variations in sentiment that might exist between individuals who tweet relatively infrequently and those who tweet more, we also categorized individual users into either low- or high-volume tweeters, which was determined by dividing the group in two at the median number of tweets for the entire sample population. When categorized, significant differences were still found between groups, with non-sleep group users having significantly more positive sentiment in both low- and high-volume groups (*P*=.002 and *P*=.03, respectively) and sleep group users showing significantly more negative sentiment in both groups (*P*=.003 and *P*=.03, respectively). Similar results were found when groups were dichotomized by the number of friends and number of followers for each user.

Sentiment was calculated by averaging ratings from two separate workers. While this approach has been used widely in the literature for AMT sentiment analysis, we also sought to determine agreement between workers. Agreement percentage and Cohen’s kappa values were calculated for the top 10 most prolific workers (out of 144 workers in total), who rated a total of 13,170 tweets, which accounted for over 36 of all jobs. Taking into account the percentage of agreement based on random chance, AMT worker agreement was 65 compared to the expected agreement of 40, with a kappa value of .420 (*P*<.001), representing moderate agreement [44]. These values are expected to be lower than the actual level of agreement, owing to the fact that we were not able to investigate the actual agreement between all workers given the sheer volume of AMT workers and because the most prolific workers are not necessarily the most “accurate” workers.

Figure 1. Proportion of statuses posted each hour by user group.

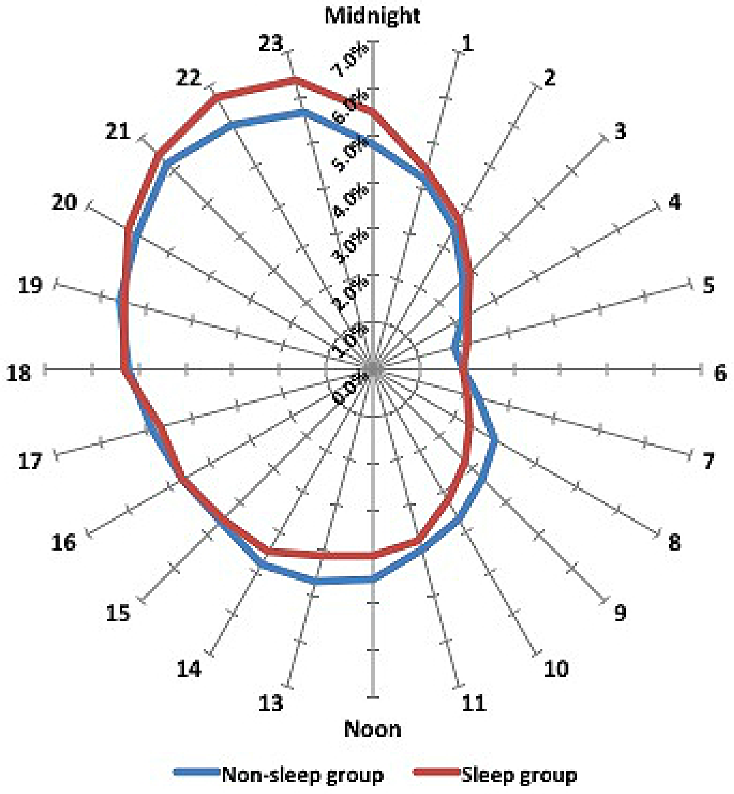
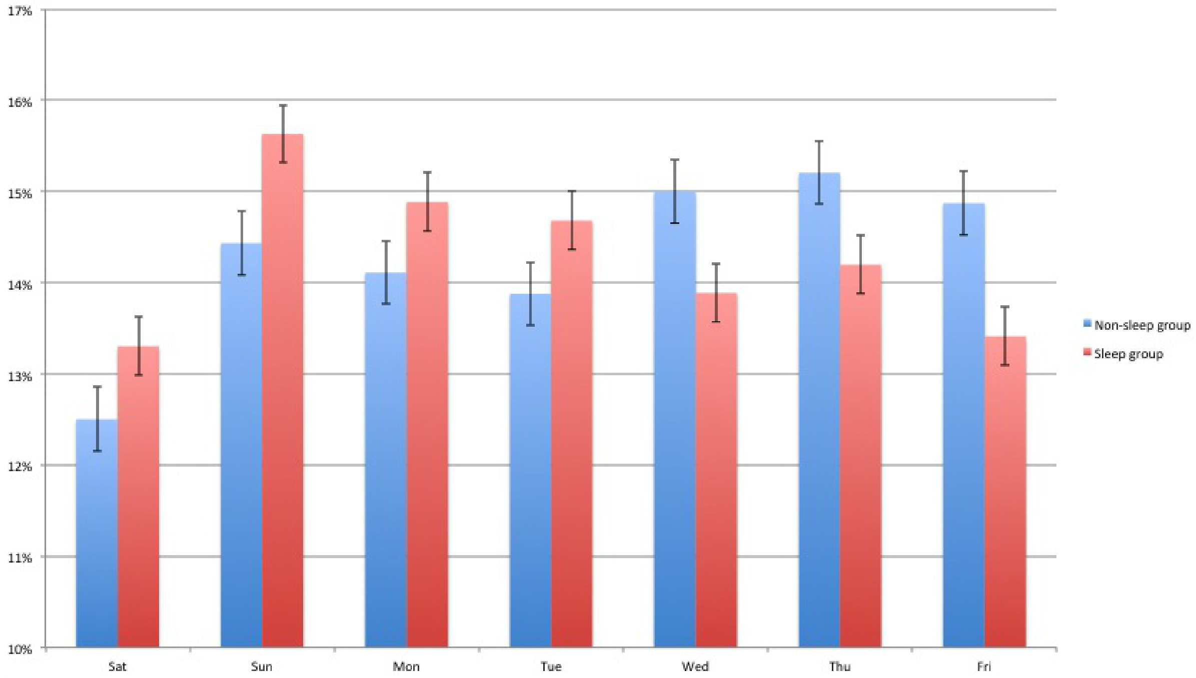


Figure 2. Proportion of statuses posted each day by user group. Y-axis begins at 10% to more clearly demonstrate differences between groups. All differences between groups were statistically significant (P<.001).



Discussion

Principal Findings

This study demonstrates introductory evidence that individuals exhibiting signs of sleep issues on Twitter were significantly less active on the social media platform than other users, but they tweet more during traditional sleeping hours and exhibited more negative sentiment in the tweets they shared.

In spite of conjecture found in the popular media [45,46], having some type of sleep issue, as it was defined in this study, did not equate to increased activity on social media. This conclusion is supported by the findings that sleep group users in our study had lower median values for number of followers, number of friends (users followed), and average number of tweets per day, which all indicate sleep group users appear to be less active on this particular social network. It is worth noting, however, that we were not able to determine how active users are on Twitter in terms of “observing” the social network, that is, reading tweets and passively tracking other users, while not actually posting tweets of their own or officially following other users. This has the potential to skew the results, as a user may appear to be relatively inactive by our definitions but could potentially be more active in ways that are not recorded by Twitter. Interestingly, while sleep group users were less active than non-sleep group users based on our definitions, it was observed that they had accounts that were significantly older (based on the date of account creation to date of tweet identification), suggesting that users with sleep issues may be more likely than those with normal sleep patterns to start using a new social media tool, even if they are less active on it. Although the phenomenon has not been scientifically investigated, there exists the possibility that users may tend to become less active on a social media account the longer they hold the account, which could explain the older account life of sleep group users, while their activity tends to be lower than others. While there are additional studies in progress that aim to elucidate these associations more clearly, these are interesting findings that may potentially warrant a reversal of how social media usage and the demographics of its users are perceived.

The finding that sleep group users posted a significantly higher proportion of their tweets during midnight and 6 am suggests that our method of sleep group determination is effective, since this is a time when most people with normal sleeping patterns would be asleep (after adjusting for time zone). It is impossible to say, without detailed investigations of all tweets from selected users, whether or not users tweeting between midnight and 6 am may actually have some reason (perhaps the user works a night shift or has some other reason for being awake during this time), but the significance of the difference between the two groups suggests the method employed to distinguish between sleep group and non-sleep group users is effective.

This study also provides introductory evidence for the argument that people suffering from insomnia and similar sleep disorders may be at increased risk of psychosocial issues. Of note, previous studies in the field of psychology and data mining have been successful in quantitatively linking online social media use, negative sentiment, and depression using automated tools [47-50]. In particular, an earlier study also found a significant relationship between Twitter users who tweeted about insomnia and a negative sentiment of those users [51]. We found that, based on AMT sentiment analysis results, Twitter users identified as potentially experiencing sleep issues had significantly lower sentiment portrayed in their tweets, suggesting this group may be experiencing some type of psychosocial disorder. Interestingly, this finding is backed up by findings that sleep group users had fewer friends, fewer followers, and fewer interactions than other users, indicating some level of decreased online social interaction for this group.

While the association is an interesting finding, it is preliminary and not conclusive, leaving much to be answered. However, these results provide an excellent starting point for a deeper investigation into the link between sleep issues, psychosocial issues, and social media usage, and warrant further investigation by more focused studies. A logical progression to further investigate these results would be to assess if the social dynamics of individuals are similar in their "real life" as in their online, social media life, or if the two areas differ significantly.

Given the nature of this study, it is worth briefly discussing the ethical, legal, and social implications of using Twitter data to conduct research on sleep disorders with potential links to psychosocial issues. Unlike other social network sites that restrict view permission of posts to approved friends, Twitter is a microblog with the sole purpose of allowing anyone to view content without prior approval. The privacy policy used by Twitter indicates that users consent to the collection, transfer, manipulation, storage, and disclosure of data that are public, while each user has the ability to change the privacy setting for their account. This study analyzed only tweets that were completely public (ie, no privacy settings were selected by the user). Thus, there was no expectation of privacy by the user. Public Twitter data are considered consistent with other existing public data sources, and as data are only passively analyzed in aggregate, this type of research is generally not considered to fall under the protections of human research. However, active data collection (eg, interaction directly with users) raises legitimate ethical, social, and legal concerns and should be conducted with appropriate caution and Institutional Review Board oversight.

Limitations

While the results presented above suggest that further research into this field is warranted, they also must be considered and interpreted in light of several potential limitations. Most importantly, due to the cross-sectional nature of the study, it is not possible to determine causality in the significant relationships found between social media usage, sleep issues, and psychosocial findings. Additionally, there were several methods pertaining to the curation of non-sleep and sleep group tweets that may merit revision when pursuing future projects. During the curation process, the curator was presented with tweets to determine whether or not the tweet was related to any sort of sleep-related disorder or not. Candidate tweets were presented to the curator in order of tweet creation, with the most recently posted tweets appearing for curation before others. This method is not optimal as the tweets a curator is reading are dependent on the time of day that curation is taking place. That is, the list of tweets the curator was working on may have been different from those that would be seen if the curator was working at 9 am compared to 9 pm To avoid this potential bias moving forward, future analyses will involve curation from a random selection of tweets stored in the database (thus randomizing the time and day of each tweet). While the method used here may potentially bias the users selected, they should not affect the analyses performed on the tweet-level data, since those analyses take all a user's tweets into consideration, and therefore the timing of the tweet identified as belonging to the sleep or non-sleep group is irrelevant.

Additional information would be useful for controlling for inherent differences in Twitter users. For instance, Twitter usage profiles may differ between users of different age, gender, or ethnicity. In future studies, it would be advantageous to collect this information in an attempt to control for these factors. This could be done either via algorithms designed to estimate these variables, by administering surveys to participants in a more interactive study, or by following a large subset of users before/after they suffer from self-described sleep issues. As well, user time zone information, which was used in conjunction with the time of tweet (recorded in UTC) to calculate the time of day a tweet was created, is a user-submitted variable and is therefore subject to potential data inaccuracies. While there are no studies that investigate the proportion of location fields that are accurately identified, we suspect it is highly likely that a user will appropriately choose their time zone (which is voluntary). However, there is the possibility that a user might indicate an incorrect time zone.

Non-sleep group users were defined by the absence of pre-defined keywords in a user's previous 10 days of tweets. As described above, this length of time was chosen to be computationally and financially achievable, while still achieving the desired amount of data. In future studies, we intend to increase the length of time a user's tweets must be free of these pre-defined keywords in order to be included in the non-sleep group. Depending on the quantity and quality of data available and the type of hypotheses involved, this may entail investigating months, years, or even a user's entire timeline of tweets, in order for group status to be designated. This will also allow us to analyze and control for specific time periods in a user's account history (eg, such as the first few months after account creation). Additionally, we may want to further characterize the sleep group population to determine if users who post "can't sleep" are different than those who post about "melatonin" or medications, for example. This finer-grain characterization may result in multiple sleep groups that should be analyzed independently.

While the information gathered in this study is interesting, and caution was taken to ensure its validity, this type of data is observational and as such no cause-and-effect relationships can be assumed. We have found significant differences between a non-sleep group and individuals who we have been categorized as having some type of sleep issue; however, we cannot be sure that those individuals who fit our definitions do in fact have a sleep issue. This is an important factor that we hope to address in further studies, potentially by directly interacting with users to help confirm our categorization methods. However, this approach raises ethical, social, and legal concerns (as mentioned above) and would need to be carefully implemented.

We also recognize that there may be inherent differences between users that can be reflected in their number of friends, followers, status update frequency, location, and other metrics that we have not accounted for. In future work, we aim to control for this by either following a large number of users for a long period of time (before and after self-described sleep issues) or by using a matching technique to more reliably compare groups.

Despite its limitations, this study and others focusing on using social media applications for addressing issues of public health concern demonstrate that this type of research can add meaningful interpretations to traditional methods. It is worth noting that while we see great promise for these new methods, they are envisioned and designed to be used alongside more traditional, highly validated methods such as the BRFSS. Both traditional and emerging ways of collecting and analyzing public health information and relationships have their strengths and shortcomings. We hope that by marrying the two types of research we can gain a more complete and accurate view of the state of health in the population.

Conclusions

This is one of the first research studies to actively investigate the relationship between social media use and sleep issues. It was found that people with apparent sleep issues were, on average, less active on Twitter and tended to be most active on the weekend and early weekdays, compared to users who did not have self-described sleep issues (based on our criteria). Additionally, we found that users with sleep issues have significantly more negative sentiment in the tweets they are posting compared to others, which may indicate a tendency for individuals identified as having a sleep issue via social media to be at a greater risk of psychosocial issues. While our findings are preliminary, they warrant further investigation and begin to provide evidence to contradict the popular belief that social media causes insomnia and other common sleep disorders. Furthermore, our current findings offer promise for expansion into the use of social media on the investigation of other health outcomes associated with sleep-related issues.

Acknowledgments

This study was funded in part by Merck & Co, Inc and in part by the NLM (T15LM007092). The authors would like to thank Dr James Fowler, University of California at San Diego, for his helpful conversations and suggestions.

Conflicts of Interest

AC, AM, TF, and SJ are employees of Merck & Co, Inc.

References

1. Morbidity and Mortality Weekly Report.: Centers for Disease Control and Prevention; 2011 Mar 04. Unhealthy Sleep-Related Behaviors - 12 States, 2009 URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6008a2.htm [accessed 2015-06-01] [WebCite Cache ID 6Yy7RBgIA]
2. Liu Y, Croft JB, Wheaton AG, Perry GS, Chapman DP, Strine TW, et al. Association between perceived insufficient sleep, frequent mental distress, obesity and chronic diseases among US adults, 2009 behavioral risk factor surveillance system. BMC Public Health 2013;13:84 [FREE Full text] [doi: 10.1186/1471-2458-13-84] [Medline: 23360346]
3. Chapman DP, Presley-Cantrell LR, Liu Y, Perry GS, Wheaton AG, Croft JB. Frequent insufficient sleep and anxiety and depressive disorders among U.S. community dwellers in 20 states, 2010. Psychiatr Serv 2013 Apr 1;64(4):385-387. [doi: 10.1176/appi.ps.201200226] [Medline: 23543168]
4. Centers for Disease Control and Prevention (CDC). Drowsy driving - 19 states and the District of Columbia, 2009-2010. MMWR Morb Mortal Wkly Rep 2013 Jan 4;61(51-52):1033-1037 [FREE Full text] [Medline: 23282860]
5. Morbidity and Mortality Weekly Report. United States: Centers for Disease Control and Prevention; 2012 Apr 27. Short Sleep Duration Among Workers — United States, 2010 URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6116a2.htm [accessed 2015-06-01] [WebCite Cache ID 6Yy7Yyets]
6. US Department of Transportation. Washington, DC: National Highway Traffic Safety Administration; 2011 Mar. Traffic Safety Facts - Drowsy Driving URL: http://www-nrd.nhtsa.dot.gov/pubs/811449.pdf [accessed 2015-06-01] [WebCite Cache ID 6Yy7jz9Ju]
7. Benca RM, Peterson MJ. Insomnia and depression. Sleep Med 2008;9:S3-S9.
8. Taylor DJ. Insomnia and depression. Sleep 2008 Apr;31(4):447-448 [FREE Full text] [Medline: 18457230]
9. Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression? J Affect Disord 2003;76(1–3):255-259.
10. Fava M. Daytime Sleepiness and Insomnia as Correlates of Depression. J Clin Psychiatry 2004;65(Suppl16):27-32.
11. Buysse DJ, Angst J, Gamma A, Ajdacic V, Eich D, Rössler W. Prevalence, course, and comorbidity of insomnia and depression in young adults. Sleep 2008 Apr;31(4):473-480 [FREE Full text] [Medline: 18457234]
12. Glidewell RN, Renn BN, Roby E, Orr WC. Predictors and patterns of insomnia symptoms in OSA before and after PAP therapy. Sleep Med 2014 Aug;15(8):899-905. [doi: 10.1016/j.sleep.2014.05.001] [Medline: 25011662]
13. Centers for Disease Control and Prevention. CDC - FMO - Budget Information URL: http://www.cdc.gov/fmo/topic/Budget [accessed 2015-05-31] [WebCite Cache ID 6Yy82TMPe]
14. Sivertsen B, Pallesen S, Glozier N, Bjorvatn B, Salo P, Tell GS, et al. Midlife insomnia and subsequent mortality: the Hordaland health study. BMC Public Health 2014;14:720 [FREE Full text] [doi: 10.1186/1471-2458-14-720] [Medline: 25024049]
15. Vozoris NT. Insomnia symptom frequency and hypertension risk: a population-based study. J Clin Psychiatry 2014 Jun;75(6):616-623. [doi: 10.4088/JCP.13m08818] [Medline: 25004185]

16. Staccini P, Douali N. Social media and patient health outcomes. Findings from the yearbook 2014 section on consumer health informatics. Yearb Med Inform 2014;9(1):195-198 [FREE Full text] [doi: 10.15265/IY-2014-0038] [Medline: 25123742]
17. Social media “likes” healthcare: From marketing to social business.: PricewaterhouseCoopers LLP PwC URL: http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml [accessed 2015-05-31] [WebCite Cache ID 6Yy8EjhvZ]
18. 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 23;15(4):e85 [FREE Full text]
19. 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]
20. Aramaki E, Maskawa S, Morita M. Twitter Catches the Flu: Detecting Influenza Epidemics Using Twitter. Proc Conf Empir Methods Nat Lang Process Internet Stroudsburg, PA. USA: Association for Computational Linguistics; 2011. URL: http://dl.acm.org/citation.cfm?id=2145432.2145600 [accessed 2015-06-01] [WebCite Cache ID 6Yy8V94Vd]
21. 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]
22. Nsoesie EO, Buckeridge DL, Brownstein JS. Guess who's not coming to dinner? Evaluating online restaurant reservations for disease surveillance. J Med Internet Res 2014;16(1):e22 [FREE Full text] [doi: 10.2196/jmir.2998] [Medline: 24451921]
23. Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. PLoS One 2013;8(4):e61373 [FREE Full text] [doi: 10.1371/journal.pone.0061373] [Medline: 23637820]
24. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
25. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf 2014 Oct;23(10):838-846 [FREE Full text] [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]
26. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
27. Paul MJ, Dredze M. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. You Are What You Tweet: Analyzing Twitter for Public Health URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880 [accessed 2015-06-01] [WebCite Cache ID 6Yy8bGilz]
28. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM. Tweeting it off: characteristics of adults who tweet about a weight loss attempt. J Am Med Inform Assoc Internet 2014 Jun 13:1032-1037 [FREE Full text]
29. Sueki H. The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. J Affect Disord 2015 Jan 1;170:155-160. [doi: 10.1016/j.jad.2014.08.047] [Medline: 25240843]
30. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]
31. Coviello L, Sohn Y, Kramer ADI, Marlow C, Franceschetti M, Christakis NA, et al. Detecting emotional contagion in massive social networks. PLoS One 2014;9(3):e90315 [FREE Full text] [doi: 10.1371/journal.pone.0090315] [Medline: 24621792]
32. Asur S, Huberman BA. Predicting the Future with Social Media. 2010 Aug 31 Presented at: Ieeewicacm Int Conf Web Intell Intell Agent Technol; 2010; Toronto, ON p. 492-499.
33. Haney C. Sentiment Analysis: Providing Categorical Insight into Unstructured Textual Data. In: Social Media, Sociality, and Survey Research. Hoboken, NJ: Soc Media Sociality Surv Res Internet John Wiley & Sons, Inc; Sep 30, 2013:c2.
34. Ceron A, Curini L, Iacus SM, Porro G. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens’ political preferences with an application to Italy and France. New Media Soc 2014 Feb 18:0466.
35. Pang B, Lee L. Opinion Mining and Sentiment Analysis. Found Trends Inf Retr 2008 Jan;2(1-2):1-135.
36. Duggan M, Ellison NB, Lampe C, Lenhart A, Madden M. Pew Research Centers. 2015 Jan 09. Demographics of Key Social Networking Platforms URL: http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/ [accessed 2015-06-01] [WebCite Cache ID 6Yy9Nm9r9]
37. Twitter. 2015 Mar 01. About Twitter URL: https://about.twitter.com/company [accessed 2015-06-01] [WebCite Cache ID 6Yy9SBWeb]
38. Bigonha CAS, Cardoso TNC, Moro MM, Almeida VAF, Goncales MA. Detecting Evangelists and Detractors on Twitter. In: Brazilian Symposium on Multimedia and the Web. 2010 Presented at: Symposium on Multimedia and the Web (WebMedia); 2010; Brazil p. 107-114 URL: http://homepages.dcc.ufmg.br/~mirella/pdf/Bigonha-Webmedia10.pdf
39. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behav Res Methods 2014 Dec;46(4):1023-1031. [doi: 10.3758/s13428-013-0434-y] [Medline: 24356996]
40. Barr J. Aws Off Blog. Get Better Results with Amazon Mechanical Turk Masters URL: https://aws.amazon.com/blogs/aws/amazon-mechanical-turk-master-workers/ [accessed 2015-06-01] [WebCite Cache ID 6Yy9eMo6I]

41. Annett M, Kondrak G. A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs. 2006 Jul 26 Presented at: Twenty-First Canadian Conference on Artificial Intelligence; 2008; Canada p. 25-35.
42. Cambria E, Schuller B, Xia Y, Havasi C. New Avenues in Opinion Mining and Sentiment Analysis. Ieee Intell Syst 2013 Mar;28(2):15-21.
43. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques. In: 2002 Proceedings of the ACL-02 conference on Empirical methods in natural language processing. USA: Association for Computational Linguistics; 2002 Presented at: Conference on Empirical methods in natural language processing; 2002; Stroudsburg, PA p. 78-86. [doi: 10.3115/1118693.1118704]
44. Dohoo I, Martin W, Stryhn H. Methods in Epidemiologic Research. 1st ed. Charlottetown, PE. Canada: VER Inc; 2012.
45. The Health Site. 2014 Apr 14. World Sleep Day 2014: Can social media give you insomnia? URL: http://www.thehealthsite.com/news/world-sleep-day-2014-can-social-media-give-you-insomnia/ [accessed 2015-06-01] [WebCite Cache ID 6YyA7YLbZ]
46. insomnicacs. Insomnia Innov. How does social media as a technology affect sleeping patterns? URL: https://insomniainnovation.wordpress.com/2013/04/26/how-does-social-media-as-a-technology-affect-sleeping-patterns/ [accessed 2015-06-01] [WebCite Cache ID 6YyA8s0WE]
47. Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. In: Li J, Cao L, Tan KC, Liu B, Pei J, editors. Trends and Applications in Knowledge Discovery and Data Mining. Berlin Heidelberg: Springer; 2013.
48. Park M, Cha C, Cha M. Depressive Moods of Users Portrayed in Online Social Networks. 2012. URL: http://wan.poly.edu/KDD2012/forms/workshop/HI-KDD12/doc/paper_16.pdf [accessed 2015-06-01] [WebCite Cache ID 6YyACgB6C]
49. DeChoudhury M, Counts S, Horvitz E. Social Media As a Measurement Tool of Depression in Populations. In: Proc 5th Annu Acm Web Sci Conf. USA: ACM; 2013 May 23 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2015; Paris, France. [doi: 10.1145/2464464.2464480]
50. DeChoudhury M, Counts S, Horvitz E, Gamon M. Predicting Depression via Social Media. Association for the Advancement of Artificial Intelligence 2013:1-1.
51. Jamison-Powell S, Linehan C, Daley L, Garbett A, Lawson S. “I Can’t Get No Sleep”: Discussing #Insomnia on Twitter. USA: ACM; 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2012; New York, NY p. 1501-1510. [doi: 10.1145/2207676.2208612]

Abbreviations

AMT: Amazon Mechanical Turk
BRFSS: Behavioral Risk Factor Surveillance System
UTC: Universal Time Code

Edited by G Eysenbach; submitted 27.03.15; peer-reviewed by F Timimi, YS Bin, C Hanson; comments to author 21.04.15; revised version received 29.04.15; accepted 24.05.15; published 08.06.15.

Please cite as:
McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, Jain SH, Brownstein JS
Characterizing Sleep Issues Using Twitter
J Med Internet Res 2015;17(6):e140
URL: http://www.jmir.org/2015/6/e140/
doi:10.2196/jmir.4476
PMID:26054530

Original Paper

Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study

Adam G Dunn[1], PhD; Julie Leask[2], MPH, PhD; Xujuan Zhou[1], PhD; Kenneth D Mandl[3,4], MD, MPH; Enrico Coiera[1], MBBS, PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[2]School of Public Health, The University of Sydney, Sydney, Australia

[3]Children’s Hospital Informatics Program, Boston Children’s Hospital, Boston, MA, United States

[4]Center for Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Adam G Dunn, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
L6 75 Talavera Rd
Sydney, 2109
Australia
Phone: 61 298502413
Fax: 61 298502499
Email: adam.dunn@mq.edu.au

Abstract

Background: Groups and individuals that seek to negatively influence public opinion about the safety and value of vaccination are active in online and social media and may influence decision making within some communities.

Objective: We sought to measure whether exposure to negative opinions about human papillomavirus (HPV) vaccines in Twitter communities is associated with the subsequent expression of negative opinions by explicitly measuring potential information exposure over the social structure of Twitter communities.

Methods: We hypothesized that prior exposure to opinions rejecting the safety or value of HPV vaccines would be associated with an increased risk of posting similar opinions and tested this hypothesis by analyzing temporal sequences of messages posted on Twitter (tweets). The study design was a retrospective analysis of tweets related to HPV vaccines and the social connections between users. Between October 2013 and April 2014, we collected 83,551 English-language tweets that included terms related to HPV vaccines and the 957,865 social connections among 30,621 users posting or reposting the tweets. Tweets were classified as expressing negative or neutral/positive opinions using a machine learning classifier previously trained on a manually labeled sample.

Results: During the 6-month period, 25.13% (20,994/83,551) of tweets were classified as negative; among the 30,621 users that tweeted about HPV vaccines, 9046 (29.54%) were exposed to a majority of negative tweets. The likelihood of a user posting a negative tweet after exposure to a majority of negative opinions was 37.78% (2780/7361) compared to 10.92% (1234/11,296) for users who were exposed to a majority of positive and neutral tweets corresponding to a relative risk of 3.46 (95% CI 3.25-3.67, *P*<.001).

Conclusions: The heterogeneous community structure on Twitter appears to skew the information to which users are exposed in relation to HPV vaccines. We found that among users that tweeted about HPV vaccines, those who were more often exposed to negative opinions were more likely to subsequently post negative opinions. Although this research may be useful for identifying individuals and groups currently at risk of disproportionate exposure to misinformation about HPV vaccines, there is a clear need for studies capable of determining the factors that affect the formation and adoption of beliefs about public health interventions.

(J Med Internet Res 2015;17(6):e144) doi:10.2196/jmir.4343

KEYWORDS

HPV vaccines; Twitter messaging; social media; public health surveillance; social networks

Introduction

In the last decade, vaccination refusal has increased in the United States and many countries have recorded substantial proportions of parents expressing concerns about the safety of vaccines [1,2]. Although variability in access to health care is an important factor influencing vaccine coverage rates, vaccination refusal also directly affects these rates and is a significant contributor to outbreaks—especially where vaccination refusal is geographically clustered and population immunity is compromised [3]. Outbreaks of pertussis and measles are known to spread through populations where rates of vaccination refusal are high [4-7].

Refusal has also been a problem for the recently introduced human papillomavirus (HPV) vaccine. The vaccine was first licensed for use in the United States in 2006 with the purpose of reducing the incidence of HPV, to which the majority of cervical cancers are attributed, as well as genital warts and some oral, anal, and penile cancers [8]. HPV vaccination in Australia has led to a marked reduction in rates of high-grade cervical abnormalities and early evidence of herd immunity [9-12]. However, uptake of HPV vaccines varies substantially across and within countries [13-16].

The introduction of HPV vaccination was hampered by controversy in some countries, where some parents attributed illness or death in their children to the vaccine despite evidence affirming the vaccine’s good safety record [17]. The quality and variety of information available online about the safety and efficacy of HPV vaccines varies [18], as does the representation of HPV vaccines in the news media [19]. Evidence from a study set in Greece suggests that the perception of risks in the community appears to have negatively influenced the intention to vaccinate [20]. More generally, there is some evidence to suggest that influence from online media and celebrities can increase vaccine risk perception and rates of vaccination refusal [21-23]. Given the importance of information sources in influencing vaccination decision making, social media platforms are seen as an opportunity for both the tracking and influencing of vaccination decision making [24].

Few studies have considered the surveillance of opinions about vaccination on social media as a precursor to vaccination decision making. Existing studies on public health surveillance applications in social media have focused primarily on finding early indicators of infectious diseases incidence [25-28]. The exceptions include examinations of responses to an influenza outbreak [29] and influenza vaccination [30]. Beyond social media, media surveillance systems have been built to track news media and other reports online [31,32]. One example considered negative sentiment in online news media and notes that systems that rely on manual classification of documents are prohibitively resource intensive [33].

Our aim was to examine the association between exposure to negative opinions about HPV vaccines and the expression of negative opinions about HPV vaccines among Twitter users. To do this, we examined sequences of messages posted on Twitter (tweets) as well as a static view of the social connections between every user that posted a tweet about HPV vaccines in a 6-month period.

Methods

Data

Tweets posted by public users were retrieved programmatically via the Application Programming Interface (API) using repeated searches of combinations of the terms human papillomavirus, HPV, vaccine, vaccination, Gardasil, and Cervarix, and labeled by Twitter as English language. These terms were fixed throughout the data collection period, which was from October 1, 2013 to April 1, 2014. We additionally collected metadata associated with the tweets, including the date and time, information about the user, related tweets such as retweets and replies, and the geo-tag (location) information if it was available. For each user who posted one or more tweets about HPV vaccines in the period, we separately used the API to retrieve the lists of users they followed and the users that followed them shortly after the first time they posted a tweet about HPV vaccines during the period.

Tweets were classified as negative if they rejected the safety or value of HPV vaccines or promoted refusal. Due to the very large number of tweets collected in the period, we used a supervised machine learning approach to classify the tweets that involved the manual labeling of a random sample of tweets, which were then used to train algorithms that recognized similar patterns in the remaining tweets. For each tweet, we determined an estimate of the likelihood of it being the expression of a negative opinion about HPV vaccines. The specific classifier we constructed was an ensemble of 4 classifiers that used the content of the tweets (the words and word combinations in the tweets themselves) or the social relations between users (the users followed by the user responsible for the tweet). A set of 2098 tweets were randomly sampled and then independently graded by 2 investigators (95% agreement, Cohen’s κ=.87), with disagreements resolved by discussion to produce the final training set. The accuracies of the 4 machine learning classifiers ranged between 87.6% and 94.0% when trained and tested in a 10-fold cross validation. The complete details of the development of the classifier are described elsewhere [34].

Analyses

To analyze population-level information exposure, we measured how users may have been exposed to tweets about HPV vaccines during the 6-month observation period. For each user that tweeted at least once about HPV vaccines during the period, we created timelines of their own tweets about HPV vaccines and the tweets about HPV vaccines posted by the users they followed. For the purpose of measuring information exposure, we handled retweets in the same way as other tweets to conserve the definition for exposure. This means that we defined an exposure as the potential flow of information between users along social connections. Not all tweets are seen by all followers, but by observing the aggregate flow of exposures through network structure, it was possible to estimate how the heterogeneous mixing of the population might affect the information to which each user is exposed.

We determined the prior exposure of a user each time they posted a tweet about HPV vaccines during the time period by compiling the list of tweets to which they were potentially exposed prior to the timestamp of the index tweet. This proportion served as an indicator of the prior exposure to negative information about HPV vaccines in the time period. To account for a potential length sampling bias (later tweets tended to be preceded by a greater number of exposures), we limited the sequence-based analysis to tweets that were preceded by at least 3 exposures.

To test our hypothesis directly, we counted how many times a user posted a negative tweet following a majority of prior negative exposures and compared that count with the number of times a posted tweet was negative when the majority of prior exposures were neutral or positive. These counts were then used to calculate the relative risk of posting a negative tweet about HPV vaccines given majority prior exposure to negative tweets. To avoid sampling biases resulting from counting the same users repeatedly, we randomly sampled only 1 tweet from each eligible user and repeated the analysis until the median proportions and relative risk measures did not change value at 3 significant figures.

Results

We identified 83,551 tweets or retweets from 30,621 users relating to HPV vaccines between the period October 1, 2013 to April 1, 2014, after eliminating tweets that were eventually deleted and tweets from users that became protected or suspended after the initial collection. Of the 83,551 tweets and retweets, 20,994 (25.13%) were classified as negative by an ensemble of supervised machine learning classifiers. Table 1 includes some examples of the different classes of tweets. There were 10 days (5.5% of 183 days) in which the number of negative tweets outnumbered the number of positive and neutral tweets about HPV vaccines (Figure 1).

Table 1. Examples of different classes of Twitter messages identified in the searches.

Classification	Twitter message text
Positive	“HPV vaccination has the potential to reduce cervical cancer deaths worldwide by as much as two-thirds. [URL removed]”
Positive	“Oral sex & male gender indep assoc with oral HPV infection: shows need for HPV vaccination of boys. #endhpv New study [URL removed]”
Neutral	“Potential of the quadrivalent human papillomavirus vaccine in the prevention and treatment of cervical cancer [URL removed]”
Negative	“Gardasil has generated nearly 30,000 adverse reaction reports to US govt, including 140 deaths [URL removed] #vaxfax”
Negative	“Lead Developer of HPV Vaccine Warns Parents Young Girls It’s a Giant Deadly Scam [URL removed]”
Negative	“Young woman’s ovaries destroyed by Gardasil: Merck ‘forgot to research’ effects of vaccine [URL removed]”

There were 30,621 users that tweeted about HPV vaccines in the period. Each user in the set posted between 1 and 1842 tweets about HPV vaccines during the period with a median of 2 tweets per user (IQR 1-2) (Figure 2). The distributions differed between users posting mostly negative tweets and users posting mostly neutral or positive tweets. Although there were more users posting neutral/positive tweets overall, the most prolific users during the time period were posting mostly negative opinions about HPV vaccines.

We defined social connections as the sets of users that followed, or were followed by, the users that tweeted about HPV vaccines. The total number of unique followers for all users that tweeted about HPV vaccines in the 6-month period was 51,397,377. The total number of followers per user varied between 0 and 5,136,595 with a median of 274 followers per user (IQR 36-996) (Figure 3, left). Considering only the connections between users that tweeted about HPV vaccines, 957,865 social connections were identified and this defined the internal network of social connections among the 30,621 users. Followers per user in this internal network varied from 0 to 10,945 with a median of 8 followers per user (IQR 2-33) (Figure 3, right). Although news organizations and magazines made up the majority of users with the greatest number of followers overall, government health organizations and academic institutions or groups were more consistently featured among the set of users with the most followers in the internal network. Practitioners and writers (books and blogs) of specific forms of alternative medicine as well as antivaccine activists and celebrities did not feature among the set of users with the most followers overall, but occupied higher ranks when counting the number of followers in the internal network.

Although only 25.13% (20,994/83,551) of tweets were classified as negative, 29.54% (9046/30,621) of users that tweeted about HPV vaccines appeared to be exposed more often to negative tweets than to neutral and positive tweets. This difference, and a visual interpretation of the network, suggests that users posting negative tweets about HPV vaccines were not evenly mixed in the network and often belonged to communities primarily consisting of users who also posted negative tweets about HPV vaccines (Figure 4).

Among the 30,621 users that tweeted about HPV vaccines, 18,657 users had timelines in which at least 1 tweet was posted after at least 3 exposures and were thus eligible for a temporal analysis of exposures and subsequent tweets. The likelihood of posting a negative tweet about HPV vaccines following a prior majority exposure to negative tweets was 37.78% (2780 of 7361 users). For users whose prior exposures were mostly neutral/positive, 10.92% (1234 of 11,296 users) subsequently posted a negative tweet. These results corresponded to a relative risk of 3.46 (95% CI 3.25-3.67, *P*<.001) indicating that users with greater prior exposure to negative opinions about HPV vaccines were more likely to express negative opinions.

To further test the association between exposure and expression within different groups of users, we undertook a post hoc subgroup analysis. Among the set of users that met the inclusion criteria and also had fewer than 1000 followers (n=11,845), we calculated the relative risk in the same way and found that the relative risk of posting a negative opinion about HPV vaccines after having been more often exposed to negative opinions about HPV vaccines was 3.61 (95% CI 3.32-3.93). For users with fewer than 500 followers (n=8790), the relative risk was 3.57 (95% CI 3.23-3.95) and for users with fewer than 300 followers (n=6521), the relative risk was 3.76 (95% CI 3.33-4.24). The results suggest that the association between previous exposure and subsequent expression was slightly stronger among Twitter users with fewer followers.

Figure 1. The number of tweets posted each day during the data collection period, including tweets rejecting the safety or value of HPV vaccines (orange) and all other HPV vaccine tweets (cyan). Gray vertical lines indicate Sundays. No corrections for time zone differences were applied.

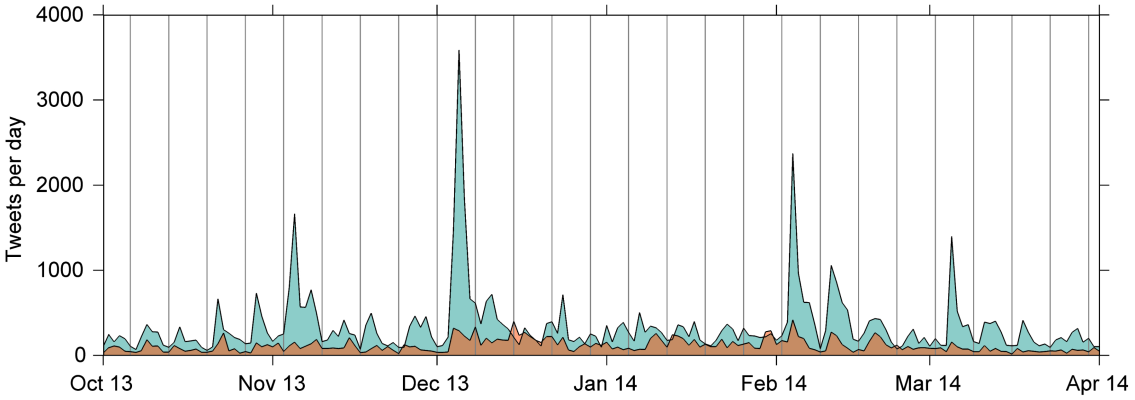


Figure 2. The ordered distribution of tweets per user related to HPV vaccines posted to Twitter between October 1, 2013 and March 31, 2014. Each user’s number of tweets is represented by a dot and illustrated separately for users that posted a majority of negative tweets (orange) and all other users (cyan).

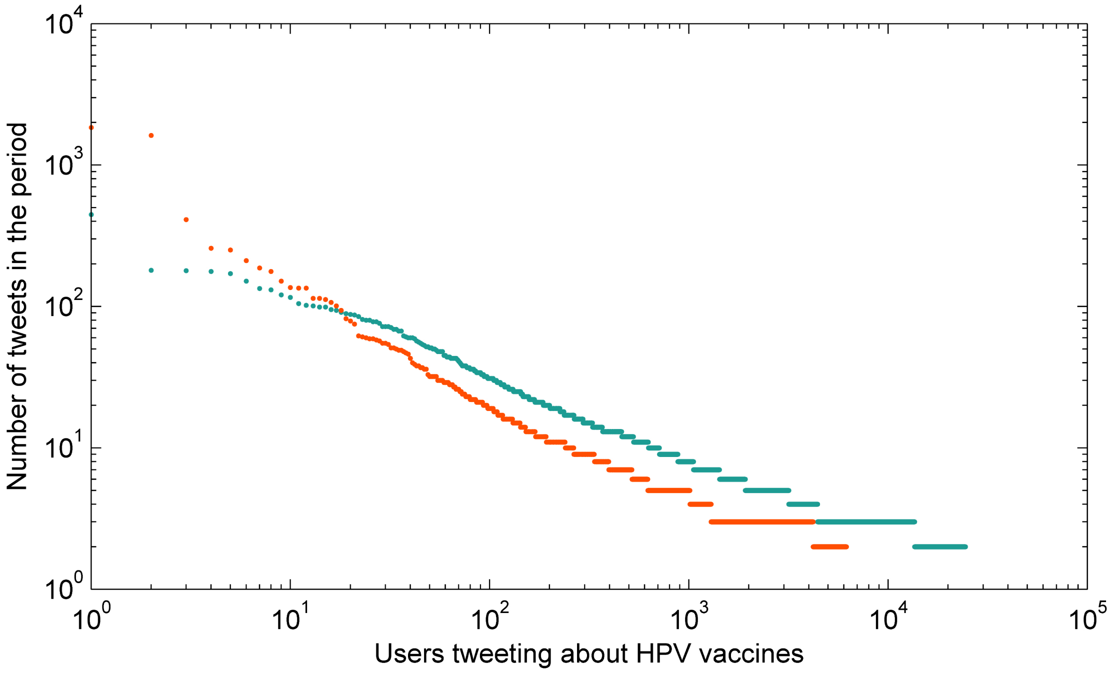


Figure 3. The ordered distribution of users according to the total follower counts (left) and follower counts within the network of 30,621 users (right). Each user is represented by a dot and colored by users that tweeted mostly negative tweets (orange) compared to all other users (cyan). The vertical axes are zero-adjusted to accommodate users that had zero followers.

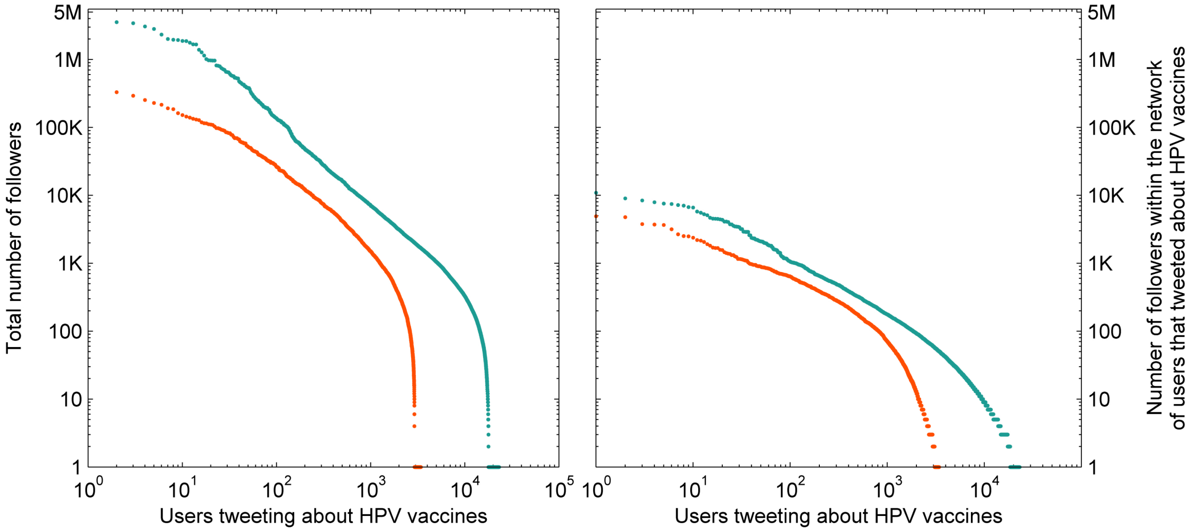
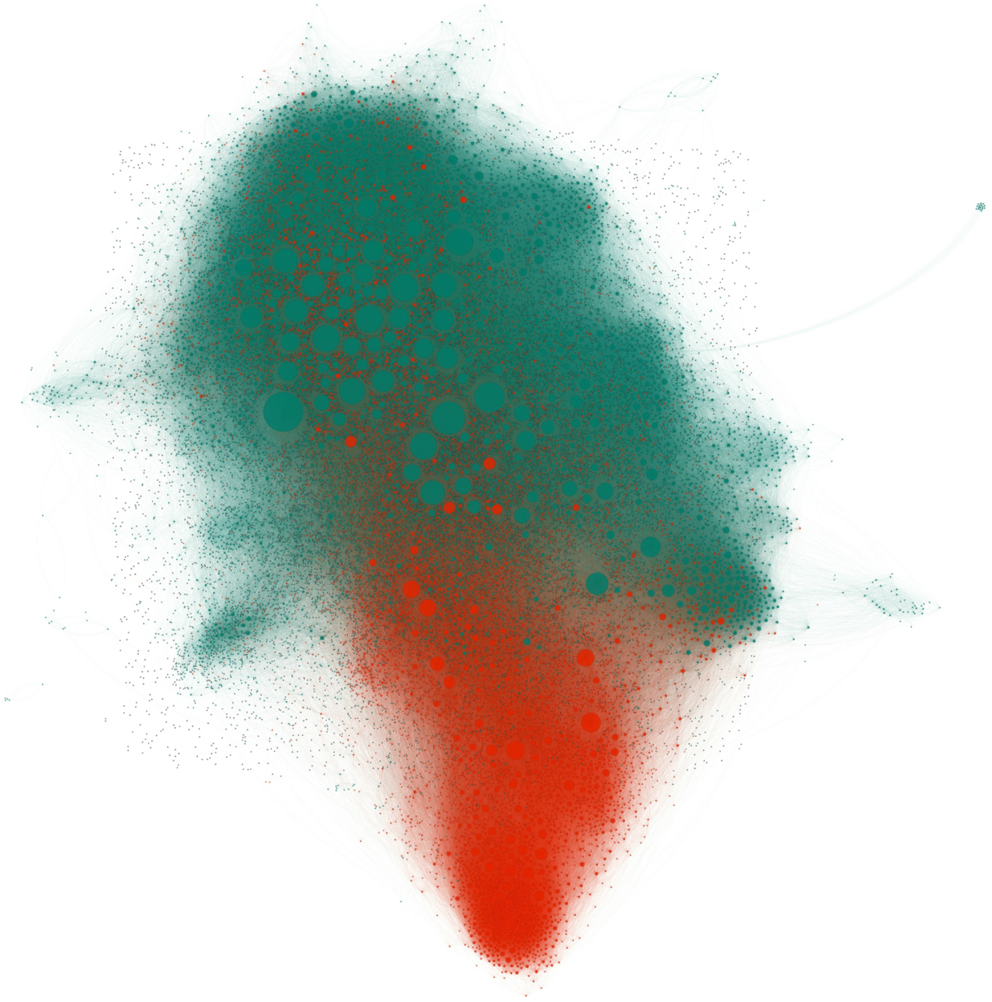


Figure 4. The network of 30,621 users that tweeted about HPV vaccines during the period between October 2013 and April 2014 organized via heuristic so that users are closer to other users with whom they are connected. The sizes of the nodes are proportional to the number of followers within the network. Users are colored according to information exposure (orange: those exposed to a majority of negative opinions; cyan: users that were exposed to mostly neutral/positive tweets; gray: users not exposed to HPV vaccine tweets).



Discussion

Principal Findings

Approximately one-quarter of the tweets about HPV vaccines that were posted in the period were critical of the safety or value of HPV vaccines or actively encouraged vaccine refusal. These tweets, which included misinformation, anecdotes, and opinions that may result in vaccine hesitancy or refusal, made up the majority of HPV vaccine-related information exposures for nearly 30% of users that tweeted about HPV vaccines in the period. Our analysis of the network of follower relationships suggests that users expressing negative opinions about HPV vaccines tended to be more closely connected to users expressing the same opinions. Our analysis of the sequences of HPV-related tweets demonstrated an association between prior exposure to negative tweets about HPV vaccines and the subsequent posting of negative tweets about HPV vaccines. Together, these results suggest that homophily or contagion may play a role in the expression of negative opinions about HPV vaccines, but the study does not help to quantify their specific contributions [35].

To the best of our knowledge, ours is the first empirical study to consider the association between information exposure and subsequent expression for vaccines on social media. Other studies have used supervised machine learning to automatically classify tweets about vaccination [30] and the frequency of tweets over time exhibits a similar temporal pattern to the one we observed. Other studies have used Twitter as a laboratory to measure the propagation of negative news content, complaints, and rumors [36-38]. Other studies that considered misinformation were specifically aimed at differentiating between credible and not credible information, the containment of misinformation, and the identification of misinformation sources [39-41].

It is important to note that the study design we used precluded conclusions about what proportions of negative opinions expressed in the period were the consequence of exposure (contagion of opinions), the consequence of users creating connections to other users who already hold similar opinions (homophily), or if other external factors caused connected users to express similar opinions [35]. Alternative study designs that measure or model contagion from observable or synthetic networks are common in other application domains and more generally in network science [42-46], including where connections between nodes change over time [47-49].

Other studies have considered the news and online media representation of vaccines in different ways. One study examining the representation of vaccines in the media identified a rate of negative opinions in media reports for vaccines generally of 31% [33], with similar percentages in a study of US and Canadian news articles about HPV vaccines [50]. In comparison, 29% of US parents have reported being unsure about the vaccines for their children or otherwise delayed or refused vaccinations [51]. In the United Kingdom, very few newspaper articles (including tabloids) were classified as negative [52], whereas 19% of parents in England responded that they would not vaccinate their children in the future [53]. An Australian study found that HPV safety concerns were present in 39% of newspaper articles between 2006 and 2009 [19]. A study examining news media in the mid-1990s found that a small number of individuals were responsible for nearly half of all the statements opposing vaccination [54]. We found a similar pattern on Twitter for HPV vaccines using data from nearly 20 years later—where a small number of individuals posting negative opinions on Twitter produced a substantial proportion of the negative opinions. Given that these proportions are much higher than the average rates of vaccination refusal recorded in registries at approximately 2% [55,56], more work is needed to understand how population-level indicators of negative opinions might relate to vaccination decision making.

Implications

Implications of this work include new avenues for understanding how community affiliation on Twitter corresponds to the exposure to misinformation, the subsequent expression of opinions, and individual decision making. The simple methods we used here may be of practical value for answering questions about how new information becomes established in different communities. For example, do the results of scientific studies demonstrating efficacy tend to spread primarily through scientific communities and not through communities of hesitant parents? Which popular news websites, influential users, or organizations are better connected to communities that are at higher risk of being exposed to, and subsequently affected by, misinformation? How often do young teenagers or their parents pass along negative opinions following encounters with misinformation or negative experiences with the vaccine process? Using new methods for classifying the location and characteristics of Twitter users [57,58], it may be possible to construct Twitter-derived indicators of skewed misinformation exposure in geographic areas and demographic strata, and these may be useful for predicting or reflecting localized shifts in decision making such as increases in refusal. From a practical perspective, this kind of information risk surveillance could be used to complement existing methods for gathering localized information (surveys, interviews, and registry analysis) and improve community engagement and public health actions by targeting resources more efficiently.

Limitations

Limitations of this study come from our inability to track social connections as they appear and disappear during the period. Due to limits in the rates at which we access this information on Twitter, the social connections associated with each user were collected only once during the period, shortly after the first time we identified a relevant tweet by the user. However, by checking the consistency of connections between users within the set, we found that 81.6% of users' connections were confirmed by the information from the other user (eg, a user's follower is confirmed as someone the user follows), so we are reasonably confident that the connection structure was relatively consistent over time.

Our search terms were fixed and although we were careful to select search terms that covered the vast majority of the discussion about HPV vaccines without collecting irrelevant tweets, we may have missed a smaller number of tweets about the topic and these tweets may not have had the same proportion of negative opinions. Query-expansion techniques used to improve search strategies over time could be applied to address this limitation in future work [59-61]. Finally, we relied on an ensemble classifier rather than manual labeling, so a small proportion of the tweets will have been misclassified. However, the imperfections in the classifier are unlikely to have affected the results because the study was across large groups, our measure of exposure was based on counting the majority across a number of tweets rather than individual tweets, and the associations were clear.

Conclusions

We found that Twitter users who were more often exposed to negative opinions about the safety and value of HPV vaccines were more likely to tweet negative opinions than users who were more often exposed to neutral or positive information. Although we were unable to determine the differential contributions of homophily, user characteristics, and contagion to this effect, the results provide a detailed view of negative opinions about HPV vaccines on Twitter in the period and indicate associations between the community structure, information exposure, and expression of negative opinions about vaccines among social media users. Ongoing surveillance of opinions about vaccination on social media may complement surveys and other public health surveillance methods to improve the efficiency and efficacy of public health communication strategies.

Acknowledgments

The National Health and Medical Research Council (NHMRC) provided funding for the Centre for Research Excellence in eHealth. The NHMRC played no role in the review or approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Omer SB, Richards JL, Ward M, Bednarczyk RA. Vaccination policies and rates of exemption from immunization, 2005-2011. N Engl J Med 2012 Sep 20;367(12):1170-1171. [doi: 10.1056/NEJMc1209037] [Medline: 22992099]
2. Leask J. Target the fence-sitters. Nature 2011 May 26;473(7348):443-445. [doi: 10.1038/473443a] [Medline: 21614055]
3. Omer SB, Pan WY, Halsey NA, Stokley S, Moulton LH, Navar AM, et al. Nonmedical exemptions to school immunization requirements: secular trends and association of state policies with pertussis incidence. JAMA 2006 Oct 11;296(14):1757-1763. [doi: 10.1001/jama.296.14.1757] [Medline: 17032989]
4. Sugerman DE, Barskey AE, Delea MG, Ortega-Sanchez IR, Bi D, Ralston KJ, et al. Measles outbreak in a highly vaccinated population, San Diego, 2008: role of the intentionally undervaccinated. Pediatrics 2010 Apr;125(4):747-755 [FREE Full text] [doi: 10.1542/peds.2009-1653] [Medline: 20308208]
5. Parker AA, Staggs W, Dayan GH, Ortega-Sánchez IR, Rota PA, Lowe L, et al. Implications of a 2005 measles outbreak in Indiana for sustained elimination of measles in the United States. N Engl J Med 2006 Aug 3;355(5):447-455. [doi: 10.1056/NEJMoa060775] [Medline: 16885548]
6. Najjar Z, Hope K, Clark P, Nguyen O, Rosewell A, Conaty S. Sustained outbreak of measles in New South Wales, 2012: risks for measles elimination in Australia. Western Pac Surveill Response J 2014;5(1):14-20 [FREE Full text] [doi: 10.5365/WPSAR.2013.4.4.002] [Medline: 25635228]
7. Atwell JE, Van OJ, Zipprich J, Winter K, Harriman K, Salmon DA, et al. Nonmedical vaccine exemptions and pertussis in California, 2010. Pediatrics 2013 Oct;132(4):624-630 [FREE Full text] [doi: 10.1542/peds.2013-0878] [Medline: 24082000]
8. Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. Vaccine 2012 Nov 20;30 Suppl 5:F12-F23. [doi: 10.1016/j.vaccine.2012.07.055] [Medline: 23199955]
9. Baldur-Felskov B, Dehlendorff C, Junge J, Munk C, Kjaer SK. Incidence of cervical lesions in Danish women before and after implementation of a national HPV vaccination program. Cancer Causes Control 2014 Jul;25(7):915-922. [doi: 10.1007/s10552-014-0392-4] [Medline: 24797870]
10. Gertig DM, Brotherton JML, Budd AC, Drennan K, Chappell G, Saville AM. Impact of a population-based HPV vaccination program on cervical abnormalities: a data linkage study. BMC Med 2013;11:227 [FREE Full text] [doi: 10.1186/1741-7015-11-227] [Medline: 24148310]
11. Tabrizi SN, Brotherton JML, Kaldor JM, Skinner SR, Liu B, Bateson D, et al. Assessment of herd immunity and cross-protection after a human papillomavirus vaccination programme in Australia: a repeat cross-sectional study. Lancet Infect Dis 2014 Oct;14(10):958-966. [doi: 10.1016/S1473-3099(14)70841-2] [Medline: 25107680]
12. Niccolai LM, Julian PJ, Meek JI, McBride V, Hadler JL, Sosa LE. Declining rates of high-grade cervical lesions in young women in Connecticut, 2008-2011. Cancer Epidemiol Biomarkers Prev 2013 Aug;22(8):1446-1450 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0272] [Medline: 23704476]

13. Brotherton JML, Liu B, Donovan B, Kaldor JM, Saville M. Human papillomavirus (HPV) vaccination coverage in young Australian women is higher than previously estimated: independent estimates from a nationally representative mobile phone survey. Vaccine 2014 Jan 23;32(5):592-597. [doi: 10.1016/j.vaccine.2013.11.075] [Medline: 24316239]
14. Tsui J, Rodriguez HP, Gee GC, Escobedo LA, Kominski GF, Bastani R. Are HPV vaccination services accessible to high-risk communities? A spatial analysis of HPV-associated cancer and Chlamydia rates and safety-net clinics. Cancer Causes Control 2013 Dec;24(12):2089-2098 [FREE Full text] [doi: 10.1007/s10552-013-0286-x] [Medline: 24043448]
15. Fisher H, Trotter CL, Audrey S, MacDonald-Wallis K, Hickman M. Inequalities in the uptake of human papillomavirus vaccination: a systematic review and meta-analysis. Int J Epidemiol 2013 Jun;42(3):896-908 [FREE Full text] [doi: 10.1093/ije/dyt049] [Medline: 23620381]
16. Gertig DM, Brotherton JML, Saville M. Measuring human papillomavirus (HPV) vaccination coverage and the role of the National HPV Vaccination Program Register, Australia. Sex Health 2011 Jun;8(2):171-178. [doi: 10.1071/SH10001] [Medline: 21592430]
17. Macartney KK, Chiu C, Georgousakis M, Brotherton JML. Safety of human papillomavirus vaccines: a review. Drug Saf 2013 Jun;36(6):393-412. [doi: 10.1007/s40264-013-0039-5] [Medline: 23637071]
18. Madden K, Nan X, Briones R, Waks L. Sorting through search results: a content analysis of HPV vaccine information online. Vaccine 2012 May 28;30(25):3741-3746. [doi: 10.1016/j.vaccine.2011.10.025] [Medline: 22019758]
19. Robbins Spring Chenoa Cooper, Pang C, Leask J. Australian newspaper coverage of human papillomavirus vaccination, October 2006-December 2009. J Health Commun 2012;17(2):149-159. [doi: 10.1080/10810730.2011.585700] [Medline: 22136302]
20. Sotiriadis A, Dagklis T, Siamanta V, Chatzigeorgiou K, Agorastos T. Increasing fear of adverse effects drops intention to vaccinate after the introduction of prophylactic HPV vaccine. Arch Gynecol Obstet 2012 Jun;285(6):1719-1724. [doi: 10.1007/s00404-011-2208-z] [Medline: 22246478]
21. Hoffman SJ, Tan C. Following celebrities' medical advice: meta-narrative analysis. BMJ 2013 Dec 17;347:f7151. [doi: 10.1136/bmj.f7151]
22. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. J Epidemiol Community Health 2000 Jun;54(6):473-474 [FREE Full text] [Medline: 10818125]
23. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
24. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 2012 May 28;30(25):3727-3733. [doi: 10.1016/j.vaccine.2012.02.025] [Medline: 22365840]
25. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
26. 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]
27. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
28. Collier N, Son NT, Nguyen NM. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011;2 Suppl 5:S9 [FREE Full text] [doi: 10.1186/2041-1480-2-S5-S9] [Medline: 22166368]
29. 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]
30. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
31. 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]
32. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 2008 Jul 8;5(7):e151 [FREE Full text] [doi: 10.1371/journal.pmed.0050151] [Medline: 18613747]
33. Larson HJ, Smith DMD, Paterson P, Cumming M, Eckersberger E, Freifeld CC, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 2013 Jul;13(7):606-613. [doi: 10.1016/S1473-3099(13)70108-7] [Medline: 23676442]
34. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information alone to robustly predict anti-vaccine opinions on social media. In: Proceedings of the 15th World Congress on Medical and Health Informatics. 2015 Presented at: MedInfo'15; August 19-23, 2015; São Paulo.
35. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociol Methods Res 2011 May;40(2):211-239 [FREE Full text] [doi: 10.1177/0049124111404820] [Medline: 22523436]

36. Hansen L, Arvidsson A, Nielsen F, Colleoni E, Etter M. Good friends, bad news - affect and virality in Twitter. Future Information Technology 2011;185:34-43. [doi: 10.1007/978-3-642-22309-9_5]
37. Pfeffer J, Zorbach T, Carley KM. Understanding online firestorms: negative word-of-mouth dynamics in social media networks. Journal of Marketing Communications 2013 Jun 10;20(1-2):117-128. [doi: 10.1080/13527266.2013.797778]
38. Jin F, Dougherty E, Saraf P, Cao Y, Ramakrishnan N. Epidemiological modeling of news and rumors on Twitter. In: Proceedings of the 7th Workshop on Social Network Mining and Analysis. 2013 Presented at: KDD'13 The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 11-14, 2013; Chicago, IL p. 1-9. [doi: 10.1145/2501025.2501027]
39. Castillo C, Mendoza M, Poblete B. Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web. 2011 Presented at: WWW '11 20th International World Wide Web Conference; March 28-April 1, 2011; Hyderabad, India p. 675-684. [doi: 10.1145/1963405.1963500]
40. Nguyen NP, Yan G, Thai MT, Eidenbenz S. Containment of misinformation spread in online social networks. In: Proceedings of the 4th Annual ACM Web Science Conference. 2012 Presented at: WebSci '12 Web Science 2012; June 22-24, 2012; Evanston, IL p. 213-222. [doi: 10.1145/2380718.2380746]
41. Shah D, Zaman T. Rumors in a network: who's the culprit? IEEE Trans Inform Theory 2011 Aug;57(8):5163-5181. [doi: 10.1109/TIT.2011.2158885]
42. Hodas NO, Lerman K. The simple rules of social contagion. Sci Rep 2014;4:4343 [FREE Full text] [doi: 10.1038/srep04343] [Medline: 24614301]
43. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 2;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
44. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev Mod Phys 2009 May 11;81(2):591-646. [doi: 10.1103/RevModPhys.81.591]
45. Weng L, Flammini A, Vespignani A, Menczer F. Competition among memes in a world with limited attention. Sci Rep 2012;2:335 [FREE Full text] [doi: 10.1038/srep00335] [Medline: 22461971]
46. Iyengar R, Van den Bulte C, Valente TW. Opinion leadership and social contagion in new product diffusion. Marketing Science 2011 Mar;30(2):195-212. [doi: 10.1287/mksc.1100.0566]
47. Holme P. Network dynamics of ongoing social relationships. Europhysics Letters 2003;64(3):427-433. [doi: 10.1209/epl/i2003-00505-4]
48. Holme P, Liljeros F. Birth and death of links control disease spreading in empirical contact networks. Sci Rep 2014;4:4999 [FREE Full text] [doi: 10.1038/srep04999] [Medline: 24851942]
49. Liu S, Baronchelli A, Perra N. Contagion dynamics in time-varying metapopulation networks. Phys Rev E 2013;87(3):032805. [doi: 10.1103/PhysRevE.87.032805]
50. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: a content analysis of Canadian and US national newspaper articles. Women Health 2009;49(5):422-440. [doi: 10.1080/03630240903238776] [Medline: 19851946]
51. Gust DA, Darling N, Kennedy A, Schwartz B. Parents with doubts about vaccines: which vaccines and reasons why. Pediatrics 2008 Oct;122(4):718-725 [FREE Full text] [doi: 10.1542/peds.2007-0538] [Medline: 18829793]
52. Hilton S, Hunt K, Langan M, Bedford H, Petticrew M. Newsprint media representations of the introduction of the HPV vaccination programme for cervical cancer prevention in the UK (2005-2008). Soc Sci Med 2010 Mar;70(6):942-950 [FREE Full text] [doi: 10.1016/j.socscimed.2009.11.027] [Medline: 20064682]
53. Stefanoff P, Mamelund SE, Robinson M, Netterlid E, Tuells J, Bergsaker MAR, VACSATC working group on standardization of attitudinal studies in Europe. Tracking parental attitudes on vaccination across European countries: The Vaccine Safety, Attitudes, Training and Communication Project (VACSATC). Vaccine 2010 Aug 9;28(35):5731-5737. [doi: 10.1016/j.vaccine.2010.06.009] [Medline: 20558250]
54. Leask JA, Chapman S. An attempt to swindle nature: press anti-immunisation reportage 1993-1997. Aust N Z J Public Health 1998 Feb;22(1):17-26. [Medline: 9599848]
55. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 2009 May 7;360(19):1981-1988. [doi: 10.1056/NEJMsa0806477] [Medline: 19420367]
56. Hull B, McIntyre P. Mapping immunisation coverage and conscientious objectors to immunisation in NSW. N S W Public Health Bull 2003;14(1-2):8-12. [Medline: 12732911]
57. Jurgens D, Finnethy T, McCorriston J, Xu YT, Ruths D. Geolocation prediction in Twitter using social networks: a critical analysis and review of current practice. 2015 Presented at: The 9th International Conference on Web and Social Media, ICWSM-15; May 26-29, 2015; Oxford, UK.
58. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: EMNLP '11; July 2011; Edinburgh, UK p. 1301.
59. Massoudi K, Tsagkias M, de Rijke M. Incorporating query expansion and quality indicators in searching microblog posts. 2011 Presented at: Advances in Information Retrieval: 33rd European Conference on IR Research; April 18-21, 2011; Dublin, Ireland p. 362-367.

60. Kumar N, Carterette B. Time based feedback and query expansion for Twitter search. In: Advances in Information Retrieval Volume 7814. Heidelberg: Springer Berlin; 2013:734-737.
61. Carpineto C, Romano G. A survey of automatic query expansion in information retrieval. ACM Comput Surv 2012 Jan 01;44(1):1-50. [doi: 10.1145/2071389.2071390]

Abbreviations

API: Application Programming Interface
HPV: human papillomavirus
IQR: interquartile range

Edited by G Eysenbach; submitted 12.02.15; peer-reviewed by C Zhang, R Briones; comments to author 15.04.15; accepted 22.04.15; published 10.06.15.

Please cite as:
Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E
Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study
J Med Internet Res 2015;17(6):e144
URL: http://www.jmir.org/2015/6/e144/
doi:10.2196/jmir.4343
PMID:26063290

Original Paper

Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture?

Rachel Lynn Kendra[1]; Suman Karki[1], MS; Jesse Lee Eickholt[1], PhD; Lisa Gandy[1], PhD

Department of Computer Science, Central Michigan University, Mount Pleasant, MI, United States

Corresponding Author:
Jesse Lee Eickholt, PhD
Department of Computer Science
Central Michigan University
413 Pearce Hall
Mount Pleasant, MI, 48859
United States
Phone: 1 989 774 3739
Fax: 1 989 774 3728
Email: eickh1jl@cmich.edu

Abstract

Background: User content posted through Twitter has been used for biosurveillance, to characterize public perception of health-related topics, and as a means of distributing information to the general public. Most of the existing work surrounding Twitter and health care has shown Twitter to be an effective medium for these problems but more could be done to provide finer and more efficient access to all pertinent data. Given the diversity of user-generated content, small samples or summary presentations of the data arguably omit a large part of the virtual discussion taking place in the Twittersphere. Still, managing, processing, and querying large amounts of Twitter data is not a trivial task. This work describes tools and techniques capable of handling larger sets of Twitter data and demonstrates their use with the issue of antibiotics.

Objective: This work has two principle objectives: (1) to provide an open-source means to efficiently explore all collected tweets and query health-related topics on Twitter, specifically, questions such as what users are saying and how messages are spread, and (2) to characterize the larger discourse taking place on Twitter with respect to antibiotics.

Methods: Open-source software suites Hadoop, Flume, and Hive were used to collect and query a large number of Twitter posts. To classify tweets by topic, a deep network classifier was trained using a limited number of manually classified tweets. The particular machine learning approach used also allowed the use of a large number of unclassified tweets to increase performance.

Results: Query-based analysis of the collected tweets revealed that a large number of users contributed to the online discussion and that a frequent topic mentioned was resistance. A number of prominent events related to antibiotics led to a number of spikes in activity but these were short in duration. The category-based classifier developed was able to correctly classify 70% of manually labeled tweets (using a 10-fold cross validation procedure and 9 classes). The classifier also performed well when evaluated on a per category basis.

Conclusions: Using existing tools such as Hive, Flume, Hadoop, and machine learning techniques, it is possible to construct tools and workflows to collect and query large amounts of Twitter data to characterize the larger discussion taking place on Twitter with respect to a particular health-related topic. Furthermore, using newer machine learning techniques and a limited number of manually labeled tweets, an entire body of collected tweets can be classified to indicate what topics are driving the virtual, online discussion. The resulting classifier can also be used to efficiently explore collected tweets by category and search for messages of interest or exemplary content.

(J Med Internet Res 2015;17(6):e154) doi:10.2196/jmir.4220

KEYWORDS

Twitter messaging; social media; Internet; Web mining; semi-supervised learning; neural network

Introduction

The development and proliferation of social media and social media networks have transformed how information is generated and shared. Participants in social media are actively engaged and are both consumers and producers of information. Toffler et al label these users “prosumers” [1]. The Pew Research Center reported in January 2014 that 74% of online adults use social networking sites. They also report that 46% of adult Internet users post original photos or videos online that they have created [2]. This content posted by users is of great utility, and several studies have demonstrated how social media and networks can be a valuable source of data. Furthermore, the nature and ubiquity of social media and how it is so interwoven in daily life means that the topics covered span the spectrum. For example, trending topics on Twitter in December 2014 were the Siege in Sydney, Australia, the antics of Rubius Gunderson, a popular prankster on YouTube, and news regarding Ross Barkley, who is a popular soccer player from the English National Team.

Twitter is a social media platform through which users post status updates called tweets [3]. A tweet can contain up to 140 characters and can be public (ie, any visitor can access and view the tweet) or protected (ie, only approved visitors can view the tweet). As of October 2012, approximately 88% of Twitter accounts were public [4]. Users can “follow” other users and thus be apprised of any new tweets that the followed users post, and a user may be followed by any number of other users. Users can also repost tweets pushed out by users they follow. This action is known as retweeting and can lead to a message “going viral”, a phenomenon that quickly spreads the reach of a message [5]. Another way that users can increase the reach of their message is through hashtags (a continuous string of characters that begin with a #). Hashtags are often used to hint at the content of a tweet and provide an additional means to tie related tweets together.

Twitter users tweet about a variety of subjects including health-related topics. Such tweets may share information about health articles or describe personal health issues. A number of approaches have been developed to extract useful health-related information from the tweets, evaluate the effectiveness of Twitter for disseminating health-related information, and determine public sentiment towards health-related topics [6-12]. Love et al did an extensive study on 6827 tweets related to vaccinations to determine the source of the information and the medical claims made [10]. They found that no particular source or medical claim dominated the content shared regarding vaccinations and that 87% of user posts were positive or neutral. Scanfeld et al manually inspected and determined a number of topic-based categories present in tweets related to antibiotics. In doing, so several examples of misunderstanding or misuse were detected based on keyword combinations (eg, “antibiotics” and “flu”) [11]. Furthermore, Vance et al explored the concept of using social media to disseminate public health information to young adults [12]. Advantages to such an approach included rapid communication and low cost. The drawbacks cited were opinions often being represented as facts, the use of blind authorship, and a lack of citations.

An additional use of Twitter data is biosurveillance [13,14]. Self-reported behaviors can be monitored and used to detect epidemics or break-outs in real time through crowd sourcing. For example, Google Flu Trends tracks the influenza rates by tracking user queries on a daily basis, and their system is usually 7-10 days faster than the Centers for Disease Control and Prevention [15]. Lampos and Cristianini found a correlation between tweets about the flu and historical data from the same time period, further strengthening the claims that Twitter can be used for biosurveillance [16].

In this work, we focused on accessing and mining topics in the Twittersphere with open source tools. In doing so, we had two objectives in mind. First, we wanted to characterize the exchange currently taking place on Twitter with respect to a particular topic. Of particular interest was examining what was dominating the virtual discussion and if it was being dominated by a small set of users. In answering these questions, we wanted to leverage as much of the virtual discussion as possible. As a result, the second objective of this work was to develop tools and workflows to access the larger Twittersphere. To this end, we developed a classifier that can be used to identify and draw out tweets pertaining to several categories. The classifier was trained using a semisupervised approach that allows it to make use of all of the collected tweets during the learning process. We also describe a number of tools and techniques for handling the larger amounts of data. In particular, Hadoop and Hive were used to query and characterize the large number of tweets that were collected. The resulting pipeline could be used as a tool for infodemiology and infoveillance, providing a means of ferreting out sources of information or specific types of messages shared through Twitter. Given the ease at which Hadoop scales and the availability of cloud computing platforms, this approach could easily be applied to much larger datasets and other topics.

As a case study, the pipeline and tools developed in this work were applied to the topic of antibiotics and antibiotic resistance. By querying 591,091 tweets with the workflow, the current discussion surrounding antibiotics could be characterized and types of misuse examined. This revealed that there were a large number of unique participants in the online discussion and that the discussion was not being dominated by a set of users but rather by a large number of users who were sharing national news stories. The topic of antibiotics was chosen given the potential economic costs associated antibiotic resistance in bacteria [17,18] and its dominance in the media (eg, Longitude Prize). Additionally, with the number of existing studies on Twitter and antibiotics, this topic provided a context in which our tools, methodology, and findings could be compared and contrasted.

Methods

Data Collection

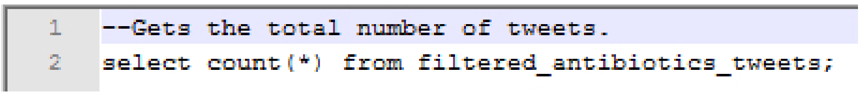
To collect tweets related to antibiotics, the Twitter Application Programming Interface (API) was used in conjunction with a list of 89 antibiotic-related terms and Apache Flume [19]. These terms included expected keywords such as antibiotic(s) as well as common abbreviations (eg, abx), names of specific antibiotics (eg, amoxicillin, penicillin), and common misspellings (eg, antiboitic). Collection of the tweets began on May 27, 2014, and ended September 11, 2014. These dates were selected to provide a minimum collection period of 3 months. Additionally, the number of tweets collected was periodically checked and an aim of collecting over 500,000 tweets was also taken into account when determining the date to stop collection. Note that the content of the tweets was not examined during the collection period and consequently did not affect the collection dates. Over this period, 591,091 tweets were collected and then subjected to a post-collection filtering process. This was needed to remove a large number of unrelated tweets that were received due to the keyword “abx”, which also referred to an active stock symbol.

Data Analysis and Associated Tools

Hadoop [20] and Apache Hive [21] were used to handle the large number of tweets. Hive is a software tool that allows for query-based processing of large amounts of data through Hadoop and accessible through HiveQL, a language similar to SQL (structured query language). All of the tweets collected were saved on a local, distributed file system in their native JSON format as provided through the Twitter API. From the tweets, a table was constructed in Hive with the columns of the table based on the properties of a tweet, such as the tweet text, user information, location information, and so forth. From there, queries were run pulling selected data from the table using HiveQL. With Hive, the queries were automatically converted to run as MapReduce tasks within Hadoop. With Hadoop and the distributed file system, it was possible to process the roughly 600,000 tweets collected in a timely matter (ie, seconds to minutes depending on the complexity of the query) on a modest Hadoop cluster (eg, 4 computing nodes containing 64 computing cores and 132 GB of RAM).

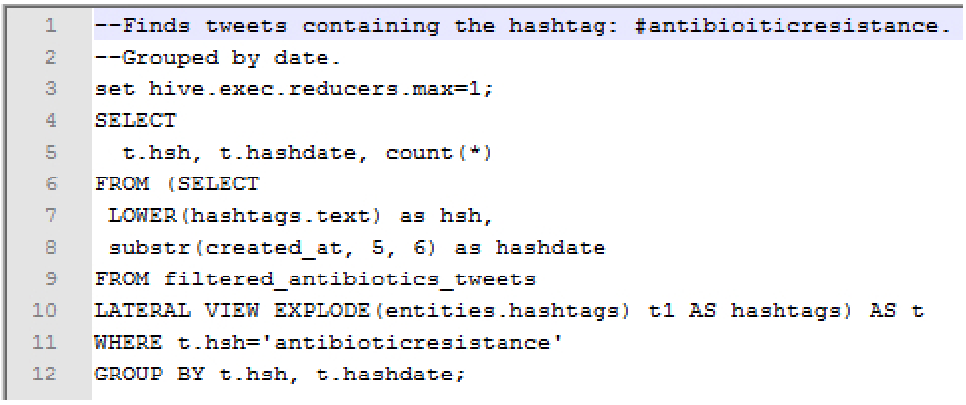
To illustrate the relative ease by which the data can be queried, two sample queries are provided. Figure 1 is an example of a simple HiveQL query. This query finds the number of total tweets within the table of filtered antibiotics tweets. Figure 2 is a more complex query and finds the number of tweets that contain the hashtag “#antibioticresistance”. It then sorts the tweets by day to get a count of the number of tweets that contained the hashtag on each day in our collection period. This query required a nested select statement, which, when converted as a MapReduce, requires two passes through the data. With Hive, large amounts of data can be efficiently queried by anyone familiar with an SQL style database.

Figure 1. HiveQL query to count the total number of tweets.



```
--Gets the total number of tweets.
select count(*) from filtered_antibiotics_tweets;
```

Figure 2. HiveQL query to determine the number of tweets containing "#antibioticresistance", sorted by date.



```
--Finds tweets containing the hashtag: #antibioiticresistance.
--Grouped by date.
set hive.exec.reducers.max=1;
SELECT
  t.hsh, t.hashdate, count(*)
FROM (SELECT
 LOWER(hashtags.text) as hsh,
 substr(created_at, 5, 6) as hashdate
FROM filtered_antibiotics_tweets
LATERAL VIEW EXPLODE(entities.hashtags) t1 AS hashtags) AS t
WHERE t.hsh='antibioticresistance'
GROUP BY t.hsh, t.hashdate;
```

Manual Classification Procedure for Tweets

An explicit aim of this work was to develop a means to classify tweets and thus provide efficient and finer-grain access to the larger pool of tweet data. Manual classification is impractical as well as inefficient for processing and categorizing a large number of tweets. As a result, a deep network was trained and used to classify antibiotic-related tweet data collected from Twitter into 9 classes. The first step in developing a topic-based classifier was to manually label a portion of the tweets. This was done by randomly sampling 1000 tweets and then manually classifying them. The categories used were inspired by Scanfeld et al [11]. The manual classification task was undertaken by 3 individuals and done independently. Fleiss’ kappa was calculated to measure the agreement between the 3 manual coders and the value was .47. After the individual classification, a tweet was added to the labeled set if at least two of manual classifications agreed. This process led to a labeled dataset of 416 tweets with an uneven distribution over the classes. Table 1 lists the categories considered and an example of each from the labeled dataset. An additional evaluation set was also constructed by randomly sampling 300 tweets and performing the same manual classification. This resulted in a set of 246 tweets and the distribution of tweets by category was similar to that of the principle training set.

Table 1. Categories considered for antibiotic-related tweets, quantities of each in the labeled dataset, and examples.

Category	Count	Example
Advertisement	21	#celebrex: Amoxicillian Antibiotic: Generic Amoxil – Antidepressant Celebrex OMITTED_URL
Advice/Information	88	Big pharma not interested in risk and low return of developing new antibiotics
Animals	28	RT @USERID: 80% of all antibiotics in the US are used on farm animals. OMITTED_URLS
General use	38	Bronchitis has got the best of me, Dr's orders to stay home rest and lots of liquids with the antibiotics.
Other	72	How to know you're in the medical field: Seeing the work 'piper' and thinking of piperacillin #nurse
Resistance	132	Antibiotic-resistant superbugs threaten return to 'dark ages'.
Side effects	16	Being put on these antibiotics did way more harm than good.. My stomach has never hurt so bad. Never again. ?
Wanting/Needing	15	Cant wait to get some antibiotics tomorrow from the doctors, can finally get back to normal!!
Misuse	7	Idk if im allowed to mix this Vicodin and this antibiotic. I forgot to ask my dr… oh well.
Total	416	

Construction of a Topic-Based Classifier

To classify all of the tweets, an in-house software package was used to train a deep network. In recent years, deep neural networks have become one of the most popular and powerful techniques in machine learning for classification tasks. Compared with other techniques such as support vector machines and neural networks, deep neural networks typically perform better [22,23] due to their ability to learn high-level abstractions and correlations present in labeled and unlabeled datasets (ie, labeled datasets are those in which each example has a known category/class). In particular, DNs are able to learn patterns present in unlabeled datasets through a layer-by-layer initialization task. This ability to use unlabeled data is particularly advantageous in this setting given the large amount of unlabeled tweets. To date, deep networks and deep learning architectures have been successfully applied in several areas such as speech recognition [24], image classification [23], protein structure prediction [25], and natural language processing [26]. In the health and medical fields, deep networks and deep learning are also gaining traction with applications in computer-aided diagnosis (eg, Alzheimer's Disease) [27], automatic segmentation of diagnostic images (eg, neurological structures in electromagnetic scans) [28,29], lymph node detection from computerized tomography scans [30], and clustering descriptions of adverse drug reactions [31].

In selecting features used to characterize a tweet, a bag-of-words approach was used along with a few global properties such as tweet length or the presence of a URL. Specifically, the text of each tweet was stemmed using a Snowball stemmer included in the Natural Language Toolkit [32] and the presence or absence of several common stems was encoded. To identify the stems used, the most common stems contained in each class (and in general) were determined using HiveQL. Up to the top 50 stems per class and the top 1000 stems overall were used to generate the bag-of-words. Additionally, 10 features were used to encode the length of the tweet (ie, one feature represented bins of 0-9, 10-19, 20-29, etc, respectively) and one binary feature was used to encode the presence of a uniform resource locator (URL) in the tweet. In total, the number of features was 1383.

The full training dataset contained 412 manually labeled examples and 150,000 randomly sampled tweets as unlabeled training examples. Again, a principle advantage of deep networks is the ability to use large amounts of unlabeled data in a semisupervised manner. The overall architecture of the deep network used consisted of 5 layers (ie, an input layer, 3 hidden layers, and an output layer of 11 nodes). The input layer consisted of 1383 features, and the three hidden nodes contained 700, 700, and 300 sigmoid nodes.

Training, Evaluation, and Application of the Classifier

To train and evaluate the deep network model, stratified 10-fold cross validation or so called rotation estimation was used. The labeled dataset was split into 10 sets using stratified sampling techniques, and 9 folds were used for training and the other held out for evaluation. In the training phase, the model was first pre-initialized using the unlabeled data and a layer-by-layer training procedure making use of Restricted Boltzmann Machines [23]. After initialization, the model was fine-tuned using standard back propagation and the labeled data. This process was repeated 10 times in order to make predictions over all the labeled dataset (ie, a prediction for each labeled example was made using 9/10 of the labeled data, which excluded the examples being evaluated). For fine-tuning of the model, a batch size of 10 was used and the refinement took place over 200 epochs. To increase the robustness of the classifier and guard against over-fitting, a dropout procedure was used [33].

To apply a classifier to all collected tweets, a final model was created using all 10 labeled datasets and the aforementioned training procedure (ie, layer-by-layer initialization with 150,000 unlabeled data followed by fine-tuning). Features were generated for all 591,091 tweets and then run through the deep network classifier. The result was a score for each tweet and each category. The total sum for the scores of a tweet across categories was 1.0 (ie, the last layer in the deep network was a multinomial node with 9 classes). For each category, a higher score corresponded to a more confident prediction for that particular category. To classify all tweets, a tweet can be assigned to the highest scoring category. To search the full dataset for tweets pertaining to a particular category, a variable threshold can be chosen (the higher the threshold, the more confident the predictions for the recovered tweets) and all the text for all tweets meeting a particular threshold can be recovered.

Results

Characterizations of Collected Tweets via Hive Queries

To begin to characterize the exchange currently taking place on Twitter with respect to antibiotics, a number of HiveQL queries were performed. First, all collected tweets were collated and counted by date posted to determine a baseline for tweet activity. There was an average of 4654.3 tweets per day. The day with the most activity had 11,365 tweets, and activity usually ranged between 3055 and 6253 (ie, mean +/- standard deviation). Figure 3 illustrates the number of tweets per day during the collection period. There were 8 days with an unusually high number of antibiotic-related tweets (ie, the *Z* score for the number of tweets >2.0). For each of these days, the tweets posted were collected, sorted, and inspected to determine what may have driven the spike in activity. A summary of these dates is contained in Table 2. By examining the most occurring words and retweeted messages by day, it was possible to describe the general cause for the increased activity. On July 2, the day with the most activity, many tweets focused on a speech given by the Prime Minister of the United Kingdom. The second and fifth most active days, September 19 and 18, had tweets related to actions made by US President Obama to battle against antibiotic resistance. On August 19, activity was inflated by an advertisement that was retweeted over 2600 times. In general, it was a news story that led to the increased amounts of tweeting but advertisements did contribute to higher than normal activity on more than one occasion. Note that the general topic for a day was determined by the contents of the tweets on these days of high activity and not by determining a specific source (eg, a particular URL or online news outlet).

Table 2. Dates with unusually high tweet count along with rationale for activity.

Date	Tweet count	General topic	Sample of most frequent words	Sample tweet
July 2	11365	Comments made by British PM regarding resistance	resistance (3411), pm (1610), cameron (1264), warns (1330), dark (965), ages (916)	Antibiotic resistance: Cameron warns of medical ‘dark ages’ URL #health #antibioticresistance #evolution
Sept 19	9489	Executive action from US president on antibiotics	resistance (2342), obama (1394), bacteria (1112), order(946), plan(780), president (747)	RT @PublicHealth: New national strategy, presidential executive order take aim at antibiotic resistance: URL
Aug 19	9188	Advertisement	antibiotics (4718), celphalexin (2657)	RT @...: Buy cephalexin URL
Sept 11	8589	News about a bee-based alternative to antibiotics	antibiotics (5093), bacteria (830), honey (713), alternative (603)	RT @Independent: Bacteria found in honeybee stomachs could be used as alternative to antibiotics
Sept 18	8412	Executive action from US president on antibiotics	resistance (2347), combat (1294), Obama (1070), bacteria (940), strategy (1050)	New Executive Actions to Combat Antibiotic Resistance and Protect Public Health: Today, the Obama admin … URL #obama
July 14	8340	Advertisements	doxycycline (2229), health (1459), treating (1360), bronchitis (1358),	RT @...: Treating Acute Bronchitis and the Use of Antibiotics URL
Sept 30	8147	Report that antibiotics increases risk for obesity	antibiotics (4394), obesity (2827), childhood (1356), study (946)	Antibiotics in infancy may be linked to childhood obesity: study URL
June 25	7871	News breaks that the focus of the Longitude Prize will be antibiotic resistance	antibiotic (2242), resistance (1619), prize (1615), longitudeprize (1089)	RT @longitude_prize: The votes have been counted and the results are in – the challenge of Longitude Prize 2014 will be …antibiotics! #longi…

To further characterize the content of the tweets collected, data on the hashtags used as well as their relative usage and distribution over time was calculated. Table 3 shows the top hashtags by usage. In total, there were roughly 27,458 distinct hashtags used a total of 228,451 times. The vast majority of hashtags (98.94%, 27,166/27,458) were used less than 100 times. Figure 4 depicts the usage of these top hashtags by day over the course of the collection period. As expected, many days that showed spikes in the overall number of antibiotic-related tweets also showed spikes in hashtag usage for #antibiotic or #antibiotics.

Table 3. Usage of the most common hashtags from collected set of antibiotic-related tweets.

Rank	Hashtag	Count (Relative frequency)
1	antibiotics	20706 (9.0%)
2	antibiotic	9329 (4.1%)
3	health	7224 (3.2%)
4	longitudeprize	2693 (1.2%)
5	penicillin	2537 (1.1%)
6	antibioticresistance	2168 (0.9%)
7	news	1942 (0.8%)
8	saveabx	1495 (0.7%)

The final query-based analysis that was performed on the entire collection of tweets was with respect to specific messages and users. In particular, the interest was in what, if any, specific messages were being shared (via retweets) and if the overall exchange taking place in the Twittersphere was being dominated by a set of users. The number of retweeted messages was substantial (27.90% of all tweets were retweets [164,973/591,091]), but most retweets had a limited reach (ie, only 11 tweets had more than 500 retweets). Many of the most retweeted messages were advertisements or concerns about mixing antibiotics and agriculture. None of the tweets went "viral". As for the source of the tweets, the collection included 327,930 different Twitter users. This value was determined through a HiveQL query to collate the tweets by screen name (ie, a unique identifier for a Twitter account) and count the number of tweets per name. Only 0.01% of these users (4255/327,930) contributed 10 or more tweets, yet this small number of users was responsible for 22.68% (134,081/591,091) of all collected tweets.

Figure 3. Number of antibiotic-related tweets collected per day.

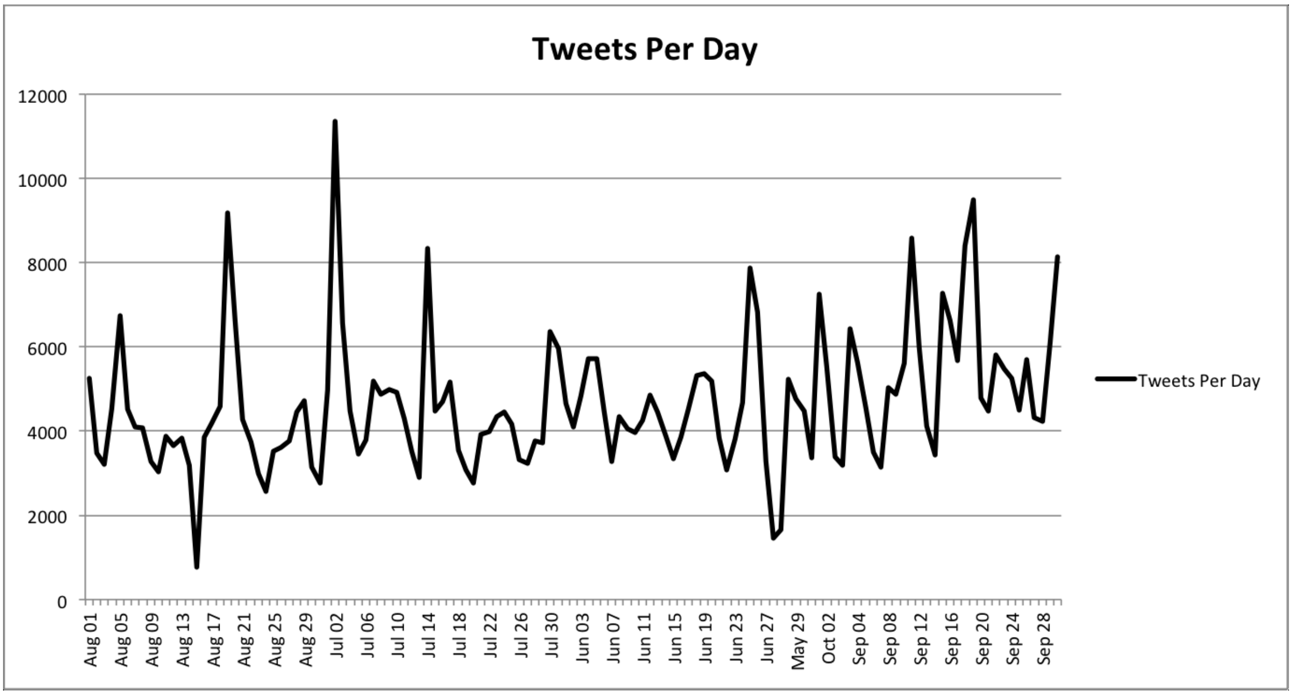
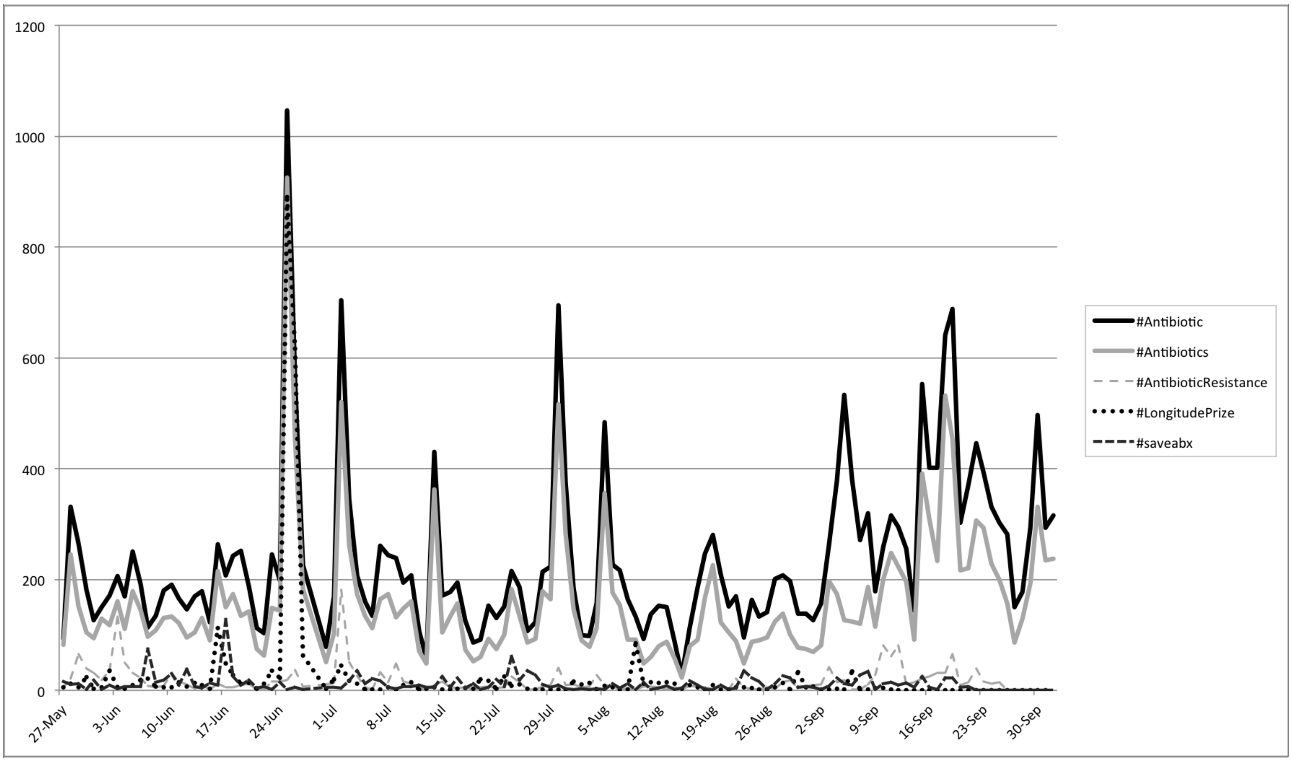


Figure 4. Usage of top overall antibiotic-related hashtags by day.



Characterizations of Tweets Through Classification

By applying the classifier to the entire dataset and assigning each tweet to the highest scoring category, it is possible to estimate the overall frequency of each category. In particular, it was determined that “Advice/Information” and “Other” were the most predicted categories, accounting for 24.47% (144,627/591,091) and 41.00% (242,318/591,091) of all tweets, respectively. “General Use” and “Resistance” each garnered about 12% (70,253/591,091 for General Use and 72,486/591,091 for resistance) of the overall number of tweets with the remaining categories each receiving less than 5%.

Evaluation of Classification

The tweet classifier was evaluated on a per-class and all-classes level using the labeled data and the previously described 10-fold cross validation. On the all-class level, a tweet was assigned to the highest scoring category. The percentage of labeled tweets that were correctly classified was 70.4% (293/416). This value is well above what would be expected by chance by a random predictor (eg, ~11% on a balanced dataset) or by a naïve predictor that always predicted the most common class (ie, a classifier that predicted every example as “resistance” would achieve ~30.5%). On a per-class level, the score for each example was used to evaluate the model with respect to a binary classification task. In this setting, the predicted score for a class (eg, “resistance”) was used in conjunction with a threshold to classify the tweet as “resistance”-related or not “resistance”-related. The recall (ie, percentage of tweets pertaining to a class recovered), precision (ie, percentage of tweets correctly predicted to be in a class), and fall-out (ie, percentage of tweets not pertaining to a class that were incorrectly predicted as pertaining to the class) were calculated for each class at varying thresholds and used to calculate the area under the receiver operating characteristic (ROC) curve (AUC). The AUC is used to characterize the effectiveness of a binary classifier regardless of threshold. Table 4 lists the AUC for each class. Note that the AUC of a random predictor would be 0.5. The break-even point (ie, precision=recall) was also calculated for several classes and shown in Table 4. A break-even point of 67, for example, would indicate that it would be expected to recover 67% of the tweets pertaining to the class “Advice/Information” and of those predicted as “Advice/Information”, 67% of the tweets would be correctly classified. Note that due to the small number of labeled examples for some of the categories (eg, “Misuse”, “Side Effects”), it was difficult determine the break-even point.

Table 4. The AUC and break-even values for per-class evaluation.

Category	AUC	Break-even point (%)
Advertisement	0.78	~45
Advice/Information	0.87	~67
Animals	0.96	~70
General use	0.95	58
Other	0.89	64
Resistance	0.96	92
Side effects	0.92	-
Wanting/Needing	0.88	29
Misuse	0.80	14

When we applied our classifier to the additional evaluation data of 246 tweets, it correctly classified 68.4% of tweets (76/111) when using a threshold of 0.75 (ie, only those predictions with a score of 0.75 or more for a particular class were considered). When considering all 246 tweets, the accuracy on this set was 48.0% (118/246). This value is lower than the 70% obtained through the 10-fold cross validation. On closer inspection of the tweets contained in the evaluation sample, it was determined that several of the tweets could not be confidently placed in one class. When considering the top 2 predicted classes, the accuracy (ie, percentage of tweets whose top 2 predicted categories corresponded to the manual classification) was 64.2% (158/246).

Discussion

Common Source and Content of Collected Tweets

Over the 3-month period during which data were collected, users posted a number of opinions, feelings, and information on many antibiotic-associated topics. Nevertheless, when considering the questions of what particular messages dominated the virtual discussion and who were the primary participants, a few trends are distinguishable. First, a commonly occurring topic during this time period was “resistance”. This is evident by the large number of overall tweets classified as relating to “resistance” (ie, ~10%) and the relative frequency and distribution of the hashtag “#antibioticresistance”. Another trend that shows through is the dominance of several news stories as an origination point for posts. Many of the days with high post counts coincide with national news events related to antibiotics. These tweets often contained a short lead-in and then a URL, indicating that they were generated by embedded Web-links aimed to help users share content through Twitter. This activity, as well as the fact that only a small number of tweets were retweeted more than 500 times and that more than 75% of all tweets collected came from users who generated fewer than 10 tweets from those collected, indicates that this discussion is not being driven by a few individuals but is more organic in nature. It is also clear that specific topics (eg, an announcement from a national leader) have a relatively short duration and with each spike in activity, the amplified number of posts greatly decreases just as quickly as it increased; most spikes lasted only one day.

Efficient Mining of Tweets

Given the performance of the constructed classifier (ie, 70% percent of the multi-class predictions were correct and reasonable values for the AUC values for the binary classification tasks), it is possible to effectively access larger amounts of tweets in a manageable fashion. In particular, it is possible to more efficiently peruse tweets by category by varying the decision threshold. Taking advantage of the classic trade-off between precision and recall, one can retrieve a small number of confident predictions or sift through a larger number of less confident predictions. Table 5 contains a sample of tweets, their predicted category, and score.

Table 5. Sample tweets that were discovered through the trained classifier: tweet text, predicted category, and score.

Category (score)	Tweet text
Need (0.85)	I’ve never been so excited to go to the doctor to get antibiotics
Need (0.83)	I hate doctors so much, I shouldn’t have to demand antibiotics ffs
Need (0.70)	Any of my local friends have antibiotics laying around they haven’t finished? Don’t have ins & can’t afford to go to the docr
Misuse (0.85)	Ive had stepthroat for a month and I will throw the biggest …… fit this doctors office has ever seen if they don’t give me antibiotics
Misuse (0.77)	I’m on antibiotics for this sinus infection which means no drinking. Guess what I’m doing?
Side Effects (0.70)	Common antibiotic may increase heart death risk URL
Side Effects (0.70)	I’ll never take an antibiotic before I go to sleep again, my body was itching all ……. night?
Side Effects (0.85)	@.... I am now. We found out that the antibiotic I was on causes sever motion sickness. Lol. I just can’t drive while I’m taking them.
Side Effects (0.95)	So sick I think I’ve turned green, never take antibiotics on an empty stomach. Learned my lesson. ??
Advice/Information (0.70)	Bacteria can evolve a biological timer to survive antibiotic treatments – Medical News Today URL
Advice/Information (0.80)	UK says …… recalls batches of Indian-made antibiotic ????
Advice/Information (0.91)	U.S. Congress urged to pass bill to speed development of antibiotics #Health care

To further investigate the use of the classifier, it was used to identify and visually inspect tweets classified as pertaining to “Misuse”. Tweets classified as pertaining to “Misuse” were collected using thresholds of 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8 which yielded 2623, 1152, 510, 205, 42, and 8 tweets respectively. Given the relatively short nature of the tweets, all tweets with a score of over 0.3 were read. In examining these tweets, a number of user messages mentioned mixing alcohol and antibiotics, missing and/or recuperating missed doses, taking antibiotics without a meal, and taking and/or looking for antibiotics for influenza.

Comparison With Existing Tools and Approaches for Mining the Twittersphere

Given the value of user-generated content, a number of enterprises have developed proprietary tools to collect, process, and provide access to Twitter data (eg, Topsy and Talkwalker). Access is typically provided through a Web interface or API and may come at a cost. A comprehensive comparison between these commercial options and the methods developed in this work is difficult due to limited access and the speed at which these tools evolve to meet market needs. Nevertheless, there are clear advantages and disadvantages to each approach that can be discussed in general terms. Commercial tools often provide access to historical Twitter data and can be used to perform retrospective studies. The Web interfaces provided make some tweets and data derived from the tweets easy to access and visualize in a manner that does not involve extensive technical knowledge (eg, to manage or query the data). The drawbacks to these tools include the monetary cost of gaining access to the tool and/or data and restricted access to the data. The open-source solution described here does provide finer-grain access to the data since all data pertaining to a tweet are stored locally. There is also no cost associated with software tools used as they are open source and liberally licensed. The principle drawback is the added technical knowledge needed to manage the data and use the tools.

Apart from the cost, the level of access to the data is perhaps the most important distinction between the two approaches. As an example, consider some of the data mining and analysis tasks performed in this study. These included counting the number of relevant tweets per day, determining the source of spikes in related traffic, and determining common contributors to the virtual discussion. These tasks could have readily been accomplished with commercial tools. By having direct access to the data and software to perform queries, it was possible to also determine what specific words and word pairings were also common in tweets relating to antibiotics. This information was used to build the classifier and also identify topics of discussion on days of peak activity. It was also possible to drill down on hashtags and hashtag usage among tweets about antibiotics. Thus, more control and more complex queries could be executed.

With regards to findings, this work confirms and complements existing approaches for mining content related to antibiotics from the Twittersphere. In a study of Twitter and antibiotics, Scanfeld et al derived categories and manually classified 1000 tweets [11]. The relative frequencies of tweets from several categories (eg, misuse, needing/wanting) were similar to those obtained in this study with the exception of resistance, which had increased since 2010. Furthermore, Scanfeld et al used several keyword pairs (eg, “extra” and “antibiotics”) to search for specific instances or details of misuse. Using the tools and approaches presented in this work, it is possible to extend the discovery process by identifying common keyword pairs (ie, query to determine what words commonly co-occur, possibly identifying new types of misuse) or searching through tweets by predicted category (ie, the tweet may lack the suspected keywords associated with misuse but be predicted as pertaining to misuse through the learned patterns). Dyar et al used a commercial tool (ie, Topsy) to study spikes in Twitter activity surrounding tweets that mentioned “antibiotics” [9]. These spikes were found to be short-lived and driven by media coverage of governmental action. This phenomenon was also seen in the aggregate collection of tweets (ie, not only those with the keyword “antibiotic” but all tweets collected for this study) and also present in hashtag usage.

Limitations

This study, and indeed analyses of large amounts of Twitter data, is not without its limitations. First, the data are very noisy. In this study for example, additional post filtering was needed to remove tweets that were related to the stock ticker “ABX”. Collecting tweets by keywords is difficult as miscellaneous or tangential data can easily be selected (eg, a joke mentioning antibiotics). These extraneous tweets can easily skew the dataset if they become popular and are retweeted several times. Additionally, the tweets are by their nature quite short and full of abbreviations, and this can make their interpretation ambiguous and difficult to interpret without the benefit of additional text to give context. In some cases, qualitative judgments had to be made as to which category a tweet belongs since the categories chosen were not mutually exclusive. This was a limitation of the machine learning approach we employed. Future works could break all needed categorizations into mutually exclusive sets and train a classifier for each set.

Another limiting factor to this study and approach is its specialization. While the approach is general enough to be applied to a number of topics and domains in the Twittersphere, a new model would need to be trained and a new feature set selected. This is because the data collected would have to be analyzed to determine what words would be useful for the bag-of-words characterization.

Conclusions

This study developed and implemented means to characterize the larger discussion taking place on Twitter with respect to a particular health-related topic (eg, antibiotics). Using tools such as Hive, Flume, Hadoop, and machine learning techniques, it is possible to collect and query large amounts of Twitter data to determine what words, phrases, or contributors were dominating the online discussion. It is also possible to identify and characterize of periods of high activity. In this study in particular, it was determined that several national actions with respect to antibiotics led to several spikes in activity. Furthermore, using newer machine learning techniques and a limited number of manually labeled tweets, the entire body of collected tweets can be classified to indicate what topics are driving the discussion. The classifier can also be used to efficiently explore collected tweets by category and search for messages of interest or exemplary content.

Authors' Contributions

JE and RK conceived of the study, and JE, RK, and SK developed and implemented the algorithms. All authors analyzed the data, wrote, edited, and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Toffler A. The third wave: the classic study of tomorrow. New York, NY: Bantam Books; 1984.
2. Social Networking Fact Sheet. 2014. URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [accessed 2014-12-09] [WebCite Cache ID 6UwIGT5fI]
3. Smith B. The Beginner’s Guide to Twitter. 2012. URL: http://mashable.com/2012/06/05/twitter-for-beginners/ [accessed 2014-12-19] [WebCite Cache ID 6UwH6K5ro]
4. Beevolve. An Exhaustive Study of Twitter Users Across the World. 2012. URL: http://www.beevolve.com/twitter-statistics/ [accessed 2014-12-19] [WebCite Cache ID 6UwI6V8rn]
5. Jenders M, Kasneci G, Naumann F. Analyzing and predicting viral tweets. In: Proceedings of the 22nd International Conference on World Wide Web Companion.: International World Wide Web Conferences Steering Committee; 2013 Presented at: 22nd Internation World Wide Web Conference; May 13-17, 2013; Rio de Janeiro, Brazil p. 657-664.
6. Gabarron E, Serrano JA, Wynn R, Lau AYS. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
7. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
8. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
9. Dyar OJ, Castro-Sánchez E, Holmes AH. What makes people talk about antibiotics on social media? A retrospective analysis of Twitter use. J Antimicrob Chemother 2014 Sep;69(9):2568-2572 [FREE Full text] [doi: 10.1093/jac/dku165] [Medline: 24862092]
10. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. Am J Infect Control 2013 Jun;41(6):568-570. [doi: 10.1016/j.ajic.2012.10.016] [Medline: 23726548]

11. 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]
12. Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. Dermatol Clin 2009 Apr;27(2):133-6, vi. [doi: 10.1016/j.det.2008.11.010] [Medline: 19254656]
13. Collier N, Son NT, Nguyen NM. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011;2 Suppl 5:S9 [FREE Full text] [doi: 10.1186/2041-1480-2-S5-S9] [Medline: 22166368]
14. Hill S, Merchant R, Ungar L. Lessons learned about public health from online crowd surveillance. Big Data 2013 Sep 10;1(3):160-167 [FREE Full text] [doi: 10.1089/big.2013.0020] [Medline: 25045598]
15. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
16. Lampos V, Bie T, Cristianini N. Flu Detector - Tracking Epidemics on Twitter. In: Balcázar JL, Bonchi F, Gionis A, Sebag M, editors. Machine Learning and Knowledge Discovery in Databases. Berlin Heidelberg: Springer; 2010:599-602.
17. Bush K, Courvalin P, Dantas G, Davies J, Eisenstein B, Huovinen P, et al. Tackling antibiotic resistance. Nat Rev Microbiol 2011 Dec;9(12):894-896 [FREE Full text] [doi: 10.1038/nrmicro2693] [Medline: 22048738]
18. Smith R, Coast J. The true cost of antimicrobial resistance. BMJ 2013;346:f1493. [Medline: 23479660]
19. Apache Software Foundation. Welcome to Apache Flume. 2014. URL: http://flume.apache.org/ [accessed 2015-01-05] [WebCite Cache ID 6VMYYuCq6]
20. Apache Software Foundation. Welcome to Apache Hadoop. 2014. URL: http://hadoop.apache.org/ [accessed 2015-01-10] [WebCite Cache ID 6VTmOWTYR]
21. Thusoo A, Sarma JS, Jain N, Shao Z, Chakka P, Anthony S, et al. Hive. In: Proc. VLDB Endow. 2009 Aug 01 Presented at: Proceedings of the VLDB Endowment; August 24-28, 2009; Lyon, France p. 1626-1629. [doi: 10.14778/1687553.1687609]
22. Hinton GE. Training products of experts by minimizing contrastive divergence. Neural Comput 2002 Aug;14(8):1771-1800. [doi: 10.1162/089976602760128018] [Medline: 12180402]
23. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006 Jul 28;313(5786):504-507 [FREE Full text] [doi: 10.1126/science.1127647] [Medline: 16873662]
24. Mohamed A, Dahl GE, Hinton G. Acoustic Modeling Using Deep Belief Networks. IEEE Trans Audio Speech Lang Process 2012 Jan;20(1):14-22. [doi: 10.1109/TASL.2011.2109382]
25. Eickholt J, Cheng J. DNdisorder: predicting protein disorder using boosting and deep networks. BMC Bioinformatics 2013;14:88 [FREE Full text] [doi: 10.1186/1471-2105-14-88] [Medline: 23497251]
26. Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. 2008 Presented at: 25th International Conference on Machine learning; July 5-9, 2008; Helsinki, Finland. [doi: 10.1145/1390156.1390177]
27. Suk H, Shen D. Deep Learning-Based Feature Representation for AD/MCI Classification. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2013. Berlin Heidelberg: Springer; 2013:583-590.
28. Ciresan D, Giusti A, Gambardella L, Schmidhuber J. Deep Neural Networks Segment Neuronal Membranes in Electron Microsopy Images. In: Pereira F, Burges C, Bottou L, editors. Advances in Neural Information Processing Systems 25. New York: Curran Associates, Inc; 2012:2843-2851.
29. Liao S, Gao Y, Oto A, Shen D. Representation Learning: A Unified Deep Learning Framework for Automatic Prostate MR Segmentation. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2013. Berlin Heidelberg: Springer; 2013:254-261.
30. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, et al. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. Med Image Comput Comput Assist Interv 2014;17(Pt 1):520-527 [FREE Full text] [Medline: 25333158]
31. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
32. Bird S, Klein E, Loper E. Natural language processing with Python. Sebastopol, Calif: O'Reilly; 2009.
33. Dahl G, Sainath T, Hinton G. Improving Deep Neural Networks for LVCSR using Rectified Linear UnitsDropout. In: Proceedings of ICASSP. 2013 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; May 26-31, 2013; Vancouver, Canada p. 8609-8613. [doi: 10.1109/ICASSP.2013.6639346]

Abbreviations

API: application programming interface
AUC: area under the ROC curve
ROC: receiver operating characteristic
SQL: structured query language
URL: uniform resource locator

Edited by G Eysenbach; submitted 10.01.15; peer-reviewed by E Castro-Sánchez, N Zhang; comments to author 26.03.15; revised version received 25.04.15; accepted 09.05.15; published 19.06.15.

Please cite as:
Kendra RL, Karki S, Eickholt JL, Gandy L
Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture?
J Med Internet Res 2015;17(6):e154
URL: http://www.jmir.org/2015/6/e154/
doi:10.2196/jmir.4220
PMID:26091775

Original Paper

The Impact of Internet Health Information on Patient Compliance: A Research Model and an Empirical Study

John Laugesen[1], MBA, PhD; Khaled Hassanein[2], MBA, PhD; Yufei Yuan[2], PhD

[1]Pilon School of Business, Sheridan College, Mississauga, ON, Canada

[2]DeGroote School of Business, Information Systems Area, McMaster University, Hamilton, ON, Canada

Corresponding Author:
John Laugesen, MBA, PhD
Pilon School of Business
Sheridan College
4180 Duke of York Blvd.
Mississauga, ON, L5B0G5
Canada
Phone: 1 905 845 9430 ext 33519
Fax: 1 905 615 7916
Email: john.laugesen@sheridancollege.ca

Abstract

Background: Patients have been increasingly seeking and using Internet health information to become more active in managing their own health in a partnership with their physicians. This trend has both positive and negative effects on the interactions between patients and their physicians. Therefore, it is important to understand the impact that the increasing use of Internet health information has on the patient-physician relationship and patients' compliance with their treatment regimens.

Objective: This study examines the impact of patients' use of Internet health information on various elements that characterize the interactions between a patient and her/his physician through a theoretical model based on principal-agent theory and the information asymmetry perspective.

Methods: A survey-based study consisting of 225 participants was used to validate a model through various statistical techniques. A full assessment of the measurement model and structural model was completed in addition to relevant post hoc analyses.

Results: This research revealed that both patient-physician concordance and perceived information asymmetry have significant effects on patient compliance, with patient-physician concordance exhibiting a considerably stronger relationship. Additionally, both physician quality and Internet health information quality have significant effects on patient-physician concordance, with physician quality exhibiting a much stronger relationship. Finally, only physician quality was found to have a significant impact on perceived information asymmetry, whereas Internet health information quality had no impact on perceived information asymmetry.

Conclusions: Overall, this study found that physicians can relax regarding their fears concerning patient use of Internet health information because physician quality has the greatest impact on patients and their physician coming to an agreement on their medical situation and recommended treatment regimen as well as patient's compliance with their physician's advice when compared to the impact that Internet health information quality has on these same variables. The findings also indicate that agreement between the patient and physician on the medical situation and treatment is much more important to compliance than the perceived information gap between the patient and physician (ie, the physician having a higher level of information in comparison to the patient). In addition, the level of agreement between a patient and their physician regarding the medical situation is more reliant on the perceived quality of their physician than on the perceived quality of Internet health information used. This research found that only the perceived quality of the physician has a significant relationship with the perceived information gap between the patient and their physician and the quality of the Internet health information has no relationship with this perceived information gap.

(J Med Internet Res 2015;17(6):e143) doi:10.2196/jmir.4333

KEYWORDS

Internet health information; information asymmetry; patient compliance; patient-physician concordance; physician quality; consumer health

Introduction

Background

The patient-physician relationship has been noted to be second only to family relationships in terms of importance. It is viewed as extremely or very important by 67%, exceeding relationships with spiritual advisors, pharmacists, coworkers, and financial advisors [1]. The benefits of this relationship and, in turn, physician advice can only be achieved if patients follow the treatment regimens relatively closely [2]. This concept, known as compliance, is important to examine because prior studies have shown that noncompliance rates can be as high as 80% and noncompliance "creates a number of serious problems: (1) for the individual in reduced quality and quantity of life, lower income due to inability to work, and higher medical costs; (2) for society, which pays higher insurance and medical costs because noncompliant patients often require more expensive and invasive health care; and (3) for corporations because they experience lower productivity from sick and absent workers" [3]. One suggested way to improve compliance is through improved patient-physician communication [3-7], collaboration and participative decision making [2,3], and better concordance between patients and physicians with respect to medical diagnoses and treatment regimens [3,4,8-11].

Patient use of the Internet in searching for and gathering health information is growing and has now become somewhat commonplace. The Pew Internet & American Life Project reports 80% of American Internet users have searched for some type of Internet health information and millions of people search for Internet health information on a typical day [12]. However, as the use of the Internet as a source for health-related information becomes more commonplace, relations between the patient and physician can become strained [1,13-15] and this strained relationship due to Internet health information could potentially impact physician-patient concordance and patient compliance. Based on the preceding discussion, it is important to understand the true impact of patients' increasing utilization of Internet health information on the patient-physician relationship and patients' compliance with their treatment regimens through theoretically rigorous and empirically validated studies. However, to the best of our knowledge, no known models have been developed and empirically validated that examine patients' compliance and concordance with their physician where Internet health information is widely available and used by patients. In fact, little research (quantitative or qualitative) on the impact of Internet health information on the patient-physician relationship and compliance has been completed. Previous studies have looked at factors that influence compliance [5,6]; however, none have looked at how patients' perceptions of Internet health information and their physician quality impacts information asymmetry, concordance with their physician, and compliance with physician advice. This new theoretical lens is important because traditionally the patient-physician relationship was subject to the influences of information asymmetry (ie, physicians having significantly more and better health-related information), but this influence may be reduced by patient use of Internet health information.

Compliance

The term compliance is the most common way to describe a patient following his/her physician's treatment instructions [16]. Numerous previous research studies and reports have identified the issue of noncompliance and the importance of compliance [4,9,16] and the global problems that noncompliance are causing for health care systems [16]. Compliance is very important to study because previous research has shown that patients who are compliant exhibit better health outcomes than those who are noncompliant [17]. Noncompliance rates range from 25% to 80% [3-5,10,18] and noncompliance is estimated in the United States to cause 125,000 deaths, 19% of all hospital admissions, and more than US $100 billion in additional health care costs per year [3,6]. Noncompliance is linked to substantial worsening of disease and death [2] and is also reported to waste resources, cause preventable morbidity and mortality, and result in the loss of health care funds and productivity [19]. Given the increasing incidence of chronic illness [20], the study of compliance becomes even more important as treatment becomes more reliant on patient self-management [19]. Therefore, understanding and improving compliance can lead to better patient health outcomes [19] and lower costs of health care.

Principal-Agent Theory and Perceived Information Asymmetry

Principal-agent theory seeks to understand and explain the association between self-interested parties who have potentially differing goals in situations where there is an imbalance of information between the parties [21]. In this theory, the principal "hires" the agent who performs some task on behalf of the principal because the principal typically has less information than the agent does (ie, information asymmetry). This theory has been applied in areas such as economics, accounting, finance, marketing, political science, organizational behavior, sociology, and buyer-seller relationships [21,22]. Previous research has applied principal-agent theory to the relationship between physicians (agents) and patients (principals) [23,24]. It is our contention that principal-agent theory applies to the patient-physician relationship (specifically in the context of Internet health information). There is a recognized asymmetry of information in the patient-physician relationship [23]. This perceived imbalance of knowledge and power has historically placed patients in a vulnerable position [25] with the flow of information between patient and physician being tenuous because of the knowledge/power gap [25]. However, the past decade (ie, the Internet health information years) has fostered a challenge to this asymmetrical model of interaction where the physician held the majority of the information and power [26]. Historically, physicians typically provided information to patients to ensure patient acceptance of the physician's diagnoses and treatments [26]; however, this is changing given the quantity and quality of Internet health information that is available to patients.

Internet Health Information

Patients receive medical information from physicians, but they also obtain medical information from a variety of other sources, such as friends, news, books, and now more frequently and conveniently, from the Internet. It is logical to assume that a patient's level of knowledge/information vis-à-vis their physician is a function of the quality of their own information (which is now mainly based on information gathered from the Internet) and the quality of their physician (an element that includes physician knowledge). Therefore, this study incorporates both Internet health information quality and physician quality as key elements in both the patient's assessment of their relative knowledge level and in the concordance between the patient and the physician. From a patient perspective, the effects of Internet health information have been shown to be both positive and negative. From a positive standpoint, the most commonly cited effect is patient empowerment, with Broom [13] indicating Internet health information can provide a sense of empowerment, purpose, and control, and patient empowerment can lead to better treatment and higher levels of patient satisfaction. Another important patient benefit from Internet health information is that it allows patient control over their rate of learning, thus reducing information overload often experienced in a physician's office [27]. Other positive effects of Internet health information are enhanced patient confidence in dealing with physicians, better health choices and decision making, improved understanding of health conditions, and improved communication with physicians [28,29]. Improved information access through Internet health information, given the information is clinically relevant, accurate, and validated, has been linked to improved outcomes [30]. From a negative standpoint, the major issue regarding Internet health information is patient concern about physician disapproval. Patients worry that this disapproval can lead to physician hostility, irritation, and lower quality of care resulting in patient anxiety, confusion, and frustration [13].

Physicians generally accept that the Internet may lead to patients becoming better informed; however, 40% of physicians believe that this may damage the patient-physician relationship [14]. Physicians worry that the use of the Internet may lead to patient confusion, unrealistic expectations, and potential increases in litigation [14]. In addition, physicians are concerned that the patient-physician relationship can be affected when they must explain to their patients that the information they have gathered from the Internet is not accurate or complete [1] and, therefore, potentially irrelevant. Physicians are concerned about potential Internet health misinformation and, more importantly, patient misinterpretation of the Internet health information [15]. However, despite this, 90% of physicians surveyed feel that providing a greater quantity of better medical information to patients is beneficial [14]. Although physician information is the most trusted source and patients report that their preference is to go to their physician first to get information, only 10.9% of patients actually go to their physician first, whereas 48.6% go online first [31], most likely because of the accessibility, convenience, and immediacy of the information.

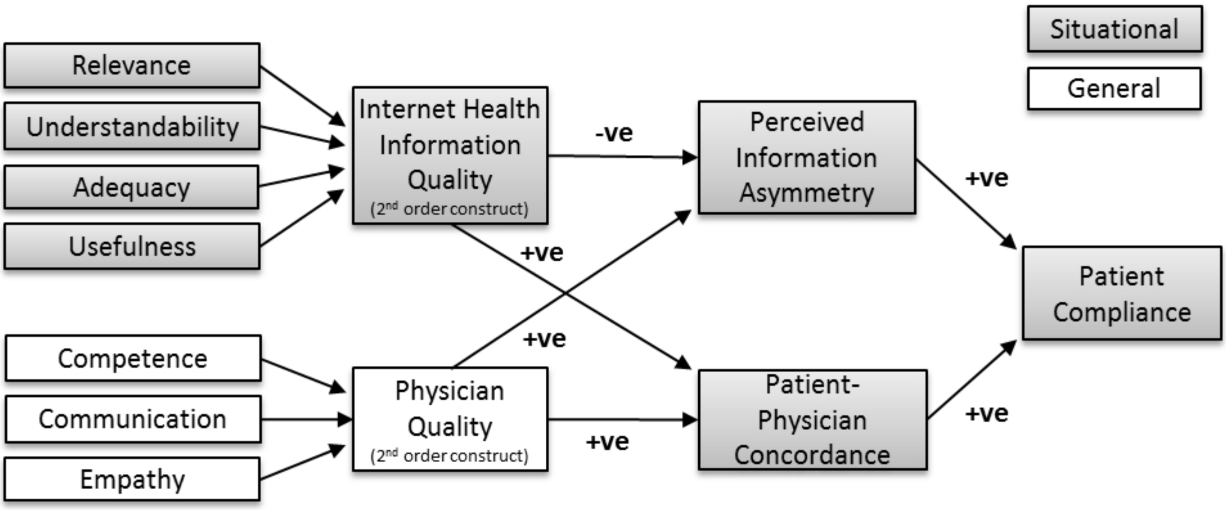
Physician Quality

Although the information a patient holds (much of which is gathered through the Internet) forms one side of the information equation, the other important element a patient considers when determining their relative (to their physician's) level of knowledge would be their perception of their physician's competence/knowledgeability and their physician's communication capabilities (because their perception of the physician's knowledge can only be derived based on communications with their physician). Therefore, both physician competence/knowledgeability and communication capabilities are essential components of physician quality [32] along with physician empathy. From a health information perspective, patients report they value their physician's knowledge more than any other health care information source, including Internet health information [27]. Therefore, it is logical to believe that physician quality plays a major role in a patient's thought process when determining information asymmetry relative to their physician and concordance with their physician's recommendation, which are 2 major elements of this research study.

Research Model and Hypotheses

We propose the theoretical model shown in Figure 1 to examine the impact of both patients' use of Internet health information combined with physician quality–related factors on patients' compliance with physician's advice in the presence of Internet health information. Although other factors may be involved in compliance, we focus on factors related to Internet health information use and physician quality and their impact on information asymmetry and concordance as antecedents to patient compliance. The majority of the constructs in the research model are assessed on a situational basis in that the survey items used referred to a specific significant health situation. However, it was not possible for one construct (ie, physician quality) to be assessed on a single situational basis (because it would be difficult for respondents to separate their general trust in the physician from the situational trust formed regarding the significant health situation) and, therefore, physician quality was assessed on an overall basis.

Figure 1. Research model.



Patient Compliance

Patient compliance involves the extent to which the behavior of a patient matches the physician's recommendations [16]. There are a number of previous studies that show compliance has relationships with a number of antecedent concepts, including the physician-patient relationship and interactions, patient-physician communication, patient knowledge and attitudes and shared decision making [3,4,6,9], concordant patient-physician relationship [4], confidence in their physician's ability to help them and satisfaction with the concern shown by physicians [10], participative decision making [3], and physician competence [33]. Previous studies have shown that patients' self-reports of compliance generally correspond with objective measures of compliance [16]; therefore, we used self-reported measures of compliance in this research.

Perceived Information Asymmetry

For the purposes of this study, we have adapted the description of perceived information asymmetry put forth by Pavlou et al [21] and define perceived information asymmetry as the patients' perception that the physician has a greater quantity and/or quality of information compared to themselves [21]. There are no known previous studies with direct theoretical support for the hypothesized relation between perceived information asymmetry and compliance. However, given that perceived information asymmetry in this research involves the information/knowledge gap between the patient and the physician, it is logical to assume that higher levels of physician information/knowledge are directly related to higher levels of information asymmetry and that patients who feel that their physician has more and/or better health-related information than they do will be more likely to comply with the physician's recommended diagnosis and treatment regimen. Previous studies have shown a relationship between physician information/knowledge levels and compliance or, conversely, a relationship between physician knowledge deficiencies and a lack of compliance [34-36]. Solem et al [36] indicated that forming an understanding of the knowledge gaps between a physician and patient may be critical to improving patient compliance. Therefore, we hypothesized that perceived information asymmetry will have a positive impact on patient compliance.

Patient-Physician Concordance

Patient-physician concordance involves agreement between a patient and their physician regarding the medical problem and treatment regimen [8]. In essence, concordance encompasses the agreement regarding the treatment whereas compliance involves whether or not the patient complied with the treatment, regardless of whether or not there was concordance. Prior studies support the relationship between the concept of patient-physician concordance and compliance [3,4,8-11] with Kerse et al [8] finding that patients reporting high levels of concordance were 33% more likely to be compliant in taking medications prescribed during the consultation. Another study found that better communication and concordance between a patient and their physician can lead to improved compliance [4]. A study of Korean patients found that patient-physician partnership, a concept very similar to concordance, had a very strong relationship with compliance, which was attributed to the Korean patient's desire for an egalitarian relationship with their physician [7]. Finally, Wroth and Pathman [10] found that patient-physician concordance is associated with medication compliance. Therefore, we hypothesized that patient-physician concordance will have a positive impact on patient compliance.

Internet Health Information Quality

Internet health information quality is a second-order construct comprised of perceptions of relevance (ie, clearness, relevance, and goodness), understandability (ie, clarity, understandability, and readability), adequacy (ie, sufficiency, completeness, necessity), and usefulness of the information on a health infomediary's website [37]. Support for the relationship between Internet health information quality and information asymmetry is provided via a previous study that examined website informativeness and perceived information asymmetry and found a significant relationship between these variables. A number of studies that include information asymmetry in the accounting and financial domains suggest that better information quality is related to lower levels of information asymmetry [38,39]. Additional studies in economics suggest that the provision of better information is a potential solution to asymmetry problems (ie, to consumers reducing the level of asymmetry) [40] and that the dissemination of information (eg, through educational programs or labeling) aims at reducing issues resulting from information asymmetry [40-43]. Finally, a previous study in the context of digital information and food traceability found a significant negative relationship between informativeness (ie, the extent to which the Internet provides participants with helpful information) and information asymmetry [44]. Therefore, we hypothesized that Internet health information quality will have a negative impact on perceived information asymmetry.

Although there are no known studies that specifically examine the relationship between Internet health information quality and patient-physician concordance, it is logical to believe that patients who have accessed high-quality information regarding their medical situation will be able to have more meaningful communication with their physician, which in turn should lead to a higher level of agreement between the patient and physician regarding the medical issue and treatment. This logic is supported through studies that report that better-informed patients can lead to enhanced communication between patients and physicians [45], and that encouraging enhanced 2-way patient-physician communication may have a positive influence on concordance [46]. Finally, a study that examined the effects of providing medical information to patients found that this led to decisions that were based on both the knowledge of the physician and the patient's preferences, which is very similar to the notion of concordance [47]. Therefore, we hypothesized that Internet health information quality will have a positive impact on patient-physician concordance.

Physician Quality

In this research, physician quality is a second-order construct comprised of perceptions of competence, empathy, and communication [32]. This representation of physician quality encompasses both professional core physician qualities along with important personal qualities of the physician [32]. Previous studies have shown support for the relationship between the individual elements of physician quality and concordance between patients and their physicians. Janz et al [48] indicate that lower levels of concordance between patient and physician regarding treatment decisions show the need for better communication between patient and clinician. Riekert et al [49] found that poor patient-physician communication and information sharing are contributing factors of nonconcordance. A study by Vermeir et al [9] indicates that physician empathy may be an essential element of patient-physician concordance. Given that individual elements of physician quality are related to concordance, it is logical to assume that the higher the patient's perception of their physician's overall quality, the more likely they are to come to an agreement about the significant health situation and recommended course of action. Therefore, we hypothesized that physician quality will have a positive impact on patient-physician concordance.

Although there is no known prior research that specifically examines the relationship between physician quality and information asymmetry, the support cited previously for the hypothesized relationship between Internet health information quality and information asymmetry also plays a role in the physician quality and information asymmetry relationship. The support noted previously shows that different amounts and quality of information on each side of the agency relationship affect information asymmetry. Given that more and better patient information should reduce information asymmetry, it is logical to assume that more and better physician information would increase information asymmetry (because this construct is the gap between the patient's and physician's information). Therefore, we hypothesized that physician quality will have a positive impact on perceived information asymmetry.

Methods

Instrument Development

This research made use of previously validated instruments to measure the constructs in the model, as per Boudreau et al [50]. Unless otherwise noted (see Multimedia Appendix 1), all items were measured using a 7-point Likert scale with ranges from strongly agree to strongly disagree. Compliance was measured using a 5-item scale adapted from Hausman [6]. Respondents were allowed to indicate "not applicable" to individual compliance questions because not all patients would be required to follow each and every one of the directions noted in the survey items (eg, some patients would not be required to take medications as part of the treatment regimen; therefore, these participants would need the ability to indicate not applicable to this question). Patient-physician concordance was measured using a 5-item scale adapted from Kerse et al [8], which was designed to assess agreement between physician and patient. Perceived information asymmetry was measured using a 4-item scale developed based on items from Pavlou et al [21] and Dunk [51]. These items were adapted to specifically address the context of the information gap between the physician and the patient regarding the significant health situation.

Physician quality was developed as a second-order construct comprised of competence, empathy, and communication. These elements of physician quality are based on Jayanti and Whipple [32] that describe physician quality as a function of listening skills (ie, communication), competence/knowledgeability, and empathy. For this second-order construct, there were no known scales that specifically addressed competence. Therefore, we adapted the validated McKnight et al [52] competence scale (which addressed competence in the legal profession) to a physician competence context. Given both contexts (ie, legal and medical) are professional ones, the McKnight et al [52] scale was deemed to be the most applicable for this research. The empathy items were adapted from Kim et al [7], a study that included an examination of the relationship between physician empathy and patient compliance. The communication items were adapted from Hausman [6], a study that examined physician communication and its relationships with both patient participation in the decision-making process (similar to concordance) and patient compliance with physician advice.

Finally, Internet health information quality was developed as a second-order construct comprised of adequacy, understandability, usefulness, and relevance. These 4 areas were each measured using 4-item scales adapted from Zahedi and Song [37]. The validated scales contained in Zahedi and Song [37] were highly applicable to this study because they specifically measured trust and quality in an online health information provider context. The final set of survey items is included in Multimedia Appendix 1.

Second-Order Constructs

Second-order constructs are used in this research to model (1) Internet health information quality because this variable is comprised of the first-order quality factors of usefulness, adequacy, relevance, and understandability [37] and (2) physician quality because this variable is comprised of competence, empathy, and communication [32]. A full statistical analysis of the second-order constructs is provided in Multimedia Appendix 2. As per Chin [53]: “Higher order latent variables are often useful if a researcher wishes to model a level of abstraction higher than those first-order constructs used in a basic [covariance-based structural equation modeling] CBSEM and [partial least squares] PLS model.” Both Internet health information quality and physician quality are structured as second-order factor models, with the direction of the relationship flowing from the first-order constructs to the second-order construct (see Multimedia Appendix 2). This model structure is characterized as reflective first-order, formative second-order as per Jarvis [54], which is the most common structure in Information Systems literature [55]. Careful consideration was given when determining to model both Internet health information quality and physician quality as second-order constructs, specifically in that the first-order factors were conceptually related to the other factors in the model and that the second-order factor fully mediated the relationships of the first-order factors in the theoretical model [53]. A number of previous Information Systems studies have made use of second-order constructs [56-59].

From a statistical perspective, this research used the indicator reuse technique proposed by Wold [60] as described in Ringle et al [55]. Specifically, “When using the PLS-SEM method for model estimation, all latent variables—which includes higher order components—must have a measurement model with at least one indicator...This approach works best when all lower order components have the same number of indicators. Otherwise, the interpretation of the relationships between the lower and the higher order components must account for the bias of unequal numbers of indicators in the lower order components” [55]. This research ensured these requirements were met with each latent variable in the model having at least one indicator and all lower order components (ie, first-order constructs) containing the same number of indicators.

Analysis Tool Selection

This research used the second-generation statistical technique of structural equation modeling (SEM), specifically PLS implemented via Smart-PLS software version 2.0.M3. As described by Gefen et al [61], “the intricate causal networks enabled by SEM characterize real-world processes better than simple correlation-based models. Therefore, SEM is more suited for the mathematical modeling of complex processes to serve both theory...and practice.” All preanalyses with respect to data screening (ie, missing data, outliers, and multivariate statistical assumptions) were completed based on well-known statistical methods [62-65]. Once the data screening process was complete, an SEM analysis comprised of both examination and assessment of the measurement model (see Multimedia Appendix 2) and structural model, as well as additional analyses (ie, common method bias, post hoc) was completed. Overall, the SEM analysis followed the guidelines set forth by SEM and PLS experts [53,66-69].

Results

Recruitment

Given that this study primarily focused on the effects that Internet health information has on medical compliance, survey participants were required to have (1) recently seen a physician regarding a recent significant health situation that they were able to clearly recall their interactions with the physician for and (2) a clear recall of their experience in a search they carried out for Internet health information regarding the significant health situation in question. The qualifying questions’ use of the phrase “significant health situation” was kept general (ie, no definition or examples of significant health situation were provided) because the interpretation of significant health situation is different for different people. The most important element of this aspect of the research is that the participant deemed the health situation to be significant. Data for this research study were collected in January 2013. Given the specific characteristics required for participants in this study, a decision was made to recruit research participants via the use of a well-known research firm (ie, Research Now). Participants were randomly selected from a pool of potential respondents contained in the database of this research firm. Ethics approval for research involving human subjects was obtained from the McMaster University Research Ethics Board (Hamilton, ON, Canada) and informed consent for all participants was obtained after the nature and possible consequences of the study were explained. All ethics requirements were enforced by Research Now and participants were compensated based on Research Now policies and procedures. A total of 234 participants were recruited.

Potential participants for this study were randomly selected from the Canadian adult population. A set of prequalifying questions was used to ensure that selected participants were able to recall a search for and use of Internet health information in the recent past for a significant health situation. In addition, selected participants were required to have recently seen a physician for this significant health situation and to be able to recall their interactions with their physician regarding the significant health situation. These prequalifying criteria were very important because this study examined how the use of Internet health information impacted elements of interactions between patients and their physicians regarding a significant health situation. A total of 1418 potential participants were contacted with 234 of these meeting the prequalifying criteria. The demographics for the research sample are shown in Table 1. The demographics of the participants in this study closely matched the demographics of the population that searched and used Internet health information (ie, higher proportion of females, younger, higher education levels, higher incomes [12,70]), thus providing confirmation that we had a representative and relevant sample for this study.

Based on an outlier analysis that examined both univariate and multivariate outliers, a total of 9 cases were removed from the dataset leaving 225 usable surveys retained for further analysis. The 9 cases removed represented less than 4% of the total cases, which can be considered an acceptable amount removed from the dataset [64]. There were no missing values for the constructs in the model and a limited number of missing values identified among the control variables. The mean imputation method was used to handle the missing control variable values as per Hair et al [63] and Meyers et al [64]. Finally, a complete multivariate statistical assumptions analysis (ie, linearity, normality, and homoscedasticity) revealed no substantive issues; therefore, the dataset was deemed viable for further statistical analysis.

Table 1. Sample demographics (N=225).

Demographic characteristic	n (%)
Gender	
Male	83 (38.4)
Female	133 (61.6)
Age	
<20	0 (0.0)
20-29	28 (12.6)
30-39	37 (16.6)
40-49	47 (21.1)
50-59	48 (21.5)
60-69	45 (20.2)
70-79	17 (7.6)
≥80	1 (0.4)
Education (highest level)	
High school	9 (4.2)
Some college/university or college/university degree	139 (64.3)
Graduate degree	68 (31.5)
Income (Can $)	
<$10,000	5 (2.6)
$10,000-$24,999	12 (6.2)
$25,000-$49,999	37 (19.2)
$50,000-$74,999	55 (28.5)
>$75,000	84 (43.5)

Statistical Analysis

A complete control variable analysis was completed prior to the analysis of the research model. This analysis showed that 4 of the control variables (ie, age, gender, income, and health knowledge) had significant relationships with 1 or more of the endogenous constructs in the model; therefore, these control variables were included in the final structural model to ensure that the effects of these extraneous variables were accounted for. The results of the structural model are shown in Figure 2.

Given the focus of PLS analysis is on prediction, an examination of the variance of the dependent measures through the R^2 results was completed. The results of this analysis showed moderate to substantial predictive powers based on the 0.19 (minimum), 0.33 (moderate) and 0.67 (substantial) thresholds [71], as shown in Table 2. In addition, an examination of the effects of the control variables was completed, indicating that the control variables had limited effects on the research model results as shown in Table 2.

Figure 2. Partial least squares structural model results.

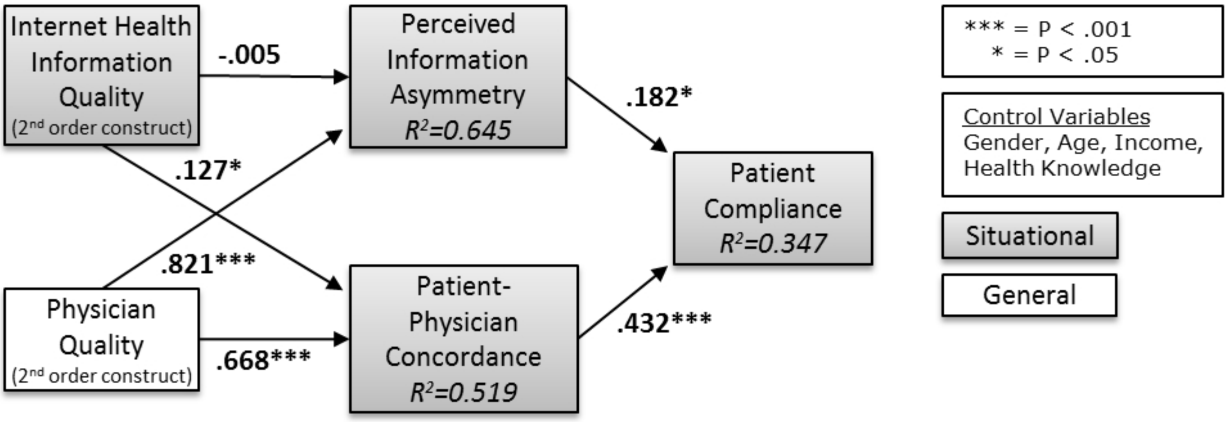


Table 2. Multivariate coefficient of determination (R^2) results.

Endogenous construct	R^2		Control variable effects		
	With control variables	Without control variables	ΔR^2	f^2	Effects
Compliance	0.347	0.333	0.014	0.021	Small
Patient-physician concordance	0.519	0.486	0.033	0.069	Small
Perceived information asymmetry	0.645	0.577	0.068	0.192	Medium

An assessment of the path estimates in the model via the magnitude and significance of the path coefficients is shown in Table 3. This assessment revealed that 5 of 6 hypotheses were fully supported with 3 of these significant at the $P<.001$ level. The *t* tests for significance were produced through the bootstrap method with the number of cases equal to the number of observations in the sample (ie, 225) and the number of samples set to 5000. For all supported hypotheses, the hypothesized algebraic sign was consistent with the path coefficient results. The lone hypotheses that was not supported (ie, Internet health information quality will have a negative impact on perceived information asymmetry) showed that the path coefficient between Internet health information quality and perceived information asymmetry was not significant and was in fact very close to zero. This perplexing result is discussed further in the Discussion section. In addition to the direct paths in the model, the 2 indirect paths (ie, Internet health information quality compliance and physician quality compliance) were also found to be significant as shown in Table 3.

Table 3. Partial least squares path analysis for direct and indirect effects.

Hypothesis	Path coefficient	t_{224}	*P*
Direct effects			
Perceived information asymmetry positively affects compliance	0.182	2.281	.02
Patient-physician concordance positively affects compliance	0.432	5.868	<.001
Internet health information quality negatively affects perceived information asymmetry	–0.005	0.121	.90
Internet health information quality positively affects patient-physician concordance	0.127	2.371	.02
Physician quality positively affects perceived information asymmetry	0.821	23.476	<.001
Physician quality positively affects patient-physician concordance	0.668	13.946	<.001
Indirect effects			
Internet health information quality affects compliance	0.054	2.139	.03
Physician quality affects compliance	0.438	6.773	<.001

Effect sizes as per Cohen’s f^2 [72] were calculated and are provided in Table 4. Effect sizes aid in evaluating the impact that the antecedent constructs have on the dependent constructs and can be assessed as small (ie, 0.02), medium (ie, 0.15), or large (ie, 0.35) effect sizes based on guidelines from Roldán and Sánchez-Franco [69]. The effect size analysis clearly shows the strong impact that physician quality had in the research model with large effect sizes on patient-physician concordance and perceived information asymmetry, whereas the effect sizes of Internet health information quality were either small or not significant. In addition, the impact of patient-physician concordance on compliance was strong with a medium effect size noted, whereas the impact of perceived information asymmetry on compliance was small.

A goodness-of-fit analysis to examine model performance for both the measurement and the structural model was calculated as per Tenenhaus et al [73]. This analysis produced a single value that can be interpreted similarly to the interpretation of effect sizes as per Wetzels et al [59]. The goodness-of-fit index for this study was 0.616, which can be considered a large effect and supports the conclusion that this model performed well.

Table 4. Partial least squares effect size analysis.

Dependent and independent constructs	R^2		ΔR^2	f^2	Effect size
	In	Out			
Compliance					
Patient-physician concordance	0.347	0.237	0.110	0.168	Medium
Perceived information asymmetry	0.347	0.327	0.020	0.020	Small
Patient-physician concordance					
Internet health information quality	0.519	0.504	0.015	0.031	Small
Physician quality	0.519	0.115	0.404	0.404	Large
Perceived information asymmetry					
Internet health information quality	0.645	0.645	0.000	0.000	NS
Physician quality	0.645	0.035	0.610	0.610	Large

Post Hoc Analysis

All additional demographic significant health situation, health status, and health knowledge variables that were captured in the survey were examined to determine if they had significant relationships with constructs in the research model. This analysis revealed that age had a significant positive relationship with both compliance and patient-physician concordance in that the older a person was, the more they tended to come to an agreement with the physician on the medical problem or need and its management and the more likely they were to comply with the physician’s instructions. Income had a significant negative relationship with patient-physician concordance in that the higher a person’s income was, the less likely they were to believe that there was agreement between themselves and the physician regarding the significant health situation. Gender had a significant relationship with perceived information asymmetry in that females were more likely to see a smaller gap in knowledge between themselves and the physician (regarding the significant health situation) than males. Finally, overall health knowledge had a significant negative relationship with perceived information asymmetry with those individuals who identified themselves as having higher overall knowledge about their health more likely to see a lower level of information asymmetry (ie, a smaller gap in knowledge between themselves and the physician with regards to the significant health situation).

Discussion

Principal Results

This research has several important theoretical contributions in the field of physician-patient relationship management and important implications for practitioners (ie, both physicians and Internet health information providers). First, we found that physician quality has the most significant impact directly on patient-physician concordance and information asymmetry as well as indirectly on compliance. This finding is similar to the results of Zolnierek and DiMatteo [5], who found that enhanced physician qualities can lead to better compliance. The implication of our finding is very important for physicians in that improved compliance can be achieved through physician quality attributes of competence, communication, and empathy. This finding is supported by Kim et al [7] who found a relationship between physician expertise and compliance, and that a patient’s assessment of physician empathy significantly influenced patient satisfaction and compliance [7]. Once the physician has established their knowledge and has the ability to effectively communicate this knowledge to the patient, patients will understand the magnitude of the physician’s knowledge and be more willing to come to an agreement regarding aspects of the patient’s medical situation (eg, diagnosis, treatment options) Once this physician quality is established, enhanced patient compliance is more likely to occur because the patient will be more likely to follow the physician’s advice.

Second, the findings regarding Internet health information quality were surprising. Although Internet health information quality has a significant positive relationship with concordance between patients and physicians, the strength of this relationship is somewhat low. This indicates that although Internet health information quality has some effect on the concordance process, physician quality has a much stronger relationship and larger effect size. The implications of this finding are important for physicians, who should focus on their personal and professional skills to improve the concordance process rather than overly concern themselves with Internet health information. Given there is a significant relationship between Internet health information quality and concordance, physicians should also encourage patients to make use of Internet health information to allow for a more concordant patient relationship. A number of previous studies support this finding, suggesting that physicians encourage Internet health information usage and share the responsibility for gathering knowledge regarding their health [15], perhaps even directing patients to reputable and relevant Internet health information websites [1]. In addition, physicians are encouraged to improve their communication skills to facilitate discussions about the Internet health information brought to them by patients [29]. The significant indirect relationship between Internet health information quality and compliance is interesting in that better Internet health information quality can lead to enhanced compliance. This finding was supported by Iverson et al [27] who also found this relationship. These results suggest that both Internet health information providers and physicians can play a role in helping to improve compliance. Internet health information developers should ensure their information is adequate, understandable, useful, and relevant to potential readers, whereas physicians can encourage patients to visit high-quality, reputable, and relevant Internet health information websites to improve compliance. It is interesting that patient-gathered high-quality Internet health information is positively related to their compliance with the physician's instructions, suggesting that better quality Internet health information most likely is in agreement with the physician's knowledge and directives.

The nonsignificant relationship between Internet health information quality and information asymmetry was very surprising and warrants a more detailed discussion. As noted previously, information asymmetry in this study is defined as the differential between the patient's perceptions regarding their own knowledge and their perceptions regarding their physician's knowledge, specifically regarding the patient's current significant health situation. Therefore, one would expect that higher levels of physician quality (including competence/knowledgeability) would increase perceived information asymmetry (a hypothesis that was supported) and that better quality Internet health information accessed by the patient should decrease the perceived level of information asymmetry between the patient and physician from the patient's perspective. However, this was not the result found in our study. Two reasons are suggested for this perplexing finding, as discussed subsequently.

Patients will most likely increase their medical knowledge from general access Internet health information websites. However, this information is often limited to the basic understanding of medical terminologies, diagnoses, and treatments. More detailed information (eg, from academic medical journals, research papers) is typically not available to the general population, especially for more rare and/or serious conditions. Therefore, although the patient may feel they have dramatically increased their level of knowledge by reading Internet health information, they will realize during their interaction with the physician that they have simply accessed basic information that their physician is already aware of and, therefore, there would be no change in information asymmetry. Although high-quality Internet health information will most likely dramatically alter the patient's perception of their knowledge level, once the patient discovers that the physician was already aware of this information and can fully explain why the information applies or potentially does not apply to the patient's condition, the patient will realize that there is still a large differential between their knowledge and the physician's knowledge. As anecdotal evidence of this phenomenon, one of the authors of this study experienced this exact situation, whereby the researcher accessed and discussed relevant high-quality Internet health information with his physician. The physician was already aware of this Internet health information and was able to clearly explain why it did not apply and how other more relevant research and medical information applied to the situation. Thus, the author left the physician's office with the perception that the gap in knowledge between himself and the physician was quite substantial even after accessing what he thought was relevant high-quality Internet health information.

Patients who access high-quality Internet health information are most likely accessing medical information on the Internet that has been developed by physicians. For example, well-known Internet health information sites such as WebMD and HealthCentral contain information either written by physicians or based on information that writers gather from physicians. Therefore, much of the information gleaned from these general access websites would be consistent with the information that the patient's physician already has and, thus, there would be no change in information asymmetry. More complex information from medical academic journal websites that the typical physician may not be aware of yet is not typically available to the general public and would most likely not have been a factor in the Internet health information research completed by patients.

Third, we found information asymmetry has some impact on compliance, but its impact is much weaker than the impact concordance has on compliance. This finding suggests that it is not the sheer volume of knowledge or the differential in physician versus patient knowledge that is most important in ensuring compliance, but rather the participative and concordant interactions between the patient and physician that will lead to compliance. This finding is important for both physicians and developers of Internet health information. For physicians, taking steps to ensure patient interactions are concordant and not confrontational can lead to compliance. This means that empathetic communication with patients while demonstrating competence is one of the keys to compliance. In addition, listening to the patient and the potential Internet health information–based knowledge they bring with them to the appointment can also enhance compliance. The findings also suggest that physicians need to be ready to have an open and honest discussion regarding the patient-researched Internet health information and not simply discount the potential knowledge and information that the patient brings to the discussion. Although concordance has a large effect on compliance, the weaker yet significant relationship between information asymmetry and compliance needs to be understood and addressed by physicians. A certain degree of information asymmetry between the patient and physician needs to be maintained to ensure the physician's advice is respected and to persuade the patient to accept the physician’s professional advice. If a physician is not viewed as an expert regarding the health situation, they may lose their professional advantage and cause patients to underestimate their need for his/her medical services.

Limitations and Future Research

As with most research, this research also has some limitations. First, this research used a cross-sectional survey that collected data from respondents at one point in time and, therefore, may not capture the full magnitude of the Internet health information use or actual compliance. In addition, cross-sectional studies do not allow definitive conclusions regarding causal inferences. We did not conduct a longitudinal research study to actually monitor the changes in patient medical knowledge, the impact of such change, and actual compliance. A longitudinal investigation may help to gain a more comprehensive understanding of both information asymmetry and compliance. It is recommended that future research include a longitudinal study to follow up with respondents on both their levels of knowledge and actual compliance. Secondly, this research did not capture partial compliance, where patients may follow physician advice but not the full course of that advice (eg, patient is prescribed medication for 10 days but stops after 8 days). The use of a 7-point Likert scale for compliance responses will have controlled much of this phenomenon because it allowed respondents to indicate if they fully, partially, or did not comply. In addition, this research included patient self-reports of compliance rather than monitoring actual compliance. Although monitoring actual compliance (and ensuring this compliance was complete and not partial) was not accomplished, previous research supports the use of compliance self-reporting and found that patient self-reports of compliance corresponded with actual compliance [16]; therefore, it is not believed that this limitation affected the results of this study. However, future research may wish to use actual monitored compliance (eg, through confirmation of follow-up appointments/treatments/tests, objective measures of taking medication) to eliminate any potential effects this may have on the results. Third, this research used patient self-reports of the significant health situation and Internet health information search recall and, therefore, there are no guarantees that patients were able to fully and accurately recall these events. However, all efforts were made to ensure only participants who were able to clearly recall the significant health situation, physician interactions, and the Internet health information search were included in this research study. In fact, a large number (ie, 84%) of potential participants were excluded from this research due to their inability to recall the required events (ie, significant health situation, physician interaction, and Internet health information search). Fourth, this research relied on patient’s assessment of the quality of the Internet health information they accessed and not the actual quality. Future research may wish to present participants with verified quality Internet health information to ensure that all participants are reporting information asymmetry, concordance, and compliance after accessing validated Internet health information. This research would allow us to isolate the effects that quality validated Internet health information has in this research model; however, this would not reflect reality because patients typically search health-related websites of varying quality. Finally, this research endeavored to match participant demographics (eg, gender, age, education, income) to the demographics of the population that search and use Internet health information. This was done to ensure that the results of this research would be generalizable to the population of people who typically use Internet health information. However, this may reduce the generalizability to the population at large. Future research may wish to recreate this study with a sample that matches the demographics of the overall population and not simply the current characteristics of typical Internet health information users.

Conclusions

Overall, it should be strongly emphasized that our findings suggest that physician quality was the most important element in our research model, with highly significant relationships and medium to large effect sizes on information asymmetry, concordance, and ultimately compliance. This suggests that physician quality dominates the impact on these factors and physicians are encouraged to spend less time distressing about the negative impacts of Internet health information and more time improving their competence, communication, and empathy characteristics. This being said, patients should also be encouraged, both by their physicians and society (perhaps via government initiatives), to seek out and make use of high-quality Internet health information in their discussions with medical professionals. By combining both of these recommendations, improved compliance and its related benefits are more likely to occur.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Measurement instruments.

[PDF File (Adobe PDF File), 130KB - jmir_v17i6e143_app1.pdf]

Multimedia Appendix 2

Statistical analysis details.

[PDF File (Adobe PDF File), 194KB - jmir_v17i6e143_app2.pdf]

References

1. Erdem SA, Harrison-Walker LJ. The role of the Internet in physician–patient relationships: The issue of trust. Business Horizons 2006 Sep;49(5):387-393. [doi: 10.1016/j.bushor.2006.01.003]
2. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
3. Hausman A. Taking your medicine: relational steps to improving patient compliance. Health Mark Q 2001;19(2):49-71. [doi: 10.1300/J026v19n02_05] [Medline: 11873456]
4. Wahl C, Gregoire JP, Teo K, Beaulieu M, Labelle S, Leduc B, et al. Concordance, compliance and adherence in healthcare: closing gaps and improving outcomes. Healthc Q 2005;8(1):65-70. [Medline: 15715337]
5. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
6. Hausman A. Modeling the patient-physician service encounter: improving patient outcomes. Journal of the Academy of Marketing Science 2004 Oct 01;32(4):403-417. [doi: 10.1177/0092070304265627]
7. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof 2004 Sep;27(3):237-251. [doi: 10.1177/0163278704267037] [Medline: 15312283]
8. Kerse N, Buetow S, Mainous AG, Young G, Coster G, Arroll B. Physician-patient relationship and medication compliance: a primary care investigation. Ann Fam Med 2004;2(5):455-461 [FREE Full text] [Medline: 15506581]
9. Vermeire E, Hearnshaw H, Van RP, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther 2001 Oct;26(5):331-342. [Medline: 11679023]
10. Wroth TH, Pathman DE. Primary medication adherence in a rural population: the role of the patient-physician relationship and satisfaction with care. J Am Board Fam Med 2006;19(5):478-486 [FREE Full text] [Medline: 16951297]
11. Cvengros JA, Christensen AJ, Hillis SL, Rosenthal GE. Patient and physician attitudes in the health care context: attitudinal symmetry predicts patient satisfaction and adherence. Ann Behav Med 2007 Jun;33(3):262-268. [doi: 10.1080/08836610701358086] [Medline: 17600453]
12. Fox S. Online Health Search 2006. Washington, DC: Pew Internet & American Life Project; 2006 Oct 29. URL: http://www.pewinternet.org/2006/10/29/online-health-search-2006/ [accessed 2015-05-31] [WebCite Cache ID 6YwBFV7X4]
13. Broom A. Virtually he@lthy: the impact of internet use on disease experience and the doctor-patient relationship. Qual Health Res 2005 Mar;15(3):325-345. [doi: 10.1177/1049732304272916] [Medline: 15761103]
14. Nwosu CR, Cox BM. The impact of the Internet on the doctor-patient relationship. Health Informatics Journal 2000 Sep 01;6(3):156-161. [doi: 10.1177/146045820000600308]
15. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
16. Horne R, Weinman J, Barber N, Elliott R, Morgan M, Cribb A. Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation Research & Development. 2005 Dec. Concordance, adherence and compliance in medicine taking URL: http://nets.nihr.ac.uk/__data/assets/pdf_file/0009/64494/FR-08-1412-076.pdf [accessed 2015-06-01] [WebCite Cache ID 6Yxk15mL7]
17. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. Arch Intern Med 1993 Aug 23;153(16):1863-1868. [Medline: 8250647]
18. Snowden A. Medication management in older adults: a critique of concordance. Br J Nurs 2008;17(2):114-119. [doi: 10.12968/bjon.2008.17.2.28140] [Medline: 18414285]
19. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [Medline: 15076819]
20. Alwan A, Armstrong T, Cowan M, Riley L. World Health Organization. 2011 Sep. Noncommunicable diseases country profiles 2011 URL: http://whqlibdoc.who.int/publications/2011/9789241502283_eng.pdf?ua=1 [accessed 2015-06-01] [WebCite Cache ID 6YxkCLcXa]
21. Pavlou P, Liang H, Xue Y. Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. MIS Quarterly 2007 Mar;31(1):105-136.
22. Eisenhardt KM. Agency theory: An assessment and review. Academy of Management Review 1989 Jan;14(1):57-74.
23. Vick S, Scott A. Agency in health care. Examining patients' preferences for attributes of the doctor-patient relationship. J Health Econ 1998 Oct;17(5):587-605. [Medline: 10185513]
24. Xie B, Dilts DM, Shor M. The physician-patient relationship: the impact of patient-obtained medical information. Health Econ 2006 Aug;15(8):813-833. [doi: 10.1002/hec.1098] [Medline: 16550612]
25. Johnson GL, Ramaprasad A. Patient-physician relationships in the information age. Mark Health Serv 2000;20(1):20-27. [Medline: 11923940]
26. Kaba R, Sooriakumaran P. The evolution of the doctor-patient relationship. Int J Surg 2007 Feb;5(1):57-65 [FREE Full text] [doi: 10.1016/j.ijsu.2006.01.005] [Medline: 17386916]

27. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]
28. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [Medline: 12617199]
29. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
30. Bader SA, Braude RM. "Patient informatics": creating new partnerships in medical decision making. Acad Med 1998 Apr;73(4):408-411. [Medline: 9580718]
31. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
32. Jayanti RK, Whipple TW. Like me … like me not: the role of physician likability on service evaluations. The Journal of Marketing Theory and Practice 2008 Mar 14;16(1):79-86. [doi: 10.2753/MTP1069-6679160106]
33. Willson P, McNamara JR. How perceptions of a simulated physician-patient interaction influence intended satisfaction and compliance. Soc Sci Med 1982;16(19):1699-1704. [Medline: 7178916]
34. Lawler FH, Viviani N. Patient and physician perspectives regarding treatment of diabetes: compliance with practice guidelines. J Fam Pract 1997 Apr;44(4):369-373. [Medline: 9108834]
35. Maiman LA, Becker MH, Liptak GS, Nazarian LF, Rounds KA. Improving pediatricians' compliance-enhancing practices. A randomized trial. Am J Dis Child 1988 Jul;142(7):773-779. [Medline: 3381783]
36. Solem C, Mody R, Stephens J, Macahilig C, Gao X. Mealtime-related dosing directions for proton-pump inhibitors in gastroesophageal reflux disease: physician knowledge, patient adherence. J Am Pharm Assoc (2003) 2014;54(2):144-153. [doi: 10.1331/JAPhA.2014.13117] [Medline: 24632930]
37. Zahedi F, Song J. Dynamics of trust revision: using health infomediaries. Journal of Management Information Systems 2008 May 12;24(4):225-248. [doi: 10.2753/MIS0742-1222240409]
38. Bartov E, Bodnar GM. Alternative accounting methods, information asymmetry and liquidity: theory and evidence. Accounting Review 1996 Jul;71(3):397-418.
39. Core JE. A review of the empirical disclosure literature: discussion. Journal of Accounting and Economics 2001 Sep;31(1-3):441-456. [doi: 10.1016/S0165-4101(01)00036-2]
40. Verbeke W. Agriculture and the food industry in the information age. European Review of Agricultural Economics 2005 Sep 01;32(3):347-368. [doi: 10.1093/eurrag/jbi017]
41. Rabinowicz E. EAAE presidential address. Redesigning the CAP to meet the challenges of EU enlargement and the WTO: what can agricultural economic research contribute? European Review of Agricultural Economics 1999 Sep;26(3):265-281.
42. Lusk JL, House LO, Valli C, Jaeger SR, Moore M, Morrow J, et al. Effect of information about benefits of biotechnology on consumer acceptance of genetically modified food: evidence from experimental auctions in the United States, England, and France. European Review of Agricultural Economics 2004 Jun;31(2):179-204.
43. Hobbs JE. Markets in metamorphosis: the rise and fall of policy institutions. In: Van Huylenbroeck GV, Verbeke W, Lauwers L, editors. Role of Institutions in Rural Policies and Agricultural Markets. Amsterdam: Elsevier; 2004:199-212.
44. Choe YC, Park J, Chung M, Moon J. Effect of the food traceability system for building trust: Price premium and buying behavior. Inf Syst Front 2008 Jul 30;11(2):167-179. [doi: 10.1007/s10796-008-9134-z]
45. Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. Curr Opin Oncol 2005 Jul;17(4):351-354. [Medline: 15933466]
46. Stevenson FA, Cox K, Britten N, Dundar Y. A systematic review of the research on communication between patients and health care professionals about medicines: the consequences for concordance. Health Expect 2004 Sep;7(3):235-245. [doi: 10.1111/j.1369-7625.2004.00281.x] [Medline: 15327462]
47. Gafni A, Charles C, Whelan T. The physician-patient encounter: the physician as a perfect agent for the patient versus the informed treatment decision-making model. Soc Sci Med 1998 Aug;47(3):347-354. [Medline: 9681904]
48. Janz NK, Wren PA, Copeland LA, Lowery JC, Goldfarb SL, Wilkins JG. Patient-physician concordance: preferences, perceptions, and factors influencing the breast cancer surgical decision. J Clin Oncol 2004 Aug 1;22(15):3091-3098 [FREE Full text] [doi: 10.1200/JCO.2004.09.069] [Medline: 15284259]
49. Riekert KA, Butz AM, Eggleston PA, Huss K, Winkelstein M, Rand CS. Caregiver-physician medication concordance and undertreatment of asthma among inner-city children. Pediatrics 2003 Mar;111(3):e214-e220. [Medline: 12612274]
50. Boudreau M, Gefen D, Straub DW. Validation in Information Systems research: a state-of-the-art assessment. MIS Quarterly 2001 Mar;25(1):1. [doi: 10.2307/3250956]
51. Dunk AS. The effect of budget emphasis and information asymmetry on the relation between budgetary participation and slack. Accounting Review 1993 Apr;68(2):400-410.
52. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust mMeasures for e-commerce: an integrative typology. Information Systems Research 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]

53. Chin WW. How to write up and report PLS analyses. In: Esposito-Vinzi V, Chin WW, Henseler J, Wang H, editors. Handbook of Partial Least Squares. Berlin: Springer-Verlag; 2010:655-690.
54. Jarvis C, MacKenzie S, Podsakoff P. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003 Sep;30(2):199-218. [doi: 10.1086/376806]
55. Ringle C, Sarstedt M, Straub D. A critical look at the use of PLS-SEM in MIS Quarterly. MIS Quarterly 2012 Mar;36(1):iii-xiv.
56. Pavlou PA, Gefen D. Psychological contract violation in online marketplaces: antecedents, consequences, and moderating role. Information Systems Research 2005 Dec;16(4):372-399. [doi: 10.1287/isre.1050.0065]
57. Agarwal R, Karahanna E. Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Quarterly 2000 Dec;24(4):665. [doi: 10.2307/3250951]
58. Rai A, Patnayakuni R, Seth N. Firm performance impacts of digitally enabled supply chain integration capabilities. MIS Quarterly 2006 Jun;30(2):225-246.
59. Wetzels M, Odekerken-Schroder G, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Quarterly 2009 Mar;33(1):177-195.
60. Wold H. Soft modeling: the basic design and some extensions. In: Jöreskog KG, Wold H, editors. Systems Under Indirect Observation: Causality, Structure, Prediction. Amsterdam: North-Holland; 1982:1-54.
61. Gefen D, Straub DW, Boudreau MC. Structural equation modeling and regression: guidelines for research practice. Communications of the Association for Information Systems 2000 Oct;4(7):1-77.
62. Hair JF, Black WC, Babin BJ, Anderson RE. Advanced Diagnostics for Multiple Regression: A Supplement to Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall; 2010.
63. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall; 2010.
64. Meyers LS, Gamst G, Guarino AJ. Applied Multivariate Research: Design and Interpretation. Thousand Oaks, CA: Sage Publications, Inc; 2006.
65. Tabachnick BG, Fidell LS. Using Multivariate Statistics. Boston: Pearson/Allyn & Bacon; 2007.
66. Götz O, Liehr-Gobbers K, Krafft M. Evaluation of structural equation models using the partial least squares (PLS) approach. In: Esposito-Vinzi V, Chin WW, Henseler J, Wang H, editors. Handbook of Partial Least Squares. Berlin: Springer-Verlag; 2010:691-711.
67. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a silver bullet. The Journal of Marketing Theory and Practice 2011 Mar 15;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
68. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
69. Roldán JL, Sánchez-Franco MJ. Variance-based structural equation modeling: guidelines for using partial least squares in Information Systems. In: Mora M, Gelman O, Steenkamp A, Raisinghani MS, editors. Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems. Hershey, PA: IGI Global; 2012:193-221.
70. Fox S, Fallows D. Internet Health Resources. Washington, DC: Pew Internet & American Life Project; 2003 Jul. URL: http://www.pewinternet.org/2003/07/16/internet-health-resources/ [accessed 2015-06-01] [WebCite Cache ID 6YxpB1YHL]
71. Chin W. The partial least squares approach to structural equation modelling. In: Marcoulides GA, editor. Modern Methods for Business Research. New York: Lawrence Erlbaum Associates; 1998:295-336.
72. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
73. Tenenhaus M, Amato S, Esposito VV. A global goodness-of-fit index for PLS structural equation modelling. 2004 Presented at: Proceedings of the XLII SIS Scientific Meeting; June 2004; Padova, Italy p. 739-742.

Abbreviations

PLS: partial least squares
SEM: structural equation modeling

Edited by G Eysenbach; submitted 12.02.15; peer-reviewed by T Hong, C Turvey; comments to author 07.03.15; revised version received 17.04.15; accepted 07.05.15; published 11.06.15.

Please cite as:
Laugesen J, Hassanein K, Yuan Y
The Impact of Internet Health Information on Patient Compliance: A Research Model and an Empirical Study
J Med Internet Res 2015;17(6):e143
URL: http://www.jmir.org/2015/6/e143/
doi:10.2196/jmir.4333
PMID:26068214

Original Paper

eHealth, Participatory Medicine, and Ethical Care: A Focus Group Study of Patients' and Health Care Providers' Use of Health-Related Internet Information

Anne Townsend[1,2], PhD; Jenny Leese[1], MA; Paul Adam[3], MSW; Michael McDonald[4], PhD; Linda C Li[1,5], PhD; Sheila Kerr[6], BApplSc (Physio); Catherine L Backman[1,2], PhD

[1]Arthritis Research Canada, Richmond, BC, Canada
[2]Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, BC, Canada
[3]Mary Pack Arthritis Program, Vancouver Coastal Health, Vancouver, BC, Canada
[4]Maurice Young Centre for Applied Ethics, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[5]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada
[6]Arthritis Patient Advisory Board, Arthritis Research Canada, Richmond, BC, Canada

Corresponding Author:
Anne Townsend, PhD
Arthritis Research Canada
5591 Number 3 Road
Richmond, BC, V6X 2C7
Canada
Phone: 1 604 207 4031
Fax: 1 604 207 4059
Email: atownsen@exchange.ubc.ca

Abstract

Background: The rapid explosion in online digital health resources is seen as transformational, accelerating the shift from traditionally passive patients to patients as partners and altering the patient–health care professional (HCP) relationship. Patients with chronic conditions are increasingly engaged, enabled, and empowered to be partners in their care and encouraged to take responsibility for managing their conditions with HCP support.

Objective: In this paper, we focus on patients' and HCPs' use of health-related Internet information and how it influences the patient-HCP relationship. In particular, we examine the challenges emerging in medical encounters as roles and relationships shift and apply a conceptual framework of relational ethics to examine explicit and nuanced ethical dimensions emerging in patient-HCP interactions as both parties make increased use of health-related Internet information.

Methods: We purposively sampled patients and HCPs in British Columbia, Canada, to participate in focus groups. To be eligible, patients self-reported a diagnosis of arthritis and at least one other chronic health condition; HCPs reported a caseload with >25% of patients with arthritis and multimorbidity. We used a semistructured, but flexible, discussion guide. All discussions were audiotaped and transcribed verbatim. Elements of grounded theory guided our constant comparison thematic analytic approach. Analysis was iterative. A relational ethics conceptual lens was applied to the data.

Results: We recruited 32 participants (18 patients, 14 HCPs). They attended seven focus groups: four with patients and three with rehabilitation professionals and physicians. Predominant themes to emerge were how use of health-related Internet information fostered (1) changing roles, (2) patient-HCP partnerships, and (3) tensions and burdens for patients and HCPs.

Conclusions: Relational aspects such as mutual trust, uncertainty, and vulnerability are illuminated in patient-HCP interactions around health-related Internet information and the negotiated space of clinical encounters. New roles and associated responsibilities have key ethical dimensions that make clear the changes are fundamental and important to understand in ethical care. When faced with tensions and burdens around incorporating health-related Internet information as a resource in clinical encounters, participants described a particular ambivalence illustrating the fundamental changes being negotiated by both patients and HCPs.

(J Med Internet Res 2015;17(6):e155) doi:10.2196/jmir.3792

KEYWORDS

relational ethics; health-related Internet information; arthritis; patient-HCP relationship; e-health; chronic illness; qualitative

Introduction

eHealth technology is widespread and wide-ranging [1] and is transforming health care delivery [2-4]. According to the World Health Organization, "eHealth is the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research" [5]. eHealth includes an array of resources and devices [6-12], with the capacity to provide patients with extensive and up-to-date information, access to medical research [13], connections to people with similar conditions, immediacy, and convenience in patient-health care professional (HCP) communications [13,14] and improved health outcomes [6,15]. The proliferation of eHealth strategies is accelerating a shift in health care from a traditional and paternalistic delivery model to a more mutual patient-HCP relationship, where informed patients are actively involved in their care and treatment decisions [16,17]. This shift supports a patient-centered model of care based on effective communication, respect for patients and shared patient-HCP responsibility in care provision and management of long-term conditions [18].

While many patients use health-related Internet information (HRII), the information may or may not be shared in medical consultations. The Pew Research Center's Internet & American Life Project reported that 35% of US adults reported going online specifically to diagnose a condition for themselves or someone they knew, and 72% of Internet users reported they looked online for health information within the past year. Over half (53%) of online diagnosers said they talked with a medical professional about what they found online [19]. While advancing patient empowerment and shared decision-making has been identified as a practical and moral necessity [20], uncertainty exists about how HRII is affecting the cornerstone of good medical care, that is, the patient-HCP relationship [21]. As Anderson et al noted over a decade ago, "While the use of the Internet does hold considerable promise for health care, it raises a host of social and ethical issues that need to be addressed. In particular, professional resistance to the use of this new technology in health care remains one of the greatest barriers to realization of the Internet's ample potential" [22]. There is growing evidence about how HRII is used in consultations [23-25]. For a range of reasons, however despite patients' sourcing HRII, it may not feature in consultation discussions [26]. When HRII is discussed in medical consultations, it can have both positive and negative impacts [27] and does not always mark a shift toward patient-centeredness [28], particularly as patient and HCP perspectives on the role of HRII can differ [29]. While patients may be motivated to spend time and effort seeking HRII about specific diseases and treatments, physicians face constraints searching relevant information for individual patients [30] and HCP responses to use varies [31]. While HRII has the capacity to advance a partnership model of care [32], it is not clear how changing roles and responsibilities for both patients and HCPs [33] are negotiated in these shifting times.

HRII has the potential to empower patients to adopt a heightened sense of individual autonomous action [34,35], and this has fundamental implications for patient-HCP relationships in both explicit and nuanced ways. The interpersonal requirements of achieving constructive and respectful partnerships [36] may be underestimated and subsequently not well examined or easily understood. As Agarwal et al noted, patient-HCP interactions in consultation discussions involve the "critical interplay" of traditional values such as trust and responsibility, and a reliance on an oversimplified model often neglects new patient-HCP dynamics and the impact of these changes on patient-HCP relationships [37].

A relational ethics lens provides a conceptual framework to better understand the relational shifts in ethical patient-HCP relationships. Relational ethics [38] builds on traditional bioethical principles of autonomy, justice, beneficence, non-maleficence [39], and shifts attention to relationships as the source of ethical action [40]. As greater emphasis is placed on how patients and HCPs negotiate communications and shared decision-making, a relational lens [41,42] provides contextual and relationship insights into health care practices [43]; this is particularly pertinent in chronic illness care [44] and patient-centered care [32,45]. Core elements of relational ethics are mutuality, engagement, respect, trust, vulnerability, uncertainty, and an interdependent environment [40] applicable to everyday experiences, practices, and interactions. Our current knowledge on the ways the relational shift (changing relationships) is happening for both patients and HCPs is limited. We need to better understand the meaning, process, and context of the clinical encounter for those involved and the potential risks and benefits of HRII to advance effective patient-HCP relationships.

In this paper, we present findings from the focus group phase of our study, which was designed to examine the influence of different types of eHealth use on patient-HCP relationships. We limit analysis here to the discussions on HRII. We make explicit the ethical dimensions underpinning the dynamics of evolving patient-HCP relationships as the use of HRII grows.

Methods

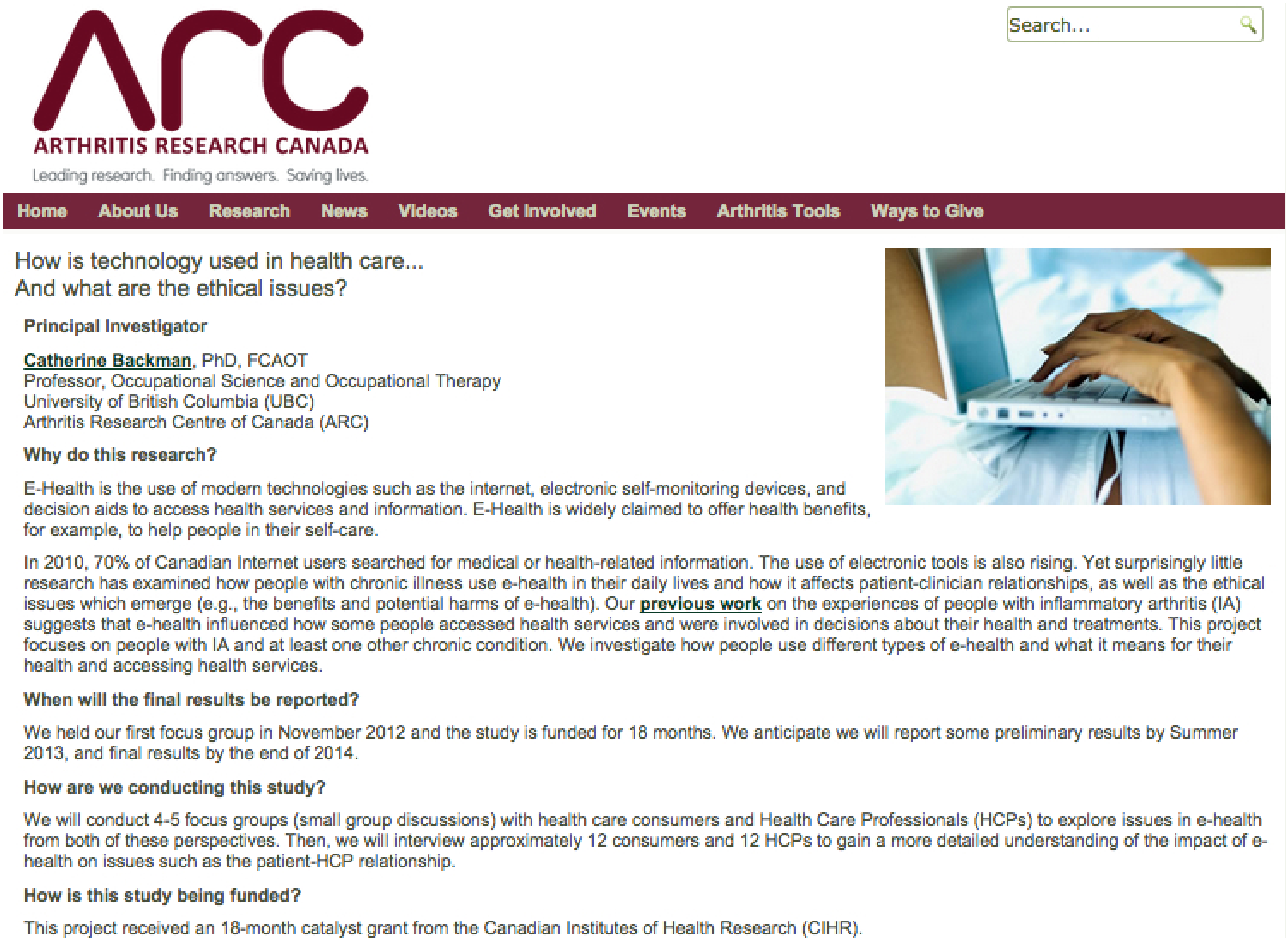
Research Design

This focus group qualitative study [1] was designed to explore relational aspects of eHealth for patients and HCPs. Our aim was to better understand how participants perceived and experienced different types of eHealth. A major focus to emerge in the discussions was how changes were occurring in patient-HCP relationships arising from use of HRII in the context of chronic illness care. Ethical relationships are integral to health care, and so a conceptual framework of relational ethics suited our aims and objectives.

Context

Being based at an arthritis research center influenced recruitment and feasibility of the study (Figure 1). Participants were recruited in two urban areas of British Columbia, Canada, between November 2012 and June 2013 for focus group discussions via online arthritis sites, for example, the Arthritis Research Center (ARC), the Arthritis Patient Advisory Board (APAB), Web and social media sites (Twitter and Facebook), posters in clinical settings, and word-of-mouth. All sessions were held in the greater Vancouver area except for one HCP group that was held in Victoria. Focus groups were held in community centers, health care centers or participants' places of work. The University of British Columbia's Behavioral Research Ethics Board granted approval. Participants gave written consent.

Figure 1. Online recruitment document.



How is technology used in health care...
And what are the ethical issues?

Principal Investigator

Catherine Backman, PhD, FCAOT
Professor, Occupational Science and Occupational Therapy
University of British Columbia (UBC)
Arthritis Research Centre of Canada (ARC)

Why do this research?

E-Health is the use of modern technologies such as the internet, electronic self-monitoring devices, and decision aids to access health services and information. E-Health is widely claimed to offer health benefits, for example, to help people in their self-care.

In 2010, 70% of Canadian Internet users searched for medical or health-related information. The use of electronic tools is also rising. Yet surprisingly little research has examined how people with chronic illness use e-health in their daily lives and how it affects patient-clinician relationships, as well as the ethical issues which emerge (e.g., the benefits and potential harms of e-health). Our **previous work** on the experiences of people with inflammatory arthritis (IA) suggests that e-health influenced how some people accessed health services and were involved in decisions about their health and treatments. This project focuses on people with IA and at least one other chronic condition. We investigate how people use different types of e-health and what it means for their health and accessing health services.

When will the final results be reported?

We held our first focus group in November 2012 and the study is funded for 18 months. We anticipate we will report some preliminary results by Summer 2013, and final results by the end of 2014.

How are we conducting this study?

We will conduct 4-5 focus groups (small group discussions) with health care consumers and Health Care Professionals (HCPs) to explore issues in e-health from both of these perspectives. Then, we will interview approximately 12 consumers and 12 HCPs to gain a more detailed understanding of the impact of e-health on issues such as the patient-HCP relationship.

How is this study being funded?

This project received an 18-month catalyst grant from the Canadian Institutes of Health Research (CIHR).

Sampling and Recruitment

The 32 participants consisted of 18 adult patients (16 female, 2 male) and 14 HCPs (11 female, 3 male) who participated in seven focus groups. Four groups were held with 4-6 patient participants, aged in their 30s-70s. Three groups were held with 4-5 HCP participants: physical and occupational therapists, a rheumatology nurse, a laboratory technician and rheumatology fellows, aged in their 30s-60s. Patients were recruited based on a self-reported diagnosis of arthritis, plus at least one other chronic condition. The majority of participants were middle class and Caucasian, which does not fully represent the diversity of the geographic region. The HCPs included physicians and rehabilitation providers who cared for people with multiple chronic conditions. We selected people with arthritis and co-conditions for this study for two reasons: pragmatism (being based in an arthritis research center) and prevalence. Arthritis is a highly prevalent and severe chronic condition globally and the leading cause of pain and disability in Canada [46]. The Canadian Community Health Survey (CCHS) (124,844 respondents, response rate 76%), based on 2007-2008 data, estimated that more than 4.2 million Canadians 15 years and older (16% of the population) had arthritis [46]. The coexistence of other chronic conditions with arthritis was reported as common by the Public Health Agency of Canada, based on the CCHS 2007-2008 data. Multimorbidity is associated with high burdens of care and cost [47]. Despite this, our knowledge and understanding of the impact of multimorbidity for patients and HCPs is poor [48]. Because eHealth (including HRII) is a vast resource for both patients and HCPs, it is vital to identify its potential benefits and harms, perhaps particularly salient for those who have multimorbidity and their HCPs who manage extensive information and encounter increasingly complex decisions [1]. A more detailed rationale is given in the published protocol [1]. All participants completed a demographic form that included data on ownership and use of digital devices. All participants used tablets, desk computers, or laptops for HRII.

Procedure

In the focus groups, eHealth was defined broadly as any digital health technologies that people used or had any knowledge of. Participants were asked to identify the range of devices and reasons they used them at the start of each discussion. This was identified on flip chart papers and guided the discussions. Ethics was framed as pros and cons of eHealth with a particular focus on patient-HCP relationships and was elaborated on depending on how the discussions unfolded. This was to avoid fixed definitions of ethics constraining the discussions. The topic guide was devised to explore patient and HCP experiences of eHealth use in chronic illness and was arranged around four key areas: (1) What types of eHealth and devices do you use? (2) What sort of things do you use eHealth for? (3) How does eHealth influence what you do regarding your health and illness conditions (patients) or practice (HCPs), including your interactions with HCPs/patients? and (4) What do you see as the benefits and drawbacks of eHealth?

We encouraged group discussion [28] where participants voiced their priorities and concerns. We attempted to tease out eHealth use in the practical circumstances of living with chronic illness or chronic illness care. We introduced relational ethics aspects by probing the nature of patient-HCP interactions and relationships, shared decision-making, and self-managing. Discussions were audiotaped, facilitated by AT or PA (authors), and lasted approximately 2 hours. Focus groups were transcribed verbatim, checked against recordings for accuracy, and de-identified.

Data Analysis

An iterative, thematic approach using constant comparative methods was applied to the data. All data were coded, with no preset categories. AT and JL (authors) independently read and annotated transcripts, identifying emerging issues for discussion as data collection proceeded. We modified and added codes in light of fresh transcripts and repeated readings. When all focus groups had been completed, PA and CB (authors) read a selection of the transcripts and shared their observations based on the original aims of the study, the topic guide, and shared interpretations as analytical discussions continued among research team members. In this way, the analysis was both inductive and deductive. We applied initial broad themes to all transcripts, and these were outlined to all research team members (including researchers, practitioners, and patients) who asked further questions and offered additional interpretations. Early common themes emerged across the patient and HCP data such as changing roles and responsibilities; collaboration, notions of trust; and teamwork. After further analysis, we agreed on higher-level themes, for example, partnerships. We then organized them into predominant themes: changing roles, with subthemes of “being prepared” and “responsibilities”; “partnerships”, which had a subtheme of “trust”; and “tensions and burdens”.

In planning the study, we agreed on conceptualizations of relational ethics as an overarching analytic conceptual framework as applicable to our research question. As in all qualitative studies, there are multiple ways to analyze and interpret data. In this primary analysis of focus group findings, the application of relational ethics helped us focus on what is valued in interactions and relationships and what is at risk, rather than specific aspects of eHealth like the nature of self-monitoring devices. Relational ethics informed the analysis in relation to how participants discussed patient-HCP roles and relationships and prized values, for example, roles and responsibilities, partnerships, mutual trust, and vulnerabilities. In extracts shared here, patient groups are notated PG, health care professionals groups are notated HCP, and within each group participants are given a number (P1 to P6).

Results

Overview

This analysis is limited to the discussions around HRII and how this influenced patient-HCP roles and relationships. The types of HRII reported by patients were Internet searches in general, health-related websites (eg, universities, health organizations, non-profit disease-oriented organizations), aimed at both patients and HCPs. To a lesser extent, patients also reported using personal websites and blogs (eg, individuals sharing personal experiences and resources), chat rooms (eg, open and closed groups on specific sites or via social media platforms like Facebook), and online links to medical test results (eg, to an eHealth record at the lab that did their blood-work). HCPs reported a similar range of Internet resources but described less frequent and extensive searches, and their resources were largely limited to health- and professional-oriented websites. HCPs did not cite use of blogs and chats as frequently as did patients. They reported a tendency to rely on a few specific online resources pertinent to their practice, some of which they recommended to patients. A few HCPs used online sites for sharing medical tests results. For both patients and HCPs, the factual content of the HRII they used was largely about research and treatment options, medication self-management strategies, and resources. Patients also sought information about diagnosis and sought the experiences of others with similar conditions/symptoms.

Both patients and HCPs relayed how HRII prompted interactive and negotiated communications. A traditional paternalistic era of care was contrasted with a partnership model illustrating patient involvement in discussions, but also tensions around change at the level of patient-HCP relationship. Here we present three predominant themes, the first two with subthemes that offer additional depth with specific examples: (1) changing roles, with subthemes of “being prepared” and “responsibilities”, (2) partnerships, with a subtheme of trust, and (3) tensions and burdens. Although for organizational purposes we identify three themes, they are overlapping and interlocking, for example, the changing roles and new responsibilities contribute to a partnership model of care, which in turn is characterized by particular tensions and burdens. We then discuss the ethical issues that emerge as patient-HCPs relationships evolve.

Changing Roles

Overview

Both patients and HCPs identified information from hospital or university sites as reliable sources of HRII. However, they also accessed other HRII and described the information as extensive, provisional, uncertain, and contradictory. Patient participants discussed being more informed but more uncertain, which prompted more questions about their health conditions and treatment options. They recognized that HCPs could not know everything, and they no longer conceived of them as the gatekeepers of knowledge. Whereas "in the old days, the doctor was God…[today] doctors don't know that much at all…I don't think any of us in the world realized that before…Before we would accept what the doctor said but in the last ten years, we've had access to the information and now we question more" (P3, PG3). Other patient focus groups concurred with the challenge to health professionals of increasing amounts of information and informed patients, given the limited time for clinical visits. This had implications for their view of the patient role. They did not rely on their HCP as their information source. Rather, in the context of ongoing illness, repeated visits to different HCPs, and increasingly expert in their conditions, they expressed the patient role as being equipped with information and enabled to ask questions and take part in an interactive dialogue about care and treatment. To participate fully, they had to be prepared for the medical encounter.

Being Prepared

Both patients and HCPs discussed the key role of HRII in preparing patients for medical consultations. They discussed potential practical and ethical benefits, for example, how more informed patients saved time and were enabled to be actively involved in discussions about treatment decisions. Reflecting experiences across the illness trajectory, patients described how they prepared for consultations with their general practitioner (GP) and specialists or allied health professionals for specific reasons (eg, to gain a referral or to get advice on treatments or exercise). Using HRII, they felt newly equipped to interact in focused and effective ways. This made "a huge difference…finding information, and what it means, before you go to the doctor so you can have an intelligent conversation…[and] ask them the right questions" (P3, PG3). Another patient compared how her consultation style had changed: "Before it was…'I ache'", whereas with the benefit of HRII, "It's allowed me to…narrow it down…'I want that, that and that clarified'…and I won't ask or talk about other things" (P2, PG2). This participant described "working" on her next visit about "what medications and how to ask for them and how to report back which ones are not working" (P2, PG2). These words indicate the need for patients to spend time and effort and develop skills in order to be fully involved in the decision making process.

Being prepared also had explicitly relational benefits for patients: "I talk to them [GP and rheumatologist] about what I feel…and what I found [online]…And it's a very respectful relationship with both of those doctors to me…the rheumatologist he is now listening to me a little more because he knows that I'm doing my homework" (P5, PG1). This is another indication of changing roles and relationships—being prepared for the consultation for this patient meant gaining respect from the HCP.

HCPs agreed that prepared patients allowed a more interactive consultation in which patients were equipped to ask relevant questions about care and treatment. One HCP described "a stellar client" who uses HRII and then emails her questions to follow-up in the consultation which "kind of gets me prepared, she's prepared" (P3, HCPG2). Another HCP reflected the positive aspects of HRII for both patient and HCP roles: "Much of it is empowering patients via education. So them coming and having done their pre-reading…gives you more time to focus on what you're going to assess, treat, and actually impact" (P1, HCPG2). HCPs noted patient preparation as key to more effective and time efficient consultations, bringing focus, and allowing the HCP more time to provide good care.

Responsibilities

Patient participants discussed how being prepared for consultations made a difference to the care they received. For some, this meant an added responsibility for their patient role. One participant noted how:

"Bringing [HRII] stuff in…makes a difference in the treatment because…it seems like I just get the basics unless I'm doing the pro-activeness…So it's just my responsibility to look into what I think might work and then ask for their opinion" (P6, PG3).

Searching for and sharing information in the consultation was framed as not only helping themselves, but their doctor, and in some cases other patients: "[Doctors are] not miracle makers…They don't think about certain things…We do have a responsibility to come to them with that [information] because not only are we helping ourselves, we're helping them to help other people" (P1, PG2).

Being responsible information-seeking patients had its risks and meant assessing the reliability of HRII for oneself. One patient participant spontaneously framed this as an ethical responsibility:

> *The ethics are…with yourself because you make the final decision…based on information that you get from all different sources…You…check the…authenticity…You try to find references to the same product or theory…By the end you have to decide for yourself. The government can't protect you from it. I don't know who else could except yourself.* [P3, PG1]

This extract illustrates one example of how some patients stated the importance of patients (and not only HCPs) to act ethically in new ways. It is clear from this patient's perspective that it is the patient's responsibility to gather the information they need to deal with health concerns and to be equipped with information in order to participate in focused discussion in the medical encounter. This implies a shift in the burdens of responsibility from the HCP to the patient. From a relational ethics perspective, it is questionable as to how reasonable and fair such a shift is for all patients. But what we note here is a perception of redistributed responsibilities. Not all patients in our study were wholly comfortable with this level of responsibility.

Partnerships

Patients, all of whom had multiple health conditions, discussed teamwork and emphasized the importance of effective communications between themselves and multiple HCPs. One patient expressed the mechanisms underlying partnering with a team of HCPs:

> *P3, P5 and P4, you touched on something very important…when we're diagnosed, we do the research even for our other un-arthritis related ailments…You're the captain of your team you take that [HRII] back to your doctor or your rheumatologist…and they know your total package. And then [you] encourage the physio, the occupational [therapists]…your pharmacist and your other specialist, your dentist…It all interacts…all parcel of…having an inflammatory disease. But they all have to be on the same page…it's your body, it's your stuff happening…you're the one that kind of brings them all together…to make sure that…there is that inter-professional [collaboration] going on.* [P2, PG1]

One patient conveyed how HRII equipped her to change the power balance, be heard, and get support from HCPs, noting "It's ammunition" (P1, PG2). Being heard, however, could be reliant on the relationship and how the HCP responds:

> *It depends…how good your doctor is…if your doctor understands that you have Internet access and that you have some understanding of what you are reading then it's a mutual respective relationship…But if your doctor doesn't understand…you probably need to look for another doctor…It's really not something you want to do when you have multiple chronic conditions. But…sometimes it's necessary because you've got to find somebody that will understand that you have this history and you understand a certain amount and you're wanting to work towards controlling and managing it. And you need his support.* [P3, PG3]

Reflecting the patient views, the HCP groups also discussed changes in the patient-HCP relationship. In some cases, patients brought HRII into the consultation and it prompted discussion, which in turn fostered a partnership form of interaction:

> *You just have to be really open to the fact that they're [patients] going to tell you things you didn't know and that's great. "Oh I hadn't seen that before. That might be useful for me with other clients". So I definitely feel it's more of a partnership…[like] P2 says it's much less didactic…Like P5 said, you just put in context what they've already brought to the table.* [P1, HCPG2]

Another HCP group discussed how "It's not the old medical hierarchy…(Now) we're negotiating...'What have you read? You've seen that, cool, let me see it.' So it's way more interactive because of the abundance of knowledge" (P2, HCPG1). It was clear that some consultations were changing to a more mutual information exchange and interactive discussion, based on HRII introduced by the patient or websites recommended by HCPs to patients: "Some of those websites have been successful in complementing and supplementing my care and helping to improve the partnership aspect of care" (P5, HCPG2).

In the context of partnership building, there was some evidence to suggest the need to negotiate a balance between ethical concerns of autonomy and beneficence. The HCPs discussed how respecting informed patients' decisions could mean avoiding "pushing" what they as HCPs saw as more medically effective (see quote below). There was a suggestion that patients are more equipped and enabled to take part in discussions. For HCPs this meant negotiating a new space, providing medical and health advice while respecting potentially differing opinions. This indicates the need for HCPs to develop new skills in consultations. One HCP described some of the consequences of informed patients, in the context of HRII and the balance between a more traditional approach and a more patient-led approach:

> *All we can do is inform people, give them our best medical opinion. And they have to make their own decision based on what they've heard from us and what other information they have. So I don't think you're always successful in convincing them to your side but I think we try pretty hard not to push our opinions on too.* [P2, HCPG3]

Participants described a negotiated element of interactions. One HCP described how new roles involved looking at information patients brought to the consultation, weighing up its relevance to the planned session and how credible the information sounds, while being respectful of the patient contribution: "It's trying to find a common path to work forward" (P3, HCPG2). Another HCP noted:

> *By giving her [patient] that list of websites and saying, "Okay this will be a partnership. I…will look for the things that might be triggering this [pain]. But I want you to look for and learn about the things that might be amplifying it. And if we agree to do this, this will be our partnership…" The websites…did enhance and improve the partnership aspect. I'm learning from her and she's learning from me…That's one of the ways of how it's affected my relationship.* [P5, HCPG2]

Trust

For patients, trust was key to sharing HRII with HCPs in consultations and encouraged them to share: "Because you trust them [HCPs], you're taking it [HRII] to them…" (P2, PG1). This sharing included learning to interact in a more open way than previously, for example, sharing concerns about medication use: "I've learned to be a bit more open with him [rheumatologist]…honest…Describing the side effects or my fears about a drug and…not continuing with it" (P2, PG2). Patients also appreciated HCPs being open about what they did not know as this built trust: "I think it's hard for some of them to say, 'I don't know that. I'll need to refer you.' Which I think is the smartest answer. That makes me have a lot of faith in a doctor" (P2, PG2).

HCPs also described how HRII could build trust:

> *If you spend that last five minutes…showing them [patients]…"This is a website that you can read too. It's got enough information but not too much and it won't overwhelm you. This is endorsed by the Canadian Arthritis Society." It kind of builds a level of trust and…adds a component of enrichment to the appointment… they read about it and I think they just feel a lot more like, empowered and cared for … equipped.* [P1, HCPG3]

In this sense, trust was based on a partnership model of care whereby the patient trusts the HCP to support her becoming informed, equipped, and empowered via guided direction to and interpretation of HRII.

Tensions and Burdens

Although the patient discussions around HRII were mostly positive, there were downsides: "It may be a burden to find all this information if you don't have the time, the energy, or the physicality" (P1, PG1). This situation raised tensions for patients: on the one hand, they saw HRII as an important resource, but many also noted negative aspects. One patient described online searching as "frustrating" and "a struggle" and stopped searching commenting: "I believe in the ostrich approach" (P4, PG4). For this patient with multiple conditions, the search for HRII became too burdensome to continue. Although patients expressed feeling informed, in control, and empowered, their HCP was a key resource to guide, verify, and advise and "have that funnel" (P2, PG1) to clarify their thinking. In this way, the role of the HCP was as an expert guide who could ease the navigation struggles through an array of HRII. Patients also noted that discussions required the HCP to "be ready" to engage with them. One patient felt that family doctors found it "threatening…if we look things up…GPs have to accept this as the new patient…We're going to…find out as much as we can…And then we're going to go in and ask the questions. And they better be ready" (P3, PG3).

HCPs agreed it was key to support patients in their Internet use, but there were associated tensions, for example, negotiating time constraints: "I have to find ways to give them [patients] as much knowledge in as short amount of time…the amount of information that's appropriate and not excessive" (P2, HCPG2). This also meant being sensitive to the needs of the patient with multiple chronic conditions and symptoms such as pain, fatigue, and "brain fog", or there was a risk of patients "falling off the system". HCPs acknowledged that patients might bring information to consultations that could be a little "out of their realm". Several HCPs described how important it was to welcome this development: "I've…decided that right upfront if somebody has clearly done way more reading into an area that I'd ever done I just say: 'Wow, you know more about that than I do'…It's really important not to feel threatened by that information because…if you [did]…that will affect your relationship" (P2, HCPG2).

Discussion

Principal Findings

Our focus group findings indicate practical and ethical implications of using HRII in chronic illness care, as care needs and patient-HCP relationships evolve. Patients and HCP discussed their changing roles and responsibilities [3] and expressed how HRII equips patients to engage in discussions with HCPs, who reflected on new ways of listening and responding to patients in light of HRII accessed by both parties. In practical terms, this mutual engagement required that patients work to prepare for consultations advancing a partnership model of care [23,49,50]. Our findings also highlight the importance of trust in the effective use of HRII in collaborative patient-HCP relationships. For example, the relational shift from a traditional hierarchical relationship to a more reciprocal relationship was associated with burdens and tensions. The presence of multimorbidity for patients involves multiple medications, risks, appointments, self-management strategies, and HCPs. These cascading complexities together with rapidly evolving HRII and increased expectations placed on patients, infuse shared decision-making. How patients and HCPs relate to one another as roles and responsibilities evolve requires careful consideration. Hence, understanding the detail of the relational aspects of medical encounters are key to ethical and effective care. Both patients and HCPs conveyed medical encounters as "negotiated spaces" where mutual vulnerabilities were revealed as patients and HCPs sought to relate to each other in a changing context of care. In this context, mutual trust could foil tensions.

Our findings resonate with previous studies of how eHealth technologies impact patient-HCP roles and relationships in a range of ways. HRII can support the decision-making process [51], yet there remains potential to provoke tensions [27,52,53] and HCPs can experience anxiety around sharing HRII with patients [54]. One literature review about the impact of HRII and patient-HCP relationships [55] concluded that a shift in the role of the patient from passive recipient to active consumer of health information prompted three types of HCP response: they felt threatened and were defensive, they collaborated in accessing the information, or they guided patients to reliable health information websites. The HCP participants in our study tended to agree on the importance of collaboration and guidance to foster care, yet acknowledged it was difficult to find the time to do so efficiently.

Another focus group study of patients with chronic illness [56] reported that patients saw the Internet as an additional resource to support valued relationships with their doctors and conveyed no desire to disrupt the existing balance of power in the consultation. We found something different. Our findings indicate patient empowerment as present and valued by patients, and that for some patients and HCPs, the balance of power in medical encounters shifted with the use of HRII. Valued relationships involved not just support and guidance from HCPs, but recognized mutual vulnerabilities. For example, patient trust in HCPs was strengthened when HCPs were open about not knowing, and patients felt confident in sharing with HCPs the HRII they had found. This finding differs from research that reported trust could be hampered due to increased patient knowledge and empowerment [57]. For the patient participants in our study, HRII prompted greater interdependence between patients and HCPs and increased opportunities for partnerships, which in turn were underpinned by key ethical values of mutual trust and respect—cornerstones of shared decision-making. This supports findings from a recent focus group study of patients with complex chronic conditions in Canada that identified the need for open two-way dialogue to improve coordination of “whole person” care [18], which may be particularly important for those with multiple conditions with holistic needs [58].

Our findings also showed how patients and HCPs learned techniques to negotiate an evolving consulting space and avoid conflict. For HCPs, this includes a dimension of risk and uncertainty about how to offer appropriate guidance or how to acknowledge that the patient may know more than them. This reflects Zufferey et al’s [28] perspective indicating a need to work at engagement by increasing professionals’ level of reflexivity—a continuous and demanding “work” to adjust one’s behavior to others. We also found that patients did work—the homework or preparation work that interactive medical encounters required resonates with previous sociological research on the unending work of chronic illness [59] including moral dimensions [60] and ethical perspectives on the high degree of patient work needed for collaborative decision-making [61]. Our findings suggest new aspects of unending work for both patients and HCPs as they navigate extensive HRII sources and negotiate new ways of interacting with each other.

Being “overloaded” with “endless” HRII has been documented elsewhere [62,63] and resonates with our findings. Taking a relational approach, we emphasize the implications of this for the roles and responsibilities of HCPs who are now required to guide those patients who may feel overwhelmed, and without relevant support may choose to take the “ostrich approach” and withdraw from HRII seeking. Our findings, however, also suggest that in some cases, patients’ use of HRII can point to open, collaborative patient-HCP relationships; some HCPs in our focus groups described how they recommend and discuss HRII with patients as part of practice. This mirrors aspects of recent research [63], which concludes that patients need skill development [63] in order to use HRII effectively. A relational ethics approach recognizes the burdens and vulnerabilities of HCPs too [40], and acknowledges that HCPs also require guidance, training, and skill development if they are to use HRII effectively.

Values that underpin health care are fundamentally ethical in nature. Core relational values identified in this study of mutuality, responsibility, honesty, and trust provide the foundation for ethically appropriate and effective care. These values were consistent across the patients and HCPs. A relational lens encourages consideration of the details of patient-HCP dynamics and reveals how interactions can support or obstruct the realization of prized values that are emerging in HRII consultations [37]. In presenting an ethical analysis, our paper contributes to the sparse literature that explicitly addresses ethical challenges of medicine and health on the Internet [64], in terms of patient-HCP relationships. This perspective also aligns with Medlock et al, who see patient empowerment as a “practical and moral necessity” [65].

Limitations and Strengths

Although we aimed to recruit a diverse sample of participants in order to examine a range of experiences of using HRII, our sample was not as varied as intended. Proportionately, we lacked views from male and younger patients, and transferability of findings is limited accordingly. We were able to compare, however, HCP and patient views and experiences of HRII and gain an understanding of the key emerging ethical and relational aspects specific to living with and caring for people with the complexity of multiple health conditions. The findings presented here are also limited to use of HRII and are not implications of other eHealth tools and devices.

Practice Implications

The relational ethics framework applied in this study may help advance evolving patient and HCP partnership roles consistent with the value of patient-centered care. Drawing on mutual respect and a more reciprocal understanding of relationships in terms of sharing knowledge, providing support, and recognizing vulnerabilities [40] should inform reflective practice and health care communications. Relational ethics then offers guidance to HCPs in their everyday clinical practice (thinking, reasoning, and decision-making) and expands the professional ethics and responsibility that already guide practice [32]. The promotion of shared decision-making is a key component of ethical patient-centered care [41] and is also regarded as a way to optimize the use of scarce resources in health care [32].

Given the complexity of managing information for multiple health conditions from multiple sources, patients reported varying levels of comfort with responsibility for shared decisions. While HCPs are generally aware of patient vulnerabilities, our findings nevertheless point out the need for negotiating this shared decision-making space—a considerable challenge in many practice contexts given time constraints and busy practices. This is highlighted when comparing more active, engaged patients with the patients who are less so in terms of self-management and shared decision-making. HCPs need to be aware of how some people with multiple chronic conditions will be “captain of their ship”, take ownership/control, and perhaps be more vigorous in shared decision-making than others. This places a burden on the HCP to understand their patients’ preferences and styles and to be aware of a range of HRII that patients may find helpful in order to engage in shared decision-making. Examining relational positioning is key to understanding changing patient-HCP relationships. Our findings offer a better understanding of the patient-HCP relationship and communications in the context of the rapid adoption of HRII.

Suggestions for Research

Research, guidelines, and theories about the adoption of HRII have not kept pace with technological developments, adoption of eHealth tools and devices, and increasing ease of access to information [13]. This presents a pressing need for a better understanding of the complexities emerging in the evolution of role and relationships of patients and HCPs including complex and everyday clinical decision-making. Relational understandings emphasize mutual vulnerability and interdependence needed for effective teamwork and encourage detailed consideration of health care interactions and patient autonomy in patient-HCP relationships and how for example HCPs can support and undermine patients' autonomy [42]. In their response to patients' preparation for visits, HCPs, perhaps unknowingly, have the potential to undermine a patient's identity and self-evaluation (which are basic to their sense of being able to function autonomously). They have the opportunity, however, to encourage patients to seek and use autonomy-supporting resources such as relevant websites or patient groups [42]. Our findings revealed that patients used the HRII to equip them with information to participate meaningfully in discussions in medical consultations. This interactive process suggested increasingly interdependent and reciprocal relationships between patients and HCPs.

Conclusion

The impact of HRII on the patient-HCP relationship will continue to grow. Considering future implications, our research offers new insights into how both patients and HCPs adopt strategies to better negotiate the changing nature of the patient-HCP relationship. Our findings add a new dimension to shared decision-making in the era of HRII by emphasizing the fundamental relational and ethical aspects, including the negotiations, vulnerabilities, and trust that are part of an ongoing process in clinical encounters for a partnership model. Relational ethics is a novel approach that in this context attends to the moral space where this care occurs.

Acknowledgments

The authors thank the patients and clinicians who participated in this study. We are also grateful for the support of patient/consumer collaborators including Mr Gordon Whitehead (Arthritis Research Center Consumer Advisory Board) and recruitment assistance from Ms Cheryl Koehn (Joint Health) and Ms Colleen Maloney (Canadian Arthritis Patient Alliance). We also thank Kam Shojania for his contributions to conceptualizing the study and recruitment. The Canadian Institutes for Health Research, Catalyst grants competition EPP-122907 funded this research in October 2012.

Conflicts of Interest

None declared.

References

1. Townsend A, Adam P, Li LC, McDonald M, Backman CL. Exploring eHealth Ethics and Multi-Morbidity: Protocol for an Interview and Focus Group Study of Patient and Health Care Provider Views and Experiences of Using Digital Media for Health Purposes. JMIR Res Protoc 2013;2(2):e38 [FREE Full text] [doi: 10.2196/resprot.2732] [Medline: 24135260]
2. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Aug;11:256-270. [doi: 10.1057/sth.2013.10)]
3. Hordern A, Georgiou A, Whetton S, Prgomet M. Consumer e-health: an overview of research evidence and implications for future policy. HIM J 2011;40(2):6-14. [Medline: 21712556]
4. Patient-Centered E-Health (Premier Reference Source). New York: Medical Information Science Reference; 2009.
5. World Health Organization. WHA58-28 on eHealth. 2005. Fifty-eighth World Health Assembly, Resolution 28, Geneva, Switzerland URL: http://apps.who.int/gb/ebwha/pdf_files/WHA58-REC1/english/A58_2005_REC1-en.pdf [accessed 2014-08-17] [WebCite Cache ID 6RsrrbRNZ]
6. Munson SA, Cavusoglu H, Frisch L, Fels S. Sociotechnical challenges and progress in using social media for health. J Med Internet Res 2013;15(10):e226 [FREE Full text] [doi: 10.2196/jmir.2792] [Medline: 24148206]
7. Swan M. Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. J Pers Med 2012;2(3):93-118 [FREE Full text] [doi: 10.3390/jpm2030093] [Medline: 25562203]
8. Lupton D. The Commodification of Patient Opinion: the Digital Patient Experience Economy in the Age of Big Data. 2014. URL: http://ses.library.usyd.edu.au/handle/2123/9063 [accessed 2014-08-17] [WebCite Cache ID 6RsrzKp4a]
9. Swan M. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. JSAN 2012 Nov 08;1(3):217-253. [doi: 10.3390/jsan1030217]
10. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
11. 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]
12. Ho K, Peter Wall Workshop Participants. Harnessing the social web for health and wellness: issues for research and knowledge translation. J Med Internet Res 2014;16(2):e34 [FREE Full text] [doi: 10.2196/jmir.2969] [Medline: 24518432]
13. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
14. 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]

15. 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;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
16. Department OH. The expert patient: a new approach to chronic disease management for the 21st century. 2014. URL: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_4006801 [accessed 2014-08-17] [WebCite Cache ID 6Rttqbd3L]
17. Schildmann J, Grunke M, Kalden JR, Vollmann J. Information and participation in decision-making about treatment: a qualitative study of the perceptions and preferences of patients with rheumatoid arthritis. J Med Ethics 2008 Nov;34(11):775-779. [doi: 10.1136/jme.2007.023705] [Medline: 18974408]
18. Steele GC, Miller D, Kuluski K, Cott C. Tying eHealth Tools to Patient Needs: Exploring the Use of eHealth for Community-Dwelling Patients With Complex Chronic Disease and Disability. JMIR Res Protoc 2014;3(4):e67 [FREE Full text] [doi: 10.2196/resprot.3500] [Medline: 25428028]
19. Pew Research Center. Health Online 2013. 2015. URL: http://pewinternet.org/Reports/2013/Health-online.aspx [accessed 2015-03-16] [WebCite Cache ID 6Xe2qgtk5]
20. Williamson L. Patient and citizen participation in health: the need for improved ethical support. Am J Bioeth 2014;14(6):4-16 [FREE Full text] [doi: 10.1080/15265161.2014.900139] [Medline: 24809598]
21. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the Internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
22. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [Medline: 12617199]
23. Gauld R, Williams S. Use of the Internet for health information: a study of Australians and New Zealanders. Inform Health Soc Care 2009 Sep;34(3):149-158. [doi: 10.1080/17538150903102448] [Medline: 19670005]
24. Wickramasinghe N, Bali RK, Suomi R, Kirn S, editors. Patient Empowerment: A Two Way Road. In: Critical Issues for the Development of Sustainable E-health Solutions (Healthcare Delivery in the Information Age). New York: Springer; 2010.
25. 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 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
26. Seçkin G. Health information on the web and consumers' perspectives on health professionals' responses to information exchange. Med 2 0 2014;3(2):e4 [FREE Full text] [doi: 10.2196/med20.3213] [Medline: 25075248]
27. Sommerhalder K, Abraham A, Zufferey MC, Barth J, Abel T. Internet information and medical consultations: experiences from patients' and physicians' perspectives. Patient Educ Couns 2009 Nov;77(2):266-271. [doi: 10.1016/j.pec.2009.03.028] [Medline: 19411157]
28. Caiata-Zufferey M, Schulz PJ. Physicians' communicative strategies in interacting with Internet-informed patients: results from a qualitative study. Health Commun 2012;27(8):738-749. [doi: 10.1080/10410236.2011.636478] [Medline: 22260422]
29. Imes RS, Bylund CL, Sabee CM, Routsong TR, Sanford AA. Patients' reasons for refraining from discussing internet health information with their healthcare providers. Health Commun 2008 Nov;23(6):538-547. [doi: 10.1080/10410230802460580] [Medline: 19089701]
30. Fujioka Y, Stewart E. How do physicians discuss e-health with patients? the relationship of physicians' e-health beliefs to physician mediation styles. Health Commun 2013;28(4):317-328. [doi: 10.1080/10410236.2012.682971] [Medline: 22716050]
31. Bowes P, Stevenson F, Ahluwalia S, Murray E. 'I need her to be a doctor': patients' experiences of presenting health information from the internet in GP consultations. Br J Gen Pract 2012 Nov;62(604):e732-e738 [FREE Full text] [doi: 10.3399/bjgp12X658250] [Medline: 23211176]
32. Elwyn G, Tilburt J, Montori V. The ethical imperative for shared decision-making. Eur J Pers Centered Healthc 2013;1(1):129-131. [doi: 10.5750/ejpch.vlil.645)]
33. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
34. Armstrong D. Actors, patients and agency: a recent history. Sociol Health Illn 2014 Feb;36(2):163-174. [doi: 10.1111/1467-9566.12100] [Medline: 24372176]
35. Henry MS. Uncertainty, responsibility, and the evolution of the physician/patient relationship. J Med Ethics 2006 Jun;32(6):321-323 [FREE Full text] [doi: 10.1136/jme.2005.013987] [Medline: 16731728]
36. Felzman H. Adherence, compliance, and concordance: an ethical perspective. Nurse Prescribing 2012;10(8):406-411.
37. Agarwal AK, Murinson BB. New dimensions in patient-physician interaction: values, autonomy, and medical information in the patient-centered clinical encounter. Rambam Maimonides Med J 2012 Jul;3(3):e0017 [FREE Full text] [doi: 10.5041/RMMJ.10085] [Medline: 23908841]

38. Sherwin S. A Relational Approach to Autonomy in Health Care in The Feminist Health Care Ethics Research Network. In: Sherwin S, editor. The Politics of Women's Health: Exploring Agency and Autonomy. Philadelphia: Temple University Press; 1998:19-47.
39. Beauchamp T, Childress J. Principles of biomedical ethics. Oxford: Oxford University Press; 1989.
40. Austin W. Texto Contexto Enferm. 2014. Engagement in contemporary practice: a relational ethics perspective URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-07072006000500015&lng=en [accessed 2014-08-16] [WebCite Cache ID 6Xe5gYvfl]
41. Fullwood C, Kennedy A, Rogers A, Eden M, Gardner C, Protheroe J, et al. Patients' experiences of shared decision making in primary care practices in the United kingdom. Med Decis Making 2013 Jan;33(1):26-36. [doi: 10.1177/0272989X12464825] [Medline: 23128580]
42. Entwistle VA, Carter SM, Cribb A, McCaffery K. Supporting patient autonomy: the importance of clinician-patient relationships. J Gen Intern Med 2010 Jul;25(7):741-745 [FREE Full text] [doi: 10.1007/s11606-010-1292-2] [Medline: 20213206]
43. Trobec I, Herbst M, Zvanut B. Differentiating between rights-based and relational ethical approaches. Nurs Ethics 2009 May;16(3):283-291. [doi: 10.1177/0969733009102689] [Medline: 19372123]
44. Townsend A, Adam P, Cox SM, Li LC. Everyday ethics and help-seeking in early rheumatoid arthritis. Chronic Illn 2010 Sep;6(3):171-182. [doi: 10.1177/1742395309351963] [Medline: 20610465]
45. Ells C, Hunt M, Chambers-Evans J. Relational Autonomy as an Essential Component of Patient-Centered Care. Engagement in Contemporary Practice: A Relational Ethics Perspective 2011;4(2):79-101. [doi: 10.1353/ijf.2011.0025]
46. Public Health Agency of CanadaLife with Arthritis in Canada. A Personal and Public Health Challenge. 2013. URL: http://www.phac-aspc.gc.ca/cd-mc/arthritis-arthrite/lwaic-vaaac-10/pdf/arthritis-2010-eng.pdf [accessed 2015-04-08] [WebCite Cache ID 6Xe6KvqTg]
47. Townsend A, Wyke S, Hunt K. Frequent consulting and multiple morbidity: a qualitative comparison of 'high' and 'low' consulters of GPs. Fam Pract 2008 Jun;25(3):168-175 [FREE Full text] [doi: 10.1093/fampra/cmn017] [Medline: 18448858]
48. Mercer S, Gunn J, Wyke S. Improving the health of people with multimorbidity: the need for prospective cohort studies. Journ Comorbid 2011;1:4-7. [doi: 10.15256/joc.2011.1.10)]
49. Dedding C, van DR, Winkler L, Reis R. How will e-health affect patient participation in the clinic? A review of e-health studies and the current evidence for changes in the relationship between medical professionals and patients. Soc Sci Med 2011 Jan;72(1):49-53. [doi: 10.1016/j.socscimed.2010.10.017] [Medline: 21129832]
50. 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]
51. Hoffmann TC, Légaré F, Simmons MB, McNamara K, McCaffery K, Trevena LJ, et al. Shared decision making: what do clinicians need to know and why should they bother? Med J Aust 2014 Jul 7;201(1):35-39. [Medline: 24999896]
52. Henderson S. Power imbalance between nurses and patients: a potential inhibitor of partnership in care. J Clin Nurs 2003 Jul;12(4):501-508. [Medline: 12790863]
53. Latter S, Maben J, Myall M, Young A. Perceptions and practice of concordance in nurses' prescribing consultations: findings from a national questionnaire survey and case studies of practice in England. Int J Nurs Stud 2007 Jan;44(1):9-18. [doi: 10.1016/j.ijnurstu.2005.11.005] [Medline: 16359677]
54. Ahluwalia S, Murray E, Stevenson F, Kerr C, Burns J. 'A heartbeat moment': qualitative study of GP views of patients bringing health information from the internet to a consultation. Br J Gen Pract 2010 Feb;60(571):88-94 [FREE Full text] [doi: 10.3399/bjgp10X483120] [Medline: 20132702]
55. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
56. Stevenson FA, Kerr C, Murray E, Nazareth I. Information from the Internet and the doctor-patient relationship: the patient perspective--a qualitative study. BMC Fam Pract 2007;8:47 [FREE Full text] [doi: 10.1186/1471-2296-8-47] [Medline: 17705836]
57. Calnan M, Rowe R. Trust and Health Care. Sociol Compass 2007;1(1):283-308. [doi: 10.1111/j.1751-9020.2007.00007.x]
58. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How Can eHealth Technology Address Challenges Related to Multimorbidity? Perspectives from Patients with Multiple Chronic Conditions. J Gen Intern Med 2015 Feb 18:-. [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
59. Corbin JM, Strauss A. Unending Work and Care: Managing Chronic Illness at Home (Jossey Bass Social and Behavioral Science Series). San Francisco: Jossey-Bass; 1988.
60. Townsend A, Wyke S, Hunt K. Self-managing and managing self: practical and moral dilemmas in accounts of living with chronic illness. Chronic Illn 2006 Sep;2(3):185-194. [doi: 10.1179/174592006X129518] [Medline: 17007695]
61. Groopman LC, Miller FG, Fins JJ. The patient's work. Camb Q Healthc Ethics 2007;16(1):44-52. [Medline: 17345966]
62. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]

63. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
64. Dyer KA. Ethical challenges of medicine and health on the Internet: a review. J Med Internet Res 2001;3(2):E23 [FREE Full text] [doi: 10.2196/jmir.3.2.e23] [Medline: 11720965]
65. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de Rooij SE, et al. Health information-seeking behavior of seniors who use the Internet: a survey. J Med Internet Res 2015;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]

Abbreviations

GP: general practitioner
HCP: health care professionals
HRII: health-related Internet information

Edited by G Eysenbach; submitted 17.08.14; peer-reviewed by D Lupton, A Hartzler, V Entwistle, M Bujnowska-Fedak; comments to author 17.12.14; revised version received 16.03.15; accepted 24.05.15; published 22.06.15.

Please cite as:
Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, Backman CL
eHealth, Participatory Medicine, and Ethical Care: A Focus Group Study of Patients' and Health Care Providers' Use of Health-Related Internet Information
J Med Internet Res 2015;17(6):e155
URL: http://www.jmir.org/2015/6/e155/
doi:10.2196/jmir.3792
PMID:26099267

Original Paper

Association Between Acute Medical Exacerbations and Consuming or Producing Web-Based Health Information: Analysis From Pew Survey Data

Risha Gidwani[1,2,3], Dr PH; Donna Zulman[2,3], MD, MS

[1]VA Health Economics Resource Center, Menlo Park, CA, United States

[2]VA Center for Innovation to Implementation, Menlo Park, CA, United States

[3]Division of General Internal Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Risha Gidwani, Dr PH
VA Health Economics Resource Center
795 Willow Road
152 MPD
Menlo Park, CA, 94025
United States
Phone: 1 650 493 5000 ext 22937
Fax: 1 650 617 2667
Email: risha.gidwani@va.gov

Abstract

Background: The Internet is an increasingly important resource for individuals who seek information from both health professionals and peers. While the demographic and health characteristics of persons who use health information technology has been well described, less is known about the relationship between these health characteristics and level of engagement with health information technology. Even less is known about whether persons who produce Web-based health information differ in health status from persons who consume such content.

Objective: We explored the health characteristics of persons who engage with the Internet for the purposes of consuming or producing Web-based health information, and specifically, whether healthier versus sicker persons engage with health information technology in different ways.

Methods: We analyzed data from the 2012 Pew Health survey, a landline and cell phone survey of 3104 adults in the United States. Using multiple logistic regression with sampling weights, we examined the association between sociodemographic and health characteristics and the consumption or production of Web-based health information. Sociodemographic variables included age, sex, race, and education. Health characteristics included self-reported health status, presence of chronic condition(s), and having an acute medical exacerbation. Acute medical exacerbations were defined as an emergency department visit, hospitalization, or other serious medical emergency in the last 12 months.

Results: The majority of the sample reported good or excellent health (79.7%), although 50.3% reported having at least one chronic condition. About a fifth (20.2%) of the sample experienced an acute medical exacerbation in the past year. Education was the sociodemographic characteristic most strongly associated with consuming Web-based health information. The strongest health-related predictors of consuming Web-based health information were an acute medical exacerbation (OR 2.39, P<.001) and having a chronic condition (OR 1.54, P=.007). Having an acute medical exacerbation was the only predictor of producing Web-based health information (OR 1.97, P=.003). All participants, regardless of health status, were most interested in Web-based health information regarding diseases or medical problems. However, persons with acute medical exacerbations were more likely to seek Web-based health information regarding medical tests, procedures, and drugs compared to persons without acute medical exacerbations.

Conclusions: Producers of Web-based health information differ from consumers of this information in important health characteristics that could skew the content of peer-generated Web-based health information and overrepresent the experiences of persons with acute medical exacerbations. Providers may have a role to play in directing patients towards high-quality, easy-to-understand online information, especially information regarding treatments and procedures.

(J Med Internet Res 2015;17(6):e145) doi:10.2196/jmir.3801

KEYWORDS

Internet; health knowledge, attitudes, practice; emergencies; hospitalization

Introduction

The Internet has become a key source of health information for many individuals. In 2012, 59% of adults looked for health information online [1]. A wide variety of sources are being searched for Web-based health information, including health care systems [2-4], government agencies [5,6], non-profit/for-profit organizations [7,8], and disease-specific advocacy groups [9-11].

Increasingly, patients are turning to other patients as trusted sources of information [12,13]. Some patients have developed websites or blogs where they share their experiences living with chronic conditions [14,15]. Others review medical treatments, procedures, health care products, and hospitals and providers online [16,17]. Individuals often seek patient-generated information after medical appointments, especially if their clinic visits generated anxiety or dissatisfaction [18]. However, while patient-generated health information can provide a unique and valuable perspective for others with shared medical issues or health care challenges, there are also concerns about the potential for dissemination of inaccurate or incomplete information [16,19]. Typically, a small number of super-users post the vast majority of content [12,13,20]. Little is known about the extent to which these individuals' sociodemographic characteristics, health conditions, and health care experiences reflect those of the majority of patients consuming information.

In order to engage patients through the Internet, it is important that Web-based health information addresses patients' specific needs and concerns and that sources are reliable, trustworthy, and relevant to the individuals seeking information. We sought to understand the circumstances under which patients consume and produce Web-based health information, and the influence of health status on these activities. Our objective was to understand the characteristics of Web-based health information consumers and especially the type of health-related information they sought online and to determine whether producers of health information represent the larger population of consumers with respect to sociodemographic and health characteristics.

Methods

We used data from the Pew 2012 Health survey, sponsored by the Pew Research Center's Internet & American Life Project, to evaluate predictors of engagement with Web-based health information. The English- and Spanish-language telephone (landline and mobile phone) survey was conducted with 3014 adults living in the United States. The survey collected self-reported data on persons' Internet use, health status, and demographic characteristics.

Our primary outcome variable was use of the Internet to consume information about diseases or health conditions, which we refer to as "health-related Internet use". A person was considered to have health-related Internet use if they reported using the Internet and reported looking online for any of the following in the previous 12 months: (1) a specific disease or medical problem, (2) trying to figure out what medical condition they or someone else might have, (3) a medical treatment or procedure, (4) how to lose or control weight, (5) a drug seen advertised, (6) drug safety or recall information, (7) any other health issue, (8) signing up for email updates regarding health or medical issues, (9) reading or watching someone else's health- or medical issue-related commentary or personal experience, (10) finding others with similar health concerns, or (11) consulting online rankings or reviews of providers, hospitals, medical facilities, drugs, or medical treatments.

We also investigated the type of engagement people had with the Internet and specifically whether they were consumers or producers of Web-based health information. A participant was considered a "consumer" if they had health-related Internet use, as described above. An Internet user was considered a "producer" if they reported (1) posting information about their own personal health experience online, or (2) posting a review of a provider, hospital, experience with a particular drug, or experience with a medical treatment online. We examined participant characteristics associated with consumer versus producer behavior and the specific types of information that were consumed and produced. In post-hoc analyses, we also evaluated the specific types of Web-based health information consumed by persons with and without acute medical exacerbations, including the source of this information (peer, professional, or both).

All analyses employed multivariate logistic regression. In each analysis, we included age, highest level of education achieved (less than high school, high school, some college, or 4-year college or above), sex, race (white, Black, Asian/Pacific Islander, mixed, or Native American), self-reported health status, presence of chronic conditions, and an acute health exacerbation within the past year as covariates. Self-reported health status was classified as poor/fair, good, or excellent. Presence of chronic conditions was indicated if the participant reported having diabetes, high blood pressure, a lung condition, a heart condition, cancer, or "another chronic condition". A participant was considered to have an acute medical exacerbation if they reported having an Emergency Department visit, a hospital visit, or a serious medical emergency in the previous 12 months. Income was not included in the model due to the high percentage of missing income data (18.7%) and concerns that income data were not missing at random.

All statistical analyses were conducted in Stata, version 12.0. Tests of significance were two-tailed and used an alpha of .05. Regression analyses employed survey sampling weights provided by Pew to account for a sample that was disproportionately stratified by race, differences in landline versus mobile phone response rates, and demographically based differential non-response.

Results

Overview

The majority of the sample was white (69.60%, 2098/3014) and insured (86.86%, 2618/3014) (Table 1). A little more than half of the sample was female (55.64%, 1677/3014); the mean age (SD) was 52.6 (19.8). Over one-third of the sample had a college degree or higher (36.99%, 1115/3014), and 8.93% (269/3014) had less than a high school education. Over a third (38.69%, 1166/3014) of the sample earned less than US $40,000 annually. However, high earners were also well represented in this sample; 13.70% (413/3014) made more than US $100,000 per year. The majority of the sample reported good or excellent health (79.66%, 2401/3014), although half (1498/3014) also reported having at least one chronic condition. About a fifth (20.21%, 609/3014) of the sample experienced an acute medical exacerbation in the past year. Of the 2392 persons who answered questions about health-related Internet use, 1717 (71.78%) reported consuming Web-based health information. Only 254 (10.62%) reported producing Web-based health information. The vast majority (97.9%) of persons who produced Web-based health information also consumed Web-based health information. There were 622 participants (20.6%) who reported neither consuming nor producing Web-based health information.

Characteristics Associated With Health-Related Internet Use

Regression analyses incorporating survey sampling weights revealed differences in consumption of Web-based health information according to a number of demographic characteristics. Consumption of Web-based health information decreased with age and increased with education (Table 2). Participants with some college education or more were significantly more likely than persons who did not finish high school to consume Web-based health information (OR 2.24, *P*<.01). Participants with a 4-year college degree or higher had 3.99 times the odds of consuming Web-based health information compared with patients who did not finish high school (*P*<.001). Consumption of Web-based health information was equivalent among persons aged 18-29 and 30-44; after age 45 people were significantly less likely to consume Web-based health information. After adjusting for other covariates, there were no differences in consumption of Web-based health information by race.

Self-reported health status was not independently associated with consuming Web-based health information. However, presence of a chronic condition was associated with significantly more consumption of Web-based health information (OR 1.54, *P*<.01). Having an acute medical exacerbation was also a significant predictor of such consumption; in fact, of all the health-status-related measures, having an acute medical exacerbation was most strongly associated with consuming Web-based health information (OR 2.39, *P*<.001).

Table 1. Demographic characteristics, unweighted (N=3014).

	Mean (SD)	n	% (unweighted)
Age	52.6 (19.8)	3014	
Race			
White		2098	69.61
Black		546	18.12
Asian/Pacific Islander		85	2.82
Mixed		80	2.65
Native American		30	1.00
Missing		175	5.80
Sex			
Female		1677	55.64
Insurance			
None		396	13.14
VA/Other		73	2.42
Medicaid		208	6.90
Medicare		710	23.56
Medicaid + Medicare		197	6.54
Private		1373	45.55
Missing		57	1.89
Education			
Less than high school		269	8.93
High school		830	27.54
Some college		778	25.81
4-year college or more		1115	37.01
Missing		22	0.73
Income			
$10,000-20,000		584	19.38
$20,001-40,000		582	19.31
$40,001-75,000		604	20.04
$75,001-100,000		267	8.86
$100,001-150,000		413	13.70
Missing		564	18.71
Health status			
Poor/Fair		604	20.04
Good		1552	51.49
Excellent		849	28.17
Chronic condition			
Yes		1498	50.32
Acute medical exacerbation			
Yes		609	20.21

Table 2. Predictors of consuming any Web-based health information.

	Odds ratio (95% confidence interval)	*P* value
Sex		
Female	2.05 (1.57-2.68)	<.001
Race (Ref: White)		
Black	0.97 (0.70-1.35)	.87
Asian/Pacific Islander	1.72 (0.84-3.52)	.14
Mixed	1.44 (0.61-3.42)	.41
Native American	1.65 (0.29-9.47)	.57
Education (ref: less than high school)		
High school	1.16 (0.66-2.06)	.60
Some college	2.24 (1.25-4.02)	<.01
4-year college or more	3.99 (2.24-7.11)	<.001
Health status (ref: Poor/Fair)		
Good	1.13 (0.74-1.71)	.57
Excellent	1.27 (0.80-2.01)	.32
Age (ref: 18-29)		
30-44	0.80 (0.53-1.22)	.30
45-64	0.60 (0.41-0.89)	.01
65+	0.28 (0.18-0.44)	<.001
Chronic disease		
Yes	1.54 (1.13-2.10)	<.01
Acute medical exacerbation		
Yes	2.39 (1.61-3.56)	<.001

Characteristics Associated With Being a Consumer Versus Producer of Health-Related Internet Content

Having an acute medical exacerbation was significantly associated with producing Web-based health information (Table 3). Persons with an acute medical exacerbation had 1.97 the odds of producing Web-based health information compared to people without an acute medical exacerbation (*P*=.003). In contrast, self-reported health status and the presence of a chronic condition were not significantly associated with being a producer of Web-based health information. Similarly, no demographic variables (including sex, race, age, and education) were associated with producing Web-based health information.

Types of Web-based Health Information Consumed and Produced by Individuals With and Without Acute Medical Exacerbations

In a post-hoc analysis, we evaluated the type of Web-based health information consumed and produced by participants with and without acute medical exacerbations. In multivariate regression, compared to participants without acute medical exacerbations, participants with acute medical exacerbations were much more likely to seek medical treatment or procedure information, drug safety information, medical test results information, information about drugs advertised, and reviews of providers or treatments (Table 4). Data indicate that regardless of whether patients had acute medical exacerbations, the type of Web-based health information they were most likely to consume was that regarding a disease or medical problem. We found no difference among patients with and without acute medical exacerbations in terms of the source of the Web-based health information they consumed (peer, professional, or both). There were also no differences among patients with and without acute medical exacerbations with respect to the type of Web-based health information they produced (personal health experience vs reviews of providers, hospitals, or treatments).

Table 3. Predictors of producing (versus consuming) Web-based health information.

	Odds ratio (95% confidence interval)	*P* value
Sex		
Female	1.07 (0.74-1.55)	.72
Race (Ref: White)		
Black	0.82 (0.50-1.35)	.44
Asian/Pacific Islander	0.78 (0.26-2.36)	.66
Mixed	1.44 (0.57-3.63)	.44
Native American	0.79 (0.21-2.97)	.56
Education (ref: less than high school)		
High school	0.83 (0.35-2.00)	.68
Some college	0.94 (0.40-2.22)	.89
4-year college or more	1.10 (0.48-2.55)	.82
Health status (ref: Poor/Fair)		
Good	1.09 (0.62-1.90)	.77
Excellent	0.80 (0.42-1.54)	.51
Age (ref: 18-29)		
30-44	1.53 (0.89-2.63)	.18
45-64	1.02 (0.59-1.76)	.93
65+	0.55 (0.29-1.05)	.07
Chronic disease		
Yes	1.42 (0.91-2.21)	.12
Acute medical exacerbation		
Yes	1.97 (1.26-3.08)	<.01

Table 4. Type of consumption among persons with acute medical exacerbations (after multivariate adjustment).[a]

Type of consumption	% of patients with acute medical exacerbations	% of patients without acute medical exacerbations	*P* value
Disease or medical problem information	95.4	92.0	.09
Medical treatment or procedure information	72.3	58.3	.001
Drug safety information (including recalls)	29.3	19.6	<.01
Medical test results information	28.5	16.9	.001
Losing or controlling weight	36.9	35.9	.81
A drug advertised	28.0	19.0	<.01
Reviews of providers or treatments	49.9	40.6	.03

[a] Multivariate regression adjusted for education, health status, chronic conditions, race, age, and sex.

Discussion

Principal Findings

Our analyses revealed a number of demographic characteristics associated with consumption of Web-based health information. Females, persons with some college education or more, and persons younger than 44 years were significantly more likely to consume Web-based health information. Conversely, no demographic variable was significantly associated with *producing* Web-based health information. These results corroborate other studies, which have found that Web-based health information-seeking typically increases with education and income and decreases with age [1,17,21,22].

In evaluating health-related variables and consumption and production of Web-based health information, we found having a chronic condition and having an acute medical exacerbation were independently associated with Web-based health information consumption, with acute medical exacerbations being the stronger predictor of this type of consumption. Having an acute medical exacerbation was the only health status–related predictor of becoming a *producer* (rather than a consumer) of Web-based health information. Therefore, having an acute medical exacerbation or a chronic condition is associated with greater engagement with consuming health information technology. For persons who already consumed Web-based health information, having an acute medical exacerbation is associated with becoming a producer of such information.

While over 90% of patients with and without acute medical exacerbations seek Web-based information about health conditions, there are some important differences about the other types of Web-based health information they consume. Persons with acute medical exacerbations in the past 12 months are significantly more likely to seek information regarding medical treatments or procedures, medications, and medical treatment results, and reviews of providers or treatments compared with persons who did not experience an acute medical exacerbation in the past 12 months.

Results indicate that having an acute medical exacerbation was the only significant predictor of producing Web-based health information. This suggests that peer-to-peer health-related Internet content may be skewed towards persons who are sicker. This is important to recognize, given that patients with stable chronic conditions are often consuming this peer-generated information. This overrepresentation of the experiences of persons with acute medical exacerbations may mean there is not enough peer-generated Web-based health information available regarding stable chronic conditions and that the information available may not be applicable to those with stable chronic conditions who wish to prevent an acute medical exacerbation.

Our analysis also revealed the discrepancy between objective evaluations of health and interpretation of one’s own health. The vast majority of our sample had a self-reported health status of good or excellent, yet over half reported having at least 1 chronic condition. Among persons with a chronic condition, 16% reported their health as excellent and half reported their health as good. Among persons with an acute medical exacerbation, 16% reported their health as excellent and less than half reported their health as good. These results suggest that patients are likely to overestimate their own health status.

Having an acute medical exacerbation may be a “window of opportunity” in which health-related online behavior changes. Patients appear to be seeking health information at this time and may be especially receptive to health information provided by their care team. It is also possible that patients are seeking Web-based health information because they are not receiving sufficient information from their providers. The desire for health information around the time of a health status change suggests a role for providers to direct patients to high-quality, easy-to-understand online information, especially information regarding treatments and procedures. This is especially important to provide to patients who have had a recent acute medical exacerbation. These online resources should augment, but not replace, the distribution of printed patient care instructions and information about relevant procedures and treatments.

The fact that patients are both more likely to consume as well as produce Web-based health information around the time of an acute medical exacerbation suggests that this time may also represent a window of opportunity regarding health behavior. Experiencing an acute exacerbation may be a time when patients take more ownership over their health status. There is some cohort-based data to support this. In a study of 253 patients followed one month after an Emergency Department (ED) visit, 12% reported abstaining from smoking for 30 days after the emergency department visit and another 38% reported that they were able to quit smoking for some period of that time. Having a smoking-related emergency department visit was the strongest predictor of abstinence/attempts to quit in these patients [23]. While this indicates that acute medical exacerbations are catalyzing health behavior change, more causal studies are needed to further explore this line of inquiry, as well as what kinds of Web-based and other health-oriented tools may be more helpful and effective at this point.

It may be important to expand the role of the health provider to educate patients not only verbally about their health and medical conditions, but also educate them as to the best Web-based sources of accurate and relevant health information. Not all Web-based health information is considered equal [24]. It can be difficult for patients to identify high-quality versus low-quality information, especially as the accuracy of a website is in fact poorly associated with its credibility [24]. However, routing all patients to the same high-quality website is likely not the optimal solution, as patients have varying abilities in understanding this information. For example, evidence indicates over 80% of post-surgical patients had difficulty understanding Web-based health information and approximately one-third reported that retrieving such information was overwhelming [25].

Limitations

This analysis was based on previously collected survey data and is subject to limitations. All data were self-reported and may be subject to recall or other bias. Data were collected via mobile phone and landline telephone surveys, and it is possible that persons who choose to respond to such surveys are systematically different from persons who choose to not participate in these surveys. Internet use was assessed as a binary variable, and it is possible the frequency or timing of Internet use varied among survey participants. These data are also cross-sectional, and we are therefore not able to tease out the chronology of events. We found that having an acute medical exacerbation was significantly associated with consuming and producing Web-based health information. It is possible that an acute medical exacerbation led to persons seeking Web-based health information, or that persons used the Internet to search for health information that led them to conclude they were having an acute medical exacerbation for which they sought medical attention.

Conclusions

This analysis provides new insight into the use of the Internet for health purposes and suggests that having an acute medical exacerbation is a time when patients’ Web-based health behavior changes and patients become more engaged with the Internet for the purposes of heath information. Practice-based implications of this research include providing high-quality Internet health links for patients around the time of an acute medical exacerbation and leveraging the window of opportunity around an acute medical exacerbation to provide patients with online tools to engage health behavior change.

Acknowledgments

Authors would like to thank Ava Wong for her assistance with literature review and manuscript preparation.

Conflicts of Interest

None declared.

References

1. Fox S. Pew Internet. 2011. The Social life of health information 2011 URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2014-08-13] [WebCite Cache ID 6Ro8R3uGC]
2. Kaiser Permanente. 2014. Kaiser Permanente: Health and Wellness URL: https://healthy.kaiserpermanente.org/health/care/consumer/health-wellness [accessed 2014-08-13] [WebCite Cache ID 6Ro5yrc9F]
3. Mayo Clinic. 2014. Patient Care and Health Information URL: http://www.mayoclinic.org/patient-care-and-health-information [accessed 2014-08-13] [WebCite Cache ID 6Ro8BhMqL]
4. US Department of Veterans Affairs. 2014. Health Topics A to Z Index URL: http://www.va.gov/health/topics/index.asp [accessed 2014-08-13] [WebCite Cache ID 6Ro88Ru4j]
5. National Institutes of Health. 2014. National Institutes of Health: Turning Discovery Into Health URL: http://www.nih.gov/ [accessed 2014-08-13] [WebCite Cache ID 6Ro850Co8]
6. Medline Plus. 2014. MedlinePlus: Trusted Health Information for You URL: http://www.nlm.nih.gov/medlineplus/ [accessed 2014-08-13] [WebCite Cache ID 6Ro81fkTZ]
7. WebMD. 2014. URL: http://www.webmd.com/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7xscuh]
8. Wikipedia. Welcome to Wikipedia: The Free Encyclopedia 2014 URL: http://en.wikipedia.org/wiki/Main_Page [accessed 2014-08-13] [WebCite Cache ID 6Ro7thl8X]
9. American Diabetes Association. 2014. URL: http://www.diabetes.org/?loc=bb [accessed 2014-08-13] [WebCite Cache ID 6Ro7qlrzJ]
10. American Heart Association. 2014. URL: http://www.heart.org/HEARTORG/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7ju7me]
11. American Cancer Society. 2014. URL: http://www.cancer.org/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7c8GgA]
12. Fox S. Pew Internet. 2011. Peer-to-peer Health Care URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7J9dCW]
13. O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F. User-generated online health content: a survey of Internet users in the United Kingdom. J Med Internet Res 2014;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]
14. DiabetesMine: A good mine of straight talk and ecouragement. 2014. URL: http://www.diabetesmine.com/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7Vgo8I]
15. DeBronkart D. e-Patient Dave. 2014. A voice of patient engagement URL: http://www.epatientdave.com/ [accessed 2014-08-13] [WebCite Cache ID 6Ro7PeRlL]
16. Verhoef LM, Van de Belt TH, Engelen LJLPG, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
17. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
18. Li N, Orrange S, Kravitz RL, Bell RA. Reasons for and predictors of patients' online health information seeking following a medical appointment. Fam Pract 2014 Oct;31(5):550-556. [doi: 10.1093/fampra/cmu034] [Medline: 24963151]
19. Pant S, Deshmukh A, Murugiah K, Kumar G, Sachdeva R, Mehta JL. Assessing the credibility of the "YouTube approach" to health information on acute myocardial infarction. Clin Cardiol 2012 May;35(5):281-285. [doi: 10.1002/clc.21981] [Medline: 22487995]
20. van MT. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
21. McInnes DK, Sawh L, Petrakis BA, Rao S, Shimada SL, Eyrich-Garg KM, et al. The potential for health-related uses of mobile phones and internet with homeless veterans: results from a multisite survey. Telemed J E Health 2014 Sep;20(9):801-809. [doi: 10.1089/tmj.2013.0329] [Medline: 25046280]

22. Lustria MLA, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
23. Boudreaux ED, Baumann BM, Camargo CA, O'Hea E, Ziedonis DM. Changes in smoking associated with an acute health event: theoretical and practical implications. Ann Behav Med 2007 Apr;33(2):189-199. [doi: 10.1080/08836610701310003] [Medline: 17447871]
24. Kunst H, Groot D, Latthe PM, Latthe M, Khan KS. Accuracy of information on apparently credible websites: survey of five common health topics. BMJ 2002 Mar 9;324(7337):581-582 [FREE Full text] [Medline: 11884323]
25. Murero M, D'Ancona G, Karamanoukian H. Use of the Internet by patients before and after cardiac surgery: telephone survey. J Med Internet Res 2001;3(3):E27 [FREE Full text] [doi: 10.2196/jmir.3.3.e27] [Medline: 11720969]

Edited by G Eysenbach; submitted 21.08.14; peer-reviewed by L Nelson, I Adeleke, S Beidas; comments to author 16.01.15; revised version received 06.02.15; accepted 27.04.15; published 23.06.15.

Please cite as:
Gidwani R, Zulman D
Association Between Acute Medical Exacerbations and Consuming or Producing Web-Based Health Information: Analysis From Pew Survey Data
J Med Internet Res 2015;17(6):e145
URL: http://www.jmir.org/2015/6/e145/
doi:10.2196/jmir.3801
PMID:

Original Paper

Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial

Tuula Karhula[1], MD, PhD; Anna-Leena Vuorinen[2], MSc; Katja Rääpysjärvi[1], BEng; Mira Pakanen[1], MSc; Pentti Itkonen[1], PhD; Merja Tepponen[1], PhD (Nursing); Ulla-Maija Junno[3], BS; Tapio Jokinen[4], MD; Mark van Gils[2], PhD; Jaakko Lähteenmäki[2], Tech Lic; Kari Kohtamäki[2], MSc; Niilo Saranummi[2], PhD

[1]Eksote, South Karelia Social and Health Care District, Lappeenranta, Finland
[2]VTT Technical Research Centre of Finland, Tampere, Finland
[3]Mawell Care Oy, Helsinki, Finland
[4]Medixine Ltd, Helsinki, Finland

Corresponding Author:
Anna-Leena Vuorinen, MSc
VTT Technical Research Centre of Finland
PO Box 1300
Tampere, FI-33101
Finland
Phone: 358 408485966
Fax: 358 207223499
Email: anna-leena.vuorinen@vtt.fi

Abstract

Background: There is a strong will and need to find alternative models of health care delivery driven by the ever-increasing burden of chronic diseases.

Objective: The purpose of this 1-year trial was to study whether a structured mobile phone-based health coaching program, which was supported by a remote monitoring system, could be used to improve the health-related quality of life (HRQL) and/or the clinical measures of type 2 diabetes and heart disease patients.

Methods: A randomized controlled trial was conducted among type 2 diabetes patients and heart disease patients of the South Karelia Social and Health Care District. Patients were recruited by sending invitations to randomly selected patients using the electronic health records system. Health coaches called patients every 4 to 6 weeks and patients were encouraged to self-monitor their weight, blood pressure, blood glucose (diabetics), and steps (heart disease patients) once per week. The primary outcome was HRQL measured by the Short Form (36) Health Survey (SF-36) and glycosylated hemoglobin (HbA1c) among diabetic patients. The clinical measures assessed were blood pressure, weight, waist circumference, and lipid levels.

Results: A total of 267 heart patients and 250 diabetes patients started in the trial, of which 246 and 225 patients concluded the end-point assessments, respectively. Withdrawal from the study was associated with the patients’ unfamiliarity with mobile phones—of the 41 dropouts, 85% (11/13) of the heart disease patients and 88% (14/16) of the diabetes patients were familiar with mobile phones, whereas the corresponding percentages were 97.1% (231/238) and 98.6% (208/211), respectively, among the rest of the patients (P=.02 and P=.004). Withdrawal was also associated with heart disease patients’ comorbidities—40% (8/20) of the dropouts had at least one comorbidity, whereas the corresponding percentage was 18.9% (47/249) among the rest of the patients (P=.02). The intervention showed no statistically significant benefits over the current practice with regard to health-related quality of life—heart disease patients: beta=0.730 (P=.36) for the physical component score and beta=-0.608 (P=.62) for the mental component score; diabetes patients: beta=0.875 (P=.85) for the physical component score and beta=-0.770 (P=.52) for the mental component score. There was a significant difference in waist circumference in the type 2 diabetes group (beta=-1.711, P=.01). There were no differences in any other outcome variables.

Conclusions: A health coaching program supported with telemonitoring did not improve heart disease patients' or diabetes patients' quality of life or their clinical condition. There were indications that the intervention had a differential effect on heart patients and diabetes patients. Diabetes patients may be more prone to benefit from this kind of intervention. This should not be neglected when developing new ways for self-management of chronic diseases.

Trial Registration: ClinicalTrials.gov NCT01310491; http://clinicaltrials.gov/ct2/show/NCT01310491 (Archived by WebCite at http://www.webcitation.org/6Z8l5FwAM).

(J Med Internet Res 2015;17(6):e153) doi:10.2196/jmir.4059

KEYWORDS

health coaching; telemonitoring; type 2 diabetes; heart disease; personal health record; health-related quality of life

Introduction

There is a strong will and need to find alternative models of health care delivery [1], driven by the ever-increasing burden of chronic diseases. To ensure adequate resources for the delivery of health care and to further improve the level of care, care-delivery models need to be changed in a way that patients themselves become more involved in their own care.

Home telemonitoring of chronic diseases seems to be a promising disease management approach with the potential to boost patients' compliance with self-care, while bringing health care services closer to patients and, thus, resulting in improved quality of life. However, the evidence of the effectiveness of telemonitoring is contradictive and is dependent on the nature of the disease [2]. In a systematic review by Pare et al [2], it was found that telemonitoring improved glycemic control of diabetics, decreased blood pressure levels of hypertensive patients, and improved peak expiratory flows of patients with asthma and symptoms associated with the illness. However, the beneficial effect of telemonitoring was not associated with heart failure and the evidence is still contradictive. Meta-analyses conducted among heart failure patients from 2009 and 2011 conclude that there are beneficial effects of telemonitoring with linkage to improved survival and decreased hospitalizations [3,4]. However, since these meta-analyses, there have been two large-scale randomized controlled trials [5,6] failing to show the effectiveness of telemonitoring as concluded by Pare et al [2]. Correspondingly, the evidence of telemonitoring on improved glycemic control is contradictive. Typically, the observed reduction in hemoglobin A1c (HbA1c) has been 0.5% [7,8], raising a question of its clinical significance. Moreover, there have been studies that show nonsignificant changes in glycemic control among diabetics [9].

In chronic diseases the condition of a patient is highly dependent on their engagement of self-care and their ability to adhere to the management recommendations long term. For successful disease management, the education of a patient is important. However, the education-based interventions are by themselves insufficient [10]. Health coaching helps the patient to clarify his motivation to initiate and maintain change, offering a variety of perspectives and recognizing that numerous factors contribute to achieving goals [11]. Promising results have been obtained among type 2 diabetes patients in health coaching conducted by telephone [11]. However, the 1-year long health coaching by telephony to support self-care in chronic diseases (TERVA) trial, in which a health coaching approach was applied, failed to achieve most of the expected improvements in clinical measures [12]. Similar findings were found by Ruggiero et al [13]. In addition to the importance of self-management, patients and health care professionals need to share complementary knowledge in health care processes, which brings challenges and responsibility from both sides [14]. Telemonitoring provides a possibility for improved interaction. The combination of telemonitoring and remote monitoring has shown promising results among hypertensive patients [15].

The purpose of this study was to assess the benefits of a structured mobile phone-based health coaching program, supported by a remote monitoring system among chronically ill patients. We expected the intervention to improve patients' engagement in self-management and to enrich the interaction between patients and health care professionals that would eventually result in improved quality of life and/or the clinical condition. Primarily, we hypothesized that we would see improved quality of life among patients suffering from heart disease or diabetes.

Methods

Study Design

The study was conducted as a two-armed randomized controlled trial (RCT) between February 2011 and December 2012 in the South Karelia Social and Health Care District (Eksote) in Finland. The trial was registered at ClinicalTrials.gov (NCT01310491). Eksote is responsible for arranging all primary and secondary health care for the inhabitants of eight municipalities, approximately 100,000 inhabitants. Patients with type 2 diabetes and patients suffering from heart disease were recruited to the study and assigned to either the control group or the intervention group. The study was approved by the Ethics Committee of the Social and Health Care District of South Karelia.

Intervention

Overview

The intervention consisted of health coaching over mobile phones and self-monitoring of health parameters with the help of a remote patient monitoring (RPM) system.

Health Coaching

Each patient in the intervention group was assigned a personal health coach who called them at regular intervals—every 4 to 6 weeks. A comprehensive evaluation of the patient's clinical, mental, and social condition was made during the first coaching call and small, achievable health behavior changes were agreed upon with the patient. A self-management plan was created based on the targeted changes. During the mobile phone calls that were planned to last for approximately 30 minutes, the health coach provided information, assistance, and support to the patients. The health coaching approach was provided by Pfizer Oy. The approach followed Wagner's Chronic Care Model [16]—one of the key foundational constructs for the approach of chronic care management—and has been developed and tested earlier. The detailed structure of the health coaching program and the behavior change techniques involved are reported elsewhere [12].

Health Coach Recruitment

Health coaches and a health coach supervisor were recruited among the personnel of Eksote. Six coaches were recruited out of 13 applicants. Four of the recruits were working in outpatient care and two in a hospital. The selected coaches continued in their regular positions and worked as health coaches 1 day a week. The health coaches were trained to obtain the needed knowledge about Pfizer's health coaching model, behavioral management skills, remote monitoring system, and trial procedures. The health coaching model was a solution-oriented working model where all patients received coaching based on their individual needs. For quality control and educational purposes, each health coach recorded some of the coaching calls, which were evaluated together with a behavioral science professional once in every 3 months. The equal quality of all health coaches was assured by continuous education and regular meetings, which all the health coaches and the trainer attended.

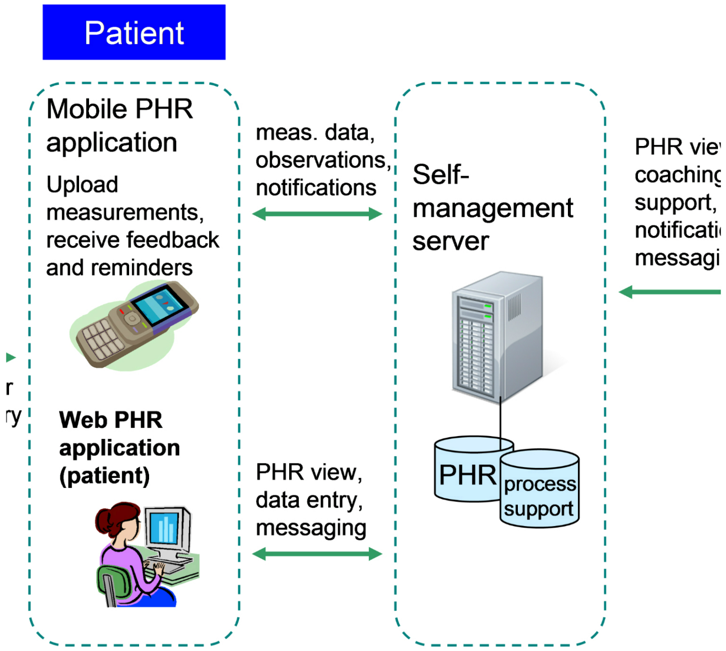
Remote Patient Monitoring

Each patient in the intervention group received a remote monitoring toolbox to be used in the trial. The toolbox consisted of a mobile phone with specific software, a mobile personal health record (PHR) app, and a set of measurement devices connected to the patient's PHR account. The mobile PHR app was needed for manual and/or automatic reporting. All patients received a blood pressure meter, which was connectable to the mobile phone via Bluetooth. When the patients measured their blood pressure, the value was automatically transferred to the PHR using a binary short message service (SMS) text message. Other health parameters to be followed were body weight, blood glucose level for diabetics, and step count for heart disease patients. The patients were instructed to measure and send these values manually via the mobile phone to the PHR once a week. The health coaches and patients were able to see the patients' measurements in the PHR and were advised to utilize them during health coaching phone calls. A self-management guide was given to the patients with the intention to increase their knowledge of their chronic disease.

Remote Patient Monitoring System

The intervention was supported by the RPM system, eClinic, provided by Medixine Ltd (Espoo, Finland) (see Figure 1). The self-management server is the central component of its architecture, providing services for the storing and accessing of information content (ie, RPM data) related to the self-management process. The RPM data included various types of information: health parameters registered by the corresponding measurement devices, personal care plan entered by the health coach in agreement with the patient, and data obtained from the electronic health record (EHR). The HTTPS protocol was used for sending all data from the mobile app to the server. The system underwent no major changes or updates during the trial.

Figure 1. Technical architecture of the health coaching system supported with remote patient monitoring.



Standard Care

Patients assigned to the control group received the care they would have received in the absence of the study. As part of standard care, patients suffering from type 2 diabetes or heart disease receive a disease management information booklet at the time of diagnosis. Standard care includes laboratory tests taken once a year and 1 appointment or phone call by a nurse or doctor. Patients can contact health care services any time they feel they need to.

Participants and Baseline Assessment

The patients' eligibility was assessed primarily based on their diagnosis. The diabetic patients were recruited based on a diagnosis of type 2 diabetes mellitus and their glycosylated hemoglobin (HbA1c) level, which needed to be above 6.5% within 1 year prior to the screening. It was required that the patients had been diagnosed with diabetes at least 3 months earlier. The heart disease group consisted of patients with a diagnosis of ischemic heart disease, heart failure, or both. Other inclusion criteria for all patients were as follows: 18 years of age or older, ability to fill in questionnaires in Finnish, ability to use the RPM system and the devices provided, having adequate cognitive capacities to participate, and being able to walk.

Potential participants were screened using the electronic health record system of Eksote. EHRs cover information about citizens living in the health care district of South Karelia who have contacted health care services at least once. Invitation letters including information about the study were sent to eligible patients. Patients willing to participate signed an informed consent form before randomization. After that, the supervisor contacted each of the patients to schedule an appointment for a baseline visit. Randomization was done after the appointment was settled.

All patients who came in for the baseline visit were asked to fill in a demographic questionnaire and the Short Form (36) Health Survey (SF-36), version 2 [17], which measures health-related quality of life. At the baseline visit, a health coach measured the patient's blood pressure, height (to the nearest 0.1 cm), weight (to the nearest 0.1 kg), and waist circumference (to the nearest 0.1 cm), and calculated their body mass index (BMI). Each patient's medical history was reviewed based on the data in the EHR system. If laboratory tests were older than 2 months, new laboratory tests (ie, HbA1c, cholesterol, triglycerides) were done. At the end of the visit, the health coach checked that the required questionnaires were returned. If not, the patient was asked to fill out the questionnaire at home and send it to a nurse on the following day.

After 1 year following the baseline visit, all patients were invited to an end-point visit. The same procedures were conducted as they were during the baseline visit.

Randomization

A stratified randomization design was used to assign patients to the control and intervention groups. Heart disease and diabetes patients were randomized into separate groups. Patients were further stratified into four subgroups according to their sex and dichotomized age—18 to 65 years versus older than 65 years. Within these subgroups, Excel-generated random numbers were produced. The allocation sequence was concealed from the research nurse by means of an opaque and sealed envelope until the baseline visit. During the baseline visit the envelope was opened and, according to its content, each patient was assigned to either group. The randomization was conducted by the Technical Research Centre of Finland (VTT).

Outcome Variables

Short Form (36) Health Survey

The primary outcome for both disease groups was self-evaluated, health-related quality of life (HRQL) assessed based on the SF-36 health survey. Eight domains of HRQL and two summary component measures of physical and mental health were analyzed. Additionally, HbA1c level was another primary outcome for the diabetes patients.

Clinical Outcomes

Secondary outcomes were as follows: blood pressure (mmHg), weight (kg), waist circumference (cm), triglycerides (mmol/l), total cholesterol (mmol/l), low-density lipoprotein (LDL) (mmol/l), and high-density lipoprotein (HDL) (mmol/l). The selection of outcome variables was based on the use of a model for assessment of telemedicine applications [18]. However, this paper examines the first three out of the seven domains concentrating on the medical perspectives. Other domains, such as organizational and economic outcomes, will be reported in other articles in the future.

Adherence

Adherence to the health coaching was measured as the number and duration of health coaching calls. The duration of a call consisted of three parts—the time a nurse needed to prepare for a call (eg, familiarize herself with the self-measurement data of a patient), the duration of the actual coaching call, and the time a nurse needed to finalize the call (eg, notes, information delivery). Another perspective of the adherence measure was based on the frequency of home telemonitoring, measured as the total number of measurements made during the study and calculation of the number of weight, blood pressure, blood sugar, and step count reports. Both pre- and postprandial measurements were included in blood glucose reports.

Statistical Analysis

We assumed we would see a difference of three points in the SF-36 scores between the intervention and control groups with a standard deviation of eight. The allocation ratio was unbalanced—approximately 2:1. The number of intervention patients was higher because we wanted to maximize the exposure to, and gain experience about, this new intervention. Defining a power of 80% and a Type I error rate of 5%, 163 intervention patients and 61 control patients were required. Predicting a dropout rate of up to 20%, at least 200 intervention patients and 75 control patients had to be randomized. The numbers were applied to both the heart disease group and diabetes group, resulting in 550 patients to be randomized in total. We used the *t* test as a basis for the power calculations, which is a conservative approach considering that repeated measures were available in the data, and thus more powerful tests could have been used.

The characteristics of dropout patients in terms of their baseline measures were explored using Student's *t* tests and chi-square tests. All analyses were conducted separately for the diabetes and heart disease groups. The analysis of covariance (ANCOVA) was used to study whether the intervention and the control groups differed in terms of their outcome variables. The analyses were done by adjusting for the corresponding baseline level by adding the baseline measure as a covariate in the regression model. The 95% CIs and the corresponding *P* values were reported. Additionally, within-group changes from baseline to postintervention were analyzed using paired *t* tests.

Analyses were conducted following the intention-to-treat principle, meaning that all patients were analyzed in their original allocation group regardless of the extent to which they followed the intervention. No imputations were made to missing values, but missing values were excluded from the analyses. All reported *P* values were two sided. Analyses were conducted using IBM SPSS Statistics version 19.

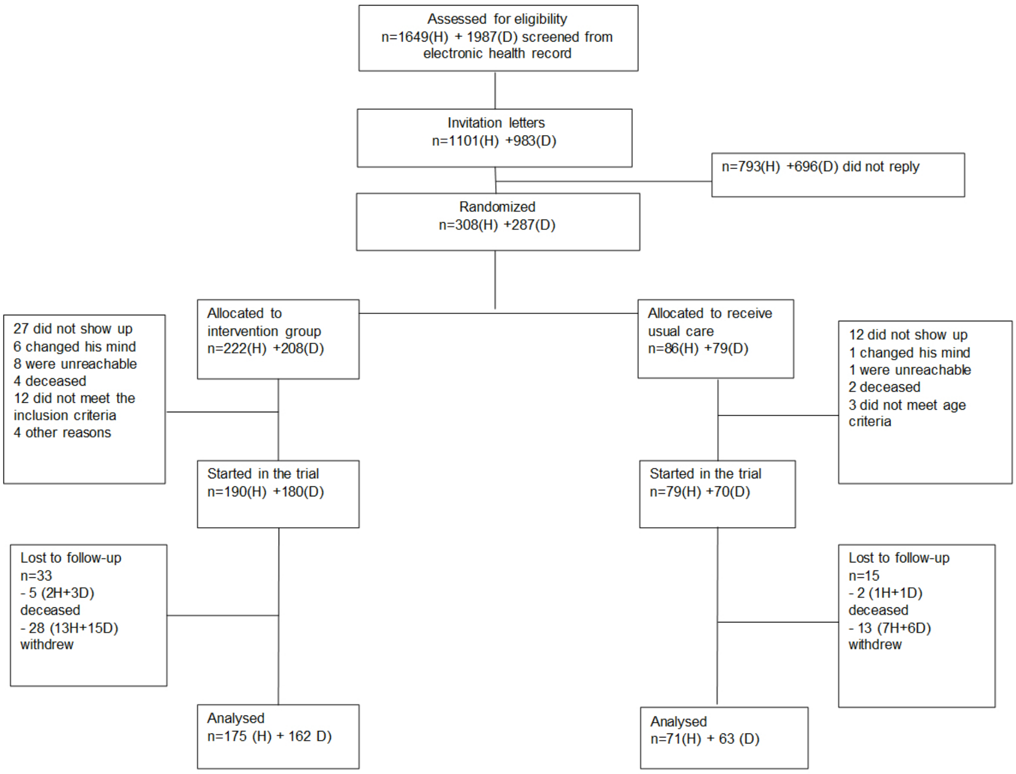
Results

Patient Flow

Figure 2 describes the progress of the trial. The electronic health records were utilized to screen patients with either heart disease or diabetes mellitus type 2. The diagnosis was either type 2 diabetes mellitus with HbA1c >6.5% or one of the following two heart diseases: ischemic heart disease or heart failure. The number of patients fulfilling the criteria was 1649 with heart disease diagnoses, and 1987 patients with diabetes diagnoses. Of these patients, 499 heart disease patients and 500 diabetes patients were randomly selected and received invitation letters in October 2010. The number of patients who refused to participate, changed their mind before the trial began, or did not show up at the baseline visit, was higher than expected. Therefore, the invitation procedure was repeated in November 2010 and August 2011 to achieve the predefined power for the pilot. In total, invitation letters were sent to 2084 patients, of which 28.02% (584) agreed to participate. Eventually, 595 patients were randomized and, of these, 519 patients (87.2%) attended the baseline visit. All participants filled out the baseline questionnaires before they were told into which group they were randomized.

There were 48 patients out of 519 (9.2%) lost to follow-up: 3 heart patients and 4 diabetes patients died, and 20 and 21 patients, respectively, withdrew from the trial without participating in the concluding visit. The baseline characteristics of the withdrawn patients were analyzed against patients who concluded the trial. Quitting was associated with the patients' unfamiliarity with mobile phones—of the dropouts in the heart disease group, 85% (11/13) were familiar with mobile phones, whereas the corresponding percentage was 97.1% (231/238) among the rest of the patients (*P*=.02). Of the dropouts in the diabetes group, 88% (14/16) were familiar with mobile phones, whereas the corresponding percentage was 98.6% (208/211) among the patients who concluded the trial (*P*=.004). Among heart patients, withdrawal was also often associated with comorbidities—40% (8/20) of the dropouts had at least one comorbidity, whereas the corresponding percentage was 18.9% (47/249) among the rest of the patients (*P*=.04). There was no difference in the dropout rate between intervention and control groups. Eventually, 246 heart disease patients and 225 diabetes patients concluded the trial.

Figure 2. The patient flow within the trial. H: patients with a diagnosis of ischemic heart disease or heart failure, D: patients with a diagnosis of diabetes mellitus type 2 and HbA1c > 6.5%.



Baseline Characteristics

Table 1 displays the baseline characteristic of patients separated according to their primary disease. Overall, patients were similar in the intervention group and in the control group in both disease groups. The mean age among heart patients was 69.1 (SD 9.1) years, and diabetes patients were slightly younger with a mean age of 66.2 (SD 8.6) years. The majority of patients were men in the heart disease group (178/269, 66.2%) and in the diabetes group (129/250, 51.6%). BMI was higher in the diabetes group than in the heart disease group, but BMI distribution was similar between the treatment arms. Over two-thirds of the patients (361/519, 69.6%) were retired. Approximately 8.1% (42/519) were smokers. The rate of missing values was clearly higher regarding smoking and alcohol questions compared to the other baseline questions. The high proportion of missing values regarding the alcohol question was explained by the fact that patients did not find a suitable option among the provided choices for answers. They told this to the nurse at the baseline visit, or it was written in the questionnaire that no proper choice was given because they did not use alcohol at all. The majority of the patients were familiar with mobile phones, and approximately half of the patients were familiar with computers. The most common comorbidities were diagnosed connective tissue disease, rheumatic disease, or chronic pulmonary disease. There were only a few patients with dementia or cerebrovascular disease.

Short Form (36) Health Survey

Tables 2 and 3 show the baseline, postintervention, and change scores of HRQL—the eight dimensions of the HRQL assessment and the two summary scores. There were no significant differences between the control and intervention arms in either of the disease groups for any of the variables.

A total of 45 patients completed the baseline questionnaire at home and later sent it to the nurse. On average, these patients posted their questionnaires 5.3 (range 1 to 7) months after they started in the trial. To exclude the bias that the late responses may have caused, the analyses of HRQL were repeated without the late responses. The level of significance of the difference between the control and intervention groups remained above .1 in all variables. Thus, no change in the interpretation was observed.

The number of respondents varied from question to question. In the diabetes group, the number of respondents varied from 146 to 159 in the intervention group and 55 to 60 in the control group, depending on the questions, which is slightly less than was assumed in the pre hoc power calculations. The lower sample size leads to a post hoc power of .76 when using the *t* test framework. However, the magnitude of .80 was reached when using the ANCOVA framework. The predefined power was reached in the heart disease group.

Table 1. Baseline characteristics of the patients in the two disease groups.

Baseline characteristic	Heart disease patients (n=269), mean (SD) or n (%)		Diabetes patients (n=250), mean (SD) or n (%)	
	Control (n=79)	Intervention (n=190)	Control (n=70)	Intervention (n=180)
Sex (female), n (%)	25 (32)	66 (34.7)	30 (43)	81 (45.0)
Age (years), mean (SD)	68.1 (9.4)	69.6 (9.1)	65.5 (9.6)	66.6 (8.2)
BMI[a](kg/m^2), mean (SD)	28.1 (4.3)	28.6 (4.7)	30.9 (5.7)	31.1 (5.4)
Education, n (%)				
Primary school or less	29 (37)	98 (51.6)	30 (43)	75 (41.7)
Secondary or high school	31 (39)	59 (31.1)	24 (34)	65 (36.1)
College/university or higher	9 (11)	24 (12.6)	12 (17)	27 (15.0)
Missing	10 (13)	9 (4.7)	4 (6)	13 (7.2)
Marital status, n (%)				
Never married	1 (1)	8 (4.2)	4 (6)	10 (5.6)
Married/cohabitating	69 (87)	133 (70)	53 (76)	120 (66.7)
Separated	3 (4)	24 (12.6)	4 (6)	25 (13.9)
Widowed	5 (6)	23 (12.1)	9 (13)	22 (12.2)
Missing	1 (1)	2 (1.1)	0 (0)	3 (1.7)
Work status, n (%)				
Working	12 (15)	34 (17.9)	11 (16)	34 (18.9)
Unemployed (able to work)	4 (5)	6 (3.2)	3 (4)	11 (6.1)
Unemployed (unable to work)	0 (0)	5 (2.6)	0 (0)	5 (2.8)
Retired	53 (67)	138 (72.6)	52 (74)	118 (65.6)
Student	0 (0)	0 (0)	0 (0)	1 (0.6)
Missing	10 (13)	7 (3.7)	4 (6)	11 (6.1)
Smoking, n (%)				
Smoker	6 (8)	14 (7.4)	6 (9)	16 (8.6)
Missing	17 (22)	27 (14.2)	14 (20)	23 (12.8)
Alcohol, n (%)				
5-7 days a week	2 (3)	6 (3.2)	2 (3)	5 (2.8)
1-4 days a week	21 (27)	40 (21.1)	13 (19)	34 (18.9)
Monthly	14 (18)	47 (24.7)	11 (16)	37 (20.6)
Less than monthly	18 (23)	52 (27.4)	23 (33)	65 (36.1)
Missing	24 (30)	45 (23.7)	21 (30)	39 (21.7)
Familiar with PC [b] use, n (%)				
Familiar	41 (52)	102 (53.7)	41 (59)	102 (56.7)
Missing	10 (13)	14 (7.4)	8 (11)	14 (7.8)
Familiar with mobile phone use, n (%)				
Familiar	69 (87)	173 (91.1)	61 (87)	161 (89.4)
Missing	8 (10)	10 (5.3)	9 (13)	14 (7.8)
Comorbidities, n (%)				
Heart diseases	79 (100)	190 (100)	15 (21)	47 (26.1)
Cerebrovascular disease	0 (0)	4 (2.1)	3 (4)	9 (5.0)

Baseline characteristic	Heart disease patients (n=269), mean (SD) or n (%)		Diabetes patients (n=250), mean (SD) or n (%)	
	Control (n=79)	Intervention (n=190)	Control (n=70)	Intervention (n=180)
Chronic pulmonary disease, including COPD[c]	8 (10)	22 (11.6)	12 (17)	19 (10.6)
Connective tissue disease or rheumatic disease	8 (10)	30 (15.8)	9 (13)	36 (20.0)
Diabetes	18 (23)	46 (24.2)	70 (100)	180 (100)
Cancer	7 (9)	17 (8.9)	4 (6)	12 (6.7)
Other[d]	49 (62)	111 (58.4)	52 (74)	135 (75.0)
No comorbidities	16 (20)	39 (20.5)	14 (9)	23 (6.2)

[a] BMI: body mass index

[b] PC: personal computer

[c] COPD: chronic obstructive pulmonary disease

[d] Hypertension is the most common “other” comorbidity.

Table 2. The baseline, postintervention, and change scores in the eight dimensions of the health-related quality-of-life assessments and in the two summary scores for heart disease patients.

Assessment	Control scores				Intervention scores				Between-group difference, beta (95% CI)	*P* [a]
	n	Baseline	Post[b]	Change (95% CI)	n	Baseline	Post	Change (95% CI)		
Physical component score	68	40.3	40.7	0.39 (-0.72, 1.49)	162	39.5	40.8	1.25 (0.29, 2.22)	0.730 (-3.00, 1.78)	.36
Mental component score	68	50.5	51.0	0.55 (-1.53, 2.58)	162	50.4	50.3	-0.05 (-1.47, 1.37)	-0.608 (-6.19, 6.26)	.62
Physical functioning (PF)	68	64.9	66.1	1.16 (-1.77, 4.09)	170	62.7	64.1	1.42 (-0.82, 3.67)	0.02 (-3.89, 3.93)	.99
Role-physical (RP)	68	60.7	63.5	2.79 (-1.84, 7.42)	168	58.9	62.1	3.16 (-0.58, 6.90)	-1.72 (-6.09, 5.75)	.95
Bodily pain (BP)	68	57.2	57.9	0.70 (-3.27, 4.66)	171	56.4	59.9	3.51 (0.58, 6.44)	2.59 (-2.34, 7.51)	.30
General health (GH)	68	48.7	49.2	0.56 (-2.93, 4.05)	171	47.7	50.3	2.60 (0.36, 4.84)	1.77 (-2.06, 5.61)	.36
Vitality (VT)	68	57.1	56.9	-0.25 (-4.71, 4.22)	165	56.3	56.8	0.48 (-2.03, 3.00)	0.52 (-4.03, 5.06)	.82
Social functioning (SF)	68	80.1	80.0	-0.18 (-4.93, 4.56)	171	78.9	79.8	0.88 (-2.15, 3.90)	0.585 (-4.44, 5.61)	.82
Role-emotional (RE)	67	72.5	75.4	2.86 (-2.63, 8.35)	168	71.2	73.0	1.74 (-1.74, 5.22)	1.54 (-7.42, 4.34)	.61
Mental health (MH)	68	77.3	77.9	0.64 (-2.92, 4.21)	164	77.4	77.2	-0.23 (-1.47, 1.37)	-0.80 (-5.00, 3.36)	.70

[a] *P* values show the level of statistical significance between the treatment arms.

[b] Postintervention score.

Table 3. The baseline, postintervention, and change scores in the eight dimensions of the health-related quality-of-life assessments and in the two summary scores for diabetes patients.

Assessment	Control scores				Intervention scores				Between-group difference, beta (95% CI)	*P* [a]
	n	Baseline	Post[b]	Change (95% CI)	n	Baseline	Post	Change (95% CI)		
Physical component score	55	41.5	42.0	0.51 (-1.19, 2.21)	146	42.6	43.2	0.53 (-0.40, 1.47)	0.875 (0.80 9, 0.95)	.85
Mental component score	56	50.1	52.0	1.84 (0.02, 3.71)	148	50.2	51.2	1.06 (-0.42, 2.53)	-0.77 (-3.15, 1.61)	.52
Physical functioning (PF)	58	64.9	66.0	1.09 (-2.87, 5.06)	157	68.1	68.2	0.17 (-1.83, 2.17)	-0.715 (-4.74, 3.13)	.73
Role-physical (RP)	58	65.2	68.4	3.23 (-2.81, 9.27)	156	65.7	68.8	3.11 (-0.45, 6.68)	-0.036 (-6.19, 6.26)	.99
Bodily pain (BP)	58	55.3	58.8	3.52 (-0.94, 7.98)	159	62.4	62.2	-0.18 (-3.05, 2.68)	-2.02 (-7.20, 3.13)	.44
General health (GH)	60	49.2	50.6	1.34 (-1.48, 4.17)	159	50.1	53.6	3.47 (1.04, 5.89)	2.34 (-1.72, 6.41)	.26
Vitality (VT)	58	52.9	58.1	5.21 (1.29, 9.19)	149	57.6	58.6	0.98 (-1.88, 3.83)	-2.98 (-7.78, 1.83)	.22
Social functioning (SF)	60	79.4	83.3	3.96 (-0.18, 8.10)	157	80.0	81.1	1.19 (-2.05, 4.44)	-2.54 (-7.70, 2.61)	.33
Role-emotional (RE)	59	74.3	78.1	3.81 (-1.72, 9.35)	157	74.7	78.7	3.93 (0.26, 7.60)	0.30 (-5.50, 6.10)	.92
Mental health (MH)	58	76.5	78.5	2.07 (-1.80, 5.93)	149	76.7	77.5	0.87 (-1.75, 3.50)	-1.12 (-5.43, 3.19)	.61

[a] *P* values show the level of statistical significance between the treatment arms.

[b] Postintervention score.

Clinical Outcomes

Tables 4 and 5 display the baseline, postintervention, and change scores in the anthropometric and laboratory measures, and the comparison between the treatment arms in both disease groups. In the heart disease group, there was no difference between the treatment arms in any of the variables. However, there was a significant within-group decrease in waist circumference (*P*=.02), systolic blood pressure (*P*<.001), and LDL-cholesterol (*P*<.001) in the intervention group. Also, in the control group, LDL-cholesterol decreased significantly (*P*<.001), as did systolic blood pressure (*P*<.001).

Among diabetics, there was a significant difference between the treatment arms in waist circumference (*P*=.01). In the intervention group, there was a significant decrease in weight (*P*=.02), waist circumference (*P*<.001), systolic blood pressure (*P*<.001), diastolic blood pressure (*P*=.007), and LDL-cholesterol (*P*<.001). In the control group, systolic blood pressure and LDL-cholesterol decreased significantly (*P*=.02 and *P*<.001, respectively).

Adherence

Out of 190 heart disease and 180 diabetes patients, 186 (97.9%) and 177 (98.3%) patients, respectively, received at least one health coach call. The average number of calls per patient was 8.7 (SD 1.6) in the heart disease patient group and 8.5 (SD 1.9) in the diabetes group. The difference between the disease groups was not significant (*P*=.40). The mean duration of a coaching call was 20.1 (SD 8.0) minutes in the heart disease group and 19.2 (SD 8.1) minutes in the diabetes group, with a significant between-group difference (*P*=.004). The mean time consumed by the nurse for the preparation of calls was 3.5 (SD 2.5) minutes in the heart disease group and 4.2 (SD 3.2) minutes in the diabetes group, and the between-group difference was significant (*P*<.001). The time consumed by the nurse after the coaching calls among heart disease and diabetes patients was 3.8 (SD 3.0) and 4.5 (SD 3.6) minutes, respectively, with a significant between-group difference (*P*<.001).

The median number of all self-measurements reported through mobile phones was 209 (interquartile range [IQR] 124-324) among heart patients and 217 (IQR 104-346) among diabetes patients. The median number for heart disease group-specific monitoring parameters per patient were the following: 18 (IQR 2-40) weight reports, 18 (IQR 4-43) step counts, 57 (IQR 36-89) blood pressure reports, and 42 (IQR 12-67) blood glucose reports—6 patients made blood glucose monitoring reports. The median number for diabetes group-specific monitoring parameters per patient were the following: 15 (IQR 3-39) weight reports, 15 (IQR 5-31) step counts, 56 (IQR 28-80) blood pressure reports, and 47 (IQR 20-89) blood glucose reports, including pre- and postprandial sugar. In the heart disease group and in the diabetes group, 174 out of 190 (91.6%) and 171 out of 180 (95.0%) patients, respectively, adhered to the self-monitoring intervention to the extent that they sent at least one report of any kind during the follow-up. Among 190 heart disease patients, 136 (71.6%) sent at least one weight measurement, 173 (91.1%) sent at least one blood pressure measurement, 6 (3.2%) sent at least one blood glucose measurement, and 118 (62.1%) sent at least one step count report. Out of 180 diabetes patients, the corresponding numbers were 119 (66.1%) for weight, 170 (94.4%) for blood pressure, 126 (70.0%) for blood glucose, and 13 (7.2%) for step count.

Table 4. Baseline, postintervention, and change scores in clinical outcomes for the heart disease group.

Clinical outcome	Control scores				Intervention scores				Between-group difference, beta (95% CI)	*P* [a]
	n	Baseline	Post[b]	Change (95% CI)	n	Baseline	Post	Change (95% CI)		
Weight	70	79.9	79.1	-0.84 (-1.85, 0.16)	170	81.4	81.5	0.04 (-0.67, 0.76)	0.934 (-0.34, 2.21)	.15
Waist	65	97.6	98.7	1.10 (-1.65, 3.85)	160	101.5	100.6	-0.88 (-1.61, -0.16)	-1.518 (-3.57, 0.53)	.15
Systolic	68	144.4	138.0	-6.36 (-10.7, -2.01)	161	145.5	140.1	-5.43 (-8.12, -2.75)	1.587 (-2.51, 5.68)	.45
Diastolic	67	81.1	80.9	-0.18 (-2.81, 2.45)	161	82.3	82.1	-0.27 (-1.95, 1.41)	0.468 (-2.24, 3.18)	.73
Total cholesterol	68	4.13	4.05	-0.08 (-0.25, 0.09)	168	4.06	4.01	-0.05 (-0.17, 0.06)	0.009 (-0.168, 0.185)	.92
HDL[c]	68	1.23	1.26	0.03 (-0.02, 0.08)	168	1.29	1.31	0.02 (-0.01, 0.06)	-0.018 (-0.086, 0.05)	.87
LDL[d]	68	2.56	2.21	-0.36 (-0.51, -0.21)	168	2.50	2.16	-0.34 (-0.43, -0.24)	-0.008 (-0.15, 0.13)	.91
Triglycerides	68	1.43	1.32	-0.12 (-0.27, 0.03)	168	1.37	1.35	-0.01 (-0.13, 0.08)	0.071 (-0.08, 0.22)	.36

[a] *P* values show the level of statistical significance between the treatment arms.

[b] Postintervention score

[c] HDL: high-density lipoprotein

[d] LDL: low-density lipoprotein

Table 5. Baseline, postintervention, and change scores in clinical outcomes for the diabetes group.

Clinical outcome	Control scores				Intervention scores				Between-group difference, beta (95% CI)	*P* [a]
	n	Baseline	Post[b]	Change (95% CI)	n	Baseline	Post	Change (95% CI)		
HbA1c[c]	61	7.20	7.36	0.18 (-0.02, 0.35)	156	7.25	7.29	0.04 (-0.09, 0.17)	-0.106 (-0.33, 0.11)	.34
Weight	60	88.9	88.6	-0.30 (-1.21, 0.60)	153	89.6	88.7	-0.90 (-1.71, -0.22)	-0.566 (-1.86, 0.73)	.39
Waist	57	107.4	107.1	-0.29 (-1.47, 0.90)	143	107.8	105.8	-2.03 (-2.76, -1.29)	-1.711 (-3.042, -0.38)	.01
Systolic	60	151.9	147.8	-4.12 (-7.43, -0.81)	148	155.4	149.3	-6.10 (-9.10, -3.09)	-0.196 (-4.57, 4.18)	.93
Diastolic	60	86.7	84.6	-2.08 (-4.50, 0.34)	148	89.2	86.6	-2.61 (-4.50, -0.72)	0.668 (-2.18, 3.52)	.65
Total cholesterol	60	4.36	4.19	-0.16 (-0.35, 0.03)	153	4.35	4.25	-0.1 (-0.23, 0.04)	0.065 (-0.15, 0.28)	.54
HDL[d]	60	1.26	1.29	0.03 (-0.05, 0.12)	156	1.24	1.26	0.02 (-0.01, 0.05)	0.005 (-0.054, 0.064)	.61
LDL[e]	60	2.66	2.27	-0.39 (-0.55, -0.23)	156	2.74	2.35	-0.40 (-0.51, -0.28)	0.037 (-0.19, 0.20)	.66
Triglycerides	59	1.78	1.89	0.11 (-0.14, 0.36)	154	1.70	1.71	0.01 (-0.10, 0.10)	-1.22 (-0.32, 0.09)	.25

[a] *P* values show the level of statistical significance between the treatment arms.

[b] Postintervention score

[c] HbA1c: hemoglobin A1c

[d] HDL: high-density lipoprotein

[e] LDL: low-density lipoprotein

Discussion

Principal Findings

This study evaluated whether health coaching, supported with home telemonitoring, improved health-related quality of life and/or the clinical condition of type 2 diabetes patients and heart disease patients after 12 months. The intervention failed to improve patients' quality of life or their clinical condition. Patients received regular health coaching calls throughout the study and the majority of the patients adhered to the home telemonitoring plan and frequently monitored at least one of the required health parameters.

The intervention showed a statistically significant difference only in waist circumference among type 2 diabetics. However, due to the lack of consistency in other variables, this finding is likely a result of multiple tests conducted in this study rather than true a difference between the study groups. Multiple testing increases the likelihood of false positive discoveries and this should be acknowledged when interpreting the findings. In addition, blood pressure and cholesterol levels showed beneficial trends for all patients. Overall, the improvements in clinical variables were more apparent in the type 2 diabetes group than in the heart disease patient group.

There were 48 out of 519 patients (9.2%) that were lost to follow-up. We found that unfamiliarity with mobile phones and poor health status measured as a result of the presence of comorbidities were associated with withdrawal. These findings highlight the importance of offering and targeting interventions to an audience with the appropriate skills. eHealth literacy is a prerequisite for the success of eHealth interventions and should be appropriately accounted for. Electronic health tools provide little value if the intended users lack the skills to effectively engage with them [19]. As suggested by Cruz et al [20], the patient skills and acceptance of the technology should be measured prior to its implementation. Appropriate skills are also required on the professional side. A recent study evaluating the use of email in the communication between the primary health care system and general practitioners showed that the easier the general practitioners thought the email system to be, the more they used it [21]. In our study, six nurses were specifically trained for health coaching and to actively utilize the RPM system as part of the care.

The positive changes in patients' clinical conditions in both study groups emphasize the well-known fact that control patients improve their lifestyles as a consequence of being involved in a trial, even if they are not subjected to the actual intervention.

Some of the control group patients were disappointed for not being randomized into the intervention group and they decided to take better care of themselves. Regarding disease-specific effects, we found that diabetes patients who received the intervention improved their health status among several health parameters. The findings were not verified by testing statistical interaction of group and disease variables, but the results in Table 5 showed significant within-group reductions in patients' weight, waist circumference, blood pressure, and LDL. We can speculate whether diabetes patients are more prone to benefit from this kind of intervention. Similarly, Pare et al reported that telemonitoring was associated with a decline in hemoglobin and better blood glucose control, but clinical effects on the condition of patients suffering from cardiac problems were not as evident [2]. Signals reflecting the state of diabetes are not apparent. Even the symptoms of the worsening condition of a patient may stay unrecognized. Therefore, the importance of self-management as a part of diabetes care should be emphasized. The utilization of self-management in health care is a good direction to take, as it was shown by Rose et al [22] that there is a risk of general practitioners, who are sensitive to patients' low self-efficacy in blood glucose monitoring, taking over the monitoring role, and inadvertently reducing self-management. Furthermore, a recent study showed that the significant improvements in HbA1c achieved during a 6-month trial of home telemonitoring, combined with active medication management, were sustained for at least that same 6 months [23].

Patients adhered to home telemonitoring in terms of measuring their blood pressure. Assuming the duration of the trial was approximately 12 months, 52 parameters were expected to be reported. Heart disease and diabetes patients respectively produced 55 and 57 blood pressure measurements on average. Across other health parameters, the monitoring frequency varied from 15 to 42. Patient groups seemed not to differ from each other in terms of monitoring frequency. Some patients had a lack of skills in using remote monitoring devices or they had technical problems, which reduced the number of remote monitoring measurements. Health coaching was realized as planned. The expected number of health coaching calls was between 9 and 12, with 4 to 6 weeks calling frequency. The number of health coaching calls was 8.7 and 8.6 in the heart disease and diabetes group, respectively. Our health coaching model was solution oriented. All coaching calls were tailored to the individual needs that affected variation to the call durations. Few patients had lengthy hospital stays, which affected the number of health coaching calls. The number and duration of health coaching calls were significantly different between the disease groups. The low level of significance was likely due to a small standard deviation in the call duration. A 1-minute difference, as seen in the call duration, has no practical relevance.

The low inclusion criteria in terms HbA1c for diabetic patients posed a limitation on this study. For inclusion, a diabetic patient was required to have an HbA1c higher than 6.5%. On average, the HbA1c levels were 7.2%, showing that there was little room for improvement.

A lack of social support was a potential factor that may have influenced the negative findings of this study. Receiving real-time social support may help people to stay engaged and feel supported, which is important in order to initiate and maintain improvements in health-related behaviors [24]. Another appealing approach to keep patients motivated, specifically those involved with self-monitoring of their health parameters, is the utilization of active assistance technology. Active assistance technology involves automatic processing of health or behavior data and delivers automatic tailored messages to users [25]. Results in this field have been promising, including work by Quinn et al [26], Charpentier et al [27], and Orsama et al [28]. As Bock et al [29] have recently shown, in order to produce successful mHealth apps with lasting effects, it is important to obtain user input throughout development. In our study, the patients were contacted every 4 to 6 weeks. An automatic feedback system, based on their self-monitored health parameters, could have kept patients motivated and informed by the delivery of individualized feedback with a coaching perspective.

Conclusions

In conclusion, this study failed to show a beneficial effect of health coaching supported by telemonitoring on patients' quality of life or their clinical status. However, we do not yet know the long-lasting benefits of the intervention. There were indications that the intervention had a differential effect on heart disease patients and diabetes patients. Diabetes patients may be more prone to benefit from this kind of intervention. This should not be neglected when developing new ways for self-management of chronic diseases.

Acknowledgments

A total of 50% of the funding of this study was received from the European Commission Information and Communication Technologies Policy Support Program (ICT PSP) 2009 of the Competitiveness and Innovation framework Programme (CIP), as part of the Renewing Health Project involving nine European countries. The other 50% of the funding was provided by Eksote.

Conflicts of Interest

Tapio Jokinen is the chairman of the board of Medixine Ltd that provided the remote patient monitoring system.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [30].

[PDF File (Adobe PDF File), 139KB - jmir_v17i6e153_app1.pdf]

References

1. Barr MS. The need to test the patient-centered medical home. JAMA 2008 Aug 20;300(7):834-835. [doi: 10.1001/jama.300.7.834] [Medline: 18714064]
2. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
3. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
4. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
5. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
6. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, TIM-HF Investigators. Telemedical Interventional Monitoring in Heart Failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in ambulatory patients with heart failure: study design. Eur J Heart Fail 2010 Dec;12(12):1354-1362. [doi: 10.1093/eurjhf/hfq199] [Medline: 21098580]
7. Health Quality Ontario. Home telemonitoring for type 2 diabetes: an evidence-based analysis. Ont Health Technol Assess Ser 2009;9(24):1-38 [FREE Full text] [Medline: 23074529]
8. Montori VM, Helgemoe PK, Guyatt GH, Dean DS, Leung TW, Smith SA, et al. Telecare for patients with type 1 diabetes and inadequate glycemic control: a randomized controlled trial and meta-analysis. Diabetes Care 2004 May;27(5):1088-1094. [Medline: 15111526]
9. Farmer A, Gibson OJ, Tarassenko L, Neil A. A systematic review of telemedicine interventions to support blood glucose self-monitoring in diabetes. Diabet Med 2005 Oct;22(10):1372-1378. [doi: 10.1111/j.1464-5491.2005.01627.x] [Medline: 16176199]
10. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
11. Wolever RQ, Dreusicke M, Fikkan J, Hawkins TV, Yeung S, Wakefield J, et al. Integrative health coaching for patients with type 2 diabetes: a randomized clinical trial. Diabetes Educ 2010;36(4):629-639. [doi: 10.1177/0145721710371523] [Medline: 20534872]
12. Patja K, Absetz P, Auvinen A, Tokola K, Kytö J, Oksman E, et al. Health coaching by telephony to support self-care in chronic diseases: clinical outcomes from The TERVA randomized controlled trial. BMC Health Serv Res 2012;12:147 [FREE Full text] [doi: 10.1186/1472-6963-12-147] [Medline: 22682298]
13. Ruggiero L, Moadsiri A, Butler P, Oros SM, Berbaum ML, Whitman S, et al. Supporting diabetes self-care in underserved populations: a randomized pilot study using medical assistant coaches. Diabetes Educ 2010;36(1):127-131 [FREE Full text] [doi: 10.1177/0145721709355487] [Medline: 20185612]
14. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
15. Walsh JM, Sundaram V, McDonald K, Owens DK, Goldstein MK. Implementing effective hypertension quality improvement strategies: barriers and potential solutions. J Clin Hypertens (Greenwich) 2008 Apr;10(4):311-316. [Medline: 18401229]
16. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4 [FREE Full text] [Medline: 10345255]
17. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JP. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ 2009;338:a3006 [FREE Full text] [Medline: 19139138]
18. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
19. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
20. Cruz J, Brooks D, Marques A. Home telemonitoring in COPD: a systematic review of methodologies and patients' adherence. Int J Med Inform 2014 Apr;83(4):249-263. [doi: 10.1016/j.ijmedinf.2014.01.008] [Medline: 24529402]
21. Karhula T, Kauppila T, Elonheimo O, Brommels M. Use of email in communication between the Finnish primary healthcare system and general practitioners. Inform Prim Care 2011;19(1):25-32. [Medline: 22118333]
22. Rose V, Harris M, Ho MT, Jayasinghe UW. A better model of diabetes self-management? Interactions between GP communication and patient self-efficacy in self-monitoring of blood glucose. Patient Educ Couns 2009 Nov;77(2):260-265. [doi: 10.1016/j.pec.2009.03.026] [Medline: 19720493]

23. Stone RA, Sevick MA, Rao RH, Macpherson DS, Cheng C, Kim S, et al. The Diabetes Telemonitoring Study Extension: an exploratory randomized comparison of alternative interventions to maintain glycemic control after withdrawal of diabetes home telemonitoring. J Am Med Inform Assoc 2012;19(6):973-979 [FREE Full text] [doi: 10.1136/amiajnl-2012-000815] [Medline: 22610495]
24. Turner-McGrievy GM, Tate DF. Are we sure that Mobile Health is really mobile? An examination of mobile device use during two remotely-delivered weight loss interventions. Int J Med Inform 2014 May;83(5):313-319 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.002] [Medline: 24556530]
25. Kennedy CM, Powell J, Payne TH, Ainsworth J, Boyd A, Buchan I. Active assistance technology for health-related behavior change: an interdisciplinary review. J Med Internet Res 2012;14(3):e80 [FREE Full text] [doi: 10.2196/jmir.1893] [Medline: 22698679]
26. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
27. Charpentier G, Benhamou P, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar;34(3):533-539 [FREE Full text] [doi: 10.2337/dc10-1259] [Medline: 21266648]
28. Orsama A, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]
29. Bock BC, Rosen RK, Barnett NP, Thind H, Walaska K, Foster R, et al. Translating behavioral interventions onto mHealth platforms: developing text message interventions for smoking and alcohol. JMIR Mhealth Uhealth 2015;3(1):e22 [FREE Full text] [doi: 10.2196/mhealth.3779] [Medline: 25714907]
30. 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

ANCOVA: analysis of covariance
BMI: body mass index
BP: bodily pain
CIP: Competitiveness and Innovation framework Programme
COPD: chronic obstructive pulmonary disease
D: patients with a diagnosis of diabetes mellitus type 2 and HbA1c > 6.5% (in Figure 2)
EHR: electronic health record
Eksote: South Karelia Social and Health Care District
GH: general health
H: patients with a diagnosis of ischemic heart disease or heart failure (in Figure 2)
HbA1c: hemoglobin A1c, glycosylated hemoglobin
HDL: high-density lipoprotein
HRQL: health-related quality of life
ICT PSP: Information and Communication Technologies Policy Support Program
IQR: interquartile range
LDL: low-density lipoprotein
MH: mental health
PC: personal computer
PF: physical functioning
PHR: personal health record
RCT: randomized controlled trial
RE: role-emotional
RP: role-physical
RPM: remote patient monitoring
SF: social functioning
SF-36: Short Form (36) Health Survey
SMS: short message service
TERVA: health coaching by telephony to support self-care in chronic diseases
VT: vitality
VTT: Technical Research Centre of Finland

Edited by G Eysenbach; submitted 24.11.14; peer-reviewed by G Paré, MM Ansari Ramandi; comments to author 22.02.15; revised version received 30.03.15; accepted 17.04.15; published 17.06.15.

Please cite as:
Karhula T, Vuorinen AL, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, Junno UM, Jokinen T, van Gils M, Lähteenmäki J, Kohtamäki K, Saranummi N
Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial
J Med Internet Res 2015;17(6):e153
URL: http://www.jmir.org/2015/6/e153/
doi:10.2196/jmir.4059
PMID:26084979

Original Paper

Using Web-Based Questionnaires and Obstetric Records to Assess General Health Characteristics Among Pregnant Women: A Validation Study

Marleen MHJ van Gelder[1,2], PhD; Naomi PE Schouten[1], BSc; Peter JFM Merkus[3,4], MD, PhD; Chris M Verhaak[5], PhD; Nel Roeleveld[1,6], PhD, MSc; Jolt Roukema[3], MD, PhD

[1]Department for Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands
[2]Radboud REshape Innovation Center, Radboud university medical center, Nijmegen, Netherlands
[3]Division of Respiratory Medicine, Department of Pediatrics, Radboudumc Amalia Children’s Hospital, Radboud university medical center, Nijmegen, Netherlands
[4]Department of Pediatrics, Canisius Wilhelmina Hospital, Nijmegen, Netherlands
[5]Department of Medical Psychology, Radboud university medical center, Nijmegen, Netherlands
[6]Department of Pediatrics, Radboudumc Amalia Children’s Hospital, Radboud university medical center, Nijmegen, Netherlands

Corresponding Author:
Marleen MHJ van Gelder, PhD
Department for Health Evidence
Radboud Institute for Health Sciences
Radboud university medical center
P.O. Box 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 24 3666126
Fax: 31 24 3613505
Email: Marleen.vanGelder@radboudumc.nl

Abstract

Background: Self-reported medical history information is included in many studies. However, data on the validity of Web-based questionnaires assessing medical history are scarce. If proven to be valid, Web-based questionnaires may provide researchers with an efficient means to collect data on this parameter in large populations.

Objective: The aim of this study was to assess the validity of a Web-based questionnaire on chronic medical conditions, allergies, and blood pressure readings against obstetric records and data from general practitioners.

Methods: Self-reported questionnaire data were compared with obstetric records for 519 pregnant women participating in the Dutch PRegnancy and Infant DEvelopment (PRIDE) Study from July 2011 through November 2012. These women completed Web-based questionnaires around their first prenatal care visit and in gestational weeks 17 and 34. We calculated kappa statistics (κ) and the observed proportions of positive and negative agreement between the baseline questionnaire and obstetric records for chronic conditions and allergies. In case of inconsistencies between these 2 data sources, medical records from the woman’s general practitioner were consulted as the reference standard. For systolic and diastolic blood pressure, intraclass correlation coefficients (ICCs) were calculated for multiple data points.

Results: Agreement between the baseline questionnaire and the obstetric record was substantial (κ=.61) for any chronic condition and moderate for any allergy (κ=.51). For specific conditions, we found high observed proportions of negative agreement (range 0.88-1.00) and on average moderate observed proportions of positive agreement with a wide range (range 0.19-0.90). Using the reference standard, the sensitivity of the Web-based questionnaire for chronic conditions and allergies was comparable to or even better than the sensitivity of the obstetric records, in particular for migraine (0.90 vs 0.40, *P*=.02), asthma (0.86 vs 0.61, *P*=.04), inhalation allergies (0.92 vs 0.74, *P*=.003), hay fever (0.90 vs 0.64, *P*=.001), and allergies to animals (0.89 vs 0.53, *P*=.01). However, some overreporting of allergies was observed in the questionnaire and for some nonsomatic conditions sensitivity of both measurement instruments was low. The ICCs for blood pressure readings ranged between 0.72 and 0.92 with very small mean differences between the 2 methods of data collection.

Conclusions: Web-based questionnaires can be used to validly collect data on many chronic disorders, allergies, and blood pressure readings among pregnant women.

(J Med Internet Res 2015;17(6):e149) doi:10.2196/jmir.3847

KEYWORDS

questionnaires; medical records; validation studies; pregnancy; chronic disease; allergens; blood pressure; Internet; PRIDE Study

Introduction

Self-reported methods of data collection are often applied in large-scale medical or biomedical studies for efficiency reasons. In these studies, it may not be feasible to conduct clinical measurements on all participants. Therefore, paper-and-pencil questionnaires or telephone interviews were traditionally used to gather information on the study variables. Nowadays, these modes of data collection are increasingly being substituted by Web-based questionnaires. However, knowledge on the validity of data collected with Web-based questionnaires is limited [1], although the quality of the data on a number of traditional epidemiologic risk factors, including body weight [2-4], smoking [5], alcohol consumption [6], and energy and macronutrient intake [7,8], is reported to be high. Medical history is included as an exposure or potential confounding factor in many studies and Web-based questionnaires may be an efficient way to collect these data in large samples of participants, if proven to be valid.

Most validation studies on medical history collected through self-reported methods has focused on chronic conditions, in particular cardiovascular diseases [9-15], diabetes [10,12-16], cancer [11,17,18], and asthma [10,13,14,19,20]. Agreement between self-reports and medical records differed among these studies and was affected by study methodology, target population, condition of interest, and the statistical analyses. In general, agreement was good for conditions that have clear diagnostic criteria, but it was low to moderate for conditions that are less serious or more complex to diagnose. Accordingly, discordance between questionnaires and biochemical measures or patch testing for allergic conditions or atopy is substantial [21-23]. Data on the validity of self-report on the results of common measurements taken during health care visits, such as blood pressure readings and hemoglobin levels, are very limited.

To the best of our knowledge, only Landkroon et al [24] compared data on medical history from a Web-based questionnaire with a "reference standard," but this study was too small (N=106) to produce robust estimates for levels of agreement. Therefore, the aim of this study was to assess the validity of a Web-based questionnaire on chronic conditions, allergies, and blood pressure readings among pregnant women by comparing the questionnaire data to obstetric records and data from general practitioners (GPs).

Methods

Setting

The Dutch prenatal care system is unique in the Western world. In the Netherlands, midwives are qualified to provide full prenatal care to all women with uncomplicated pregnancies and deliveries. The first prenatal care visit, which may be scheduled without referral of a general practitioner, usually takes place in gestational weeks 8 to 10 and frequent contacts are scheduled throughout pregnancy. Women are referred to a secondary or tertiary midwife or gynecologist in case of risk factors or complications. In 2013, 85% of pregnant women started their prenatal care in a primary care setting [25].

Study Population

We used data from the PRegnancy and Infant DEvelopment (PRIDE) Study, an ongoing, prospective cohort study that enrolls Dutch women early in pregnancy. The PRIDE Study started enrollment in July 2011 in the Nijmegen region and aims at including more than 150,000 pregnancies to study a broad range of research questions pertaining to maternal and child health. Details on the study design are described elsewhere [26]. Briefly, pregnant women aged 18 years and older were invited to participate in the PRIDE Study by their midwife or gynecologist just before or during their first prenatal care visit. They were asked to complete Web-based questionnaires at baseline, in gestational weeks 17 (questionnaire 2) and 34 (questionnaire 3), as well as 2 and 6 months after the estimated date of delivery. The baseline questionnaire was completed between weeks 6 and 16 of gestation. Researchers from various medical disciplines selected, modified, and tailored existing, validated paper-based questionnaires or parts thereof to fit our Web-based application. Paper-based questionnaires were available for women who could not or did not want to participate through the Internet (n=1; excluded from this study). Questions were asked on demographic factors, reproductive history, maternal health, lifestyle factors, and occupational exposures. Furthermore, consent was asked for review of medical records to enrich the PRIDE Study database with detailed clinical information.

Data Collection

Through the baseline questionnaire, data on medical history were collected. Women were asked gateway questions to assess chronic conditions ("Do you have a chronic or long-term illness that was diagnosed by a medical doctor" followed by some examples of chronic conditions) and allergies ("Do you have an allergy or eczema?"). These questions were followed by multiple-choice questions with blank options to specify the chronic condition or allergy among those who answered positively to the relevant gateway question. Chronic conditions reported in other parts of the baseline questionnaire (eg, as causes for subfertility or as indications for medication use) were included in the analysis as well. In each prenatal questionnaire, we asked for the date of the most recent prenatal care visit, whether blood pressure was measured during this visit, and if so, for the systolic and diastolic blood pressure readings in mm

Hg. A screenshot of the relevant parts of the questionnaires is provided in Multimedia Appendix 1.

A pretested, standardized case report form (CRF) was used to abstract data from the obstetric records of women who gave consent for medical record review. For logistical reasons, obstetric records were only reviewed in participating study centers in the Nijmegen region (7 midwifery practices and 1 academic hospital). Using the CRF, 2 medically trained abstracters collected data from the obstetric records on medical history, including chronic conditions, allergies, and pregnancy history, the pregnancy itself, anthropometrical measures including blood pressure taken during pregnancy, and pregnancy outcome, not all of which were included in this validation study.

Preexisting medical conditions are self-reported by the pregnant woman during the first prenatal care visit and are usually only recorded in the obstetric record by the prenatal care provider if deemed important for the course of pregnancy or the delivery [27]. As a consequence, obstetric records may not be a suitable reference standard for self-reported chronic conditions and allergies. Therefore, information on the diagnosis of chronic conditions and allergies was obtained from the woman’s GP in case of inconsistencies between the questionnaire and the obstetric record for reasons of efficiency.

Chronic conditions were classified and coded according to the World Health Organization’s *International Classification of Diseases, Tenth Revision* [28]. Allergies were ordered into 6 mutually exclusive categories: (1) inhalation allergies (hay fever, allergies to animals, and house dust mite allergy), (2) food allergies, (3) allergic contact dermatitis (allergies to metal, fragrance hypersensitivity, plaster allergy, and latex allergy), (4) insect sting allergy, (5) medication allergies, and (6) other allergies.

Statistical Analysis

Only PRIDE Study participants with complete information on chronic conditions, allergies, and blood pressure during the most recent prenatal care visit in the baseline questionnaire who gave consent to review their medical records were included in this validation study. For chronic conditions and allergies with at least 5 cases in either the questionnaire or the obstetrical record, we calculated kappa statistics (κ) to quantify agreement between the baseline questionnaire and the obstetric record regarding chronic conditions and allergies. We also calculated the observed proportions of positive and negative agreement (p_{pos} and p_{neg}, respectively) because kappa is strongly affected by imbalances in marginal totals (ie, a low kappa despite a high level of agreement) [29]. The calculation of p_{pos} and p_{neg} is shown in Figure 1 [30].

To determine which method of data collection was most valid to collect information on chronic conditions and allergies among pregnant women, sensitivity and specificity were calculated with GP data until the date of completion of the baseline questionnaire as our reference standard. When GP data were unavailable, pharmacy records were screened for diagnoses of chronic conditions or allergies and for medication dispensed that was indicative for chronic conditions or allergies. In addition to the discordant questionnaire–obstetric record pairs, women with positive scores on both the Web-based questionnaire and the obstetric record were included in these calculations as true positives. Likewise, women with negative scores on both methods were included as true negatives. We assessed potential differences in sensitivity and specificity between the questionnaires and the obstetric records using chi-square tests.

For the validity analyses regarding blood pressure readings, only women with an exact match between the date of the most recent prenatal care visit reported in any of the prenatal questionnaires and a visit date recorded in the obstetric record were included to be certain that both data sources referred to the same measurement. Intraclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) for systolic blood pressure (SBP) and diastolic blood pressure (DBP) were calculated using 2-way mixed effects models (single measure). To assess absolute agreement and potential differences in bias within the SBP and DBP range, we plotted the difference in blood pressure readings between the questionnaire and the obstetric record (y-axis) against the mean of the 2 methods of data collection (x-axis) according to the Bland-Altman technique [31]. In secondary analyses, we included all women who reported the most recent prenatal care visit date in the questionnaire within 5 days of a visit date recorded in the obstetric record. All statistical analyses were performed using IBM SPSS version 20 (IBM Corp, Armonk, NY, USA), except for p_{pos} and p_{neg}, which were calculated in Microsoft Office Excel 2007 (Microsoft Corp, Redmond, WA, USA).

Figure 1. Calculation of positive and negative agreement between two tests.

	Test 2		
Test 1	Positive	Negative	Total
Positive	a	b	g_1
Negative	c	d	g_2
Total	f_1	f_2	N

$$p_{pos} = \frac{\text{\# positive readings that both tests agree on}}{\text{average number of positive tests}} = \frac{a}{\left(\frac{f_1 + g_1}{2}\right)} = \frac{2a}{f_1 + g_1} = \frac{2a}{N + (a - d)}$$

$$p_{neg} = \frac{\text{\# negative readings that both tests agree on}}{\text{average number of negative tests}} = \frac{d}{\left(\frac{f_2 + g_2}{2}\right)} = \frac{2d}{f_2 + g_2} = \frac{2d}{N - (a - d)}$$

Results

Women enrolled in the PRIDE Study between July 2011 through November 2012 were eligible for this study (N=725). The overall participation rate in the PRIDE Study was 42.90% (725/1690) during this time period. Figure 2 shows the flow of participants. Of the 725 women enrolled during the study period, 22 (3.0%) only completed a few sections of the baseline questionnaire, mostly because of technical issues in the first weeks of enrollment. Among those with complete baseline questionnaires, 24.8% (174/703) did not give consent for medical record review. Furthermore, 10 women were excluded because their obstetric records were not available (n=9) or they participated with multiple pregnancies in the PRIDE Study (n=1). Therefore, 519 women were included in this validation study. Compared with the women who did not give consent to obtain medical records, women participating in this validation study were more likely to have a lower level of education (*P*=.03) and to be obese (*P*=.06; Table 1). Furthermore, women who did not give consent for medical record review were more likely to have completed the baseline questionnaire before their first prenatal care visit compared to women included in the validation study (*P*=.02). We did not observe substantial differences in maternal age, country of birth, gravidity, and gestational age at inclusion between these 2 groups. Regarding the blood pressure readings, follow-up information was not available for all participants for several reasons: (1) they did not reach the gestational week for administration of questionnaire 2 or 3 yet at the date of obstetric record review; (2) they had a miscarriage, stillbirth, termination of pregnancy (TOP), or very preterm birth; or (3) they skipped questionnaire 2 or 3, were lost to follow-up, or changed prenatal care provider resulting in incomplete obstetric records.

Figure 2. Flow chart of study participants.

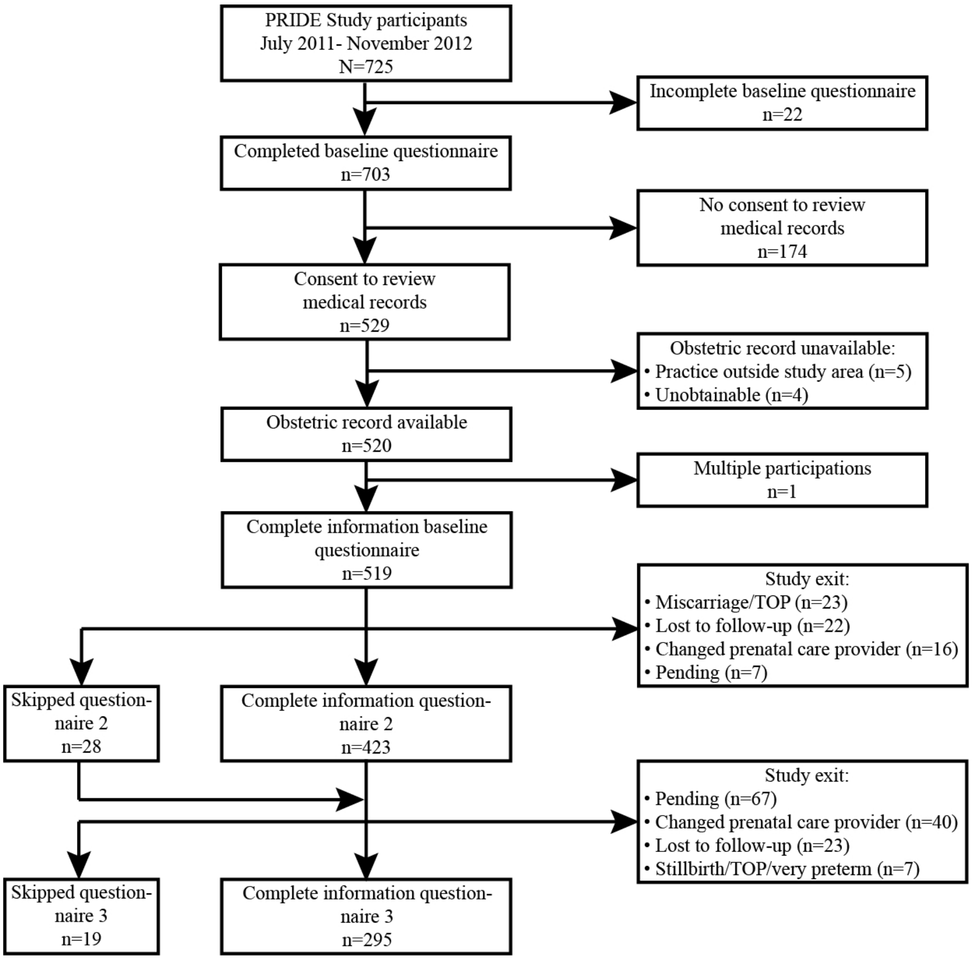


Table 1. Characteristics of PRIDE Study participants included in this validation study and participants who did not give consent for review of medical records.

Characteristic	Participants in validation study, n (%) (N=519)[a]	No consent for medical record review, n (%) (N=174)[a]	P [b]
Maternal age at inclusion (years)			.45
<25	15 (2.9)	2 (1.1)	
25-29	172 (33.1)	52 (29.9)	
30-34	238 (45.9)	88 (50.6)	
≥35	94 (18.1)	32 (18.4)	
Maternal country of birth			.20
Netherlands	483 (93.1)	154 (88.5)	
Other	31 (6.0)	15 (8.6)	
Maternal level of education [c]			.03
Low/intermediate	107 (20.6)	22 (12.6)	
High	407 (78.4)	147 (84.5)	
Gravidity			.62
0	232 (44.7)	74 (42.5)	
≥1	287 (55.3)	100 (57.5)	
BMI before pregnancy [d]			.06
Underweight (<18.5 kg/m^2)	15 (2.9)	2 (1.1)	
Normal (18.5-24.9 kg/m^2)	336 (64.7)	122 (70.1)	
Overweight (25.0-29.9 kg/m^2)	90 (17.3)	27 (15.5)	
Obese (≥30 kg/m^2)	33 (6.4)	3 (1.7)	
Height/weight unknown	45 (8.7)	20 (11.5)	
Gestational age at inclusion (weeks)			.17
<8	96 (18.5)	41 (23.6)	
8-10	178 (34.3)	48 (27.6)	
>10	245 (47.2)	85 (48.9)	
Timing of baseline questionnaire			.02
Completed before first prenatal care visit	131 (25.2)	60 (34.5)	
Completed after first prenatal care visit	388 (74.8)	114 (65.5)	

[a] Numbers may not add up to total group size due to missing values.

[b] Difference between the 2 groups using chi-square tests.

[c] High level of education: completed higher vocational education or university.

[d] Body mass index (BMI) derived from self-reported height and weight.

Of the 519 participants, 118 (22.7%) women reported having a chronic condition in the baseline questionnaire, whereas chronic conditions were recorded in the obstetric records of 105 (20.2%) women. Overall, agreement between the Web-based questionnaire and the obstetric record was substantial for any chronic condition (κ=.61; Table 2) with a higher p_{neg} (0.92) than p_{pos} (0.69). Level of agreement differed between the specific chronic conditions with relatively high levels of agreement for endocrine, nutritional, and metabolic diseases (κ=.72) and in particular for thyroid disease (κ=.90), epilepsy (κ=.89), and diseases of the genitourinary tract (κ=.72). However, for a number of conditions, including migraine (κ=.30), diseases of the circulatory system (κ=.25), and irritable bowel syndrome (κ=.39), agreement between the questionnaire and the obstetric record was poor. For all specific conditions, the p_{neg} was high (range 0.98-1.00), but the p_{pos} followed a pattern comparable to the kappa statistic.

Table 2. Agreement between data from the Web-based questionnaire and obstetric record for chronic conditions (n=519).

Chronic condition	Questionnaire positive, n		Questionnaire negative, n		κ	p_{pos}[a]	p_{neg}[b]
	Record positive	Record negative	Record positive	Record negative			
Any chronic condition	77	41	28	373	.61	0.69	0.92
Diseases of the blood and blood-forming organs	2	0	0	517	—	—	—
Thalassemia	1	0	0	518	—	—	—
Immunodeficiency	1	0	0	518	—	—	—
Endocrine, nutritional, and metabolic diseases	15	5	6	493	.72	0.73	0.99
Thyroid disease	9	1	1	508	.90	0.90	1.00
Polycystic ovarian syndrome	5	3	5	506	.55	0.56	0.99
Hypercholesterolemia	0	1	0	518	—	—	—
Periodic fever syndrome	1	0	0	518	—	—	—
Mental and behavioral disorders	7	6	4	502	.57	0.58	0.99
Depression/anxiety	5	5	4	505	.52	0.53	0.99
Posttraumatic stress disorder	1	1	0	517	—	—	—
ADD/ADHD[c]	1	1	0	517	—	—	—
Diseases of the nervous system	12	17	4	486	.51	0.53	0.98
Multiple sclerosis	1	0	0	518	—	—	—
Epilepsy	4	0	1	514	.89	0.89	1.00
Migraine	4	17	1	497	.30	0.31	0.98
Tension-type headache	1	0	1	517	—	—	—
Chronic fatigue syndrome	2	0	1	516	—	—	—
Diseases of the circulatory system	1	1	5	512	.25	0.25	0.99
Hypertension	0	0	3	516	—	—	—
Cardiac arrhythmia	0	1	1	517	—	—	—
Raynaud syndrome	1	0	1	517	—	—	—
Diseases of the respiratory system	19	17	5	478	.61	0.63	0.98
Asthma	19	17	5	478	.61	0.63	0.98
Diseases of the digestive system	8	4	6	501	.61	0.62	0.99
Crohn disease	2	0	1	516	—	—	—
Ulcerative colitis	3	0	0	516	—	—	—
Irritable bowel syndrome	3	4	5	507	.39	0.40	0.99
Diseases of the skin and subcutaneous tissue	4	7	1	507	.49	0.50	0.99
Psoriasis	4	5	0	510	.61	0.62	1.00
Rosacea	0	2	1	516	—	—	—
Diseases of the musculoskeletal system and connective tissue	9	4	4	502	.68	0.69	0.99
Rheumatoid arthritis	2	2	0	515	—	—	—
Sjögren syndrome	1	0	0	518	—	—	—
Ankylosing spondylitis	1	0	1	517	—	—	—
Hypermobility	1	1	3	514	—	—	—
Fibromyalgia	2	2	1	514	—	—	—
Complex regional pain syndrome	1	0	0	518	—	—	—
Diseases of the genitourinary tract	4	1	2	512	.72	0.73	1.00

Chronic condition	Questionnaire positive, n		Questionnaire negative, n		κ	p_{pos}[a]	p_{neg}[b]
	Record positive	Record negative	Record positive	Record negative			
Endometriosis	1	1	1	516	—	—	—
Lichen sclerosis	3	0	1	515	—	—	—

[a] Observed proportion of positive agreement.

[b] Observed proportion of negative agreement.

[c] ADD: attention deficit disorder; ADHD: attention deficit hyperactivity disorder.

Allergies were reported by 229 of 519 (44.1%) women in the baseline questionnaire and recorded in the obstetric record of 168 (32.4%) women. In Table 3, agreement between the Web-based questionnaire and the obstetric record is shown for the mutually exclusive groups of allergies and selected specific allergies. For any allergy, agreement between the questionnaire and the obstetric record was moderate (κ=.51) with a p_{pos} and p_{neg} of 0.70 and 0.81, respectively. The kappa values for the groups of allergies ranged between 0.21 (insect sting allergy) and 0.66 (drug allergies) and between 0.33 (fragrance hypersensitivity) and 0.73 (latex allergy) for the specific types of allergies. House dust mite allergy, latex allergy, and drug allergies were more often reported in the obstetric record than in the questionnaire. Again, the p_{neg} (range 0.81-1.00) was higher than the p_{pos} (range 0.19-0.73) for all groups of allergies or specific allergies included.

Table 3. Agreement between data from a Web-based questionnaire and obstetric record for allergies (n=519).

Allergy	Questionnaire positive, n		Questionnaire negative, n		κ	p_{pos}[a]	p_{neg}[b]
	Record positive	Record negative	Record positive	Record negative			
Any allergy	138	91	30	260	.51	0.70	0.81
Inhalation allergies	80	83	10	346	.53	0.63	0.88
Hay fever	53	53	3	410	.60	0.65	0.94
Allergies to animals	21	45	7	446	.40	0.45	0.94
House dust mite allergy	15	15	22	467	.41	0.45	0.96
Food allergies	15	39	1	464	.40	0.43	0.96
Allergic contact dermatitis	33	73	11	402	.36	0.44	0.91
Allergies to metal	11	37	3	468	.33	0.35	0.96
Fragrance hypersensitivity	6	51	0	462	.17	0.19	0.95
Plaster allergy	9	16	7	487	.42	0.44	0.98
Latex allergy	4	0	3	512	.73	0.73	1.00
Insect sting allergy	3	20	1	495	.21	0.22	0.98
Drug allergies	26	9	15	469	.66	0.68	0.98
Other allergies	2	3	0	514	.57	0.57	1.00

[a] Observed proportion of positive agreement.

[b] Observed proportion of negative agreement.

Regarding the 254 women with an inconsistency between the Web-based questionnaire and the obstetric record for chronic conditions or allergies, complete GP data were obtained for 194 (76.4%) women; the GP was unknown for 12 women, 21 women were not registered with the GP whose name was provided, the GP did not respond to our multiple data requests for 25 women, and GP records were incomplete for 2 women. For 7 women lacking GP data, the diagnosis of a chronic disorder was ascertained from their pharmacy records. Generally, sensitivity was better for the Web-based questionnaire than for the obstetric record when compared to GP data (Table 4), specifically for migraine (0.90 vs 0.40, *P*=.02), asthma (0.86 vs 0.61, *P*=.04), any allergy (0.96 vs 0.85, *P*=.007), inhalation allergies (0.92 vs 0.74, *P*=.003), hay fever (0.90 vs 0.64, *P*=.001), and allergies to animals (0.89 vs 0.53, *P*=.01). For a number of chronic conditions, including mental and behavioral disorders, depression/anxiety, and irritable bowel syndrome, sensitivity of both measurement instruments was low. Overall, specificity of the Web-based questionnaire and the obstetric record was high. However, specificity of the questionnaire was slightly lower than specificity of the obstetric record for a number of (groups of) allergies, including any allergy (0.74 vs 0.85, *P*=.009), inhalation allergies (0.83 vs 0.93, *P*<.001), hay fever (0.91 vs 0.96, *P*=.001), allergies to animals (0.91 vs 0.97, *P*<.001), food allergies (0.92 vs 0.98, *P*<.001), allergic contact dermatitis (0.84 vs 0.94, *P*<.001), allergies to metal (0.94 vs 0.99, *P*<.001), and fragrance hypersensitivity (0.91 vs 0.99, *P*<.001).

Table 4. Validity comparisons of chronic conditions and allergies among pregnant women: Web-based questionnaires and obstetric records compared to GP records.

Condition or allergy	n	Sensitivity			Specificity		
		Questionnaire	Record	*P* [a]	Questionnaire	Record	*P* [a]
Any chronic condition	496	0.83	0.74	.14	0.93	0.92	.79
Endocrine, nutritional, and metabolic diseases	515	0.76	0.70	.73	1.00	0.99	.41
Thyroid disease	518	0.77	0.69	.66	1.00	1.00	>.99
Polycystic ovarian syndrome	516	0.67	0.83	.52	1.00	0.99	.41
Mental and behavioral disorders	516	0.45	0.27	.39	0.99	0.99	.78
Depression/anxiety	516	0.50	0.30	.37	0.99	0.99	.48
Diseases of the nervous system	513	0.85	0.65	.15	0.99	1.00	.09
Epilepsy	519	0.80	1.00	.32	1.00	1.00	>.99
Migraine	514	0.90	0.40	.02	0.99	1.00	.03
Diseases of the respiratory system	511	0.86	0.61	.04	0.99	0.99	.76
Asthma	511	0.86	0.61	.04	0.99	0.99	.76
Diseases of the digestive system	517	0.58	0.67	.68	0.99	0.99	.74
Irritable bowel syndrome	517	0.33	0.33	>.99	0.99	0.99	.74
Diseases of the skin and subcutaneous tissue	519	0.69	0.38	.12	1.00	1.00	.16
Psoriasis	519	0.80	0.40	.08	1.00	1.00	.32
Diseases of the musculoskeletal system and connective tissue	515	1.00	0.82	.15	1.00	1.00	.16
Diseases of the genitourinary tract	518	0.80	1.00	.32	1.00	1.00	.32
Any allergy	494	0.96	0.85	.007	0.74	0.85	.009
Inhalation allergies	496	0.92	0.74	.003	0.83	0.93	<.001
Hay fever	508	0.90	0.64	.001	0.91	0.96	.001
Allergies to animals	509	0.89	0.53	.01	0.91	0.97	<.001
House dust mite allergy	512	0.58	0.65	.58	0.98	0.97	.44
Food allergies	511	1.00	0.86	.32	0.92	0.98	<.001
Allergic contact dermatitis	502	1.00	1.00	>.99	0.84	0.94	<.001
Allergies to metal	508	1.00	1.00	>.99	0.94	0.99	<.001
Fragrance hypersensitivity	507	1.00	1.00	>.99	0.91	0.99	<.001
Plaster allergy	511	1.00	1.00	>.99	0.96	0.97	.36
Drug allergies	515	0.84	0.95	.30	0.97	0.96	.32

[a] Difference between the 2 modes of data collection using chi-square tests.

For 4 chronic conditions, additional self-reports were identified in the questions about causes of subfertility that preceded the chronic condition question (polycystic ovarian syndrome [n=7] and endometriosis [n=2]) and through medication use (rosacea [n=2] and lichen sclerosis [n=1]). When these women were considered as not having reported these chronic conditions, agreement between the Web-based questionnaire and the obstetric record decreased, except for skin diseases. Furthermore, it decreased the sensitivity of the questionnaire, especially for endocrine diseases (0.67), polycystic ovarian syndrome (no true positive subjects), and diseases of the genitourinary tract (0.67).

Analyses on the validity of the Web-based questionnaires for blood pressure readings could not be conducted on the complete study sample. At baseline, 123 of 519 (23.7%) women did not have a prenatal care visit yet and, therefore, no valid blood pressure measurement (Table 5). Among women with a prenatal care visit, no match on visit date was established for 91 of 396 (23.0%), 65 of 423 (15.4%), and 32 of 295 (10.8%) women for the baseline questionnaire, questionnaire 2, and questionnaire

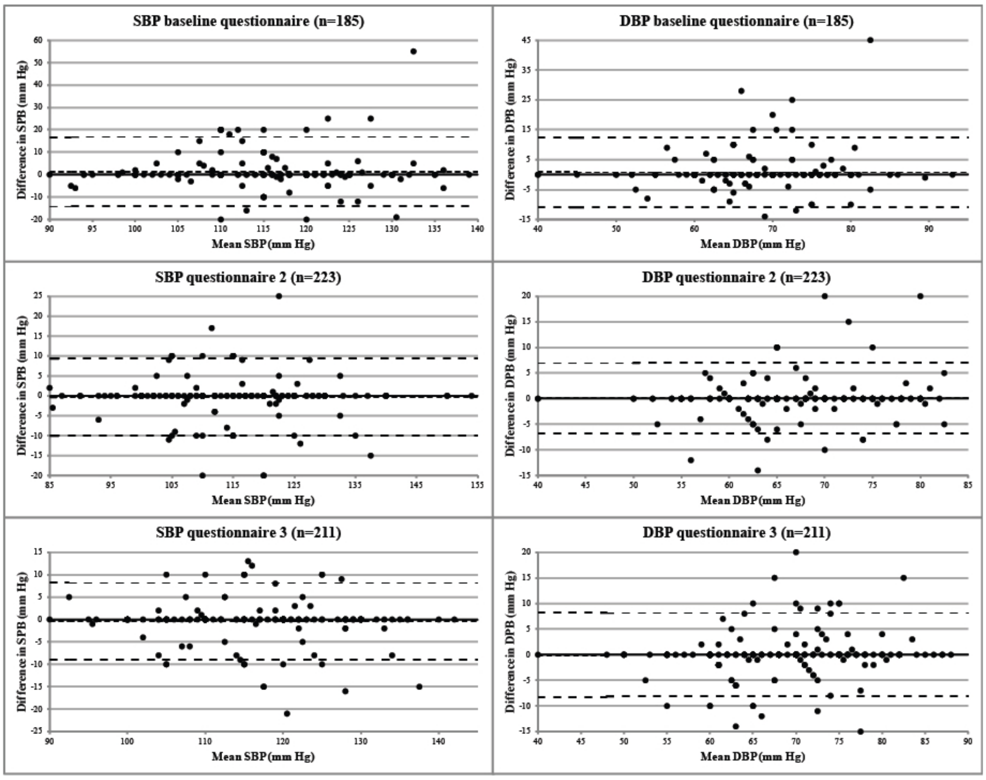
3, respectively. Furthermore, a substantial proportion of women whose blood pressure was measured could not remember the blood pressure readings (baseline questionnaire: 27.9%, 76/272; questionnaire 2: 28.4%, 93/328; questionnaire 3: 19.1%, 50/262). Of the women included at baseline and eligible for the reliability analyses of the follow-up questionnaires, 78.6% (121/154) and 84.6% (88/104) were included for questionnaires 2 and 3, respectively. Out of the 142 women included for questionnaire 2 and eligible for the analysis of questionnaire 3, 128 (90.1%) were included for questionnaire 3.

Table 5. Validity analyses comparing Web-based questionnaires and obstetric records for systolic and diastolic blood pressure readings: sample description and intraclass correlation coefficients.

Characteristic	Baseline questionnaire (n=519)	Questionnaire 2 (n=423)	Questionnaire 3 (n=295)
Did not have prenatal care visit yet, n	123	0	0
No match date prenatal care visit, n	91	65	32
Questionnaire: blood pressure not measured, n	33	30	1
Questionnaire: blood pressure unknown, n	76	93	50
Obstetric record: blood pressure not recorded, n	11	12	1
Included in validity analyses, n	185	223	211
Blood pressure, ICC (95% CI)			
Systolic	0.72 (0.65-0.79)	0.92 (0.89-0.94)	0.90 (0.88-0.93)
Diastolic	0.79 (0.73-0.84)	0.91 (0.88-0.93)	0.89 (0.86-0.91)

At baseline, the ICCs for SBP and DBP were 0.72 (95% CI 0.65-0.79) and 0.79 (95% CI 0.73-0.84), respectively. In the follow-up questionnaires, ICCs were substantially higher, ranging between 0.89 (95% CI 0.86-0.91; DBP in questionnaire 3) and 0.92 (95% CI 0.89-0.94; SBP in questionnaire 2). The Bland-Altman plots (Figure 3) also showed good agreement between the 2 methods of data collection with very small mean differences, ranging between 1.26 mm Hg (SD 7.72) for SBP in the baseline questionnaire and –0.04 (SD 4.09) for DBP in questionnaire 3. No trends in bias within the SBP and DBP ranges were observed. The secondary analyses, in which the date of the prenatal care visit was allowed to differ up to 5 days between the questionnaire and the obstetric record, yielded similar results (data not shown).

Figure 3. Bland-Altman plots showing the differences in reported systolic blood pressure (SBP) and diastolic blood pressure (DBP) between the 3 Web-based questionnaires and the obstetric record plotted against the mean of the 2 methods of data collection. Each data point shows one participant. The short dashed line shows the mean difference. The long dashed lines show the 95% limits of agreement (mean difference ±2 SD).



Discussion

Principal Findings

Web-based questionnaires are increasingly being used as a method of data collection in medical research. The results from the present study show that data on many chronic conditions and allergies can be validly collected among pregnant women using Web-based questionnaires with sensitivities comparable to or even higher than obstetric records. However, some overreporting of allergies was observed and absence of disease was more accurately reported than presence of disease. In addition, pregnant women were able to reliably recall blood pressure readings from the most recent prenatal care visit, especially in the follow-up questionnaires, but a substantial proportion of women could not remember their blood pressure readings at all.

Strengths and Limitations

In addition to the relatively large sample size, the use of GP records as a reference standard to validate the Web-based questionnaire and obstetric records for chronic conditions and allergies is a major strength of this study. In the Netherlands, inhabitants are obligatory listed with one GP, who coordinates access to specialized care and always receives all relevant medical information about the patient [32]. Therefore, GP records should contain the most complete information, although inaccuracies in registration of diagnoses cannot be excluded. Other strengths of this validation study include the high consent rate (75.2%) to review medical records, the high retrieval rate of obstetric and GP records (98.3% and 76.4%, respectively), and the high willingness of PRIDE Study participants to complete questionnaires through the Internet despite the study's mixed-mode design.

Women participating in the PRIDE Study represent a highly educated population, potentially limiting the generalizability of our results. However, women included in the validation study had a lower level of education compared to women who did not give consent for review of medical records. Previous studies on the association between maternal level of education and recall sensitivity of pregnancy-related events showed inconsistent results [33-36], indicating that imbalances in this baseline characteristic may or may not be a major threat to external validity.

Validity could not be determined reliably for a number of specific chronic conditions due to their low prevalence rates in our study population or in strata based on baseline characteristics. However, it was not feasible to increase the size of the study population because medical record abstraction is a labor-intensive process. Moreover, during the time frame of this study, only one secondary/tertiary care facility participated in the PRIDE Study. Women with certain medical conditions, including preexisting hypertension or diabetes and rheumatoid arthritis, are often referred to these facilities for prenatal care in the Netherlands. Reassuringly, only a small proportion of pregnant women (15%) start prenatal care in a secondary or tertiary care setting, mainly because of complications in a previous pregnancy [25].

Comparison With Prior Work

For many chronic conditions that were included in our analyses, data on the validity of self-report are scarce due to differences in study populations between this study among pregnant women and previous studies, which often selected an older population with higher prevalences of cardiovascular diseases, diabetes, and cancer. However, the general pattern of a better agreement for chronic conditions that have clear diagnostic criteria than for conditions that are less well-defined observed previously [11,37,38] was also visible in our study. We observed high sensitivities and specificities for somatic diseases, but low levels of agreement for a number of nonsomatic diseases, including mental and behavioral disorders and irritable bowel syndrome. This was not only the case for data from the Web-based questionnaire, but also for data from the obstetric records. Possible causes for this variability include poor communication between the patient and the health care provider, limited health literacy of the patient, or self-diagnosis in the absence of a satisfactory medical explanation for the symptoms [39].

Surprisingly, sensitivity of the Web-based questionnaire was substantially higher for asthma (0.86) and migraine (0.90) compared to the obstetric record, whereas the specificities were comparable. The traditional self-reported modes of data collection have a sensitivity ranging between 0.55 and 0.95 (median 0.72) for asthma [10,13,14,19,40] and between 0.35 and 0.67 (median 0.51) for migraine [41-43], suggesting that Web-based questionnaires might be more suitable for detecting subjects with these conditions in epidemiologic studies than paper-based questionnaires, interviews, and obstetric records. However, future studies should confirm these findings, also taking into account the manner in which the questions about these conditions are posed.

With regard to allergies, the Web-based questionnaire also seemed to be more sensitive than the obstetric record, but at the expense of its specificity indicating that overreporting occurs with the use of the Web-based questionnaire and underreporting is present when using obstetric records. However, participants with allergic symptoms who manage their symptoms with over-the-counter medication may not be registered as allergic in GP records, resulting in a lower specificity (increased number of false positives). Therefore, skin-prick tests or serum-specific immunoglobulin E levels may be a more appropriate reference standard. In comparison with previous studies in different populations [20-23], allergies were somewhat more accurately reported in our Web-based questionnaire compared to the other self-reported modes of data collection.

Research interests in changes in blood pressure over time in relation to disease outcomes is growing (eg, [44,45]), but obtaining data on individual blood pressure readings may be challenging. Alonso et al [46] observed a low correlation between self-reported and directly observed information on SBP and DBP among 127 university graduates with an ICC of 0.35 (95% CI 0.09-0.55 and 95% CI 0.16-0.51, respectively). We are not aware of other studies reporting on the validity of self-reported blood pressure readings. In our longitudinal study, we observed a learning effect; the ICC for SBP and DBP was higher for the follow-up questionnaires than for the baseline questionnaire. Once women reported a blood pressure reading, they were very likely to report blood pressure readings in follow-up questionnaires as well. In addition, the proportion of women who could not remember their blood pressure readings decreased. As a future alternative to self-reports of blood pressure measurements conducted in health care settings, home blood pressure telemonitoring may be used to collect data on blood pressure changes over time. In addition, dedicated applications may be developed in which pregnant women could record their blood pressure readings directly after every prenatal care visit.

Conclusions

We showed that Web-based questionnaires can validly collect data on many chronic disorders, including asthma, migraine, and thyroid disease, and also allergies among pregnant women with equal or better data quality compared to obstetric records. Although a substantial proportion of women could not remember their blood pressure readings, pregnant women who did recall the readings, recalled them well. This indicates that accurate data on general health characteristics may be collected using Web-based questionnaires in this population.

Acknowledgments

This work was supported by a grant from the Lung Foundation Netherlands and grants to the PRIDE Study from the Netherlands Organisation for Health Research and Development (ZonMw).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the Web-based questionnaire.

[PDF File (Adobe PDF File), 289KB - jmir_v17i6e149_app1.pdf]

References

1. Van Gelder MMHJ, Pijpe A. E-epidemiology: a comprehensive update. OA Epidemiology 2013 Jun;1(1):5. [doi: 10.13172/2053-079X--1-584]
2. Bonn SE, Trolle Lagerros Y, Bälter K. How valid are Web-based self-reports of weight? J Med Internet Res 2013 Apr;15(4):e52 [FREE Full text] [doi: 10.2196/jmir.2393] [Medline: 23570956]
3. Lassale C, Péneau S, Touvier M, Julia C, Galan P, Hercberg S, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Santé study. J Med Internet Res 2013 Aug;15(8):e152 [FREE Full text] [doi: 10.2196/jmir.2575] [Medline: 23928492]
4. West R, Gilsenan A, Coste F, Zhou X, Brouard R, Nonnemaker J, et al. The ATTEMPT cohort: a multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. Addiction 2006 Sep;101(9):1352-1361. [doi: 10.1111/j.1360-0443.2006.01534.x] [Medline: 16911736]
5. Brigham J, Lessov-Schlaggar CN, Javitz HS, Krasnow RE, McElroy M, Swan GE. Test-retest reliability of web-based retrospective self-report of tobacco exposure and risk. J Med Internet Res 2009 Mar;11(3):e35 [FREE Full text] [doi: 10.2196/jmir.1248] [Medline: 19674962]
6. Graham AL, Papandonatos GD, Bock BC, Cobb NK, Baskin-Sommers A, Niaura R, et al. Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation. Nicotine Tob Res 2006 Dec;8 Suppl 1:S49-S57 [FREE Full text] [Medline: 17491171]
7. Christensen SE, Möller E, Bonn SE, Ploner A, Wright A, Sjölander A, et al. Two new meal- and web-based interactive food frequency questionnaires: validation of energy and macronutrient intake. J Med Internet Res 2013 Jun;15(6):e109 [FREE Full text] [doi: 10.2196/jmir.2458] [Medline: 23739995]
8. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
9. St Sauver JL, Hagen PT, Cha SS, Bagniewski SM, Mandrekar JN, Curoe AM, et al. Agreement between patient reports of cardiovascular disease and patient medical records. Mayo Clin Proc 2005 Feb;80(2):203-210. [Medline: 15704775]
10. Heliövaara M, Aromaa A, Klaukka T, Knekt P, Joukamaa M, Impivaara O. Reliability and validity of interview data on chronic diseases. The Mini-Finland Health Survey. J Clin Epidemiol 1993 Feb;46(2):181-191. [Medline: 8437034]
11. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. Am J Epidemiol 1986 May;123(5):894-900. [Medline: 3962971]

12. Alves E, Lunet N, Correia S, Morais V, Azevedo A, Barros H. Medical record review to recover missing data in a Portuguese birth cohort: agreement with self-reported data collected by questionnaire and inter-rater variability. Gac Sanit 2011 Jun;25(3):211-219 [FREE Full text] [doi: 10.1016/j.gaceta.2010.10.013] [Medline: 21466908]
13. Muggah E, Graves E, Bennett C, Manuel DG. Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. BMC Public Health 2013 Jan;13:16 [FREE Full text] [doi: 10.1186/1471-2458-13-16] [Medline: 23302258]
14. Oksanen T, Kivimäki M, Pentti J, Virtanen M, Klaukka T, Vahtera J. Self-report as an indicator of incident disease. Ann Epidemiol 2010 Jul;20(7):547-554. [doi: 10.1016/j.annepidem.2010.03.017] [Medline: 20538198]
15. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol 2004 Oct;57(10):1096-1103. [doi: 10.1016/j.jclinepi.2004.04.005] [Medline: 15528061]
16. Ramadhani TA, Canfield MA, Waller DK, Case AP. Medical records vs. interview responses: a comparative analysis of selected variables for linked birth defect cases. Birth Defects Res A Clin Mol Teratol 2004 Sep;70(9):592-596. [doi: 10.1002/bdra.20068] [Medline: 15368558]
17. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. Am J Epidemiol 1998 Mar;147(6):556-562 [FREE Full text] [Medline: 9521182]
18. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. Am J Epidemiol 2001 Feb;153(3):299-306 [FREE Full text] [Medline: 11157418]
19. Weakley J, Webber MP, Ye F, Zeig-Owens R, Cohen HW, Hall CB, et al. Agreement between obstructive airways disease diagnoses from self-report questionnaires and medical records. Prev Med 2013 Jul;57(1):38-42. [doi: 10.1016/j.ypmed.2013.04.001] [Medline: 23597657]
20. Linet MS, Harlow SD, McLaughlin JK, McCaffrey LD. A comparison of interview data and medical records for previous medical conditions and surgery. J Clin Epidemiol 1989;42(12):1207-1213. [Medline: 2585011]
21. Delclos GL, Arif AA, Aday L, Carson A, Lai D, Lusk C, et al. Validation of an asthma questionnaire for use in healthcare workers. Occup Environ Med 2006 Mar;63(3):173-179 [FREE Full text] [doi: 10.1136/oem.2005.021634] [Medline: 16497858]
22. Josefson A, Färm G, Meding B. Validity of self-reported nickel allergy. Contact Dermatitis 2010 May;62(5):289-293. [doi: 10.1111/j.1600-0536.2010.01702.x] [Medline: 20536476]
23. Hoppin JA, Jaramillo R, Salo P, Sandler DP, London SJ, Zeldin DC. Questionnaire predictors of atopy in a US population sample: findings from the National Health and Nutrition Examination Survey, 2005-2006. Am J Epidemiol 2011 Mar;173(5):544-552 [FREE Full text] [doi: 10.1093/aje/kwq392] [Medline: 21273397]
24. Landkroon AP, de Weerd S, van Vliet-Lachotzki E, Steegers EAP. Validation of an internet questionnaire for risk assessment in preconception care. Public Health Genomics 2010;13(2):89-94. [doi: 10.1159/000228980] [Medline: 19609073]
25. The Netherlands Perinatal Registry. Perinatal care in the Netherlands, 2013. Utrecht: The Netherlands Perinatal Registry; 2014. URL: http://www.perinatreg.nl/uploads/150/153/PRN_jaarboek_2013_09122014.pdf [accessed 2015-03-23] [WebCite Cache ID 6XEwgoVa8]
26. Van Gelder MMHJ, Bretveld RW, Roukema J, Steenhoek M, van Drongelen J, Spaanderman MEA, et al. Rationale and design of the PRegnancy and Infant DEvelopment (PRIDE) Study. Paediatr Perinat Epidemiol 2013 Jan;27(1):34-43 [FREE Full text] [doi: 10.1111/ppe.12023] [Medline: 23215710]
27. Ravelli ACJ, Tromp M, van Huis M, Steegers EAP, Tamminga P, Eskes M, et al. Decreasing perinatal mortality in The Netherlands, 2000-2006: a record linkage study. J Epidemiol Community Health 2009 Sep;63(9):761-765. [doi: 10.1136/jech.2008.080440] [Medline: 19416928]
28. World Health Organization. International Classification of Diseases. 2013. URL: http://www.who.int/classifications/icd/en/ [accessed 2013-10-18] [WebCite Cache ID 6KSW32ney]
29. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [Medline: 2348207]
30. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43(6):551-558. [Medline: 2189948]
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb;1(8476):307-310. [Medline: 2868172]
32. Van Weel C, Schers H, Timmermans A. Health care in the Netherlands. J Am Board Fam Med 2012 Mar;25 Suppl 1:S12-S17 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110212] [Medline: 22403245]
33. Tate AR, Dezateux C, Cole TJ, Davidson L, Millennium Cohort Study Child Health Group. Factors affecting a mother's recall of her baby's birth weight. Int J Epidemiol 2005 Jun;34(3):688-695 [FREE Full text] [doi: 10.1093/ije/dyi029] [Medline: 15737964]
34. Coolman M, de Groot CJM, Jaddoe VW, Hofman A, Raat H, Steegers EAP. Medical record validation of maternally reported history of preeclampsia. J Clin Epidemiol 2010 Aug;63(8):932-937. [doi: 10.1016/j.jclinepi.2009.10.010] [Medline: 20189760]

35. Bat-Erdene U, Metcalfe A, McDonald SW, Tough SC. Validation of Canadian mothers' recall of events in labour and delivery with electronic health records. BMC Pregnancy Childbirth 2013;13 Suppl 1:S3 [FREE Full text] [doi: 10.1186/1471-2393-13-S1-S3] [Medline: 23445768]
36. Van Gelder MMHJ, van Rooij IALM, de Walle HEK, Roeleveld N, Bakker MK. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. Drug Saf 2013 Jan;36(1):43-54. [doi: 10.1007/s40264-012-0004-8] [Medline: 23315295]
37. Wada K, Yatsuya H, Ouyang P, Otsuka R, Mitsuhashi H, Takefuji S, et al. Self-reported medical history was generally accurate among Japanese workplace population. J Clin Epidemiol 2009 Mar;62(3):306-313. [doi: 10.1016/j.jclinepi.2008.04.006] [Medline: 18774692]
38. Bergmann MM, Jacobs EJ, Hoffmann K, Boeing H. Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. Eur J Epidemiol 2004;19(5):411-416. [Medline: 15233312]
39. Smith B, Chu LK, Smith TC, Amoroso PJ, Boyko EJ, Hooper TI, Millennium Cohort Study Team. Challenges of self-reported medical conditions and electronic medical records among members of a large military cohort. BMC Med Res Methodol 2008 Jun;8:37 [FREE Full text] [doi: 10.1186/1471-2288-8-37] [Medline: 18644098]
40. Baumeister H, Kriston L, Bengel J, Härter M. High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. J Clin Epidemiol 2010 May;63(5):558-565. [doi: 10.1016/j.jclinepi.2009.08.009] [Medline: 19959329]
41. Rasmussen BK, Jensen R, Olesen J. Questionnaire versus clinical interview in the diagnosis of headache. Headache 1991 May;31(5):290-295. [Medline: 1860786]
42. Valentinis L, Valent F, Mucchiut M, Barbone F, Bergonzi P, Zanchin G. Migraine in adolescents: validation of a screening questionnaire. Headache 2009 Feb;49(2):202-211. [doi: 10.1111/j.1526-4610.2008.01261.x] [Medline: 19222594]
43. Hagen K, Zwart JA, Aamodt AH, Nilsen KB, Bråthen G, Helde G, et al. The validity of questionnaire-based diagnoses: the third Nord-Trøndelag Health Study 2006-2008. J Headache Pain 2010 Feb;11(1):67-73 [FREE Full text] [doi: 10.1007/s10194-009-0174-7] [Medline: 19946790]
44. Wills AK, Lawlor DA, Muniz-Terrera G, Matthews F, Cooper R, Ghosh AK, FALCon Study Team. Population heterogeneity in trajectories of midlife blood pressure. Epidemiology 2012 Mar;23(2):203-211 [FREE Full text] [doi: 10.1097/EDE.0b013e3182456567] [Medline: 22249241]
45. Allen NB, Siddique J, Wilkins JT, Shay C, Lewis CE, Goff DC, et al. Blood pressure trajectories in early adulthood and subclinical atherosclerosis in middle age. JAMA 2014 Feb;311(5):490-497 [FREE Full text] [doi: 10.1001/jama.2013.285122] [Medline: 24496536]
46. Alonso A, Beunza JJ, Delgado-Rodríguez M, Martínez-González MA. Validation of self reported diagnosis of hypertension in a cohort of university graduates in Spain. BMC Public Health 2005;5:94 [FREE Full text] [doi: 10.1186/1471-2458-5-94] [Medline: 16156889]

Abbreviations

BMI: body mass index
CRF: case report form
DBP: diastolic blood pressure
GP: general practitioner
ICC: intraclass correlation coefficient
pneg: observed proportion of negative agreement
ppos: observed proportion of positive agreement
PRIDE: PRegnancy and Infant DEvelopment
SBP: systolic blood pressure
TOP: termination of pregnancy

Edited by G Eysenbach; submitted 05.09.14; peer-reviewed by S Farag, Y Jayasinghe; comments to author 19.02.15; revised version received 02.04.15; accepted 18.04.15; published 16.06.15.

Please cite as:
van Gelder MMHJ, Schouten NPE, Merkus PJFM, Verhaak CM, Roeleveld N, Roukema J
Using Web-Based Questionnaires and Obstetric Records to Assess General Health Characteristics Among Pregnant Women: A Validation Study
J Med Internet Res 2015;17(6):e149
URL: http://www.jmir.org/2015/6/e149/
doi:10.2196/jmir.3847
PMID:26081990

Original Paper

Comparing Patients' Opinions on the Hospital Discharge Process Collected With a Self-Reported Questionnaire Completed Via the Internet or Through a Telephone Survey: An Ancillary Study of the SENTIPAT Randomized Controlled Trial

Berengere Couturier[1,2,3], RN, MSc Epi; Fabrice Carrat[1,2,3], MD, PhD; Gilles Hejblum[2,3], PhD; SENTIPAT Study Group[1]

[1]Assistance Publique - Hôpitaux de Paris, Unité de Santé Publique, Hôpital Saint Antoine, Paris, France
[2]Sorbonne Universités, UPMC Univ Paris 06, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France
[3]INSERM, U1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France

Corresponding Author:
Berengere Couturier, RN, MSc Epi
Assistance Publique - Hôpitaux de Paris
Unité de Santé Publique
Hôpital Saint Antoine
184 rue du Faubourg Saint Antoine
Paris, 75571
France
Phone: 33 (0)149283228
Fax: 33 (0)149283233
Email: berengere.couturier@aphp.fr

Abstract

Background: Hospital discharge, a critical stage in the hospital-to-home transition of patient care, is a complex process with potential dysfunctions having an impact on patients' health on their return home. No study has yet reported the feasibility and usefulness of an information system that would directly collect and transmit, via the Internet, volunteer patients' opinions on their satisfaction concerning the organization of hospital discharge.

Objective: Our primary objective was to compare patients' opinions on the discharge process collected with 2 different methods: self-questionnaire completed on a dedicated website versus a telephone interview. The secondary goal was to estimate patient satisfaction.

Methods: We created a questionnaire to examine hospital discharge according to 3 dimensions: discharge logistics organization, preplanned posthospital continuity-of-care organization, and patients' impressions at the time of discharge. A satisfaction score (between 0 and 1) for each of those dimensions and an associated total score were calculated. Taking advantage of the randomized SENTIPAT trial that questioned patients recruited at hospital discharge about the evolution of their health after returning home and randomly assigned them to complete a self-questionnaire directly online or during a telephone interview, we conducted an ancillary study comparing satisfaction with the organization of hospital discharge for these 2 patient groups. The questionnaire was proposed to 1141 patients included in the trial who were hospitalized for ≥2 days, among whom 867 eligible patients had access to the Internet at home and were randomized to the Internet or telephone group.

Results: Of the 1141 patients included, 755 (66.17%) completed the questionnaire. The response rates for the Internet (39.1%, 168/430) and telephone groups (87.2%, 381/437) differed significantly (*P*<.001), but their total satisfaction scores did not (*P*=.08) nor did the satisfaction subscores (*P*=.58 for discharge logistics organization, *P*=.12 for preplanned posthospital continuity-of-care organization, and *P*=.35 for patients' impressions at the time of discharge). The total satisfaction score (median 0.83, IQR 0.72-0.92) indicated the patients' high satisfaction.

Conclusions: The direct transmission of personal health data via the Internet requires patients' active participation and those planning surveys in the domain explored in this study should anticipate a lower response rate than that issued from a similar survey conducted by telephone interviews. Nevertheless, collecting patients' opinions on their hospital discharge via the Internet proved operational; study results indicate that conducting such surveys via the Internet yields similar estimates to those obtained via a telephone survey. The results support the establishment of a permanent dedicated website that could also be used to obtain users' opinions on other aspects of their hospital stay and follow-up.

Trial Registration: Clinicaltrials.gov NCT01769261; http://clinicaltrials.gov/ct2/show/NCT01769261 (Archived by WebCite at http://www.webcitation.org/6ZDF5bdQb).

(J Med Internet Res 2015;17(6):e158) doi:10.2196/jmir.4379

KEYWORDS

hospital information systems; patient-centered care; patient discharge; patient satisfaction; quality of health care

Introduction

Hospital discharge constitutes a pivotal step in a hospitalized patient's management between the hospital and the return home. Discharge organization conditions the subsequent continuity of care, at least in part. Hospital discharge is a complex process that requires the participation of different actors (including the patient and his/her entourage) and the use of some documents and tools (eg, checklist and discharge package/brochure, patient records, discharge summaries, nursing discharge notes, medical prescriptions). The complexity of the process can explain the occurrence of organizational dysfunctions during its course that could potentially affect the health of individuals far after the hospitalization [1].

The opinion of health professionals on how discharge is organized has been the topic of several studies [2-6] and some of those studies also obtained patients' opinions [3-6]. Hence, the patient has become a major player and a key partner at the center of the health care system [7]. Collecting his/her opinions on the organization of care seemed to be a relevant way to evaluate the quality of the process, for example, to ascertain its perceived quality as experienced by the patient [8,9]. Some studies evaluated the quality of the hospital discharge process based on patients' opinions [10-15]. These studies involved various specific tools, administered at different times in regards to the hospital discharge moment, and based on various modes of administration: telephone, face-to-face interview, or self-administered paper questionnaires. None of those investigations sought the participation of patients via the Internet, even though its use by patients has increased markedly [16].

In this context, we undertook a study aimed at estimating the contribution of a system based on the direct transmission via a dedicated website of volunteer patients' self-reported experiences on their own hospital discharge process. In particular, we wanted to explore the feasibility of such an information collection method by examining patients' response rates and determining whether the information collected via the Internet was of similar quality as that obtained during a telephone interview, which is more difficult to conduct and more expensive to put in place.

First, we created a questionnaire concerning the hospital discharge process according to 3 dimensions: discharge logistics organization, preplanned posthospital continuity-of-care organization, and the patients' impressions during discharge. Then, we collected patients' responses to the questionnaire according to 2 different methods requiring the patient's more- or less-active participation in reporting his/her opinions: patient's direct transmission of information on a dedicated website or a classical telephone interview. We took advantage of the multicenter, randomized SENTIPAT trial (described in Methods) that had already randomized patients at discharge with Internet access at home to transmit personal information via the Internet or by telephone and a third group without Internet at home who were included to determine the representativeness of the randomized sample. For our ancillary study focused on the organization of hospital discharge, the primary objective was to compare the satisfaction of internet and telephone groups, hypothesizing no significant difference according to the data-collection method. The secondary goal was to analyze patients' opinions on the different components of discharge process.

Methods

Overview

This investigation was conceived as an ancillary study of the multicenter, randomized SENTIPAT trial [17]. We took advantage of the trial to examine patients' opinions on the organization of their hospital discharge. SENTIPAT participants were also asked to describe their experience with this process.

General Description of the SENTIPAT Trial

This multicenter (5 adult acute care units in a Parisian teaching hospital participated voluntarily: digestive and general surgery, gastroenterology, hepatology, infectious diseases, and internal medicine), randomized trial focused on the evolution of patients' health on returning home posthospitalization (follow-up duration: 6 weeks). The principal objective was to determine whether the information directly transmitted by the volunteer patients via a dedicated website was comparable with that obtained during a telephone interview. It was a noninferiority trial (the main judgment criterion was the percentage of patients reporting at least one clinically significant adverse event occurring during the 42 days after hospital discharge).

Consecutive patients with Internet access at home were eligible for inclusion. They were enrolled the day of hospital discharge and randomized into 2 groups (stratified by department): Internet or telephone follow-up. Patients not eligible (ie, with the same characteristics as those randomized but without Internet access at home) were also included at a ratio of 1:4 of noneligible to eligible patients.

Lastly, 2550 patients (510 from each unit) were initially planned. Between February 25, 2013 and September 8, 2014, we enrolled 2090 patients who were not cognitively impaired and did not have a behavioral disorder, who spoke and wrote French, and were returning home after an acute care hospitalization, regardless of the type of stay—standard hospitalization (scheduled or not) on weekdays only (maximum Monday to Friday or any combination thereof) or outpatient hospitalization (1 day)—and not opposed to participating in the trial.

Characteristics of the Ancillary Study Focusing on the Discharge Process

Patients

This study concerned the 1141 patients included whose hospitalization lasted 2 or more days. The results of patients whose hospitalization lasted only 1 day (n=949) are reported in Multimedia Appendix 1; the organization of the discharge process after these very short stays was logically analyzed independently of those of longer duration.

Questionnaire Structure

Several tools have been developed to collect patients' opinions on their hospital-to-home transition, including The Care Transition Measure [11], The Patient Continuity of Care Questionnaire [13], the Brief PREPARED instrument [12], or the Readiness for Hospital Discharge Scale [15]. None of these explore the 3 hospital discharge-related dimensions of interest to us. Therefore, we constructed a specific questionnaire, based on French national recommendations [18-20], and an international literature review (BC, FC, and GH, unpublished data, 2015).

The questionnaire explored 3 hospital discharge dimensions (henceforth referred to as 3 items) addressed in 17 questions (Table 1): discharge logistics organization (Q1, Q2, Q3, Q4, Q5, and Q11C-E), henceforth referred to as item 1; preplanned posthospital continuity-of-care organization (Q6, Q7, Q8, Q9, and Q10), henceforth referred to as item 2; and the patients' impressions during discharge (Q11A, B, F, and G), henceforth referred to as item 3. Several questions (Q1, Q6 and Q8) specifically attempt to document specific aspects of the hospital-discharge process; the corresponding responses and a general discussion of the questionnaire are given in Multimedia Appendix 2.

Questionnaire Administration

For the telephone and noneligible patients, the hospital discharge questionnaire was administrated during a telephone interview with a clinical research technician 7 days after discharge (the appointment was scheduled the day of discharge), with a maximum of 3 attempts to contact them. For the Internet group, the same questionnaire was available on the dedicated website on the day of discharge (D0) and was completed directly online by the patient, who had been given oral and written instructions (information sheet) to connect for the first time 7 days postdischarge. "Reminders" were sent once weekly for 6 weeks after discharge to potential responders (of the Internet group) who had not completed the discharge questionnaire yet.

Statistical Analyses

A patient satisfaction score concerning the hospital discharge process was calculated for each of the 3 items (see scores accorded in Table 1). Each subscore was calculated with a normalized sum of all the item questions (range 0-1), each question within a given item having the same weight, and the total score was computed (mean of the 3 subscores, range 0-1). Questions not concerning the patient (eg, Q11D did not concern patients who used public transportation) were not included in the calculations.

The global distribution of the responses for each discharge questionnaire question was assessed. Between-group qualitative variable (eg, sex) differences were compared with Fisher's exact test and quantitative variables (eg, scores, age, hospitalization duration, and level of education that was considered as a 4-level ordinal variable) were compared with either the Mann-Whitney-Wilcoxon test or, for matched-paired data, Wilcoxon's signed rank test. A *P* value ≤.05 defined significance of usual comparisons. However, the significance threshold was lower when a Bonferroni correction was applied for multiple comparisons, as indicated where appropriate. Missing data were taken into account as follows: nonresponding patients and incomplete questionnaires were excluded from the analyses. Also, for questions not concerning the patient, the response "nonapplicable" was used and they were not included in the analyses. All statistical computations used the R program.

Table 1. Discharge questionnaire and satisfaction scoring.

Item and question	Response choices	Score
Discharge logistics organization		
Q1. Who decided when you would be discharged from the hospital?[a]	Staff physician of the unit in which you were hospitalized	
	Yourself	
	Your entourage	
	Your primary care physician	
	Other	
Q2. Were you informed by a doctor or nurse of the modalities of your discharge (date, time, transportation: taxi, ambulance...)?	Yes	1
	No	0
Q3. Were you consulted for the choice of discharge date and time?	Yes	1
	No	0
Q4. Were the discharge date and time compatible with your return home and/or your entourage?	Yes	1
	No	0
Q5. Was the scheduled discharge time respected?	Yes	1
	No	0
Q11C. At discharge, what did you think of the time needed to obtain your medical and administrative documents?	Reasonable	1
	Too long	0
Q11D. At discharge, what did you think of the time needed for your transportation to arrive?	Reasonable	1
	Too long	0
	Not concerned	NA
Q11E. At discharge, did you have any difficulties dealing with the administrative discharge formalities?	Yes	0
	No	1
	Not concerned	NA
Preplanned posthospital continuity-of-care organization		
Q6. At discharge, what documents were you given concerning your subsequent care? Check all that apply[a]	Prescription(s)	
	Discharge summary	
	Letter for primary care physician	
	Nursing discharge notes	
	Information booklet(s)	
	Appointment for a next hospitalization	
	Appointment for your next consultation	
	Appointment for your complementary test(s)	
	Other, specify	
	None	
Q7. What did you think about the information provided by the medical or nursing staff when you received those discharge documents?	Highly satisfied	1
	Satisfied	0.75
	Poorly satisfied	0.25
	Not at all satisfied	0

Item and question	Response choices	Score
	No information given	0
	Not concerned	NA
Q8. Did you meet with a social worker during your hospitalization to discuss the organization of your return home?[a]	Yes	
	No	
Q9. Was your primary care physician informed of your hospitalization?	Yes	1
	No	0
	I don’t know	0
	I don’t have a primary care physician	NA
Q10. Did you have the phone number of the unit in which you were hospitalized (should you need it)?	Yes	1
	No	0
Patient’s impressions of the hospital discharge process		
Q11A. At discharge, what did you think about its organization?	Well planned	1
	A sense of haste, upheaval	0
Q11B. At discharge, what did you think about returning home?	Relieved	1
	Anxious	0
Q11F. At discharge, what did you think about the information provided?	Sufficient	1
	Insufficient	0
Q11G. At discharge, what did you think about the health care team’s availability and listening to you?	Sufficient	1
	Insufficient	0

[a] Question intended to document the situation but not to be a score component.

Results

Patient Characteristics and Response Rates

A total of 755 (66.17%) completed discharge questionnaires were collected from the 1141 patients included after a hospital stay of 2 or more days (Figure 1). The relative contribution of each unit ranged from 13.41% (153/1141) to 35.58% (406/1141) and their response rates did not significantly differ from one unit to another (*P*=.08) and ranged from 60.3% (132/219) to 70.0% (284/406). Patients’ median age was 55 (IQR 39-66) years and 591 of 1141 (51.80%) were women. Hospitalization lasted a median 6 (IQR 3-10) days (median 7, IQR 4-11 and median 3, IQR 2-3 days for standard and weekday-only hospitalizations, respectively) (Table 2). Responders were significantly older than nonresponders (*P*<.001) for comparable sex distributions, level of education, and hospitalization durations. Internet, telephone, and noneligible group patients completed the questionnaire within median 6 (IQR 3-16), median 7 (IQR 7-9), and median 7 (IQR 7-8) days postdischarge, respectively, with respective Internet and telephone response rates of 39.1% (168/430) and 87.2% (381/437, *P*<.001). Noneligible patients were significantly older than telephone patients were (*P*<.001) and their response rate was significantly lower (75.2%, 206/274 vs 87.2%, 381/437, *P*<.001).

Table 2. Characteristics of patients with a hospital stay of 2 or more days.

Characteristic	Total	Responders	Nonresponders	*P*
Group, n (%)				
All	1141	755 (66.17)	386 (33.83)	
Internet	430	168 (39.1)	262 (60.9)	<.001[a]
Telephone	437	381 (87.2)	56 (12.8)	
Noneligible	274	206 (75.2)	68 (24.8)	<.001[b]
Sex (male/female), n				
All	550/591	365/390	185/201	.90
Internet	204/226	79/89	125/137	
Telephone	228/209	198/183	30/26	
Noneligible	118/156	88/118	30/38	
Age (years), median (IQR)				
All	55 (39-66)	56 (41-67)	51 (34-64)	<.001
Internet	51 (36-63)	55 (38-63)	48 (34-62)	
Telephone	52 (34-64)	52 (36-64)	47 (30-66)	
Noneligible	65 (52-74)	66 (55-75)	62 (49-72)	
Length of hospital stay (days), median (IQR)				
All	6 (3-10)	6 (3-10)	6 (3-9)	.92
Internet	6 (3-9)	5.5 (3-9)	6 (3-9)	
Telephone	6 (3-10)	6 (3-10)	8 (3-10)	
Noneligible	7 (3-11)	7 (3-12)	7 (3-9)	
Level of education, [c] **n (%)**				
All				.95
Level 1	223 (19.54)	152 (20.1)	71 (18.4)	
Level 2	365 (31.99)	233 (30.9)	132 (34.2)	
Level 3	150 (13.15)	103 (13.6)	47 (12.2)	
Level 4	401 (35.14)	267 (35.4)	134 (34.7)	
Do not wish to answer	2 (0.18)	0 (0.0)	2 (0.5)	
Internet				
Level 1	48 (11.2)	17 (10.1)	31 (11.8)	
Level 2	138 (32.1)	51 (30.4)	87 (33.2)	
Level 3	61 (14.2)	25 (14.9)	36 (13.7)	
Level 4	182 (42.3)	75 (44.6)	107 (40.8)	
Do not wish to answer	1 (0.2)	0 (0.0)	1 (0.4)	
Telephone				
Level 1	47 (10.8)	38 (10.0)	9 (16)	
Level 2	139 (31.8)	121 (31.8)	18 (32)	
Level 3	68 (15.6)	62 (16.3)	6 (11)	
Level 4	182 (41.6)	160 (42.0)	22 (39)	
Do not wish to answer	1 (0.2)	0 (0.0)	1 (2)	
Noneligible				
Level 1	128 (46.7)	97 (47.1)	31 (46)	

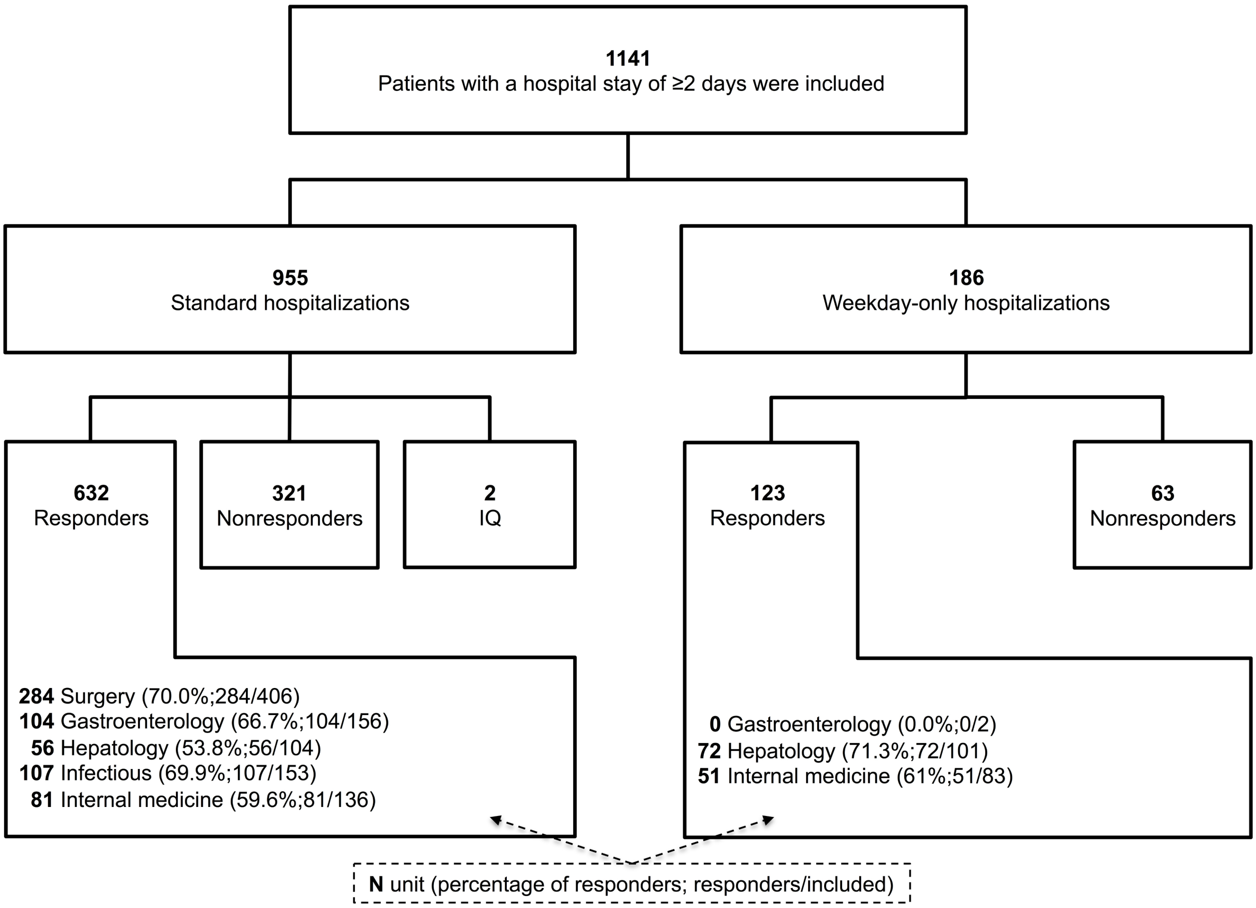
Characteristic	Total	Responders	Nonresponders	*P*
Level 2	88 (32.1)	61 (29.6)	27 (40)	
Level 3	21 (7.7)	16 (7.8)	5 (7)	
Level 4	37 (13.5)	32 (15.5)	5 (7)	
Do not wish to answer	0 (0.0)	0 (0.0)	0 (0)	

[a] Internet vs telephone.

[b] Noneligible vs telephone.

[c] The levels of education were coded as follows: level 1, at most junior high school; level 2, high school; level 3, college; level 4, bachelor's degree or above.

Figure 1. Flowchart of patients with a hospital stay of ≥2 days included in the SENTIPAT trial who responded or not to the discharge questionnaire according to the type of hospitalization and the recruitment unit. IQ: incomplete questionnaires; surgery: general and digestive surgery; infectious: infectious and tropical diseases.



Comparisons of Internet and Telephone Groups' Satisfaction Scores

Box plot comparisons between the Internet and telephone groups for each of the 3 items or their total scores (Figure 2) revealed no important differences as confirmed by the corresponding statistical comparison results; all were associated with nonsignificant *P* values (Table 3). Moreover, the telephone and noneligible groups did not differ significantly for the total score or its 3 subscores.

Table 3. Distribution of satisfaction scores (percentiles) of the 755 responders with a hospital stay of 2 or more days according to group.

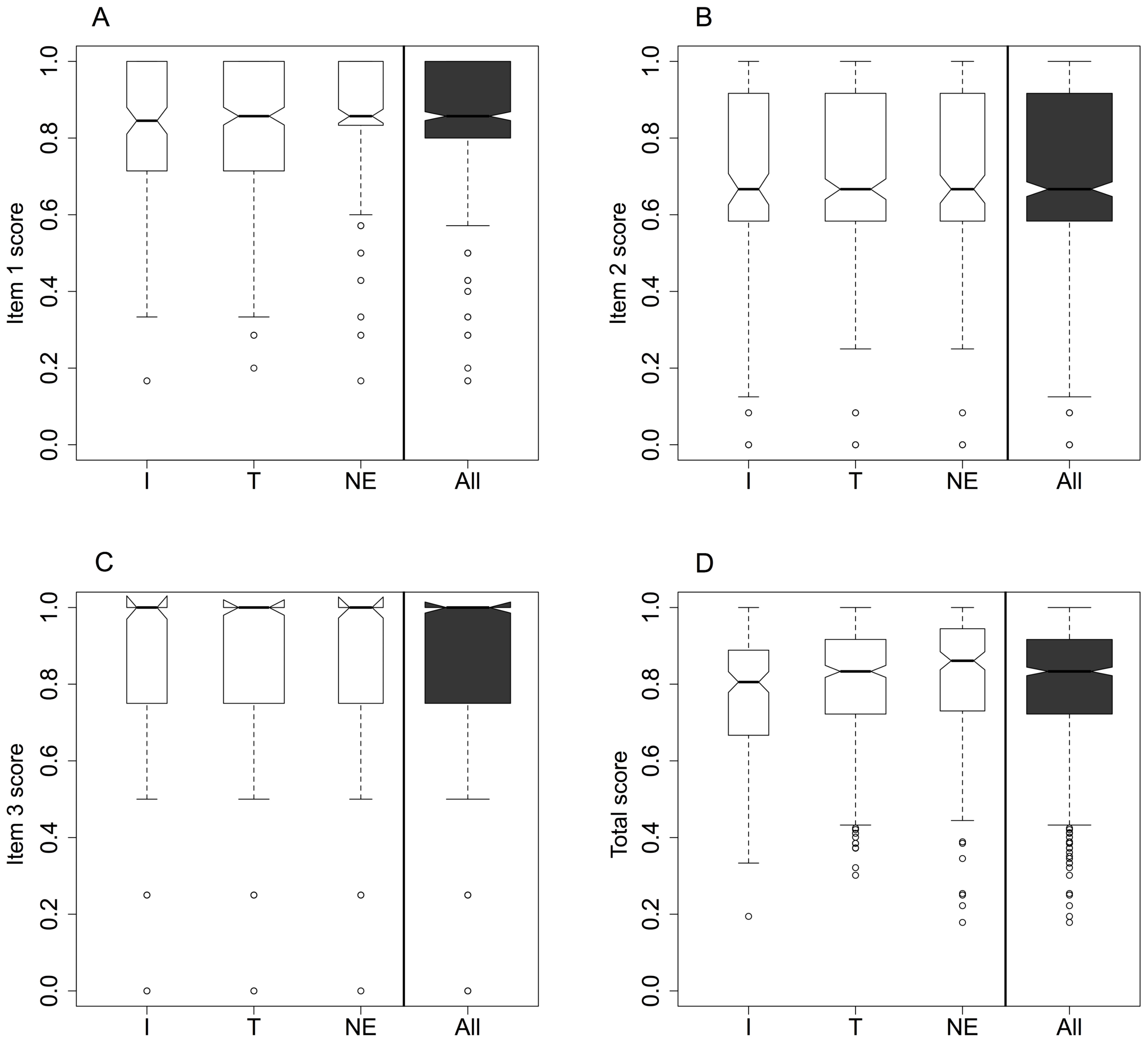
Score	Percentile							P [a]
	5%	10%	25%	50%	75%	90%	95%	
Total								
All	0.47	0.56	0.72	0.83	0.92	0.97	0.97	
Internet	0.46	0.51	0.67	0.81	0.89	0.97	0.97	.08[b]
Telephone	0.48	0.56	0.72	0.83	0.92	0.97	1	
Noneligible	0.50	0.61	0.73	0.86	0.94	0.97	0.97	.03[c]
Discharge logistics organization								
All	0.50	0.57	0.80	0.86	1	1	1	
Internet	0.45	0.57	0.71	0.85	1	1	1	.58[b]
Telephone	0.50	0.57	0.71	0.86	1	1	1	
Noneligible	0.50	0.60	0.83	0.86	1	1	1	.06[c]
Preplanned posthospital continuity-of-care organization								
All	0.33	0.33	0.58	0.67	0.92	1	1	
Internet	0.25	0.33	0.58	0.67	0.92	1	1	.12[b]
Telephone	0.33	0.33	0.58	0.67	0.92	1	1	
Noneligible	0.33	0.33	0.58	0.67	0.92	1	1	.45[c]
Patient’s impressions of the hospital discharge process								
All	0.25	0.50	0.75	1	1	1	1	
Internet	0.25	0.43	0.75	1	1	1	1	.35[b]
Telephone	0.25	0.50	0.75	1	1	1	1	
Noneligible	0.50	0.50	0.75	1	1	1	1	.09[c]

[a] Mann-Whitney-Wilcoxon tests with the corresponding Bonferroni correction for 2 comparisons: the telephone group was compared with the Internet and noneligible groups; all comparisons yielded nonsignificant *P* values.

[b] Internet vs telephone.

[c] Noneligible vs telephone.

Figure 2. Box plots of score distributions according to Internet (I), telephone (T), or noneligible (NE) group. Item 1: discharge logistics organization; item 2: preplanned posthospital continuity-of-care organization; item 3: patient's impressions of the hospital discharge process. The bold horizontal line is the median, the bottom and top borders of the boxes are 25th and 75th percentiles, respectively; the T-bar below and above the boxes represent 2.5th and 97.5th percentiles, respectively; the small white circles are outliers of the latter limits.



Satisfaction Scores for All Responders

Overview

The total satisfaction score was median 0.83 (IQR 0.72-0.92), with respective items 1-3 subscores of median 0.86 (IQR 0.8-1), median 0.67 (IQR 0.58-0.92), and median 1 (IQR 0.75-1). The box plots (Figure 2) for the entire population of responders differed in shape from one item to another. The Wilcoxon signed rank test confirmed that item 2 was significantly less well-rated than item 1 (*P*<.001), which was less well-rated than item 3 (*P*<.001) (Figure 2). The main characteristics associated with each of the 3 items are presented subsequently.

Item 1: Discharge Logistics Organization

Figure A3-1 in Multimedia Appendix 3 shows that 87.9% (664/755) of patients were informed of the modalities (eg, date and time, transportation) of their discharge; 53.1% (401/755) declared not having been consulted for the discharge date and time, and 92.6% (699/755) considered that those choices did not pose a problem. For 90.7% (685/755), the time was respected, 91.7% (676/737 excluding 18/755 patients not concerned) deemed the waiting time for medical and administrative discharge documents satisfactory, and 90.3% (650/720 excluding 35/755 patients not concerned) did not encounter difficulties completing administrative discharge procedures.

Item 2: Preplanned Posthospital Continuity-of-Care Organization

The explanations provided by the medical and/or nursing staff to accompany the document delivered at discharge (36/755 patients not concerned) were considered poorly satisfactory or totally unsatisfactory for 11.1% (80/719) and 16.0% (115/719) declared having received no explanations (Figure A3-2, Multimedia Appendix 3). According to 23.4% (177/755) of patients, their primary care physicians were not informed of their hospitalizations and 16.6% (125/755) did not know if they had been informed or not. In addition, 89.4% (675/755) had the telephone numbers of the unit in which they were hospitalized, if needed.

Item 3: Patients' Impressions During the Discharge Process

Figure A3-3 (see Multimedia Appendix 3) describes patients' impressions of the discharge process. Notably, 85.0% (642/755) thought that their discharge had been well planned. Those with the opposing opinion were significantly younger (*P*<.001). In addition, patients anxious about their return home (13.8%, 104/755) were significantly younger than those relieved at the idea of going home (*P*<.001). Moreover, 20.3% (153/755) of the patients thought they lacked information at discharge. Finally, 9.5% (72/755) of the patients judged hospital caregivers insufficiently available and that they listened insufficiently to the patient.

Discussion

Main Results and Comparison With Previous Findings

This study was designed to investigate patients' opinions of the hospital discharge process with a questionnaire administered either by self-reporting directly online or by a traditional telephone interview. Response rates to our questionnaire showed that patients are willing to assume an active partnership role—87.2% for the telephone group (with Internet access at home) and 75.2% for noneligible patients (without Internet access at home)—values close to the average response rate for 13 studies that included a telephone inquiry to obtain patients' impressions of hospital care delivered (ie, mean 70%, range 24%-91%) [21]. However, such interviews are more cumbersome and expensive to organize (eg, interviewer, scheduling of calls) and implement as routine practice over the long term. Responders were significantly older than nonresponders, with a similar observed pattern in the Internet and the telephone groups, and this might reflect individuals' greater availability or interest in health care questions, which increases globally with age.

Given the rising availability and utilization of the Internet in all populations, this easy, low-cost approach as a means of collecting patients' opinions might be an attractive alternative to telephone calls. In our study, the Internet patients' response rate was lower (39.1%, 168/430) than the telephone patients' response rate (87.2%, 381/437). In many respects, it is not surprising. For example, ignoring an invitation to actively enter personal data on a website is much easier than ignoring a live person who reaches another by phone. Nevertheless, the 39% response rate observed in our study exceeded that usually reached with Internet surveys, according to a meta-analysis of 39 studies (median 27%, range 14.5%-51%) [22]. Nonetheless, the information reported by online patients did not differ significantly from those collected by phone. Obtaining patients' opinions on the discharge process is in-line with current initiatives to achieve a patient-centered health care system [23-25]. Our observations suggest that long-term implementation of an information system, similar to that developed for this study, would enable patients to directly transmit their hospital discharge experiences. The scores observed for the 3 dimensions of the discharge process explored herein indicate an overall positive patient perception: discharge logistics organization (median 0.86, IQR 0.8-1), preplanned posthospital continuity-of-care organization (median 0.67, IQR 0.58-0.92), and patients' impressions of the process (median 1, IQR 0.75-1). Discharge logistics organization, in particular, was judged globally satisfactory even though approximately half of patients were not involved in the scheduling of their discharge date and time.

Pertinently, our results also identify several difficulties, notably hospital transmission of information to primary care physicians and the patient, and thereby also indicate how to potentially improve performance. Only 20.3% (153/755) of patients declared having left the hospital with a discharge summary and/or letter for their primary care physician. These observations agree with those previously reported by authors investigating discharge summary availability at the time of discharge for health care professionals responsible for posthospitalization continuity of care [26-33]. For example, in their review, Kripalani et al [28] indicated that only a median 14.5% (range 9%-20%) and median 52% (range 51%-77%) of primary care physicians had received discharge summaries 1 and 4 weeks after discharge, respectively.

In addition, more than a quarter of patients deemed the medical and/or nursing explanations of their discharge documents as poor or unsatisfactory, or had received none. Moreover, one-fifth of patients reported a lack of information at discharge. This absence of information and/or delivery of information not corresponding to patient expectations was also noted previously [13,27,34,35]. Other than the strict enumeration of the hospital discharge instructions provided to the patient by health professionals at discharge, the patient's understanding of them is not always optimal [36,37], notably concerning medical treatments to be pursued [38-41], and can be underestimated by health professionals [42,43]. The findings of Horwitz et al [37] are particularly interesting because despite the demonstration of a gap between the information given to the patients and their understanding of it, the patients "were uniformly positive in their assessment of discharge care" as in our study. In a 2014 systematic revue [44] of 36 studies targeting patients' opinions on quality of care, only 2 addressed the quality of the discharge process [45,46] and they reported globally positive impressions. However, 2 other studies [10,14] examined the association between patient satisfaction with the discharge process and the hospital readmission rate within 30 days, an important health outcome measure, and found it to be significant suggesting this impression reflects, at least in part, the quality of hospital care. Nevertheless, the associated performance differences were relatively modest, thereby suggesting that improving patient satisfaction with discharge organization would also have a minor impact on health in terms of solid outcome measures.

Limitations

Our study has several limitations. Inherent to studies requiring an active participation, responders constitute an intrinsic biased selection sample of patients. An example of a demographic status significantly associated with the responder status is age. Moreover, differences are introduced by the questionnaire administration mode (Internet vs telephone). The response rate observed in the Internet group is somewhat disappointing, but another study on patient satisfaction also reported similar response rates: 34% and 78% in the Internet and telephone groups, respectively [47]. Nevertheless, even if a similar response rate had been observed in the telephone and Internet groups, this would not exclude different selection biases from one group to the other (eg, inherent to the mode of administration). The major result of the study is that despite the biases of this study (potential or not), the estimates issued from the 2 groups are very similar; therefore, an Internet-based survey in the domain investigated should be considered as a useful alternative to a "reference" telephone survey. Nonetheless, collecting patients' opinions via the Internet, as done in this study, has numerous advantages. First, unlike the telephone interview that inserts a third person and a potential information bias (survey subjectivity), resorting to the Internet allows self-administration of the questionnaire, triggering the patient's active participation. Finally, this method of data collection is less costly than managing a telephone cohort and yields comparable information. However, the similar scores in the telephone and Internet mode of administration observed in this study are based on a particular newly developed questionnaire deployed in a given patient population; therefore, this limits the generalizability (external validity) of the results.

Conclusions

The results of this study advocate for establishing a permanent information system that would enable volunteering patients to express their opinions concerning hospital discharge. Such an information system could also be used for other management issues related to health care organization. Those planning to design similar surveys via the Internet should anticipate a response rate comparable to that observed in the present study. Nevertheless, the concept of sentinel patient explored in this study could constitute, in the future, an essential tool in a patient-centered approach to the organization of care.

Acknowledgments

The SENTIPAT study group Scientific Committee: Fabrice Carrat, Bérengère Couturier, Gilles Hejblum, Morgane Le Bail, Alain-Jacques Valleron, and the subsequently mentioned heads and physicians within the poles and units in which patients were recruited.

Personnel of the study group within the poles and units concerned: heads: Marc Beaussier, Jean-Paul Cabane, Olivier Chazouillères, Jacques Cosnes, Jean-Claude Dussaule, Pierre-Marie Girard, Emmanuel Tiret, and Dominique Pateron; additional corresponding physicians: Laurent Beaugerie, Laurence Fardet, Laurent Fonquernie, François Paye, and Laure Surgers; and nursing supervisors: Françoise Cuiller, Catherine Esnouf, Hélène Haure, Valérie Garnier, Josselin Mehal-Birba, Nelly Sallée, and Sylvie Wagener.

The authors are indebted to the excellent technical team of the study, including the clinical research technicians: Élodie Belladame, Azéline Chevance, Magali Girard and Laurence Nicole, who included patients, collected baseline data, and interviewed the patients followed-up by telephone; and the software staff, especially Pauline Raballand, but also Frédéric Chau and Frédéric Fotré, who created and maintained the trial's dedicated website.

We thank all the medical and nursing and administrative staff of the General and Digestive Surgery (including Ambulatory Surgery), Gastroenterology, Hepatology, Infectious Diseases, and Internal Medicine departments of Hôpital Saint-Antoine; and we thank all patients who participated in the study.

The Assistance Publique-Hôpitaux de Paris (Département de la Recherche Clinique et du Développement) was the trial sponsor. The SENTIPAT study was funded by grant AOM09213 K081204 from Programme Hospitalier de Recherche Clinique 2009 (Ministère de la Santé). This work was supported by a grant from Assistance Publique-Hôpitaux de Paris, program "Recherche Infirmière." The sponsor and the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

Study conception and design: BC, FC, and GH; data acquisition: BC and GH; analysis and interpretation of data: BC, FC, and GH; first draft of the paper: BC and GH; all authors approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analysis of discharge-questionnaire responses of patients with 1-day hospital stays.

[PDF File (Adobe PDF File), 2MB - jmir_v17i6e158_app1.pdf]

Multimedia Appendix 2

Additional information about the discharge questionnaire.

[PDF File (Adobe PDF File), 68KB - jmir_v17i6e158_app2.pdf]

Multimedia Appendix 3

Other figures supplementing the main manuscript.

[PDF File (Adobe PDF File), 455KB - jmir_v17i6e158_app3.pdf]

Multimedia Appendix 4

CONSORT-EHEALTH checklist V1.6.1 [48].

[PDF File (Adobe PDF File), 878KB - jmir_v17i6e158_app4.pdf]

References

1. Neale G, Woloshynowych M, Vincent C. Exploring the causes of adverse events in NHS hospital practice. J R Soc Med 2001 Jul;94(7):322-330 [FREE Full text] [Medline: 11418700]
2. Graumlich JF, Grimmer-Somers K, Aldag JC. Discharge planning scale: community physicians' perspective. J Hosp Med 2008 Dec;3(6):455-464. [doi: 10.1002/jhm.371] [Medline: 19084895]
3. Graumlich JF, Novotny NL, Nace GS, Aldag JC. Patient and physician perceptions after software-assisted hospital discharge: cluster randomized trial. J Hosp Med 2009 Jul;4(6):356-363. [doi: 10.1002/jhm.565] [Medline: 19621342]
4. Grimmer K, Moss J. The development, validity and application of a new instrument to assess the quality of discharge planning activities from the community perspective. Int J Qual Health Care 2001 Apr;13(2):109-116 [FREE Full text] [Medline: 11430660]
5. Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010 May;48(5):482-486. [doi: 10.1097/MLR.0b013e3181d5feae] [Medline: 20393364]
6. Weiss ME, Costa LL, Yakusheva O, Bobay KL. Validation of patient and nurse short forms of the Readiness for Hospital Discharge Scale and their relationship to return to the hospital. Health Serv Res 2014 Feb;49(1):304-317 [FREE Full text] [doi: 10.1111/1475-6773.12092] [Medline: 23855675]
7. Laine C, Davidoff F. Patient-centered medicine. A professional evolution. JAMA 1996 Jan 10;275(2):152-156. [Medline: 8531314]
8. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. N Engl J Med 2008 Oct 30;359(18):1921-1931. [doi: 10.1056/NEJMsa0804116] [Medline: 18971493]
9. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]
10. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Manag Care 2011 Jan;17(1):41-48 [FREE Full text] [Medline: 21348567]
11. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. Med Care 2005 Mar;43(3):246-255. [Medline: 15725981]
12. Graumlich JF, Novotny NL, Aldag JC. Brief scale measuring patient preparedness for hospital discharge to home: Psychometric properties. J Hosp Med 2008 Dec;3(6):446-454. [doi: 10.1002/jhm.316] [Medline: 19084894]
13. Hadjistavropoulos H, Biem H, Sharpe D, Bourgault-Fagnou M, Janzen J. Patient perceptions of hospital discharge: reliability and validity of a Patient Continuity of Care Questionnaire. Int J Qual Health Care 2008 Oct;20(5):314-323 [FREE Full text] [doi: 10.1093/intqhc/mzn030] [Medline: 18635587]
14. Jha AK, Orav EJ, Epstein AM. Public reporting of discharge planning and rates of readmissions. N Engl J Med 2009 Dec 31;361(27):2637-2645. [doi: 10.1056/NEJMsa0904859] [Medline: 20042755]
15. Weiss ME, Piacentine LB. Psychometric properties of the Readiness for Hospital Discharge Scale. J Nurs Meas 2006;14(3):163-180. [Medline: 17278337]
16. Ohno-Machado L. Focusing on the patient: mHealth, social media, electronic health records, and decision support systems. J Am Med Inform Assoc 2014 Dec;21(6):953. [doi: 10.1136/amiajnl-2014-003341] [Medline: 25301806]
17. ClinicalTrials.gov. Sentinel Patients: Value of an Information System Collecting Patient's Own Report on His Healthcare Management (SENTIPAT) URL: https://clinicaltrials.gov/show/NCT01769261 [accessed 2015-01-16] [WebCite Cache ID 6VcoY4d9S]
18. Agence Nationale d'Accréditation et d'Evaluation en Santé. Préparation de la sortie du patient hospitalisé. 2001. URL: http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-08/preparation_de_la_sortie_du_patient_hospitalise_guide_2001.pdf [accessed 2015-01-16] [WebCite Cache ID 6VcouS5ZU]

19. Haute Autorité de Santé. Rapport de l'expérimentation nationale. Audit clinique ciblé appliqué à la préparation de la sortie du patient hospitalisé. 2006. URL: http://www.has-sante.fr/portail/upload/docs/application/pdf/psp_-_rapport_final.pdf [accessed 2015-01-16] [WebCite Cache ID 6Vcozukl4]
20. Haute Autorité de Santé. Manuel de certification des établissements de santé V2010. 2009. URL: http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-12/20081217_manuel_v2010_nouvelle_maquette.pdf [accessed 2015-01-16] [WebCite Cache ID 6Vcp95RvA]
21. Castle NG, Brown J, Hepner KA, Hays RD. Review of the literature on survey instruments used to collect data on hospital patients' perceptions of care. Health Serv Res 2005 Dec;40(6 Pt 2):1996-2017 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00475.x] [Medline: 16316435]
22. Shih TH, Fan X. Comparing response rates from Web and mail Surveys: a meta-analysis. Field Methods 2008 Mar 03;20(3):249-271. [doi: 10.1177/1525822X08317085]
23. Davis K, Schoenbaum SC, Audet AM. A 2020 vision of patient-centered primary care. J Gen Intern Med 2005 Oct;20(10):953-957 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0178.x] [Medline: 16191145]
24. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood) 2010 Aug;29(8):1489-1495 [FREE Full text] [doi: 10.1377/hlthaff.2009.0888] [Medline: 20679652]
25. Stewart M. Towards a global definition of patient centred care. BMJ 2001 Feb 24;322(7284):444-445 [FREE Full text] [Medline: 11222407]
26. Bell CM, Schnipper JL, Auerbach AD, Kaboli PJ, Wetterneck TB, Gonzales DV, et al. Association of communication between hospital-based physicians and primary care providers with patient outcomes. J Gen Intern Med 2009 Mar;24(3):381-386 [FREE Full text] [doi: 10.1007/s11606-008-0882-8] [Medline: 19101774]
27. Kripalani S, Jackson AT, Schnipper JL, Coleman EA. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. J Hosp Med 2007 Sep;2(5):314-323. [doi: 10.1002/jhm.228] [Medline: 17935242]
28. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
29. Li JY, Yong TY, Hakendorf P, Ben-Tovim D, Thompson CH. Timeliness in discharge summary dissemination is associated with patients' clinical outcomes. J Eval Clin Pract 2013 Feb;19(1):76-79. [doi: 10.1111/j.1365-2753.2011.01772.x] [Medline: 21999337]
30. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002 Mar;17(3):186-192 [FREE Full text] [Medline: 11929504]
31. van Walraven C, Seth R, Laupacis A. Dissemination of discharge summaries: not reaching follow-up physicians. Can Fam Physician 2002 Apr;48:737-742 [FREE Full text] [Medline: 12046369]
32. van Walraven C, Taljaard M, Bell CM, Etchells E, Stiell IG, Zarnke K, et al. A prospective cohort study found that provider and information continuity was low after patient discharge from hospital. J Clin Epidemiol 2010 Sep;63(9):1000-1010. [doi: 10.1016/j.jclinepi.2010.01.023] [Medline: 20656194]
33. van Walraven C, Taljaard M, Etchells E, Bell CM, Stiell IG, Zarnke K, et al. The independent association of provider and information continuity on outcomes after hospital discharge: implications for hospitalists. J Hosp Med 2010 Sep;5(7):398-405. [doi: 10.1002/jhm.716] [Medline: 20845438]
34. Charles C, Gauld M, Chambers L, O'Brien B, Haynes RB, Labelle R. How was your hospital stay? Patients' reports about their care in Canadian hospitals. CMAJ 1994 Jun 1;150(11):1813-1822 [FREE Full text] [Medline: 8199958]
35. Soler RS, Canal DJ, Noguer CB, Poch CG, Motge NB, Gil MDG. Continuity of care and monitoring pain after discharge: patient perspective. J Adv Nurs 2010 Jan;66(1):40-48. [doi: 10.1111/j.1365-2648.2009.05136.x] [Medline: 20423435]
36. Albrecht JS, Gruber-Baldini AL, Hirshon JM, Brown CH, Goldberg R, Rosenberg JH, et al. Hospital discharge instructions: comprehension and compliance among older adults. J Gen Intern Med 2014 Nov;29(11):1491-1498. [doi: 10.1007/s11606-014-2956-0] [Medline: 25015430]
37. Horwitz LI, Moriarty JP, Chen C, Fogerty RL, Brewster UC, Kanade S, et al. Quality of discharge practices and patient understanding at an academic medical center. JAMA Intern Med 2013 Oct 14;173(18):1715-1722 [FREE Full text] [doi: 10.1001/jamainternmed.2013.9318] [Medline: 23958851]
38. Kerzman H, Baron-Epel O, Toren O. What do discharged patients know about their medication? Patient Educ Couns 2005 Mar;56(3):276-282. [doi: 10.1016/j.pec.2004.02.019] [Medline: 15721969]
39. Makaryus AN, Friedman EA. Patients' understanding of their treatment plans and diagnosis at discharge. Mayo Clin Proc 2005 Aug;80(8):991-994. [doi: 10.4065/80.8.991] [Medline: 16092576]
40. Pullar T, Roach P, Mellor EJ, McNeece J, Judd A, Feely M, et al. Patients' knowledge concerning their medications on discharge from hospital. J Clin Pharm Ther 1989 Feb;14(1):57-59. [Medline: 2921303]
41. Ziaeian B, Araujo KL, Van Ness PH, Horwitz LI. Medication reconciliation accuracy and patient understanding of intended medication changes on hospital discharge. J Gen Intern Med 2012 Nov;27(11):1513-1520 [FREE Full text] [doi: 10.1007/s11606-012-2168-4] [Medline: 22798200]

42. Calkins DR, Davis RB, Reiley P, Phillips RS, Pineo KL, Delbanco TL, et al. Patient-physician communication at hospital discharge and patients' understanding of the postdischarge treatment plan. Arch Intern Med 1997 May 12;157(9):1026-1030. [Medline: 9140275]
43. Reiley P, Iezzoni LI, Phillips R, Davis RB, Tuchin LI, Calkins D. Discharge planning: comparison of patients and nurses' perceptions of patients following hospital discharge. Image J Nurs Sch 1996;28(2):143-147. [Medline: 8690431]
44. Mohammed K, Nolan MB, Rajjo T, Shah ND, Prokop LJ, Varkey P, et al. Creating a patient-centered health care delivery system: a systematic review of health care quality from the patient perspective. Am J Med Qual 2014 Jul 31:1-10. [doi: 10.1177/1062860614545124] [Medline: 25082873]
45. Bakan I, Buyukbese T, Ersahan B. The impact of total quality service (TQS) on healthcare and patient satisfaction: an empirical study of Turkish private and public hospitals. Int J Health Plann Manage 2014;29(3):292-315. [doi: 10.1002/hpm.2169] [Medline: 23494819]
46. Clark PA, Drain M, Gesell SB, Mylod DM, Kaldenberg DO, Hamilton J. Patient perceptions of quality in discharge instruction. Patient Educ Couns 2005 Oct;59(1):56-68. [doi: 10.1016/j.pec.2004.09.010] [Medline: 16198219]
47. Harewood GC, Wiersema MJ, de Groen PC. Utility of Web-based assessment of patient satisfaction with endoscopy. Am J Gastroenterol 2003 May;98(5):1016-1021. [doi: 10.1111/j.1572-0241.2003.07400.x] [Medline: 12809822]
48. Eysenbach G, Consort- E. 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]

Edited by G Eysenbach; submitted 13.03.15; peer-reviewed by K Engel; comments to author 09.04.15; revised version received 30.04.15; accepted 24.05.15; published 24.06.15.

Please cite as:
Couturier B, Carrat F, Hejblum G, SENTIPAT Study Group
Comparing Patients' Opinions on the Hospital Discharge Process Collected With a Self-Reported Questionnaire Completed Via the Internet or Through a Telephone Survey: An Ancillary Study of the SENTIPAT Randomized Controlled Trial
J Med Internet Res 2015;17(6):e158
URL: http://www.jmir.org/2015/6/e158/
doi:10.2196/jmir.4379
PMID:26109261

Original Paper

Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why

Sonja Bidmon[1], PhD; Ralf Terlutter[1], PhD

[1] Department of Marketing and International Management, Alpen-Adria-Universitaet Klagenfurt, Klagenfurt am Woerthersee, Austria

Corresponding Author:
Sonja Bidmon, PhD
Department of Marketing and International Management
Alpen-Adria-Universitaet Klagenfurt
Universitaetsstrasse 65-67
Klagenfurt am Woerthersee, 9020
Austria
Phone: 43 463 2700 ext 4048
Fax: 43 463 2700 994048
Email: sonja.bidmon@aau.at

Abstract

Background: Many studies have shown that women use the Internet more often for health-related information searches than men, but we have limited knowledge about the underlying reasons. We also do not know whether and how women and men differ in their current use of the Internet for communicating with their general practitioner (GP) and in their future intention to do so (virtual patient-physician relationship).

Objective: This study investigates (1) gender differences in health-related information search behavior by exploring underlying emotional, motivational, attitudinal as well as cognitive variables, situational involvement, and normative influences, and different personal involvement regarding health-related information searching and (2) gender differences in the virtual patient-physician relationship.

Methods: Gender differences were analyzed based on an empirical online survey of 1006 randomly selected German patients. The sample was drawn from an e-panel maintained by GfK HealthCare. A total of 958 usable questionnaires were analyzed. Principal component analyses were carried out for some variables. Differences between men (517/958) and women (441/958) were analyzed using t tests and Kendall's tau-b tests. The survey instrument was guided by several research questions and was based on existing literature.

Results: Women were more engaged in using the Internet for health-related information searching. Gender differences were found for the frequency of usage of various Internet channels for health-related information searches. Women used the Internet for health-related information searches to a higher degree for social motives and enjoyment and they judged the usability of the Internet medium and of the information gained by health information searches higher than men did. Women had a more positive attitude toward Web 2.0 than men did, but perceived themselves as less digitally competent. Women had a higher health and nutrition awareness and a greater reluctance to make use of medical support, as well as a higher personal disposition of being well-informed as a patient. Men may be more open toward the virtual patient-physician relationship.

Conclusions: Women have a stronger social motive for and experience greater enjoyment in health-related information searches, explained by social role interpretations, suggesting these needs should be met when offering health-related information on the Internet. This may be interesting for governmental bodies as well as for the insurance and the pharmaceutical industries. Furthermore, women may be more easily convinced by health awareness campaigns and are, therefore, the primary target group for them. Men are more open to engaging in a virtual relationship with the GP; therefore, they could be the primary target group for additional online services offered by GPs. There were several areas for GPs to reinforce the virtual patient-physician relationship: the fixing of personal appointments, referral to other doctors, writing prescriptions, and discussions of normal test results and doctor's notes/certificates of health.

(J Med Internet Res 2015;17(6):e156) doi:10.2196/jmir.4127

KEYWORDS

gender differences; physician-patient relations; information seeking behavior; general practitioners; statistical factor analysis

Introduction

General Background

The Internet is one of the most important sources of health information, no longer only for a small segment of Internet users, but now also for the “general public” [1]. According to the Pew Internet & American Life Project [2], the most prominent opinion poll about health-related information searches on the Internet, gender is one of the most important predictors of seeking health information on the Internet [1,3]. Over a vast span of empirical studies (eg, [2-8]), it has been demonstrated that women are more likely than men to look for health information on the Internet. However, the research so far has focused on the frequency of health information searching rather than on the underlying constructs that may help to explain such differences in Internet health information searching. Thus, this paper goes beyond existing literature by analyzing possible reasons for gender differences in Internet health information search behavior. The paper addresses gender differences in motives, emotions, cognitions, situational, and personal involvement with regard to health-related information searching.

Furthermore, the second part of the paper deals with whether and how men and women differ with regard to the virtual patient-physician relationship. In our paper, we define the virtual patient-physician relationship as communication between a patient and the physician (or the physician’s surgery or office) via the Internet. Examples include emailing, making an appointment online to see the doctor, and a virtual meeting with the doctor (eg, via Skype). We address current communication as well as future intention to communicate with the general practitioner (GP) via the Internet in general and with regard to different areas of treatment (eg, routine treatments, acute disorders, discussion of health test results, referrals to other physicians).

Theoretical Background and Research Questions

There are many approaches and models that aim at explaining why individuals search for information. For instance, Marton and Choo [1] analyzed 4 theoretically grounded quantitative studies of health information seeking on the Internet and found that the multidisciplinary frameworks differ substantially. In addition, information technology research has yielded many different competing models of users’ acceptance of new technologies (see for an overview Venkatesh et al [9]), each with a different focus and a different set of antecedents of technology adoption. With regard to the main focus of this paper to investigate gender differences in health-related information search behavior on the Internet as well as in the virtual patient-physician relationship, 3 models seem to be particularly appropriate as a theoretical basis: the theory of planned behavior (TPB) [10,11], the technology acceptance model (TAM) [12-14], and the functional theory of media use [15].

An extension of the theory of reasoned action (TRA) [16,17], TPB proposes a causal relationship of the exogenous variables attitude toward behavior, subjective norm, and perceived behavioral control with the endogenous variable behavioral intention [10,11]. Behavioral intention and perceived behavioral control together influence behavior. Based on the TPB, Venkatesh et al [18] found that there are clear gender differences in the salience of factors influencing the decision to adopt a new technology in the workplace. According to these authors, the “role of gender in technology adoption and usage behavior is crucial” [18]. These differences could be observed even when controlling several confounding variables, such as income, education, or digital literacy. Men were more strongly influenced by attitudes, whereas women were influenced more heavily by subjective norms and perceived behavioral control. Interestingly, subjective norms and perceived behavioral control had no significant impact on the decision regarding technology adoption among men at all. The gender differences were strengthened by older age [18]. Based on the TPB, the variables subjective norms and perceived behavioral control were included in the empirical study by analyzing participants’ perceived digital competence.

Based on the TRA [16,17], TAM is an applied and widely used model for describing and predicting the acceptance and use of a new information technology [12-14]. The TAM conceptualizes 2 central beliefs about a new technology that influence the intention to use it: perceived usefulness and perceived ease of use [14,19-22]. *Perceived usefulness* is defined as “the user’s perception of the degree to which using a particular system will improve his or her job performance” (eg, [22]), whereas *perceived ease of use* is defined as the “user’s perception of the extent to which using a particular system will be free of physical and mental effort” (eg, [22]). The TAM has been supported by many studies and has been applied in different contexts including the area of health information websites [23]. Various versions and extensions of the TAM have been developed. Bruner and Kumar [24] developed a “consumer technology acceptance model” (c-TAM) and demonstrated that in the consumer context, the fun of using a device was a more powerful predictor of attitude toward usage than the perceived usefulness of the device. Based on the TAM and its extension c-TAM, the variables perceived ease of use, perceived usefulness, and fun of use were included in the empirical study.

According to Dutta-Bergman [15], the functional theory of media use assumes that the use of a certain medium is motivated by different reasons and that communication behavior is goal-directed. In her study, motivation was a crucial factor in determining use of media (ie, the Internet). In her opinion, searching the Internet for health-oriented information is a reflection of health information orientation and is influenced by health consciousness and health awareness. There is relatively stable empirical evidence for a higher nutrition and health consciousness of women (eg, [25-33]). Therefore, the functional theory of media use lends support to the decision to focus on health and nutrition consciousness, as well as on motives for using the Internet as a source of health-related information and investigating the usage of different channels. All these variables were included in the empirical study.

Insights into gender differences in the virtual patient-physician relationship can also be drawn from the consumer behavior literature. According to Solomon [34], consumers' reactions to stimuli depend on psychographic variables, which can be classified into activating, emotional, motivational, and cognitive processes, and they also depend on social influence variables (eg, normative and situational antecedents).

Based on the aforementioned concepts, the objectives of the paper are as follows:

1. Investigate differences in health-related information searching on the Internet in part 1 of the paper, especially by investigating gender differences in using the Internet for health-related information searching. This will be done by (1) analyzing gender differences in feelings toward the Internet and Web 2.0 for health-related information searching (emotional perspective); (2) analyzing gender differences in perceived behavioral control, which we conceptualize as perceived digital competence (cognitive perspective); (3) analyzing gender differences in the underlying motives for using the Internet for health-related information searching (motivational perspective); (4) analyzing gender differences in health and nutrition awareness (attitudinal perspective); (5) analyzing gender differences in the personal disposition of being well-informed as a patient (personal involvement perspective); and (6) analyzing gender differences in the importance of situational circumstances, which foster the usage of Internet health information searching as well as differences in the importance of normative pressure on the usage of the Internet for health-related information searching (situational involvement and/or a normative perspective).
2. Analyze gender differences in the present and future virtual patient-physician relationship in part 2 of the paper.

Methods

Participant Recruitment

An online survey of 1006 German patients was conducted in September 2012. The term "patients" in this paper refers to individuals who visited a physician at least once in the previous 3 months. The sample was drawn from an e-panel maintained by GfK HealthCare, a leading survey research company in Nuremberg, Germany. It was based on a randomly generated set of users who had visited a GP at least once during the 3 months before the beginning of the survey. Originally, 1561 individuals were contacted; 555 persons could not participate because they did not fulfill this criterion. The recruitment rate was 64.45% (1006/1561) [35]. In all, 20 participants were excluded from the analysis because of an extremely short answer time and/or inconsistent answer patterns (eg, flatliners, contradictions). Another 28 respondents were excluded because the number of missing values exceeded the limit of 30% in scale items [36]. The final sample consisted of 958 participants. Small monetary incentives were offered for survey completion.

Questionnaire

The survey was designed by the researchers based on the existing literature and was guided by the research questions. All items apart from categorical variables (sociodemographic variables) and ordinal variables (frequency variables) were measured with 7-point rating scales. Most of the items had a "no answer" category as an alternative. Existing scales and items from the literature were used where applicable. Data were analyzed using SPSS version 22 (IBM Corp, Armonk, NY, USA).

Measurement of Sociodemographic and Psychographic Variables

Multimedia Appendix 1 provides an extract of the questionnaire and refers to the corresponding literature for items. The original questionnaire was an online questionnaire in German; English translation is merely for the purpose of this paper. In the following section, the measurement of variables included in the present study will be explained. The denomination of items (F1_1 to F42_9, D1 to D8) in brackets refers to Multimedia Appendix 1.

Sociodemographic Variables

Age (D2_1), gender (D1), the highest educational level attained (D4), family status (D5), household size (D6_1), and the categorical monthly household net income were measured (D8).

Part 1: Psychographic Variables

Feelings Toward the Internet and Other Web-Based Applications

Feelings toward the Internet and other Web-based applications in general were included in the questionnaire and measured by an item derived from Porter and Donthu [7,19] (F1_1).

Digital Literacy

Digital literacy is the ability to effectively and critically use a range of digital technologies. Literate individuals are able to make responsible choices and to access information and ideas in the digital world and to share information with others. In-line with previously published studies, digital literacy was measured with an item based on Norman and Skinner [7,8,37-39] (F2_1). In reference to the gender differences focus of this study, it has to be underlined that the construct digital literacy should be interpreted in the sense of perceived digital competence in order to do justice to the fact that especially in the area of technological knowledge it seems that women "are perhaps as susceptible to the belief in their own lack of technological ability as men are likely to delight in their own supposed superiority" [40]. Hence, our item measures perceived digital competence rather than real digital literacy.

Daily Internet Use

Respondents were asked about their daily Internet use, especially how many hours they spent on the Internet for private purposes on average on a daily, weekly, or monthly basis (total private use) (F3_1 to F3_3), and on average searching for health-related information (total private use for health-related information) (F4_1 to F4_3). We then calculated the total private Internet use and the total private Internet use for health-related information for each respondent on a daily basis.

Importance of Different Sources for Health-Related Information

For the purpose of this investigation, the importance of different sources (family, friends, physician, pharmacist, insurance agent, Internet, books/journals, other sources) was examined using items adopted from Moorman and Matulich [41] and Kummervold et al [42] (F6_1 to F6_8). The possible sources were listed in the questionnaire and the respondents had to rate the importance of each of the information sources.

Frequency of Using Different Channels on the Internet for Health-Related Information

For the purpose of investigating different search methods in the use of the Internet for health-related information, participants were asked to indicate how often they used the following channels on the Internet for health-related information searches: search engines (eg, Google), wikis (eg, Wikipedia), electronic databases and electronic papers as well as scientific papers and studies (eg, www.bmj.com), email, social networks/microblogs/networks (eg, Facebook), health forums (eg, www.imedo.de), podcasts (eg, YouTube), instant messaging/chat (eg, Skype, ICQ), and apps [43] (F7_1 to F7_10). Frequency was measured on a 6-point ordinal scale.

Motives of Using the Internet for Health-Related Information Searching

Concerning the motives of using the Internet for health-related information searching, different items from literature were used (F11_1 to F11_18). Perceived ease of use and perceived usefulness of the Internet to gain health-related information were measured by existing multi-item scales derived and adapted from Davis et al [44,45] and Venkatesh et al [9,13,46] and other authors investigating the motivational side of information searching [21,47-50]. Items measuring fun to use were adapted from Shih [51]. Additional items were developed after an extensive literature review in the health information search literature to measure the motives of saving time, of managing time flexibly, of the social component of sharing knowledge and/or making contact with someone easily, of being anonymous, and of being up-to-date.

Personal Disposition of Being Well-Informed as a Patient

According to Cacioppo and Petty [52] and Petty et al [53], the amount of information a person is seeking as well as the amount of cognitive effort and elaboration an individual is willing to devote to a specific task can be seen as individually varying personal disposition. In the area of health information searching, this means that some patients are inclined to prepare themselves for visiting a doctor and search for health-related information extensively, whereas others do this to a lesser extent [39]. Thus, some patients value health-related knowledgeability more highly because they may believe that being well-informed leads to better patient-physician communication or that the physician offers more time to well-informed patients. These individuals are more inclined to make significant health decisions on the basis of health-related information found on the Internet [7,53]. They even decide whether professional medical care is needed or not and alternatively rely on self-treatment based on their online findings [54]. For the purpose of investigating this personal disposition of being well-informed as a patient, a scale of 9 items (F20_1 to F20_9) was developed by the researchers. Some of the items were adapted from the health information orientation scale by Dutta-Bergman [15], from Simon et al [55], and from Wilson and Lankton [56].

Nutrition and Health Awareness and Attitude Toward Medical Support

Attitude is conceptualized by Solomon [34] as “a lasting, general evaluation of people (including oneself), objects, or issues” that merges into a system of values influencing the individual. The construct of health awareness primarily refers to the extent to which a person takes care of his/her own health [57-59]. We decided to denominate the construct health awareness instead of health consciousness because this sounds less clinical. Concerning nutrition and health awareness, 9 items were developed by the researchers based on a literature review and were partially adapted from the health consciousness attitude scale by Dutta-Bergman [15,60] and others [41,61] (F42_1 to F42_9).

Part 1: Situational and Normative Influences on Health-Related Information Searching on the Internet

Eight additional items were developed and integrated into the questionnaire in recognition of the fact that using the Internet could not only be due to a reason lying in the respondent himself or herself, but rather because of normative or situational reasons (F12_1 to F12_8). Therefore, after literature reviews, some complementary items measuring situational and normative influences were derived and adapted from the TAM and the TPB [9,11,17,62-67] to represent these normative or situational reasons for using the Internet for health-related information searching.

Part 2: Present and Future Communication With the General Practitioner and Internet-Based Treatment

For the purpose of investigating the present usage and future intention to communicate with the GP on the Internet and to partially replace personal communication with and treatment by a GP by the Internet, some additional items were developed by the researchers as shown in Multimedia Appendix 1. The frequency of using the Internet for communicating with the GP at present was measured by a single item on a 6-point ordinal scale (F13). Future intention to use the Internet for communicating with the GP was measured on a 7-point rating scale (F15_1). Additionally, the researchers measured which fields might conceivably be replaced by listing different areas of treatment along the virtual patient-physician relationship chain (F17_1 to F17_14). Finally, respondents were asked whether it was important to them to be able to use online treatments as well (F18_1) and how willing they would be to pay additionally for online treatment (F19_1).

Results

Sample Characteristics

A comparison of the sample used in the current study and German Internet users in 2012 (the German online population) [68] revealed that the sample represented the German online population quite well with regard to our most important variable gender (Table 1). Gender distribution in our sample (male: 54.0%, 517/958; female: 46.0%, 441/958) reflects the distribution among German Internet users (51.8% males, 48.2% females). Regarding age, participants in our sample were slightly older than those in the German online population. However, it should be noted that the comparable German Internet user basis were aged 10 years and older, whereas our study was based on respondents with a minimum age of 18 years. Another reason for this deviation probably lies in the selection criterion for participation; to qualify for our study, participants must have visited a GP at least once in the previous 3 months. With regard to education, the percentage of respondents with higher education was larger in our sample than in the German online population [68], which could be at least partially explained by the minimum age requirement of 18 years respectively the minimum age of 10 years in the comparison of the 2 databases. There were no comparable data in the German online population regarding marital status, household size, or household net income. Table 1 displays the characteristics of the sample.

Table 1. Characteristics of study sample compared to the German Internet population in 2012.

Variable and category	Female n=441	Male n=517	Total N=958	German Internet users[a] N=57,045,000
Gender, n (%)				
Men	0	517 (100.00)	517 (53.97)	29,553,000 (51.81)
Women	441 (100.00)	0	441 (46.03)	27,492,000 (48.20)
Age (years), mean (SD)	41.21 (13.39)	45.88 (12.40)	43.73 (13.04)	
Age range (years)	18-70	18-70	18-70	>10
Age categories (years), n (%)	441 (100.00)	517 (100.00)	958 (100.00)	
<24	56 (12.70)	25 (4.84)	81 (8.45)	12,552,000 (22.00)
25-44	192 (43.54)	198 (38.30)	390 (40.71)	20,344,000 (35.60)
45-64	177 (40.14)	254 (49.13)	431 (44.99)	18,799,000 (32.96)
>65	16 (3.64)	40 (7.74)	56 (5.85)	5,348,000 (9.38)
Education, n (%) [b]	437 (100.00)	514 (100.00)	951 (100.00)	52,589,000 (100.00)[b]
Without school qualification	2 (0.46)	2 (0.39)	4 (0.42)	
Secondary general school	8 (1.83)	5 (0.97)	13 (1.37)	9,487,000 (18.04)[c]
Polytechnic secondary school	43 (9.84)	77 (14.98)	120 (12.62)	
Intermediate secondary school	142 (32.49)	127 (24.71)	269 (28.28)	29,467,000 (56.03)[d]
Matura examination or higher	242 (55.38)	303 (58.95)	545 (57.31)	13,635,000 (25.93)[e]
Number in household, n (%)	439 (100.00)	517 (100.00)	956 (100.00)	
1	90 (20.50)	117 (22.63)	207 (21.65)	
2	169 (38.49)	194 (37.52)	363 (37.97)	
3	87 (19.82)	113 (21.86)	200 (20.92)	
4	83 (18.91)	72 (13.93)	155 (16.21)	
>4	10 (2.28)	21 (4.06)	31 (3.24)	
Marital status, n (%)	439 (100.00)	509 (100.00)	948 (100.00)	
Single	92 (20.95)	108 (21.22)	200 (21.10)	
Close-partnered	110 (25.06)	105 (20.63)	215 (22.68)	
Married	194 (44.19)	266 (52.26)	460 (48.52)	
Divorced	36 (8.20)	28 (5.50)	64 (6.75)	
Widowed	7 (1.60)	2 (0.39)	9 (0.95)	
Monthly household net income (€), n (%)	347 (100.00)	429 (100.00)	776 (100.00)	
<1500	77 (22.19)	52 (12.12)	129 (16.63)	
1500-2500	97 (27.95)	105 (24.47)	202 (26.03)	
2501-3500	94 (27.09)	134(31.24)	228 (29.38)	
3501-4500	53 (15.28)	68 (15.85)	121 (15.59)	
>4500	26 (7.49)	70 (16.32)	96 (12.37)	

[a] Rounded to 1000 people. Projected number of Germans who have used the Internet in the last 3 months. Age limit for questions concerning education and occupation: 16 years.

[b] For the German Internet users, low education corresponds with levels 0, 1, and 2 of the ISCED classification system (up to secondary general school), medium education corresponds with levels 3 and 4 of the ISCED classification system (up to university entrance qualification), and high education corresponds with levels 5 and 6 of the ISCED classification system (higher than matura examination respectively university entrance qualification).

[c] low education

[d] medium education

[e] high education

Part 1: Health-Related Information Searching on the Internet

Gender Differences in Health Information Search Behavior on the Internet, Emotions, and Cognitions

Table 2 provides the corresponding results of unrelated *t* tests for the psychographic variables feelings toward the Internet and other Web-based applications, perceived digital competence, daily Internet use, importance of different sources for health-related information, and the frequency of using different search methods on the Internet for health-related information between men and women.

There was a significant difference between the 2 groups in terms of their perceived digital competence (t_{899}=3.91, *P*<.001). Male respondents ascribed a higher level of perceived digital competence to themselves than female respondents did. When the participants were asked to evaluate the importance of different sources for health-related information, women rated friends (t_{944}=–3.08, *P*=.002), books or journals (t_{920}=–2.64, *P*=.009), the Internet (t_{951}=–2.36, *P*=.02), and pharmacists (t_{936}=–2.52, *P*=.012) more highly than men did (see Table 2 for details). The groups did not differ in their daily Internet use measured by the daily hours spent online for private use, or in their feelings toward the Internet and other Web-based applications in general. However, female respondents revealed a higher frequency of using the Internet for health-related information, but this difference did not meet statistical significance (t_{572}=–1.76, *P*=.08). There were some differences between female and male respondents in the frequency of usage of different channels on the Internet for health-related information searches. Women reported a higher frequency of using health forums and blogs (Kendall’s tau-b=–0.06, *P*=.03). Women revealed a higher frequency of usage of search engines (eg, Google, Bing, or Yahoo!) for health-related information searching (Kendall’s tau-b=–0.06, *P*=.045). Men, on the other hand, revealed a higher frequency of using apps for health-related information searching (Kendall’s tau-b=0.07, *P*=.02).

To do justice to the relatively large sample size, which lead to a higher probability of differences becoming significant between the 2 groups, we added the effect size of Hedges’ *g* to evaluate the group differences in all the subsequent tables. The estimates of effect size can be used to determine the practical and/or theoretical relevance of an effect and the power of an analysis [69]. There are different ways to calculate effect sizes, the most often applied being Cohen’s *d* [69]. However, we decided to apply Hedges’ *g* [70-72]. While Cohen’s *d* favors identical sample sizes, Hedges’ *g* allows for different sample sizes, which we have in our study. In contrast to Cohen’s *d*, in Hedges’ *g* the population standard deviation is replaced by the pooled sample standard deviations, calculated by using a denominator of n-1 (see the detailed formula in Multimedia Appendix 2) [69,73,74]. All the differences in the following tables will be complemented by the report of Hedges’ *g*. We are aware of the potentially misleading influences of sample size and of the risk of overvaluing observed effects because of their significance [69]; therefore, we will interpret our results in the discussion section in the light of significance and magnitude of effect sizes [75].

Gender differences in the specified psychographic variables relating to health-related information searching are reported in the next section. Because of the large number of subsequent psychographic variables, we decided to summarize the motivational, attitudinal, and personal involvement items that might contribute to the explanation of gender differences in health-related information searching. Therefore, for each group of psychographic variables (motivational, attitudinal, and personal involvement processes underlying Internet health information searching) and the group of normative and situational influences, exploratory factor analyses (EFAs) were calculated for the total sample. Only those subsets of variables were factor analyzed, which were measured on an interval scale level (statistical precondition) and which could be assigned to a specific psychographic construct or to the group of normative and situational influences. This procedure was chosen to reduce the complexity versus the alternative of a large number of group differences on a single item level. The number of factors for each of the subscales was determined by the eigenvalue criterion; principal component analyses were used with a subsequent varimax rotation with Kaiser normalization. Items with low loadings and with loadings greater than 0.45 on more than 1 factor were removed. The variances extracted were reported only for the purified scales. The factor loadings of the purified scales were used for subsequent calculation of weighted means of factor sum scores. One advantage of this method is that items with the highest loadings on the factor have the largest effect on the factor score [76]. Afterwards, *t* tests were calculated for the weighted means of factor sum scores between male and female respondents and Hedges’ *g* scores were added. The differences are described in detail in the following section.

Table 2. Gender differences in Internet health information search behavior, emotions, and cognitions influencing Internet health-related information searching.

Variables	Female (n=441)		Male (n=517)		Total (N=958)		*t* (*df*)	Kendall's tau-b	*P*	Hedges' *g*
	n	Mean (SD) or median	n	Mean (SD) or median	n	Mean (SD) or median				
Feelings toward the Internet and other Web-based applications in general[a]	431	5.75 (1.04)	514	5.80 (1.16)	954	5.78 (1.11)	0.63 (943)		.53	0.05
Perceived digital competence[b]	441	5.72 (1.11)	517	5.99 (1.0)	958	5.87 (1.06)	3.91 (899)		<.001	0.26
Daily Internet use (hours)										
Total private use	441	3.18 (2.52)	517	3.02 (2.07)	958	3.10 (2.29)	–1.05 (853)		.30	–0.07
Total private use for health-related information	441	0.53 (2.05)	517	0.35 (0.86)	958	0.43 (1.53)	–1.76 (572)		.08	–0.12
Importance of different sources for health-related information [c]										
Family	437	4.85 (1.71)	511	4.85 (1.73)	948	4.85 (1.72)	0.04 (946)		.97	0.00
Friends	436	4.36 (1.70)	510	4.01 (1.74)	946	4.17 (1.73)	–3.08 (944)		.002	–0.20
Physician	440	6.44 (0.90)	515	6.41 (0.98)	955	6.42 (0.95)	–0.40 (953)		.69	0.03
Pharmacist	432	5.15 (1.54)	506	4.89 (1.59)	938	5.01 (1.57)	–2.52 (936)		.012	–0.17
Insurance agent	405	1.75 (1.34)	486	1.80 (1.34)	891	1.78 (1.34)	0.63 (889)		.53	0.04
Internet	437	4.73 (1.44)	516	4.51 (1.43)	953	4.61 (1.44)	–2.36 (951)		.02	–0.15
Books/journals	425	4.44 (1.64)	497	4.15 (1.70)	922	4.29 (1.68)	–2.64 (920)		.009	–0.17
Other sources	280	3.02 (1.8)	352	2.81 (1.79)	632	2.90 (1.79)	–1.5 (630)		.14	–0.12
Frequency of usage of different channels on the Internet for health-related information searches, median [d]										
Search engines	441	3	517	4	958			–0.06	.045	
Wikis online encyclopedia	441	4	517	4	958			–0.02	.41	
Electronic databases/journals	441	5	517	5	958			0.03	.39	
Email	441	5	517	5	958			0.03	.38	
Social network/microblogging	441	6	517	6	958			–0.03	.27	
Health forums/blogs	441	5	517	5	958			–0.06	.03	
Podcasts	441	6	517	6	958			–0.03	.35	
Videoconferences	441	6	517	6	958			0.02	.55	
Instant messaging/chat	441	6	517	6	958			–0.04	.24	
Apps	441	6	517	6	958			0.07	.02	

[a] 1=very negative, 7=very positive.

[b] 1=not literate at all, 7=very literate.

[c] 1=not important at all, 7=very important.

[d] 1=daily, 2=weekly, 3=less often than weekly, 4=monthly, 5=less often than monthly, 6=never.

Gender Differences of Weighted Means of Factor Sum Scores for Motives Influencing Internet Health Information Searching: Exploratory Factor Analysis 1

Strong evidence was found for the existence of different motives when using the Internet for health-related information searching. Because the same procedure for the EFA was executed for all the groups of variables (attitudinal, personal involvement, situational, and normative perspective), it is only described in detail for the EFA 1. Detailed information for the other EFAs are included in the respective tables in Multimedia Appendix 2. An EFA of the 18 items measuring the underlying motives for Internet health information searching lead to a 3-factor solution of the purified scale explaining 66.69% of variance (2 items were excluded from further analysis due to low factor loadings.). As required, the Kaiser-Meyer-Olkin (KMO) measure of the appropriateness of the sample was not significant (*P*=.93) and the Bartlett-Test of sphericity was significant (χ^2_{120}=8345.2, *P*<.001). The reduced scale lead to a 3-factor solution for the motivational variables underlying Internet health information searches. The first factor (eigenvalue=7.28) consisted of 7 items featuring the social motive and enjoyment of Internet health information searching, the second factor (eigenvalue=2.38) comprised 6 items representing perceived usefulness of the Internet as a medium for health information searching, and the third factor (eigenvalue=1.01) was construed by 3 items focusing on the usefulness of the information gained from the Internet for health information searching. Table A in Multimedia Appendix 2 shows the fully rotated factor component matrix. For all the remaining 16 variables, 3 weighted means of factor sum scores were calculated (see Table E in Multimedia Appendix 2 for details of the formula) and *t* tests were calculated (see Table 3).

Women used the Internet to a greater extent than men did due to a social motive and enjoyment of Internet health information searching (t_{835}=–2.31, P=.02). Additionally, women judged the usefulness of the information gained from the Internet health information searching more highly than men did (t_{943}=–3.16, P=.002). There was a difference between men and women according to the perceived usefulness of the Internet as a medium for health information searching, but these differences did not meet statistical significance ($t_{908.55}$=–1.94, P=.05).

Table 3. Gender differences of weighted means of factor sum scores for motives influencing Internet health information searching on an aggregate level.

Factors	Female (n=441)		Male (n=517)		Total (N=958)		*t* (*df*)	*P*	Hedges' *g*
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)			
Social motive and joyousness of Internet health information searching	387	4.50 (1.50)	450	4.27 (1.48)	837	4.37 (1.49)	–2.31 (835)	.02	–0.15
Perceived usefulness of the Internet for health information searching	417	6.08 (0.93)	494	5.94 (1.12)	911	6.0 (1.04)	–1.94 (908.55)	.05	–0.14
Usefulness of the information gained from Internet health information searching	434	5.37 (1.17)	510	5.12 (1.26)	944	5.23 (1.22)	–3.16 (942)	.002	–0.21

Gender Differences of Weighted Means of Factor Sum Scores for Attitudes Influencing Internet Health Information Searching: Exploratory Factor Analysis 2

An EFA of the 9 items measuring the attitudinal influences deriving from different health and nutrition awareness and proneness to use medical support lead to a 2-factor solution for the purified scale explaining 61.14% of variance (see Table B in Multimedia Appendix 2 for details of the analysis). For all the remaining 6 variables, 2 weighted means of factor sum scores were calculated (see Table E in Multimedia Appendix 2 for details of the formula) and *t* tests were executed between the 2 weighted means of factor sum scores of the subsamples of female and male respondents.

As is shown in Table 4, there were significant differences in both areas between female and male respondents. Women had higher health and nutrition awareness on an aggregate level than men (t_{953}=–3.07, P=.002) and a greater reluctance to make use of medical support (t_{951}=–2.58, P=.01).

Table 4. Gender differences of weighted means of factor sum scores for attitudes influencing Internet health information searching on an aggregate level.

Factors	Female (n=441)		Male (n=517)		Total (N=899)		*t* (*df*)	*P*	Hedges' *g*
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)			
Health and nutrition awareness	440	5.24 (1.08)	515	5.02 (1.15)	955	5.12 (1.12)	–3.07 (953)	.002	–0.20
Reluctance to make use of medical support	439	4.79 (1.56)	514	4.52 (1.52)	953	4.64 (1.54)	–2.58 (951)	.010	–0.18

Gender Differences of Weighted Factor Sum Scores for the Personal Disposition of Being Well-Informed as a Patient: Exploratory Factor Analysis 3

An EFA of the 9 items measuring the personal disposition of being well-informed as a patient lead to a single factor solution explaining 52.93% of variance (see Table C of Multimedia Appendix 2 for details of the analysis). For all 9 variables, 1 weighted factor sum score was calculated (see Table E in Multimedia Appendix 2 for details of the formula) and a *t* test for the weighted mean of the factor sum score was calculated between the 2 subsamples of female and male respondents. There was no significant difference in this single factor of personal disposition of being well-informed as a patient in general. Therefore, it was decided to investigate the single items too (see Table 5).

Table 5. Gender differences of weighted factor sum scores for the personal disposition of being well-informed influencing Internet health information search behavior on an aggregate as well as on a basis level.

Factors/Variables	Female (n=441)		Male (n=517)		Total (N=958)		*t* (*df*)	*P*	Hedges' *g*
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)			
Personal disposition of being well-informed as a patient	413	4.05 (1.41)	486	3.94 (1.37)	899	3.99 (1.39)	–1.19 (897)	.24	–0.08
Different aspects of the personal disposition of being well-informed as a patient [a]									
It is important to me to be well-informed when consulting a physician.	436	4.70 (1.70)	514	4.72 (1.72)	955	4.71 (1.71)	0.12 (953)	.91	0.01
When I obtain health-related information from the Internet, I need to talk about this information with my physician.	436	3.99 (1.90)	514	4.12 (1.82)	950	4.06 (1.86)	1.11 (948)	.27	0.07
When a therapy is prescribed for me, I look for alternative therapies on the Internet.	439	4.30 (1.92)	514	4.33 (1.85)	953	4.32 (1.88)	0.23 (951)	.82	0.02
Sometimes I have the feeling that I am better informed about my medical condition than my physician.	439	3.78 (2.01)	513	3.55 (1.92)	952	3.66 (1.96)	–1.79 (950)	.08	–0.12
If the patient is informed, the communication with the physician is improved.	437	4.74 (1.70)	510	4.68 (1.72)	947	4.71 (1.71)	–0.51 (945)	.61	–0.04
I only decide whether a consultation with a physician is really necessary, once I have conducted some health information searches on the Internet.	437	3.18 (1.97)	512	2.93 (1.88)	949	3.05 (1.92)	–1.98 (947)	.048	–0.13
If some medicines have been prescribed, I look for information about them on the Internet.	440	4.25 (2.04)	515	3.99 (2.01)	955	4.11 (2.03)	–2.01 (953)	.045	–0.13
If the patient is informed, the physician allows more time for the treatment.	427	3.32 (1.92)	502	3.42 (1.88)	929	3.37 (1.90)	0.75 (927)	.46	0.05
The physician is more likely to prescribe a requested medicine, if the patient is informed.	418	3.66 (1.98)	500	3.58 (1.92)	918	3.61 (1.95)	–0.62 (916)	.54	–0.04

[a] 1=strongly disagree, 7=strongly agree.

As shown in Table 5, women actually differed to a certain extent in this facet of personality, but only in some distinctive aspects. Women seemed to decide on the basis of Internet health information whether to consult a physician or not to a greater extent than men did (t_{947}=–1.98, *P*=.048) and they seemed to inform themselves more than men about suggested remedies on the Internet (t_{953}=–2.01, *P*=.045). Additionally, female patients sometimes felt better informed about their medical state than their physician to a greater degree in comparison to male patients, but the difference did not meet statistical significance (t_{950}=–1.79, *P*=.08).

Gender Differences of Weighted Means of Factor Sum Scores for Situational and Normative Variables Influencing Internet Health Information Searching Exploratory Factor Analysis 4

An EFA of the 5 items measuring the underlying situational and involvement influences on Internet health information searching lead to a 2-factor solution explaining 78.88% of variance (see Table D in Multimedia Appendix 2 for details of the analysis). For all 5 variables, 2 weighted means of factor sum scores were calculated (see Table E in Multimedia Appendix 2 for details of the formula) and *t* tests were executed (see Table 6).

Table 6. Gender differences of weighted means of factor sum scores for situational and normative variables influencing Internet health information searching on an aggregate level.

Factors	Female (n=441)		Male (n=517)		Total (N=958)		*t (df)*	*P*	Hedges' *g*
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)			
Situational influences on Internet health information searching	434	5.54 (1.23)	512	5.35 (1.27)	946	5.44 (1.25)	–2.25 (944)	.03	–0.15
Normative influences on Internet health information searching	384	3.48 (1.88)	460	3.33 (1.85)	844	3.40 (1.86)	–1.20 (842)	.23	–0.08

Women seemed to be caught in a crossfire of situational, but not normative influences, to a greater extent than men which reinforced the usage of the Internet for health-related information searches. The factor including situational influences had a higher mean score for women than for men (t_{944}=–2.25, *P*=.03). The most striking result to emerge from the data was that neither women nor men were exposed to a normative influence when using the Internet or Web 2.0 for health-related information searching. The mean was comparably low for both groups (see Table 6). There was no difference between the 2 subgroups in reference to the normative influence of important individuals or individuals whom the respondents looked up to who might recommend the usage of the Internet for health-related information searching. Comparing the results between women and men, situational influences were predominantly important for women, and to a lesser extent for men, whenever they used the Internet for health-related information searching.

Part 2: The Virtual Patient-Physician Relationship

Gender Differences in Present Communication With the General Practitioner on the Internet

For the purpose of establishing whether there are gender differences in the present virtual patient-physician relationship, several unrelated *t* tests and Kendall's tau-b tests were executed. In reference to the actual use of the Internet for communicating with the GP at present, there were significant differences between the 2 groups. Given that respondents were asked about the frequency of present use of online communication with the GP on an ordinal scale (1=daily, 6=never), Kendall's tau-b was calculated to investigate gender differences. Men reported a higher frequency of communicating online with the GP than women at present (Kendall's tau-b=0.07, *P*=.02).

Gender Differences in Future Intention to Replace Personal Communication With the General Practitioner and Treatment by the Internet

In reference to the future behavioral intention of using the Internet for communication with the GP, male respondents were more prone to replace personal communication with the GP and treatment by the Internet (see Table 7). Male respondents had a higher intention to use the Internet for communicating with the GP in general than female respondents did (t_{905}=4.15, *P*<.001) and they were more ready to pay additionally for online treatment (t_{941}=2.24, *P*=.03). Gender differences were found with regard to the importance of being able to additionally use online treatment, but this did not meet statistical significance (t_{946}=1.88, *P*=.06). To see if the subsamples categorized possible areas of the physician-patient relationship in reference to their imagination of being replaced by the Internet differently, additional unrelated *t* tests were executed for each of the listed areas in the questionnaire (see Multimedia Appendix 1). As shown in Table 7, men had a higher intention to replace personal communication with the GP by the Internet for the fixing of personal appointments (t_{841}=2.13, *P*=.03), the supervision of chronically ill people (t_{943}=2.45, *P*=.01), and for routine treatments (sore throat, head cold, etc) (t_{944}=2.45, *P*=.01) than women did. Gender differences with regard to the discussion of critical test results were found, but did not meet statistical significance (t_{947}=1.85, *P*=.07). By looking at the ranking of the means of the total sample and the 2 subsamples (see Table 7), the following aspects of the virtual physician-patient relationship were the most conceivable in terms of being replaced by the Internet in the future: (1) fixing of personal appointments (female: mean 6.21, SD 1.56; male: mean 6.41, SD 1.26), (2) referrals to other doctors (female: mean 5.99, SD 1.66; male: mean 5.86, SD 1.69), (3) writing of prescriptions (female: mean 5.60, SD 1.97; male: mean 5.68, SD 1.83), (4) discussion of normal results of a test (female: mean 5.07, SD 2.17; male: mean 4.95, SD 2.15), and (5) secondary effects of drugs (female: mean 4.74, SD 2.14; male: mean 5.00, SD 2.00).

Table 7. Gender differences for future intention to replace personal communication and treatment by the Internet.

Variables	Female (n=441)		Male (n=517)		Total (N=958)		*t* (*df*)	*P*	Hedges' *g*
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)			
Intention of using the Internet more often in the future for communicating with the GP[a]	438	4.05 (2.31)	513	4.66 (2.17)	951	4.38 (2.54)	4.15 (905)	<.001	0.27
Importance of being able to use online treatment as well[b]	436	3.47 (1.99)	512	3.71 (2.04)	948	3.60 (2.02)	1.88 (946)	.06	0.12
Willingness to pay a certain amount additionally for online-treatment[c]	436	2.15 (1.74)	515	2.42 (1.87)	951	2.30 (1.81)	2.24 (941)	.03	0.15
For which of the following areas could you imagine the replacement of personal communication with your GP through Internet communication in the future? [a]									
Fixing of personal appointments	440	6.21 (1.56)	514	6.41 (1.26)	954	6.32 (1.41)	2.13 (841)	.03	0.14
Preliminary advice	436	4.75 (2.19)	511	4.86 (2.10)	947	4.81 (2.14)	0.80 (945)	.42	0.05
Writing of prescriptions	439	5.60 (1.97)	512	5.68 (1.83)	951	5.64 (1.90)	0.60 (902)	.55	0.04
Doctor's notes/certificates of health	434	4.58 (2.32)	504	4.67 (2.27)	938	4.63 (2.29)	0.62 (936)	.54	0.04
Referrals to other doctors	438	5.99 (1.66)	513	5.86 (1.69)	951	5.92 (1.68)	–1.17 (949)	.24	–0.08
Discussion of "normal" test results	437	5.07 (2.17)	514	4.95 (2.15)	951	5.01 (2.15)	–0.87 (949)	.39	–0.06
Discussion of "critical" test results	437	2.61 (2.01)	512	2.85 (2.05)	949	2.74 (2.04)	1.85 (947)	.07	0.12
Follow-up checks after treatment	436	3.13 (2.13)	506	3.29 (2.05)	942	3.21 (2.09)	1.21 (940)	.23	0.08
Supervision of chronically ill people	435	3.90 (2.16)	510	4.25 (2.17)	945	4.09 (2.17)	2.45 (943)	.01	0.16
Secondary effects of drugs	438	4.74 (2.14)	512	5.00 (2.00)	950	4.88 (2.07)	1.95 (902)	.052	0.13
Routine treatments (eg, sore throat, head cold)	436	3.96 (2.25)	510	4.31 (2.15)	946	4.15 (2.20)	2.45 (944)	.014	0.16
Psychotherapy	435	2.42 (2.00)	505	2.48 (1.95)	940	2.45 (1.97)	0.50 (938)	.62	0.03
Mental health problems	438	2.59 (2.03)	504	2.74 (1.95)	942	2.67 (1.99)	1.21 (940)	.23	0.08
Acute disorders (eg, chest pains)	438	2.42 (2.04)	505	2.56 (2.01)	941	2.50 (2.03)	1.11 (939)	.27	0.07

[a] 1=highly unlikely, 7=very likely.

[b] 1=not important at all, 7=very important.

[c] 1=I would not be willing at all, 7=I would be willing.

Discussion

Principal Findings

In reviewing the literature, only scarce empirical evidence was found on the underlying emotional, motivational, normative and situational, attitudinal, cognitive, and personal involvement variables, which may explain gender differences in Internet health-related information searching and on gender differences in the virtual patient-physician relationship. Therefore, the aim of the current investigation was to shed light on gender differences in these areas.

In order to do justice to the large sample size, we added the effect size Hedge's *g* for all *t* test values in the Results section. According to Cohen [77,78], a measure of 0.2 reflects a small effect, 0.5 reflects a medium effect, and a score greater than 0.8 reflects a large effect. Bortz and Döring [79] classify effect sizes greater than 0.50 as large, effect sizes between 0.50 and 0.30 as medium, effect sizes between 0.30 and 0.10 as small, and those less than 0.10 as trivial, the latter indicating low practical relevance. However, according to Fröhlich et al [74,80], effect sizes have to be specified according to the research field and should be interpreted dynamically (ie, in the light of the methods applied or in comparison to other extant results reported in similar research). The design of the study may also influence effect size [75]. From the point of view of effect sizes, experimentation is desirable because of the possibility of causality inference and because effect sizes seem to be more accurate. According to McCartney and Rosenthal [75], experiments in the field are likely to cause larger effects, whereas effect sizes from nonrandomized and quasi-experimental designs are likely to be affected by possible confounding variables that may interfere with the interesting variables.

The effect sizes in our study are mostly small, but exceeded the limit of 0.1 as suggested by Bortz and Döring [79] in most cases. However, we did not manipulate conditions or interventions to investigate gender differences in an experimental setting, but investigated gender differences in a real field research setting on an exploratory basis. In addition, to the best of our knowledge, comparable reports of measures of effect sizes in the literature in the area of gender differences in health-related information search behavior and the virtual patient-physician relationship are lacking, further obstructing the comparison of our effect sizes against other research findings. In reference to McCartney and Rosenthal [75], “no criterion can be developed to separate small, useless effects from small, useful ones; researchers need to evaluate effect sizes using logic and argument.” Therefore, we discuss our results with the gender differences and the effect sizes in the light of the exploratory nature of our study.

Part 1: Health-Related Information Searching on the Internet

In reference to behavioral variables the study is by trend in-line with studies reporting that women are more frequent users of the Internet for health-related information searches [2-5,8,48], but the respective gender differences found in our study did not meet statistical significance. Additionally, it was demonstrated that women and men differ in their frequency of usage of different channels on the Internet for health-related information searching. In comparison to men, women report a higher frequency of using health forums and blogs and search engines (according to Kendall’s tau-b test) as well as search engines, but the latter does not meet statistical significance. Friends, pharmacists, books and journals, and the Internet are more important sources for health-related information searching for women than for men. Male respondents, conversely, use apps more often than women for health-related information searching. This is in-line with research demonstrating that men consistently show higher levels of mobile Internet and app usage than women do (eg, [81]). For instance, the German Digitalbarometer, a telephone survey conducted 2012 in cooperation between TNS Emnid, IP Deutschland, and the trade magazine *Werben & Verkaufen* among 1142 Germans between 14 and 64 years [82] reported that 36% of men and 18% of women used apps. One important explanation for the higher usage of mobile devices and apps by men is given by Or and Karsh [83], who report that women have higher computer anxiety and less perceived behavior control. This argument is in-line with the fact that, in our study, men ascribe themselves higher perceived digital competence (cognitive perspective). Technological competence refers to sexual identity and Cockburn argues that femininity seems to be incompatible with technological competence and women who feel technologically competent perceive themselves as being more manly [40]. Therefore, being comfortable with technology contributes more or less to some kind of male gender identity [40,84]. However, men ascribing themselves higher perceived digital competence may not correspond to real differences in digital literacy because differences were not measured by observation, but were based on self-reported answers. Differences in self-ascribed digital competence may simply reflect differences in culturally evolved gender identity. Nevertheless, a higher perceived digital competence may also prevent computer anxiety and may correspond with higher behavior control in the area of Internet information searching.

The current study found that there are no differences between the female and male respondents in their feelings toward the Internet and other Web-based apps in general.

The next question in this research was whether women and men differ in their motivations to use the Internet for health-related information searching. The most interesting finding was that women use the Internet for health-related information searching to a higher degree than men for social reasons and for pleasure. They evaluate it as a more useful medium and they perceive the gained information as more useful than men do. When looking at the differences on the level of the items, the Internet is attractive for women because it is an efficient method of searching (easy, quick, always available, capable of enhancing search success) because of its social dimensions (offering different formats, getting in contact with other people easily) and its entertainment potential. These results can be explained from a social role perspective. Due to the multitasking agenda of women, especially those of middle age, who play key roles as health managers and family caregivers [1,85,86], efficiency is very important. On the other hand, the Internet offers a new way of getting in contact with other people at times when the children are asleep, for example. Therefore, when women are responsible for young children, they have to overcome more obstacles when they want to meet other people in person. Thus, the social dimensions of the Internet may be more attractive for women than for men and the entertaining dimensions of the Internet may be of higher importance for them than for men.

With regard to the question of how situational involvement differs between women and men in relation to health-related information searching on the Internet, this study found that situational influences are predominantly important for women, and to a smaller extent for men, whenever they use the Internet for health-related information searching. Surprisingly, normative influences seem to make no contribution to gender differences in usage of the Internet for health-related information searching. A possible explanation for this might be that women, especially middle-aged women, sometimes work part-time because of their manifold roles and therefore have only limited access to and limited time for the Internet. This may cause a higher dependency on situational circumstances and a higher situational involvement with the Internet and Web 2.0. Nevertheless, this explanation must be interpreted with caution, because there are many middle-aged women who work full time in spite of possible manifold roles. Therefore, this interpretation cannot be extrapolated to all women; hence, there is room for many other complementary root cause analyses.

From an attitudinal perspective, the results are consistent with those of other studies revealing that women show higher nutrition and health awareness across different countries and settings (eg, [28-33]) and prefer homeopathic remedies to a higher extent, which was also found consistently in studies from different countries (eg, [87,88]). However, the findings of this study do not support the results from a recent study by Cho et al [57]; they found that men had higher health consciousness.

This study found that women are more reluctant to visit a physician than men. This result is contrary to a recent study from Smith et al [29], who found that men have a higher reluctance than women to visit a doctor for minor mental health concerns, but seem to seek help once a problem reaches a specific threshold. In our study, the items were formulated in a more general manner and did not focus on mental health problems. Therefore, our results in this context may be explained partially by the personal disposition of being well-informed as a patient, which is higher for women than for men. Women may often decide to visit a physician only once they have conducted some Internet health information searches. They are also more prone to look for information about prescribed remedies. Altogether, they seem to value being well-informed as a patient more highly than men and they strive to be better informed through the search for health-related information on the Internet. Thus, because of their need to be well-informed about their symptoms, they hesitate to consult a GP more than men in the case of illness. However, social role interpretations are not only useful in explaining the frequency differences between men and women in Internet health information searches, but also in explaining the underlying motives and attitudes toward Internet health information searching.

Part 2: The Virtual Patient-Physician Relationship

At present, men report a higher frequency of communicating online with the GP and they are also more willing than women are to replace personal communication with the GP and treatment by the Internet in the future. Men can imagine fostering the virtual patient-physician relationship in the areas of making personal appointments, the supervision of chronically ill people, and for routine treatments (eg, sore throat, head cold). Additionally, they are more willing to pay a certain amount of extra money for online treatment. We see 2 main explanations for these findings. First, and as outlined previously, women perceive themselves as less digitally literate than men and, therefore, may feel a higher level of unease with regard to replacing the relatively intimate personal face-to-face GP consultation by a virtual one, which is probably rated as being less intimate. Secondly, from a social role perspective, women visit GPs not only for themselves, but also in their role as caregiver to their children. Hence, the replacement of a personal consultation by a virtual consultation may be perceived as being even more difficult if women are acting on behalf of someone else, especially their own children.

Hence, the replacement of the personal dimension through the Internet may be more difficult for women than it is for men. Reduced willingness to pay additionally for online treatment may also be explained by women's smaller amount of disposable income. Comparing the household net income of the female and the male subsample, in-line with the census data, it was shown that the household net income was higher for the male subsample. Therefore, it may be more affordable for men to pay a certain amount of extra money for online treatment.

Limitations

The study is not without limitations. There is the possibility of selection bias among respondents, although random selection out of the database was held to minimize its likelihood. The recruitment rate of 64% for this online panel sample also indicates that selection bias among respondents is probably low. A demographic comparison showed that our sample reflects the German online population relatively well. However, in the subsample of male respondents, the age category of older men (45 years and older) was overrepresented and there were also more respondents with higher education than in the general online population for both of the subgroups. Future studies may try to make use of a larger randomized sample of the average online population.

The questionnaire was very comprehensive because of the many variables that were addressed, which might raise the issue of fatigue among the respondents. However, the exact duration of the survey completion was automatically measured and saved in a control variable offering the possibility to control for answer duration and to exclude participants with an extremely short answer time from the analysis. In addition, data were also analyzed for inconsistent answer patterns (eg, flatliners, contradictions). Several multi-item scales were aggregated using EFAs. However, such data treatment for the sake of complexity reduction always leads to a loss of variance of the individual items. Our measurement of daily Internet use by asking respondents for their average usage may have been challenging for participants, especially for individuals with an intermittent usage pattern. An alternative would have been to ask respondents for their duration of Internet usage in the previous week (or month). However, such alternative measurement faces the problem that the previous week (month) might not be representative of the average duration. The construct digital literacy may face a special problem for a gender-specific research focus. The problem is that men and women perceive digital competence differently with men being, in general, more self-confident in this area and women facing less self-ascribed digital affinity. These interpretations may follow differences in self-identity as has been elaborated previously. For this reason, the results conveying gender differences for the construct digital literacy were interpreted as differences in perceived digital competence from a gender identity perspective.

Our study can be categorized as being exploratory in nature, delivering some pioneer knowledge in investigating reasons for gender differences in health-related information search behavior and the virtual patient-physician relationship. Although the *t* tests and Kendall's tau-b tests demonstrated significant differences in many areas, the effect sizes (Hedges' *g*) were relatively low (however, low effect sizes are not necessarily a limitation). It seems possible that the small effect sizes may be traced back to the field research paradigm instead of experimental design. Nevertheless, due to the exploratory nature of the study, we think that the results deliver interesting insights into gender differences in health-related information search behavior and the underlying psychographic, situational, and normative variables. Results also shed light on the virtual patient-physician relationship.

Another limitation of our study is that gender differences are likely to be bounded to the respective cultural background, especially when they are interpreted from a social role perspective. Although we believe that the findings are generalizable beyond the German population to a certain extent (eg, to other German-speaking countries), comparable studies in other countries would bring forward the generalizability of our results.

It would also be interesting to investigate the research questions and validate our results on gender differences by using other methods of inquiry, samples, and countries in the future.

Practical Implications

The first implication that can be derived from our study is one from a more general gender perspective. Results from this survey are mostly in-line with previous studies demonstrating that women ascribe themselves a lower degree of digital competence than men. The current study delivers an additional argument from the health sector, namely that the government might want to be more proactive in enabling and encouraging women to be interested in technology and in technical devices from an early age.

Our study delivered the interesting finding that women have a higher social motive for health-related information searches and value the enjoyment of Internet health information searching to a higher degree than men do. Hence, measures to increase the pleasure of health information searching may be especially beneficial to women. This may be interesting for government institutions (eg, for health consciousness campaigns), but it is also of interest to the pharmaceutical industry wanting to promote their products. For instance, advergames targeted at female virtual players could be a means to reinforce health consciousness (educational advergames) or brand knowledge and brand awareness of pharmaceutical products or dietary supplements [89].

The lower health and nutrition awareness of men could be interesting for GPs, for the government, for the insurance industry, and for entrepreneurs developing apps. Men have a shorter life expectancy, which may be influenced to a certain degree by their lower health and nutrition awareness. Because men have a higher tendency to use apps for health-related information searching, men could be an interesting target group for health-promoting apps and/or fitness apps, which have been booming in recent years. These apps could also be interesting for the insurance industry and the government, which is confronted with ever-increasing expenditures in the health sector.

The fact that men are also more interested in fostering the virtual patient-physician relationship may be of special interest for GPs. For example, if a GP wants to reduce waiting times and operate more efficiently (eg, through Internet communication for administrative purposes), men may be more easily convinced than women.

Aside from gender, there are several areas for GPs in which the virtual patient-physician relationship could be reinforced: the fixing of personal appointments, referrals to other doctors, the writing of prescriptions, discussions of normal test results, and doctor’s notes/certificates of health. If a GP intends to foster her/his customer orientation, she/he may think about reducing waiting times by offering more online services in the preceding areas. An important step here would be to clarify the legal framework conditions for implementing an enhancement of the virtual patient-physician relationship. Yet it will be necessary to segment the patient base according to their individual disposition toward fostering the virtual patient-physician relationship, which may be influenced by gender.

Acknowledgments

The authors are grateful to Martina Moick for her contribution in developing the questionnaire and to GfK HealthCare Nuremberg, Germany, in particular Dr Susanna Meyer and Norbert Schell, for their contributions and for collecting the data for this analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Extract of the questionnaire and justification of items.

[PDF File (Adobe PDF File), 127KB - jmir_v17i6e156_app1.pdf]

Multimedia Appendix 2

Additional tables (A-F) and methodological details of exploratory factor analyses 1- 4.

[PDF File (Adobe PDF File), 115KB - jmir_v17i6e156_app2.pdf]

References

1. Marton C, Wei Choo C. A review of theoretical models of health information seeking on the web. Journal of Documentation 2012 Apr 20;68(3):330-352. [doi: 10.1108/00220411211225575]
2. Fox S. The social life of health information. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2013-11-20] [WebCite Cache ID 6LGpMcwCG]
3. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28 [FREE Full text] [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
4. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711 [FREE Full text] [Medline: 19075034]

5. Umefjord G, Sandström H, Malker H, Petersson G. Medical text-based consultations on the Internet: a 4-year study. Int J Med Inform 2008 Feb;77(2):114-121. [doi: 10.1016/j.ijmedinf.2007.01.009] [Medline: 17317292]
6. Beck F, Richard JB, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
7. Moick M, Terlutter R. Physicians' motives for professional internet use and differences in attitudes toward the internet-informed patient, physician-patient communication, and prescribing behavior. Med 2 0 2012;1(2):e2 [FREE Full text] [doi: 10.2196/med20.1996] [Medline: 25075230]
8. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
9. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterley 2003;27(3):425-478.
10. Ajzen I. From Intentions to Actions: A Theory of Planned Behavior. Berlin: Springer; 1985.
11. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
12. Davis FD. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. (PhD dissertation). Cambridge, MA: MIT Sloan School of Management; 1986. URL: http://dspace.mit.edu/handle/1721.1/15192 [accessed 2014-05-29] [WebCite Cache ID 6Pvwb5Y2O]
13. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb 01;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
14. Venkatesh V, Davis FD. A model of the antecedents of perceived ease of use: development and test. Decision Sciences 1996 Sep;27(3):451-481 [FREE Full text] [doi: 10.1111/j.1540-5915.1996.tb00860.x]
15. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among Internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [doi: 10.2196/jmir.6.2.e15] [Medline: 15249264]
16. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood-Cliffs, NJ: Prentice-Hall; 1980.
17. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley Pub Co; 1975.
18. Venkatesh V, Morris MG, Ackerman PL. A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. Organ Behav Hum Decis Process 2000 Sep;83(1):33-60. [doi: 10.1006/obhd.2000.2896] [Medline: 10973782]
19. Porter CE, Donthu N. Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. Journal of Business Research 2006 Sep;59(9):999-1007. [doi: 10.1016/j.jbusres.2006.06.003]
20. Davis FD, Venkatesh V. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. International Journal of Human-Computer Studies 1996 Jul;45(1):19-45 [FREE Full text] [doi: 10.1006/ijhc.1996.0040]
21. Subramanian GH. A replication of perceived usefulness and perceived ease of use measurement. Decision Sciences 1994 Sep;25(5-6):863-874 [FREE Full text] [doi: 10.1111/j.1540-5915.1994.tb01873.x]
22. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. International Journal of Man-Machine Studies 1993 Mar;38(3):475-487 [FREE Full text] [doi: 10.1006/imms.1993.1022]
23. Kim D, Chang H. Key functional characteristics in designing and operating health information websites for user satisfaction: an application of the extended technology acceptance model. Int J Med Inform 2007;76(11-12):790-800. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]
24. Bruner II GC, Kumar A. Explaining consumer acceptance of handheld Internet devices. Journal of Business Research 2005 May;58(5):553-558. [doi: 10.1016/j.jbusres.2003.08.002]
25. Furnham A, Kirkcaldy B. Age and sex differences in health beliefs and behaviours. Psychol Rep 1997 Feb;80(1):63-66. [doi: 10.2466/pr0.1997.80.1.63] [Medline: 9122353]
26. Schätzer M, Rust P, Elmadfa I. Fruit and vegetable intake in Austrian adults: intake frequency, serving sizes, reasons for and barriers to consumption, and potential for increasing consumption. Public Health Nutr 2010 Apr;13(4):480-487. [doi: 10.1017/S136898000999142X] [Medline: 19807933]
27. Dehghan M, Akhtar-Danesh N, Merchant AT. Factors associated with fruit and vegetable consumption among adults. J Hum Nutr Diet 2011 Apr;24(2):128-134. [doi: 10.1111/j.1365-277X.2010.01142.x] [Medline: 21332835]
28. Johnson RL. Gender differences in health-promoting lifestyles of African Americans. Public Health Nurs 2005;22(2):130-137. [doi: 10.1111/j.0737-1209.2005.220206.x] [Medline: 15860068]
29. Smith KLW, Matheson FI, Moineddin R, Dunn JR, Lu H, Cairney J, et al. Gender differences in mental health service utilization among respondents reporting depression in a national health survey. Health 2013 Oct;05(10):1561-1571 [FREE Full text] [doi: 10.4236/health.2013.510212]

30. Green CA, Pope CR. Gender, psychosocial factors and the use of medical services: a longitudinal analysis. Social Science & Medicine 1999 May;48(10):1363-1372. [doi: 10.1016/S0277-9536(98)00440-7] [Medline: 10369437]
31. Hoefkens C, Verbeke W, Van Camp J. European consumers' perceived importance of qualifying and disqualifying nutrients in food choices. Food Quality and Preference 2011 Sep;22(6):550-558. [doi: 10.1016/j.foodqual.2011.03.002]
32. Ahmed F, Al-Radhwan L, Al-Azmi GZS, Al-Beajan M. Association between stress and dietary behaviours among undergraduate students in Kuwait: gender differences. Journal of Nutrition and Health Sciences 2014 May;1(1):14-16 [FREE Full text] [doi: 10.15744/2393-9060.1.104]
33. Sharma B, Harker M, Harker D, Reinhard K. Youth transition to university in Germany and Australia: an empirical investigation of healthy eating behaviour. Journal of Youth Studies 2010 Jun;13(3):353-367. [doi: 10.1080/13676260903447510]
34. Solomon MR. Consumer Behavior: Buying, Having and Being. Engelwood Cliffs, NJ: Prentice Hall; 2014.
35. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
36. Wirtz M. Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann [On the problem of missing data: how to identify and reduce the impact of missing data on findings of data analysis]. Rehabilitation 2004 Apr;43(2):109-115. [doi: 10.1055/s-2003-814839] [Medline: 15100920]
37. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]
38. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
39. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
40. Wilson M. A conceptual framework for studying gender in information systems research. J Inf Technol 2004 Mar;19(1):81-92. [doi: 10.1057/palgrave.jit.2000008]
41. Moorman C, Matulich E. A model of consumers' preventive health behaviors: the role of health motivation and health ability. J Consum Res 1993 Sep;20(2):208-228. [doi: 10.1086/209344]
42. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
43. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23. [doi: 10.1111/j.1471-1842.2007.00701.x]
44. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
45. Davis FD, Bagozzi RP, Warshaw PR. Extrinsic and intrinsic motivation to use computers in the workplace. J Appl Soc Psychol 1992 Jul;22(14):1111-1132. [doi: 10.1111/j.1559-1816.1992.tb00945.x]
46. Venkatesh V, Speier C, Morris MG. User acceptance enablers in individual decision making about technology: toward an integrated model. Decision Sciences 2002 Mar;33(2):297-316 [FREE Full text] [doi: 10.1111/j.1540-5915.2002.tb01646.x]
47. 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]
48. Sechrest RC. The internet and the physician-patient relationship. Clin Orthop Relat Res 2010 Oct;468(10):2566-2571 [FREE Full text] [doi: 10.1007/s11999-010-1440-3] [Medline: 20574803]
49. Bennett NL, Casebeer LL, Zheng S, Kristofco R. Information-seeking behaviors and reflective practice. J Contin Educ Health Prof 2006;26(2):120-127. [doi: 10.1002/chp.60] [Medline: 16802313]
50. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [Medline: 12617199]
51. Shih HP. Extended technology acceptance model of Internet utilization behavior. Information & Management 2004 Jul;41(6):719-729. [doi: 10.1016/j.im.2003.08.009]
52. Cacioppo JT, Petty RE. The need for cognition. Journal of Personality and Social Psychology 1982 Jan;42(1):116-131 [FREE Full text] [doi: 10.1037/0022-3514.42.1.116]
53. Petty RE, Briñol P, Loersch C, McCaslin MJ. The need for cognition. In: Handbook of Individual Differences in Social Behavior. New York: Guilford; 2009:318-329.
54. Goonawardene N, Jiang J, Tan SSL, Jiang Z(. Pacific Asia Conf Inf Syst. 2013. Online Health Information Seeking and Adolescents' Intention Towards Health Self-Management URL: http://www.pacis-net.org/file/2013/PACIS2013-174.pdf [accessed 2015-06-12] [WebCite Cache ID 6ZPbW2KZF]
55. Simon D, Kriston L, Loh A, Spies C, Scheibler F, Wills C, et al. Confirmatory factor analysis and recommendations for improvement of the Autonomy-Preference-Index (API). Health Expect 2010 Sep;13(3):234-243. [doi: 10.1111/j.1369-7625.2009.00584.x] [Medline: 20579122]
56. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]

57. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
58. Dutta-Bergman MJ. Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Health Commun 2004;16(3):273-288. [doi: 10.1207/S15327027HC1603_1] [Medline: 15265751]
59. Basu A, Dutta MJ. The relationship between health information seeking and community participation: the roles of health information orientation and efficacy. Health Commun 2008 ;23(1):70-79. [doi: 10.1080/10410230701807121] [Medline: 18443994]
60. Dutta-Bergman MJ. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]
61. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. J Acad Mark Sci 1998;26(1):6-15. [doi: 10.1177/0092070398261002]
62. Mathieson K. Predicting user intentions: comparing the Technology Acceptance Model with the Theory of Planned Behavior. Information Systems Research 1991 Sep;2(3):173-191. [doi: 10.1287/isre.2.3.173]
63. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterley 1989;13(3):319-340.
64. Venkatesh V, Morris MG. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS Quarterly 2000 Mar;24(1):115-139. [doi: 10.2307/3250981]
65. Pai FY, Huang KI. Applying the Technology Acceptance Model to the introduction of healthcare information systems. Technol Forecast Soc Change 2011 May;78(4):650-660. [doi: 10.1016/j.techfore.2010.11.007]
66. Taylor S, Todd P. Understanding information technology usage: a test of competing models. Information Systems Research 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]
67. Taylor S, Todd P. Assessing IT usage: the role of prior experience. MIS Quarterly 1995 Dec;19(4):561-570. [doi: 10.2307/249633]
68. Statistisches B. Wirtschaftsrechnungen. Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien. 2012. URL: https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400127004.pdf?__blob=publicationFile [accessed 2014-03-28] [WebCite Cache ID 6OPPC4HiD]
69. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18 [FREE Full text] [doi: 10.1037/a0024338] [Medline: 21823805]
70. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational and Behavioral Statistics 1981 Jan 01;6(2):107-128. [doi: 10.3102/10769986006002107]
71. Hedges LV. Estimation of effect size from a series of independent experiments. Psychological Bulletin 1982 Sep;92(2):490-499. [doi: 10.1037/0033-2909.92.2.490]
72. Hedges LV. Issues in meta-analysis. Rev Res Educ 1986;13:353-398.
73. Rosnow RL, Rosenthal R. Effect Sizes. Zeitschrift für Psychologie / Journal of Psychology 2009 Feb 3;217(1):6-14. [doi: 10.1027/0044-3409.217.1.6]
74. Fröhlich M, Emrich E, Pieter A, Stark R. Outcome effects and effects sizes in sport sciences. Int J Sports Science and Engineering 2009;3(3):175-179 [FREE Full text]
75. McCartney K, Rosenthal R. Effect size, practical importance, and social policy for children. Child Dev 2000 Feb 01;71(1):173-180.
76. Distefano C, Zhu M, Mîndrila D. Understanding and using factor scores: considerations for the applied researcher. Pract Assessment, Res Eval 2009;14(20):1-11 [FREE Full text]
77. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977.
78. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
79. Bortz J, Döring N. Forschungsmethoden und Evaluation: für Human-und Sozialwissenschaftler [Research Methods and Evaluation: for Human- and Social Scientists]. Berlin: Springer; 2006.
80. Fröhlich M, Pieter A. Cohen's Effektstärken als Mass der Bewertung von praktischer Relevanz - Implikationen für die Praxis [Cohen’s effect sizes as measure for practical relevance - implications for practice]. Schweizerische Zeitschrift für Sportmedizin und Sporttraumatologie 2006;57(4):139-142.
81. van Eimeren B, Frees B. Media Perspektiven. 2013. Rasanter Anstieg des Internetkonsums - Onliner fast drei Stunden täglich im Netz URL: http://www.media-perspektiven.de/uploads/tx_mppublications/0708-2013_Eimeren_Frees_01.pdf [accessed 2013-11-21] [WebCite Cache ID 6LHZaWeHA]
82. TNS Emnid. DigitalBarometer 1/2012: Mobiles Internet. 2012 Jun. URL: http://www.tns-emnid.com/presse/pdf/presseinformationen/DigitalBarometer2012_TNS_Emnid.pdf [accessed 2013-11-21] [WebCite Cache ID 6LHZWmmMn]
83. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 ;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]

84. Wajcman J. Feminist theories of technology. Cambridge Journal of Economics 2009 Jan 08;34(1):143-152. [doi: 10.1093/cje/ben057]
85. Young R. The household context for women's health care decisions: Impacts of UK policy changes. Social Science & Medicine 1996 Mar;42(6):949-963. [doi: 10.1016/0277-9536(95)00192-1] [Medline: 8779006]
86. Yoo EY, Robbins LS. Understanding middle-aged women's health information seeking on the web: A theoretical approach. J Am Soc Inf Sci 2008 Feb 15;59(4):577-590. [doi: 10.1002/asi.20766]
87. Menniti-Ippolito F, Gargiulo L, Bologna E, Forcella E, Raschetti R. Use of unconventional medicine in Italy: a nation-wide survey. Eur J Clin Pharmacol 2002 Apr;58(1):61-64. [doi: 10.1007/s00228-002-0435-8] [Medline: 11956675]
88. Wilkinson JM, Jelinek H. Complementary medicine use among attendees at a rural health screening clinic. Complement Ther Clin Pract 2009 May;15(2):80-84. [doi: 10.1016/j.ctcp.2008.12.005] [Medline: 19341985]
89. Terlutter R, Capella ML. The gamification of advertising: analysis and research directions of in-game advertising, advergames, and advertising in social network games. Journal of Advertising 2013 Apr 03;42(2-3):95-112. [doi: 10.1080/00913367.2013.774610]

Abbreviations

EFA: exploratory factor analysis
GfK: Gesellschaft für Konsumforschung
GP: general practitioner
TAM: technology acceptance model
TPB: theory of planned behavior
TRA: theory of reasoned action

Edited by G Eysenbach; submitted 09.12.14; peer-reviewed by P Resch, M Emmert, C Urquhart; comments to author 12.02.15; revised version received 11.04.15; accepted 23.04.15; published 22.06.15.

Please cite as:
Bidmon S, Terlutter R
Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why
J Med Internet Res 2015;17(6):e156
URL: http://www.jmir.org/2015/6/e156/
doi:10.2196/jmir.4127
PMID:26099325

Original Paper

The Invisible Work of Personal Health Information Management Among People With Multiple Chronic Conditions: Qualitative Interview Study Among Patients and Providers

Jessica S Ancker[1], MPH, PhD; Holly O Witteman[2,3,4], PhD; Baria Hafeez[1], MS; Thierry Provencher[3]; Mary Van de Graaf[1], MPH; Esther Wei[5], NP

[1]Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medical College, New York, NY, United States
[2]Department of Family and Emergency Medicine, Faculty of Medicine, Laval University, Québec City, QC, Canada
[3]Office of Education and Continuing Professional Development, Faculty of Medicine, Laval University, Québec City, QC, Canada
[4]Research Centre of the CHU de Québec, Québec City, QC, Canada
[5]Department of Medicine, Weill Cornell Medical College, New York, NY, United States

Corresponding Author:
Jessica S Ancker, MPH, PhD
Division of Health Informatics
Department of Healthcare Policy and Research
Weill Cornell Medical College
425 E 61st St, Suite 301
New York, NY, 10065
United States
Phone: 1 646 962 9404
Fax: 1 646 962 0105
Email: jsa7002@med.cornell.edu

Abstract

Background: A critical problem for patients with chronic conditions who see multiple health care providers is incomplete or inaccurate information, which can contribute to lack of care coordination, low quality of care, and medical errors.

Objective: As part of a larger project on applications of consumer health information technology (HIT) and barriers to its use, we conducted a semistructured interview study with patients with multiple chronic conditions (MCC) with the objective of exploring their role in managing their personal health information.

Methods: Semistructured interviews were conducted with patients and providers. Patients were eligible if they had multiple chronic conditions and were in regular care with one of two medical organizations in New York City; health care providers were eligible if they had experience caring for patients with multiple chronic conditions. Analysis was conducted from a grounded theory perspective, and recruitment was concluded when saturation was achieved.

Results: A total of 22 patients and 7 providers were interviewed; patients had an average of 3.5 (SD 1.5) chronic conditions and reported having regular relationships with an average of 5 providers. Four major themes arose: (1) Responsibility for managing medical information: some patients perceived information management and sharing as the responsibility of health care providers; others—particularly those who had had bad experiences in the past—took primary responsibility for information sharing; (2) What information should be shared: although privacy concerns did influence some patients' perceptions of sharing of medical data, decisions about what to share were also heavily influenced by their understanding of health and disease and by the degree to which they understood the health care system; (3) Methods and tools varied: those patients who did take an active role in managing their records used a variety of electronic tools, paper tools, and memory; and (4) Information management as invisible work: managing transfers of medical information to solve problems was a tremendous amount of work that was largely unrecognized by the medical establishment.

Conclusions: We conclude that personal health information management should be recognized as an additional burden that MCC places upon patients. Effective structural solutions for information sharing, whether institutional ones such as care management or technological ones such as electronic health information exchange, are likely not only to improve the quality of information shared but reduce the burden on patients already weighed down by MCC.

(J Med Internet Res 2015;17(6):e137) doi:10.2196/jmir.4381

KEYWORDS

consumer health information technology; electronic medical records; electronic patient portals; personal health records (PHRs); chronic disease; qualitative research; diabetes; information management

Introduction

Some 90 million individuals in the United States are affected by more than one chronic disease simultaneously, and the number of people with “multiple chronic conditions” (MCC) continues to grow as the population ages [1]. The designation of MCC is a broad one that has been defined by the US Department of Health and Human Services as any combination of conditions that last at least one year and that require ongoing medical attention or limit activities of daily living [1]. Individuals can be described as having MCC if they have two or more of any physical or mental conditions (heart disease, depression, anxiety, diabetes, asthma, arthritis, HIV, chronic obstructive pulmonary disorder, chronic pain, etc). Medical care for individuals with MCC is challenging as the evidence base about specific combinations of conditions may be weak or absent, and the therapies and management strategies for a particular condition might be contraindicated by another condition [1].

Dealing with the health care system is potentially very challenging for patients with MCC, as they typically consult more doctors and have more medical appointments than patients with single conditions [1]. One critical problem for patients who see multiple health care providers is the issue of communication among those providers. Clinicians and policymakers have long recognized that critical patient data is often missing at clinical encounters even in medically straightforward situations [2], and that the chances of missing data increase with care transitions [3,4]. Such missing data contributes to lack of care coordination, low quality of care, and medical errors [2-6].

Potential health information technology (HIT) solutions have been focused primarily on facilitating provider-to-provider information sharing, including interoperable electronic health records (EHRs) and health information exchange (HIE) systems [7-9]. However, in addition, a number of consumer technologies offer patients the opportunity to transfer their own records across care settings, a process known as consumer-mediated or patient-mediated HIE [10]. These include patient-controlled personal health records (PHRs), electronic patient portals managed by health care organizations, and Blue Button functionalities that allow patients to export medical record information for personal use [11-16]. Consumer surveys frequently find strong public support for the concept of patient-mediated HIE [17-20]. Yet concerns have been expressed about whether all patients will be sufficiently engaged or informed to serve as stewards of their own data, whether patients might suppress or alter sensitive information [10], and whether socioeconomically disadvantaged and elderly patients will have adequate computer access or skills to use these technologies [21]. Recent data shows that patient use of portals and PHRs is beginning to climb, but these tools are still reaching only a minority of the public [14].

From the patient perspective, the tasks involved in collecting and managing personal medical information have been called “personal health information management” *(*PHIM) [22-25]. PHIM encompasses a variety of activities conducted largely outside the medical encounter: examples include tracking health data, seeking information, and organizing it [22], creating personal histories, and planning medical activities [23], and providing records to doctors [26]. As these are all effortful, directed activities to attain goals, it is appropriate to recognize them as *work* [22,27-29]. Most PHIM activities fall in the category of “illness work”, that is, the activities involved with managing an illness, such as taking medicines, getting information, and using technologies such as blood glucose meters [27-29]. Other PHIM activities constitute “articulation work”, in other words, the planning and managing tasks that allow people to complete other types of work, whether illness work or everyday life work [27,28]. Articulation work might include such essential tasks as keeping a family calendar or organizing transportation to medical appointments.

A rich PHIM literature is developing. Some work has focused on healthy individuals and families [22,23,26,30] and on computer-literate participants [26]. Another body of work is developing on patients with cancer [31-34]. As part of a broader project on potential applications of consumer HIT and barriers to its use, we sought to explore PHIM conducted by patients with MCC, whose long-term complex medical situations would be expected to result in heavy demands for information management. Our qualitative study focused on the management of medical information and medical records. Our research questions were: How do patients with MCC manage their medical records and medical information sharing with medical providers? How do they perceive their role in managing their medical information? Management of information was defined broadly to include information transfers across the patient’s network of current providers as well as during care transitions from one provider to another.

Methods

Participants

As described in the companion piece to this paper [35], we recruited adult English-speaking patients with MCC, as well as health care providers with experience providing care for patients with MCC. Patients and providers were recruited independently from the same settings but not specifically to represent patient-provider pairs. One researcher (JSA) also attended six 90-minute sessions of a diabetes education support group in order to triangulate themes arising from patient interviews. The diabetes group was chosen because diabetes is prevalent among patients with MCC, because the majority of the patients in the diabetes group had at least one comorbid condition, and also because of availability (we found no local group education programs focusing on MCC).

The primary focus of the study was on the patient perspective. Provider interviews were used to triangulate themes arising in patient interviews, explore situations in which provider perspectives contrasted with patient perspectives, and fact-check medical concepts.

Settings

Participants were recruited from Weill Cornell Physicians (a multispecialty academic medical practice in Manhattan), New York-Presbyterian Hospital (the largest academic hospital in Manhattan), and the Institute for Family Health (a federally qualified health center serving New York City). We distributed promotional flyers at the three institutions, and also elicited referrals from physicians and nurse practitioners at outpatient clinics in internal medicine and endocrinology. Patient interviews were conducted in conference rooms or spare offices at the three locations, usually immediately before or after a clinical visit. Provider interviews were conducted in provider offices.

Interview Methods

We developed a semistructured interview instrument about PHIM (the focus of the current manuscript) as well as the related topic of personal health information tracking (reported elsewhere [35]). The interview guide included questions about: (1) how patients perceived their level of knowledge about their medical conditions, (2) times they had looked up or done research on health topics, (3) whether they tracked or logged information about their personal health or their medical care (probe questions asked about types of information such as medications, diet and exercise, personal medical data such as blood glucose, records of doctor's visits or of surgical procedures, etc), and (4) information or documents they typically brought to share with their doctor or nurse, including information that they brought when moving from one doctor or medical center to another. A follow-up probe question asked if they had ever looked at their medical chart as a Web portal, via a phone, or as a paper record. Interviews were conducted in person, audiorecorded, and professionally transcribed. The interviewer (JSA) also took field notes and photographed artifacts or documents such as log sheets used to record blood glucose values.

Analysis Methods

Qualitative analysis was conducted by our multidisciplinary team, which included members with training in journalism, public health, informatics, psychology, human factors, nursing, and diabetes education. Two of the researchers (HOW and EW) also brought personal experience of chronic disease or multiple chronic disease. Following Strauss and Corbin's grounded theory methods [36], we conducted open coding (allowing codes to emerge from the data), axial coding (identifying relationships), and selective coding (developing underlying themes and theory). Each transcript and photograph was reviewed by at least two researchers (the 1st author and one or more additional team members), who coded independently and then met to reach consensus. Interrater reliability was not calculated as coding was finalized during consensus meetings. A total of 47 open codes were developed, which were grouped via axial coding into six broader concepts before the final themes were identified.

Analysis was conducted concurrently with recruitment, which was halted when saturation was achieved (ie, no new concepts were arising from new interviews) [37]. Fewer providers were recruited than patients because provider perspectives proved more homogeneous in the analysis.

Member checking [38] was conducted by (1) discussing emergent concepts and themes with new informants, and by (2) presenting the final list of themes in a 90-minute session of the diabetes education group. Two members of the diabetes education group had previously participated in an individual interview as part of the study. During the member check, the themes appeared to resonate strongly with the participants, many of whom offered additional anecdotes and personal experiences. In the post-member check meeting, the researchers concluded that all of the new comments and anecdotes were congruent with the existing themes.

Ethics Approval

This study was approved by the Institutional Review Boards of Weill Cornell Medical College and the Institute for Family Health. All interview participants gave written informed consent. Members of the diabetes education group provided oral consent at each session that the researcher attended. During individual interviews, permission was asked to take photographs of patient artifacts that excluded identifying information; participants reviewed each photograph as it was taken and decided whether it would be deleted or saved.

Results

Participants

Interviews were conducted with 22 patients and 7 health care providers. Slightly more than half of the patients (13/22, 59%,) had a relationship with one of the providers who was interviewed, and the rest did not. Conversely, 4 of the 7 providers had patients who were included in the study.

Patients had an average of 3.5 (SD 1.5) chronic conditions, including type 2 diabetes, hypertension, heart disease, chronic pain, depression, asthma, HIV, and hepatitis C. Several reported taking anticoagulants, although they did not all explain what condition they had. Participants mentioned regular relationships with an average of 5 different providers, including primary care physicians or nurse practitioners; medical and surgical specialists; allied health providers including physical therapists, dietitians, and diabetes educators; pharmacists; and dentists or oral surgeons. In addition, many of the patients had had recent visits to an emergency department or urgent care center for urgent conditions, which included diverticulitis, flu, appendicitis, burns, and other physical injuries.

The patient sample was half men (11/22, 50%) and half women (11/22, 50%); 7 of the 22 patients (32%) were black. The average age was 64 years (range 37-89). Two-thirds (15/22, 68%) were not currently married; 8 of 22 (36%) used English as a second language. One-third (7/22, 32%) had Medicare (US public insurance for those over age 65), one-third (7/22, 32%) had Medicaid (US public insurance for low income individuals), and the remainder had commercial insurance (8/22, 36%).

The health care providers were 2 nurse practitioners, 2 internists, 2 family medicine physicians, and an emergency medicine physician (4 women and 3 men).

Major themes pertaining to PHIM are summarized in Table 1 and presented in detail below.

Table 1. Major themes in Personal Health Information Management.

Themes	Summary	Representative quotes
A. Responsibility for managing medical information across organizational settings	Some patients perceive medical records management as the health care system's responsibility, whereas others perceive it as their own.	"[The doctors] are supposed to have all the information. They're supposed to look it up."
B. What medical information should be shared?	Patients make frequent judgments about what data is relevant to their health and therefore should be shared or reported.	"The things that [the dermatologists] were doing really wasn't, you know, something that [my primary care doctor] needed to know."
C. Methods, tools, artifacts	Patients who took an active role in managing their records used electronic tools, paper, and memory	"I keep it in my head... I know the dosage, the day, for what is this medicine and how many times I [take it] daily."
D. Managing medical information as "invisible work"	Managing transfers of medical information to solve problems such as health insurance denials is a tremendous amount of work that largely goes unrecognized.	"It's hard enough when you're healthy and you're with it, and you're feeling good… When you're not feeling well at all, it's difficult."

Theme A: Responsibility For Sharing Medical Information Across Organizations

We found a range of opinions about who—patients, providers, or both—had primary responsibility for sharing medical information and records.

Patients' Responsibility

Many of the people with MCC felt strong responsibility for sharing their medical information and records across their networks of providers. "It's up to you [to keep track of that information], really," said one. Some of the patients with this perspective had developed their approach because of previous negative experiences in which important information from one provider had failed to reach another provider. These individuals often recorded or memorized their own information, brought documents from one provider to another, or requested transferal of lab results, records, and imaging studies from one provider to another. It was very common for patients to maintain a written, printed, or memorized medication list because they knew that they might be asked to provide it to a new doctor or to emergency room staff. Sometimes, the responsibility was assumed by a family member (often a female family member such as a patient's wife or an elderly patient's adult daughter).

Responsibility of the Medical System

However, other patients perceived medical information management to be primarily the responsibility of the health care system. "They're supposed to have all the information. They're supposed to look it up," said one individual, who seemed surprised to be asked about it. Patients sometimes expressed a preference for going to a hospital where "they know me" because of previous records. Two even mentioned the shared electronic health record as a reason why they sought primary and specialty care within the same institution. Even a few patients who did not themselves use computers knew that EHRs were being used to capture and share their information: one described his chart as being "on the terminal"' and another called it "the modern thing". Providers confirmed that some patients did not take a very active role in informing their providers about their other ongoing relationships with physicians or their previous records.

Confidence and Trust

Among the patients, beliefs about responsibility for information management appeared to be closely linked to feelings of confidence in doctors and health care organizations. One patient said it was important to ask questions and collect records because, "I believe you have to keep the doctors honest." Conversely, another patient, when asked whether he would be interested in accessing his medical record via the portal, said, "Why should I ask for it? I'm being seen on a regular basis for everything."

Providers' Pragmatism

Providers were unanimous that they needed easy access to their patients' information from other institutions in order to make the best decisions about their care. They were pragmatic, saying that any way of getting information was preferable to not having information. Methods of obtaining previous information and records about a patient included searching the institution's own records, interviewing the patient, interviewing the patient's family, asking the patient to bring copies of records from previous institutions, and calling other physicians and health care organizations to obtain oral reports or faxes. Only one provider reported having a patient who logged in to an electronic patient portal to retrieve information from a previous institution. Despite patients' confidence in computers, the providers recognized that data in electronic format was not necessarily shareable. The providers interviewed relied heavily upon EHRs within their organizations but also complained about lack of interoperability between different health care organizations and sometimes even different divisions within the same organization. It also was common for them to express frustration with patients who could not clearly report their own history. In some cases, providers suspected patients were trying to conceal information. “Whether it’s they don’t want us to contact [the previous doctor] or they really just don’t remember is an issue.”

Theme B: What Medical Information Should Be Shared?

Those patients who were instrumental in information sharing across physicians and health care organizations made judgments about what information was important to share with whom. These judgments were linked to their understanding of their medical conditions, their understanding of the health care system, and privacy issues.

Understanding of Disease

Patients generally wanted to share information relevant to their care. However, patients’ decisions about which information was relevant were influenced by concepts of health, disease, and relationships between diseases. For example, one patient judged that there was no need for information sharing between a dermatologist treating a scar and her other physicians: “[It] really wasn’t, you know, something that they needed to know.” Almost all the patients seemed familiar with the concept of medication interactions as the justification for providing their complete medication list to all of the physicians that they saw. Yet most, when asked, said they did not tell their medical providers about herbal treatments, dietary supplements, or dental visits, and many said dentists had never asked about their medical conditions.

Providers also talked about patients’ selective reporting of information, generally ascribing it to limited health literacy. For example, physicians talked about needing to instruct a patient to obtain previous laboratory results or medical records. According to one physician’s anecdote, a patient failed to report partial loss of vision in one eye while being examined for possible multiple sclerosis; the physician believed the omission was motivated by denial, but an alternative explanation is that the patient had no idea it might be relevant.

Privacy Concerns

In only a few cases, we encountered patients who were concerned that medical information would be used against them. A woman with a previous psychiatric diagnosis believed her history had been misused by ambulance personnel who “put my name in the computer” and diverted her to psychiatric care instead of the medical emergency care she was seeking. Another individual was concerned about how doctors interpreted the history of sexually transmitted infection in his medical record. One woman was strongly motivated to conceal her diabetes from her insurer because she was concerned the company would raise her premiums.

Understanding of the Health Care System

In addition, patients’ decisions about sharing medical information were shaped by their experience with and understanding of health care systems. Patients with diabetes who saw multiple health care providers generally learned that they would be asked about their hemoglobin A1c results by all of them. One woman explained why she knew to bring her medication list to a hospital appointment: “Well, I’ve been in the hospital before, or even another doctor’s appointment, ‘what medicines are you taking?’ And they always want you to fill it out again.” One woman explained that she didn’t think that dental information was relevant to doctors because she had never had a doctor ask about it. A large number of the patients recognized that their pharmacist was likely to check their medication list for potential interactions (or “clashes”, in the words of one woman). Misconceptions about the health care system could also play a large role in patient decision-making. One patient said that she did not need to bring x-rays from one hospital to another nearby one because the doctors could see each other’s computer systems. (A post-interview fact check showed this was not the case.) The woman (described above in the Privacy section) who was concerned about insurance rate increases believed her insurer learned about its patients through the billing history for medications, and was confident the insurer would not know she had diabetes as long as she continued to avoid the need for medications by controlling her diabetes through diet.

Physicians often recognized that the patient’s understanding of the health care system influenced the way that they shared medical records. Several reported that patients attempting to bring medical records to their doctor mistakenly brought hospital bills or even generic patient information printouts.

Theme C: Methods And Tools For Information Sharing

The patients who actively managed their own medical information had a variety of strategies for doing so, all of which were described by both patients and providers.

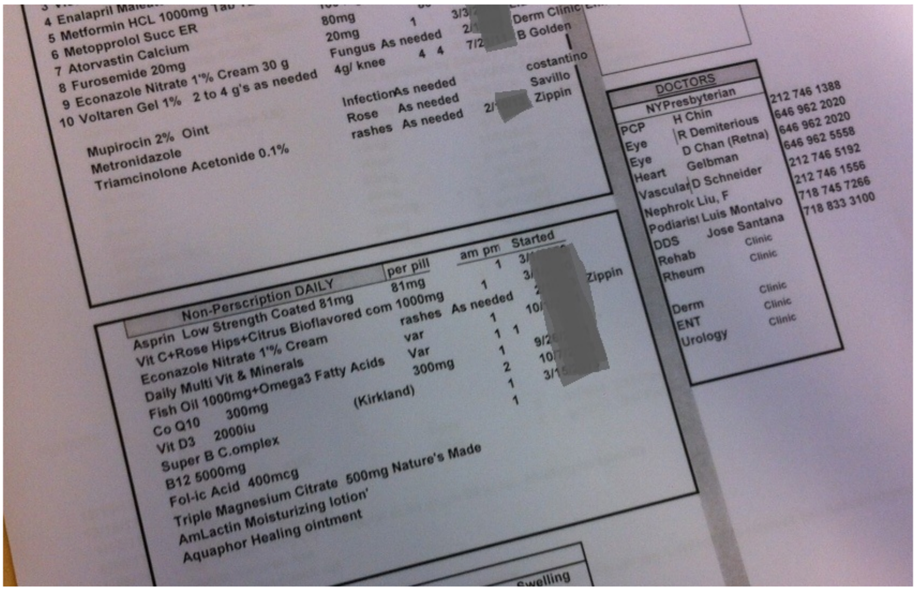
Memorization

Some kept track of lab values in their head. Most of the patients with diabetes were accustomed to giving an oral report to the primary care provider about their most recent podiatry, ophthalmology, and dental visits; only rarely were records from these visits transferred. Many informants were confident that they had memorized their medication lists. “I keep it in my head. I drink more than 20 medicines daily, and I know the dosage, the day, for what is this medicine and how many times I drink daily, how many dose.” (Many of the Spanish-speaking informants used the phrase “drink medicine”, even for pills and tablets, as the English translation of the Spanish expression “tomar la medicina”.)

Personal Electronic or Paper Records

Keeping or developing paper or electronic documents was less common. Some kept folders containing medical bills, reports, and test results. Two individuals created detailed spreadsheets of past medical history, current medications, physician and personal contact information, recent lab results, and other information which they regularly updated and carried with them (Figure 1). Others used handwritten lists (Figure 2). One who primarily relied on memorization of his medication list used a paper list as a backup. “Sometimes if you’re sick with... pain, the memory don’t work the same.”

Figure 1. Portion of a 3-page personally tracked record by an individual with multiple chronic diseases. This patient regularly updated the Excel spreadsheet with medications, dates of medical appointments and events, contact information, etc. Dates have been masked.



4 Enalapril Maleate
5 Metformin HCL 1000mg Tab
6 Metopprolol Succ ER 80mg 1 3/3[redacted] Derm Clinic
7 Atorvastin Calcium 20mg 2/[redacted] B Golden
8 Furosemide 20mg Fungus As needed 7/2[redacted]
9 Econazole Nitrate 1'% Cream 30 g 4 4
10 Voltaren Gel 1% 2 to 4 g's as needed 4g/ knee costantino

Infection As needed Savillo
Rose As needed 2/[redacted] Zippin
rashes As needed

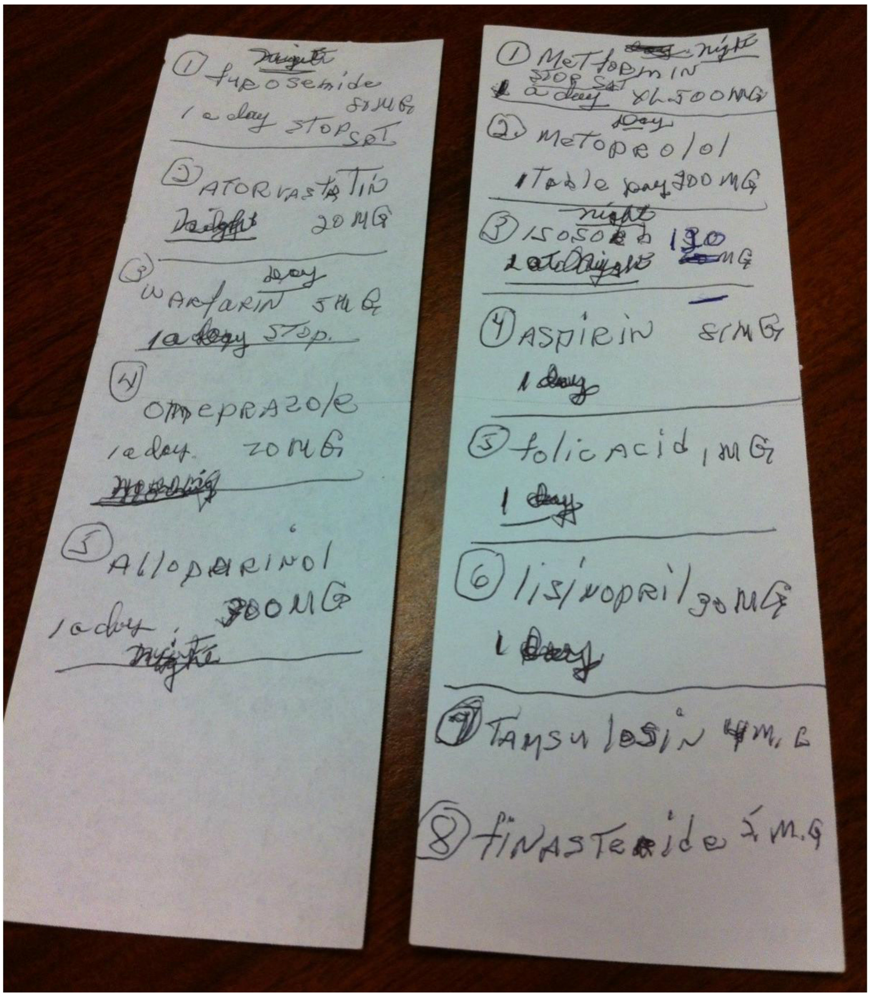
Mupirocin 2% Oint
Metronidazole
Triamcinolone Acetonide 0.1%

Non-Perscription DAILY	per pill	am	pm	Started	
Asprin Low Strength Coated 81mg	81mg		1	3/[redacted]	
Vit C+Rose Hips+Citrus Bioflavored com 1000mg		1		3/[redacted]	Zippin
Econazole Nitrate 1'% Cream	rashes	As needed		2[redacted]	
Daily Multi Vit & Minerals		1		10/[redacted]	
Fish Oil 1000mg+Omega3 Fatty Acids	var	1	1		
Co Q10 300mg	Var	1		9/2[redacted]	
Vit D3 2000iu	300mg	2		10/[redacted]	
Super B C.omplex	(Kirkland)	1		3/1[redacted]	
B12 5000mg		1			
Fol-ic Acid 400mcg					
Triple Magnesium Citrate 500mg Nature's Made					
AmLactin Moisturizing lotion'					
Aquaphor Healing ointment					

DOCTORS		
NYPresbyterian		212 746 1388
PCP	H Chin	646 962 2020
Eye	R Demiterious	646 962 2020
Eye	D Chan (Retna)	646 962 5558
Heart	Gelbman	212 746 5192
Vascular	D Schneider	212 746 1556
Nephrolc	Liu, F	718 745 7266
Podiarist	Luis Montalvo	718 833 3100
DDS	Jose Santana	
Rehab	Clinic	
Rheum	Clinic	
	Clinic	
Derm	Clinic	
ENT	Clinic	
Urology		

Swelling

Figure 2. Portion of medication list used by a patient to track 13 medications. Originally, he had designed his system so all daytime medications were on one sheet and all nighttime medications on the other, but as the medication regimen changed, he updated his notations. The patient kept the lists in a plastic grocery bag which he brought to medical appointments.



① ~~night~~
furosemide
1 a day STOP SAT 80MG

② ATORVASTATIN
~~Daylight~~ 20 MG

③ day
WARFARIN 5MG
1 a ~~day~~ STOP.

④ OMEPRAZOLE
1 a day. 20MG
~~morning~~

⑤ ALLOPURINOL
1 a day 300MG
~~night~~

① METFORMIN ~~day~~ night
STOP SAT
~~1~~ a day XL 500MG

② Day
METOPROLOL
1 Table pay 100 MG

③ night
ISOSORBID 120 MG
~~1 at night~~

④ ASPIRIN 81MG
1 ~~day~~

⑤ folic Acid 1 MG
1 ~~day~~

⑥ lisinopril 30 MG
1 ~~day~~

⑦ TAMSULOSIN 4 M.G

⑧ FINASTERIDE 5 M.G

Original Artifacts

Some patients kept track of information about themselves and their providers by saving objects provided by the health care system, including business cards and empty pill bottles with prescription labels.

Electronic Patient Portals

A very small number of patients had experience with patient portals. One had separate portal accounts with his outpatient physician, his hospital, and the Veterans Administration, and used them to help inform his doctors about what went on in the other health care systems (for example, requesting that a colonoscopy report be sent to his primary care physician). One patient used her account to print her medication list for a surgical admission. Although familiarity with computers was more common among younger patients, we observed cases of older, well-educated patients using computers easily and younger patients in less affluent circumstances being unfamiliar with computers.

Theme D: Managing Medical Information As “Invisible Work”

When patients tracked clinical data such as their own blood pressure, weight, blood glucose, or medication administration, their work was apparent and therefore visible to their doctors and nurses. By contrast, the work they performed to manage records or correct their information was generally invisible to their health care providers. This invisibility raised new challenges: patients found these tasks interfering with their regular illness work and felt they had nowhere to turn for assistance. “Nobody wants to help you,” said one. Medical providers were sometimes aware, in general terms, of the challenges these tasks posed to by their patients, but often found out about the problems long after the patient had already put in substantial effort. Real-time assistance, when it was available, came from people outside the health care system who had previous experience with some of these problems, such as family members or pastors.

The most common event triggering invisible PHIM work was an error in information. Two patients had similar stories about pharmacies substituting their prescription for an extended-release equivalent that their physician had previously decided was inappropriate for them. Other examples included erroneous information in the medical record discovered through the electronic patient portal, important laboratory results missing prior to surgery, delivery of home medical equipment without instructions for use, and mistaken denial of insurance coverage and errors about co-pays and deductibles.

These incidents launched patients into lengthy projects to find relevant information, often accompanied by a search for the correct party to whom to deliver it. Frequently, multiple attempts were needed to resolve the problem. For example, the man who was trying to correct information in his electronic patient portal account got referred from a technical support phone line to the doctor’s front desk staff to a technical support email address and then back again without getting the problem resolved. Another patient recounted: “So my primary doctor did an authorization for it, and I just got the letter that they denied me. Now they’re saying in the letter that was because she didn’t put enough information. So on the 12th I have to go to her with the letter and then I got to ask her what is the information that I need that has to comply with what they’re asking for.” One man seeking instructions for using his medical device said different offices had referred him to different places: “They keep giving you the runaround.”

Several patients developed preemptive procedures to deal with what they expected would be errors. One man said with exasperation that he double-checked the status of every lab result after a situation in which lab tests required prior to his surgery had been lost. “I would follow up… Every time. Not just once. Every time.” Another said he called his insurer before trying to fill any new prescription. “I say, ‘I’m getting this, this, this, this, and this medicine. You cover? You sure? [Give] me your name,’ I say, and I write down the name. And when I had a problem, I called to insurance and I say, ‘somebody with this name gave me this information. Why [with] you now it’s different?’” A third said he routinely stockpiled extra pills before getting a refill because he expected to encounter mistakes about his co-pay, and the extras would give him time to sort out the error before he ran out of medicine.

This work frequently felt frustrating, exhausting, and unfair. Many of our participants became angry when discussing it. Another said it made her “so tired,” and another began crying. One woman dealing with an insurance denial said, “Sometimes I would like to hear a human voice that will be able to reassure me and tell me this is what’s going on.” One woman trying to resolve a disagreement with a home health agency said, “I couldn’t sleep one single minute yesterday.”

One man explained why he had not followed up on a potential route to get insurance coverage for the shingles vaccine. “Who wants to go through all that? Who has the time and energy to continue the struggle, especially someone who is chronically trying to deal with everything else they’ve got to deal with? … It’s hard enough when you’re healthy and you’re with it, and you’re feeling good… When you’re not feeling well at all, it’s difficult. I don’t have the energy. I don’t have the time. I don’t feel good. I don’t want to deal with it.” The same man later said, “It’s hard to be on top of everything. I mean I’m not a computer. I’m a person, you know.”

Unfairness was a frequent theme. A woman seeking to renew a medication for hepatitis C said, “I shouldn’t have to be the one who straightens it out with the insurance company, because that’s their job.… It’s a lot of my time that I’d rather spend with other things.” The unfairness could be linked to the power difference between the insurer who had access to resources and information and the patient who did not. “They just send a [denial] letter and you’re stuck with the rest of the mess when you’re not even familiar with the plan.”

Discussion

Principal Findings

Patients with multiple chronic conditions have relationships with complex and changing networks of physicians and other care providers, pharmacies, allied health providers, and insurers. Providers consider it essential that information flows freely across institutional boundaries to help them take care of their patients. Some patients with MCC rely heavily on the health care system itself to maintain up-to-date records and make sure relevant information is accessible to health care professionals who might need it. Yet many other patients take an extremely active role in collecting, monitoring, and transferring their medical records across organizational settings. These patients use a variety of tools and methods to accomplish these tasks, ranging from memorization to requesting documents to using electronic patient portals.

We find that some patients were concerned about the privacy of their medical information because of ways it could be, or had been, used against them. However, privacy concerns came up relatively rarely in our interviews. Instead, patients' choices about what information to collect and share are strongly shaped by their understanding of health and disease and what information was relevant for specific medical conditions. Medical information is often left out because the patient did not see its relevance (for example, a patient who judged that her dermatology treatment was not relevant to her primary care). Previous work on mental models of disease [39-41] demonstrates that patients may create multiple internally coherent representations (or mental models) of the same disease, and these representations have varying degrees of similarity or difference from the biomedical model promoted by their physicians and nurses. We also find that patients' decisions about records management were also influenced by their understanding of how the health care system worked. For example, patients did not request records transfers between institutions if they thought doctors at each institution could access the other's records, and at least one was making medical decisions in part on the basis of whether they would reveal information to her insurer.

We also find that one of the biggest issues facing patients is the enormous amount of work involved in fixing errors, many of which arise from the complexities of seeking care across different institutions or, even more frequently, from complexities in health insurance. This work can be exhausting, upsetting, and frustrating, especially in light of the demands patients already face because of their illness work. Because this work is conducted outside of the relationship with any individual health care provider, it is often invisible to providers.

Limitations

Our sampling approach focused on English-speaking patients with multiple chronic conditions who were in regular medical care in a major urban area in the United States. The resulting sample was economically diverse but contained few advanced users of information technology. This may put some limits on generalizability to rural patients, people of other cultures, or more experienced users of information technology. Many of the informational challenges reported by our patients arose from negotiating the interface between health care organizations and health insurance companies, and results may not be fully generalizable to the patients of very different health care delivery models, such as integrated delivery systems in the United States or national health care systems in other countries. However, a member of our research team (HOW) found that these themes resonated with her similar experiences in two Canadian provinces. Attending the diabetes education group for ongoing triangulation and relying on the group for the member check could have made the final themes more representative of patients with diabetes than of patients with other chronic conditions. Our focus on information being used by patients in their interactions with the medical system also means that the types of medical information being discussed was probably narrower than the broader range described in some other PHIM literature [22,24,26].

Comparison With Prior Research

Our findings are highly congruent with perspectives from the sociology of illness. In their landmark 1985 work, Corbin and Strauss described the experience of being diagnosed with chronic disease as ushering in a series of new tasks and responsibilities as illness work [18,19]. Corbin and Strauss focused on activities such as following medication regimens and using home medical equipment, and others have since extended this concept to include the work of managing personal health information [18,19,22,27,28].

Our work contributes to a growing body of work on personal health information management or PHIM [22-25]. To date, much of this work has been performed with generally healthy individuals and families and in cancer. In extending this research to patients with multiple chronic disease, we found many similarities. For example, we found that individuals use both custom-made tools (such as electronic patient portals) as well as paper and pencil and a variety of other artifacts. Others have noted that patients use commercial calendars to track medical appointments, post medication checklists on refrigerator doors, or intermingle pediatric immunization records with memorabilia about the child's milestones [22,24,26].

However, in many other ways our patients provided a different perspective on PHIM. An earlier study of generally healthy individuals showed that many people rejected the idea that activities such as sharing health records or investigating medical options was *work*, and instead preferred terms such as management [30]. By contrast, one of the most striking findings from our interviews is that those with MCC see many of these activities as work. In particular, our patients had frequent experience with addressing informational errors within and across health care institutions, triggering tasks that were effortful, time-consuming, and emotionally draining, especially in light of the burdens of their existing illnesses. Unruh and Pratt [42] described very active patient work in detecting, preventing, and recovering from medical errors in outpatient cancer treatment. By contrast, our focus on outpatient chronic care meant that most of the errors described by our patients were information errors (such as the failure to transfer laboratory results to a surgeon, leading to the postponement of surgery). These sorts of information errors rarely involved medical errors, although they certainly appeared to have the potential to trigger medical errors.

We propose that these types of work fall in the category of *invisible work*. *Visible work*, such as care provided by doctors and nurses, is recognized, valued, and sometimes compensated. Other examples of *visible work* are informational tasks directly related to disease management that patients take on in collaboration with their providers, such as tracking blood pressure, blood glucose, or diet. By contrast, the concept of *invisible work* [29,30] describes necessary tasks that go unrecognized because they take place outside of the public sphere, require a degree of effort that may not be fully understood by others, or are conducted by people who are not seen as important [43]. Our patients describe tasks that are invisible because they take place almost entirely in the spaces between institutions, such as between health care providers and insurance companies. Health care providers may hear about this work from their patients, or even assist by (for example) providing letters to combat insurance denials. But they are not involved in the day-to-day "struggle" (in the words of one patient) and may hear about these challenges only after patients have already put in considerable time and effort trying to resolve them.

Consistent with what has been found previously by others, our informants were often selective in deciding when and with whom to share medical information [44]. Privacy concerns, although a factor, did not appear to be the most important issue raised. Instead, we find that many patients base these judgments on their own understanding of both health care organizations and disease processes, which may not coincide with their providers' views.

Although very few of our participants had ever used electronic patient portals, our findings are highly relevant to this rapidly evolving field. Portals are being offered by more and more health care organizations seeking to comply with the "meaningful use" regulations. California Health care Foundation/National Partnership for Women and Families find sharply increasing interest in and access to electronic health records via patient portals, with an estimated half of Americans having access to their electronic records via portals as of 2014 [45,46]. Patients newly exposed to the concept of the patient portal may be enthusiastic about its potential [47]. Those who already have portal access report that it helps them share data with health care providers, find and correct errors in medical records, and avoid having to fill out the same forms repeatedly, which were some of the common tasks mentioned by patients in our study [45,46]. Experiments with sharing the entire medical record (including often-hidden elements such as notes) have had positive results, with many patients feeling empowered and better informed about their care [48,49]. Nevertheless, some patients have reported negative feelings about seeing their electronic records, such as their diagnosis information or their lab results [50]. To date, few high-quality controlled studies have been conducted to assess the effects of portal-based interventions, and relatively few of these have reported positive findings on patient outcomes [15].

Conclusions and Implications

Because of their complex medical situations, patients with multiple chronic conditions maintain relationships with multiple health care providers, usually spanning several medical institutions. Providers, and many patients, recognize the need for easy information flow across these medical settings. In the absence of seamless health information exchange processes or technologies, patients and providers use a wide variety of workaround approaches, sharing information through combinations of memorization, paper, fax, and electronic tools. Privacy concerns were not universal, but a minority of patients had serious concerns about the ways their medical information might be used. Perhaps more importantly, patients made decisions on the basis of their mental model of their health and disease, providing information that they believed was relevant and omitting other facts that they considered irrelevant. Patients also made decisions on the basis of their understanding of the health care system, including the way that insurance works. To the extent that their mental models fail to coincide with their health care providers' models, this may lead to significant information gaps or suboptimal decisions. One of the biggest issues facing patients is the enormous amount of difficult, frustrating, and emotionally tiring work involved in addressing informational errors. Because this work is conducted outside of the relationship with any individual health care provider, it is often invisible to their health care providers. Furthermore, because this invisible work arises from complexities in medical care and medical coverage, it seems likely to fall most heavily on those with the most encounters with the medical system, constituting a systemically regressive tax on illness.

Effective structural solutions for information sharing are likely to not only improve the quality of information shared but also reduce the burden on patients already weighed down by MCC. Types of information technologies that might help resolve these problems include health information exchange (HIE) and personal health records (PHRs). Traditionally, HIE technologies are provider-centered, allowing doctors to look up communitywide data on their patient or push an individual patient record to a fellow physician [7,51,52]. By contrast, PHRs are designed for patients to keep and manage their own medical information in electronic form, accessible on the Web or mobile devices [53-55]. In the United States, the "meaningful use" regulations [9] are promoting adoption of "tethered" PHRs, which allow patients to view or export their medical records from a single institution [12]. Tethered PHRs are now routinely offered by many US health insurance companies as well, giving patients access to their insurance claims information and supporting patient education materials. By contrast, "untethered" PHRs give patients full control over collecting, tracking, annotating, and sharing data from multiple institutions or information of their own [56]. Examples include Microsoft HealthVault and Google Health (discontinued in 2011 for lack of adoption.)

Our findings suggest that both provider-centered and patient-centered information technologies will continue to be needed. Each has the potential to support patients in many of the most problematic aspects of health information management, but neither is likely to resolve all problems. Electronic PHRs are likely to hold the most appeal for patients who already take active roles in collecting, managing, and sharing medical information across their fragmented networks of care. These tools are gaining traction [14,21], and yet many still pose barriers related to less than optimal usability, lack of patient-centeredness in both vocabulary and functionality, and lack of integration with devices [57,58]. Furthermore, tethered PHRs offer access only to one institution's data. They can be enormously helpful for exporting medical records or findings such as lab results, but this alone cannot address all of the between-institution informational gaps that arose in our interviews. In addition, it is critical to acknowledge that, as we found, many patients do not take an active role in managing their own information and even those patients who did manage their information sharing tended to choose what to share based upon lay mental models of health and health care. We and others have also previously found that patients are broadly supportive of provider-facing HIE technologies [20,59]. For all these reasons, it seems likely that patient-centered and provider-centered approaches should be considered complementary, fulfilling different functions for different stakeholder groups. Ultimately, an ideal health information management technology would allow patient data to flow easily across organizational boundaries and also be fully accessible to that subset of patients who wish to view or manage their data [60].

Acknowledgments

JSA is supported by the Agency for Healthcare Research and Quality (K01 HS021531). HOW is supported by a Research Scholar Junior 1 Career Development Award by the Fonds de Recherche du Québec - Santé. We thank the interview participants for their candor and support of this project.

Conflicts of Interest

None declared.

References

1. Multiple chronic conditions - a strategic framework: Optimum health and quality of life for individuals with multiple chronic conditions. Washington, DC: US Department of Health and Human Services; 2010. URL: http://www.hhs.gov/ash/initiatives/mcc/mcc_framework.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqLxpuse]
2. Smith PC, Araya-Guerra R, Bublitz C, Parnes B, Dickinson LM, Van VR, et al. Missing clinical information during primary care visits. JAMA 2005 Feb 2;293(5):565-571. [doi: 10.1001/jama.293.5.565] [Medline: 15687311]
3. Coleman EA. Falling through the cracks: challenges and opportunities for improving transitional care for persons with continuous complex care needs. J Am Geriatr Soc 2003 Apr;51(4):549-555. [Medline: 12657078]
4. Coleman EA, Berenson RA. Lost in transition: challenges and opportunities for improving the quality of transitional care. Ann Intern Med 2004 Oct 5;141(7):533-536. [Medline: 15466770]
5. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
6. Cwinn MA, Forster AJ, Cwinn AA, Hebert G, Calder L, Stiell IG. Prevalence of information gaps for seniors transferred from nursing homes to the emergency department. CJEM 2009 Sep;11(5):462-471 [FREE Full text] [Medline: 19788791]
7. Kuperman GJ. Health-information exchange: why are we doing it, and what are we doing? J Am Med Inform Assoc 2011;18(5):678-682 [FREE Full text] [doi: 10.1136/amiajnl-2010-000021] [Medline: 21676940]
8. National eHealth Collaborative. Health information exchange roadmap: The landscape and the path forward. Washington, DC: Health Information Management Systems Society; 2012 Apr. URL: http://www.himss.org/nehc [accessed 2015-05-27] [WebCite Cache ID 6YqMrWvsV]
9. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
10. Cimino JJ, Frisse ME, Halamka J, Sweeney L, Yasnoff W. Consumer-mediated health information exchanges: the 2012 ACMI debate. J Biomed Inform 2014 Apr;48:5-15 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.009] [Medline: 24561078]
11. The Markle Foundation. Attitudes of Americans regarding personal health records and nationwide electronic health information exchange. 2005. URL: http://www.markle.org/publications/951-attitudes-americans-regarding-personal-health-records-and-nationwide-electronic-hea [accessed 2015-05-27] [WebCite Cache ID 6YqNXogO8]
12. Conn J. Blue Button gains fans, apps. Simple tech from VA puts interoperability to work. Mod Healthc 2012 Oct 15;42(42):14. [Medline: 23163087]
13. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
14. Ancker JS, Silver M, Kaushal R. Rapid growth in use of personal health records in New York, 2012-2013. J Gen Intern Med 2014 Jun;29(6):850-854. [doi: 10.1007/s11606-014-2792-2] [Medline: 24519102]
15. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The Impact of Electronic Patient Portals on Patient Care: A Systematic Review of Controlled Trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238]
16. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
17. Patel VN, Dhopeshwarkar RV, Edwards A, Barrón Y, Sparenborg J, Kaushal R. Consumer support for health information exchange and personal health records: a regional health information organization survey. J Med Syst 2012 Jun;36(3):1043-1052. [doi: 10.1007/s10916-010-9566-0] [Medline: 20703633]

18. Dimitropoulos L, Patel V, Scheffler SA, Posnack S. Public attitudes toward health information exchange: perceived benefits and concerns. Am J Manag Care 2011 Dec;17(12 Spec No):SP111-SP116 [FREE Full text] [Medline: 22216769]
19. Patel VN, Dhopeshwarkar RV, Edwards A, Barron Y, Likourezos A, Burd L, et al. Low-income, ethnically diverse consumers' perspective on health information exchange and personal health records. Inform Health Soc Care 2011 Dec;36(4):233-252. [doi: 10.3109/17538157.2011.554930] [Medline: 21851182]
20. Ancker JS, Edwards AM, Miller MC, Kaushal R. Consumer perceptions of electronic health information exchange. Am J Prev Med 2012 Jul;43(1):76-80. [doi: 10.1016/j.amepre.2012.02.027] [Medline: 22704751]
21. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
22. Moen A, Brennan PF. Health@Home: the work of health information management in the household (HIMH): implications for consumer health informatics (CHI) innovations. J Am Med Inform Assoc 2005;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.M1758] [Medline: 16049230]
23. Civan A, Skeels M, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
24. Zayas-Cabán T. Health information management in the home: a human factors assessment. Work 2012;41(3):315-328. [doi: 10.3233/WOR-2012-1306] [Medline: 22398501]
25. Agarwal R, Khuntia J. Personal health information in the design of consumer health information technology: Background report. (Prepared by Insight Policy Research under contract number HHSA 290200710072T). Rockville, MD: Agency for Healthcare Research and Quality; 2009. URL: http://healthit.ahrq.gov/sites/default/files/docs/citation/09-0075-EF.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqOEZdjd]
26. Piras E, Zanutto A. Prescriptions, X-rays and Grocery Lists. Designing a Personal Health Record to Support (The Invisible Work Of) Health Information Management in the Household. Comput Supported Coop Work 2010 Nov 3;19(6):585-613. [doi: 10.1007/s10606-010-9128-5]
27. Corbin J, Strauss A. Managing chronic illness at home: Three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
28. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10. [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
29. Strauss AL, Fagerhaugh S, Suczek B, Wiener C. The work of hospitalized patients. Soc Sci Med 1982;16(9):977-986. [Medline: 7112176]
30. MacGregor JCD, Wathen CN. 'My health is not a job': a qualitative exploration of personal health management and imperatives of the 'new public health'. BMC Public Health 2014;14:726 [FREE Full text] [doi: 10.1186/1471-2458-14-726] [Medline: 25030501]
31. Unruh KT, Pratt W. The invisible work of being a patient and implications for health care: "[the doctor is] my business partner in the most important business in my life, staying alive". 2008 Presented at: Ethnographic Praxis in Industry Conference; November 2008; San Jose, CA p. 40-50. [doi: 10.1111/j.1559-8918.2008.tb00093.x]
32. Unruh KT, Pratt W. Barriers to organizing information during cancer care: "I don't know how people do it.". AMIA Annu Symp Proc 2008:742-746 [FREE Full text] [Medline: 18999223]
33. Klasnja P, Hartzler AC, Unruh KT, Pratt W. Blowing in the Wind: Unanchored Patient Information Work during Cancer Care. CHI Conf Proc 2010:193-202 [FREE Full text] [doi: 10.1145/1753326.1753355] [Medline: 20697453]
34. Unruh KT, Skeels M, Civan-Hartzler A, Pratt W. Transforming Clinic Environments into Information Workspaces for Patients. CHI Conf Proc 2010:183-192 [FREE Full text] [doi: 10.1145/1753326.1753354] [Medline: 21654895]
35. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. Challenges to personal data tracking among patients with multiple chronic conditions: "You get reminded you"re a sick person". Under review 2015:1-2 (forthcoming).
36. Corbin JM, Strauss AL. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications, Inc; 2007.
37. Morse J. The Significance of Saturation. Qualitative Health Research 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]
38. Morse J, Barrett M, Mayan M, Olson K, Speirs J. Verification strategies for establishing reliability and validity in qualitative research. International Journal of Qualitative Methods 2002;1(2):1-19 [FREE Full text]
39. Meyer D, Leventhal H, Gutmann M. Common-sense models of illness: the example of hypertension. Health Psychol 1985;4(2):115-135. [Medline: 4018002]
40. Leventhal H, Leventhal EA, Breland JY. Cognitive science speaks to the "common-sense" of chronic illness management. Ann Behav Med 2011 Apr;41(2):152-163. [doi: 10.1007/s12160-010-9246-9] [Medline: 21136224]
41. Grzywacz JG, Arcury TA, Ip EH, Chapman C, Kirk JK, Bell RA, et al. Older adults' common sense models of diabetes. Am J Health Behav 2011 May;35(3):318-333 [FREE Full text] [Medline: 21683021]
42. Unruh KT, Pratt W. Patients as actors: the patient's role in detecting, preventing, and recovering from medical errors. Int J Med Inform 2007 Jun;76 Suppl 1:S236-S244. [doi: 10.1016/j.ijmedinf.2006.05.021] [Medline: 16829180]

43. Daniels AK. Invisible Work. Social Problems 1987 Dec;34(5):403-415. [doi: 10.2307/800538]
44. Rassin M, Zilcha L, Berger M, Silner D. Personal medical documents management--how patients perceive, keep and manage their medical documents: a qualitative study. Int J Nurs Stud 2007 Aug;44(6):862-868. [doi: 10.1016/j.ijnurstu.2006.03.005] [Medline: 16678180]
45. Making IT meaningful: How consumers value and trust health IT. Washington, DC: National Partnership for Women & Families; 2012. URL: http://go.nationalpartnership.org/site/DocServer/HIT_Making_IT_Meaningful_National_Partnership_February_2.pdf?docID=9783 [accessed 2015-05-27] [WebCite Cache ID 6YqSBLhFF]
46. Engaging patients and families: How consumers value and use health IT. Washington DC: National Partnership for Women & Families URL: http://www.nationalpartnership.org/research-library/health-care/HIT/engaging-patients-and-families.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqSFCD2e]
47. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
48. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
49. Leveille SG, Walker J, Ralston JD, Ross SE, Elmore JG, Delbanco T. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. BMC Med Inform Decis Mak 2012;12:32 [FREE Full text] [doi: 10.1186/1472-6947-12-32] [Medline: 22500560]
50. Wright A, Feblowitz J, Maloney FL, Henkin S, Ramelson H, Feltman J, et al. Increasing patient engagement: patients' responses to viewing problem lists online. Appl Clin Inform 2014;5(4):930-942. [doi: 10.4338/ACI-2014-07-RA-0057] [Medline: 25589908]
51. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010;17(3):288-294 [FREE Full text] [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
52. Campion TR, Ancker JS, Edwards A, Patel V, Kaushal R. Push and pull: physician usage of and satisfaction with health information exchange. AMIA Annu Symp Proc 2012;2012:77-84 [FREE Full text] [Medline: 23304275]
53. Steinbrook R. Personally controlled online health data--the next big thing in medical care? N Engl J Med 2008 Apr 17;358(16):1653-1656. [doi: 10.1056/NEJMp0801736] [Medline: 18420496]
54. Kaelber DC, Shah S, Vincent A, Pan E, Hook J, Johnston D, et al. The Value of Personal Health Records. Charlestown, MA: Center for Information Technology Leadership, Healthcare Information and Management System Society (HIMSS); 2008.
55. Connecting for Health: A Public-Private Collaborative.: The Markle Foundation; 2003. URL: http://www.markle.org/health/connecting-health [accessed 2015-05-30] [WebCite Cache ID 6YvSTAjfH]
56. Lohr S. The New York Times. 2011 Jun. Google to end health record service after it fails to attract users URL: http://www.nytimes.com/2011/06/25/technology/25health.html?_r=1 [accessed 2015-05-27] [WebCite Cache ID 6YqSrickq]
57. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc 2009;16(5):660-669 [FREE Full text] [doi: 10.1197/jamia.M3154] [Medline: 19567793]
58. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]
59. O'Donnell HC, Patel V, Kern LM, Barrón Y, Teixeira P, Dhopeshwarkar R, et al. Healthcare consumers' attitudes towards physician and personal use of health information exchange. J Gen Intern Med 2011 Sep;26(9):1019-1026 [FREE Full text] [doi: 10.1007/s11606-011-1733-6] [Medline: 21584839]
60. Ancker JS, Miller M, Patel V, Kaushal R. Sociotechnical challenges to developing technologies for patient access to health information exchange data. J Am Med Inform Assoc 2014;21(4):664-670. [doi: 10.1136/amiajnl-2013-002073] [Medline: 24064443]

Abbreviations

HIE: health information exchange
HIT: health information technology
MCC: multiple chronic conditions
PHIM: personal health information management
PHR: personal health record

Edited by G Eysenbach; submitted 26.02.15; peer-reviewed by A Moen, A Civan; comments to author 18.03.15; revised version received 11.04.15; accepted 17.04.15; published 04.06.15.

Please cite as:
Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E
The Invisible Work of Personal Health Information Management Among People With Multiple Chronic Conditions: Qualitative Interview Study Among Patients and Providers
J Med Internet Res 2015;17(6):e137
URL: http://www.jmir.org/2015/6/e137/
doi:10.2196/jmir.4381
PMID:26043709

Original Paper

Informing the Design of Direct-to-Consumer Interactive Personal Genomics Reports

Orit Shaer[1], PhD; Oded Nov[2], PhD; Johanna Okerlund[1]; Martina Balestra[2], MSc; Elizabeth Stowell[1]; Laura Ascher[1]; Joanna Bi[1]; Claire Schlenker[1]; Madeleine Ball[3], PhD

[1]Human-Computer Interaction Lab, Computer Science Department, Wellesley College, Wellesley, MA, United States
[2]Department of Technology Management and Innovation, School of Engineering, New York University, New York, NY, United States
[3]PersonalGenomes.org, Boston, MA, United States

Corresponding Author:
Oded Nov, PhD
Department of Technology Management and Innovation
School of Engineering
New York University
5 Metrotech Center
Brooklyn, New York
New York, NY, 11201
United States
Phone: 1 6462077864
Fax: 1 7182603874
Email: onov@nyu.edu

Abstract

Background: In recent years, people who sought direct-to-consumer genetic testing services have been increasingly confronted with an unprecedented amount of personal genomic information, which influences their decisions, emotional state, and well-being. However, these users of direct-to-consumer genetic services, who vary in their education and interests, frequently have little relevant experience or tools for understanding, reasoning about, and interacting with their personal genomic data. Online interactive techniques can play a central role in making personal genomic data useful for these users.

Objective: We sought to (1) identify the needs of diverse users as they make sense of their personal genomic data, (2) consequently develop effective interactive visualizations of genomic trait data to address these users' needs, and (3) evaluate the effectiveness of the developed visualizations in facilitating comprehension.

Methods: The first two user studies, conducted with 63 volunteers in the Personal Genome Project and with 36 personal genomic users who participated in a design workshop, respectively, employed surveys and interviews to identify the needs and expectations of diverse users. Building on the two initial studies, the third study was conducted with 730 Amazon Mechanical Turk users and employed a controlled experimental design to examine the effectiveness of different design interventions on user comprehension.

Results: The first two studies identified searching, comparing, sharing, and organizing data as fundamental to users' understanding of personal genomic data. The third study demonstrated that interactive and visual design interventions could improve the understandability of personal genomic reports for consumers. In particular, results showed that a new interactive bubble chart visualization designed for the study resulted in the highest comprehension scores, as well as the highest perceived comprehension scores. These scores were significantly higher than scores received using the industry standard tabular reports currently used for communicating personal genomic information.

Conclusions: Drawing on multiple research methods and populations, the findings of the studies reported in this paper offer deep understanding of users' needs and practices, and demonstrate that interactive online design interventions can improve the understandability of personal genomic reports for consumers. We discuss implications for designers and researchers.

(J Med Internet Res 2015;17(6):e146) doi:10.2196/jmir.4415

KEYWORDS

genomics; genetic testing; interactive visualizations; personal electronic health records; direct-to-consumer genetic testing

Introduction

Overview

Recent years are seeing a dramatic growth in the availability of personal genomic data to end users. Consumers with varying levels of relevant education who seek genomic testing services are confronted with an unprecedented amount of sensitive information about themselves [1], often online and in interactive forms [2]. These consumers are not necessarily experts in genetics. They span the gamut from curious or concerned laypeople, to educated early adopters, to experts in genetics. Although genetic testing is available to these diverse populations, the resulting data reports can be difficult to understand without specialized training. Furthermore, the inherent complexity of genomic data is compounded by the frequency with which research in genetics is updated.

Consequently, questions about how consumers understand and engage with their personal genomic information are not only of paramount importance for society and policy makers, but are also a pressing issue for human-computer interaction (HCI) researchers. Specifically, the highly personal and dynamic nature of personal genomic information raises the following questions: What are the functional requirements for supporting meaningful engagement of consumers with varying levels of relevant knowledge with personal genomic information? How can we design effective interaction with personal genomic information? How can we evaluate the effectiveness of interactions with personal genomic information? Addressing these questions, this paper explores the roles HCI can play in helping consumers understand and engage with personal genomics.

We present findings from three complementing research activities:

1. Study 1: Understanding Users. Study 1 consists of a qualitative study with early adopters to understand users' motivations, needs, and information practices when engaging with their personal genomic information.

2. Study 2: Informing Users. Study 2 involves a design workshop with early adopters in which the current state-of-the-art genomic reports are evaluated and various existing and possible features for interactive reports are explored.

3. Study 3: Probing Users. Study 3 designs and tests alternative interactive reports informed by the needs and practices identified in the prior qualitative studies. The designs, using different visualizations, were tested using online experiments with Amazon Mechanical Turk users to investigate how variations in interface design and data visualization affect users' understanding of, as well as preference and attitude toward, online personal genomic reports.

Taken together, these studies contribute toward understanding and improving the ways people engage with and understand personal genomics information.

Background

Personal Genomics

The Human Genome Project (HGP) published the full reference sequence of the human genome in April 2003. This international, collaborative research program, whose goal was the complete mapping and understanding of all the human genes, lasted 13 years and cost US $2.7 billion. The HGP DNA sequence is a composite derived from the DNA of several anonymous volunteers. The first individual's genome was sequenced in 2007. Since then, many more individuals have had their genome, or part of it, sequenced anonymously for research, but until June 2013, only about 500 individuals had ever had their full results returned to them [3]. The cost of sequencing a single human genome has dropped from US $2.7 billion in 2003 to about US $5000 in 2013, a cost drop far faster than the rate of Moore's law [3]. Decreasing sequencing costs and technological advances offer the promise of personalized medicine to the masses, with genomic information integrated into medical care to provide individualized risk assessment, tailored lifestyle change recommendations, and medications to reduce risk [4].

Online Interaction With Personal Genomics

The precipitous decline in the costs of DNA sequencing has led to widespread access of personal genomic data. An increasing number of large-scale efforts, representing millions of people combined, are already underway. For example, the government of England recently announced their plan to sequence and return whole personal genomes to 100,000 British citizens by 2017. In the United States, the Veterans Administration is pursuing an effort that aims to enroll 1 million veterans in a research study that incorporates genetic profiling.

At the other end of the spectrum from large centralized efforts, several companies currently offer services directly to consumers. For example, Illumina provides consumers (with prescription) genome sequencing services. Direct-to-consumer genetic testing (DTCGT) is a relatively new and developing online service, which enables individuals to acquire genetic information without the mandatory involvement of a health care provider by sending a saliva sample to a DTCGT company, at the cost of a few hundred dollars. To date, DTCGT does not typically offer whole genome or exome sequencing, rather, these tests use the single nucleotide polymorphism (SNP)-chip technique, which looks at thousands of very short sections of DNA known to vary across populations [1]. Results are delivered through online interactive reports. Several popular DTCGT services additionally offer interactive online reports of nonhealth-related information including traits and ancestry information (eg, AncestryDNA [5] and Family Tree DNA [6]). The service 23andMe [7] also provided risk assessment results for about 250 conditions, however, as of December 2013 the reporting of health-related information directly to consumers has been stopped while it is undergoing US Food and Drug Administration (FDA) review, which seeks to determine whether test results are accurate and are adequately communicated to, and understood by, consumers [8].

Traditionally, medical genetic testing targets individual loci and is performed for specific medical contexts (eg, when investigating a suspected genetic condition). Results are returned in a verbal process, mediated by a medical expert. The decreased cost of genome-scale tests combined with their application to an increasingly broad scope of individuals means the number of possible genetic test results has become several orders of magnitude larger than traditional context. While an expert advisor may remain as a part of the initial communication of data results, it becomes infeasible to present results in the same verbal manner. Thus, interactive computer-mediated presentation of this data to individuals has become a core aspect of giving individuals access to their genome-scale test results. For example, Illumina's genome sequencing service provides initial genetic counseling upon the return of results, but data has also been returned to consumers digitally on an Apple iPad using the Illumina MyGenome app, which allows users to browse their genome, compare it to a reference genome, and review a health report that provides risk assessment for about 250 conditions.

Additionally, individuals will increasingly have ongoing access to extensive genetic test data. In the United States, patients now have a legal right to directly access clinical test data [9]. In addition, to date, all of the DTCGT services mentioned above also return raw genotyping data to users, who in turn can actively engage with their personal genomic data, for example, by learning about specific gene variants or conditions of interest. Indeed, consumers of genomic data have been observed transporting their data between services to capitalize on different features that allow them to engage more deeply with their data. For example, 23andMe users may export their data to AncestryDNA for genealogy, or to the Personal Genome Project (PGP) database—discussed in the proceeding section—to share with people of interest. Because this data is inherently digital, and because its interpretation gets updated frequently based on new research findings, we anticipate increased focus on the development of online interactive report methods that perform automatic reanalysis.

In summary, given recent advances in the field of personal genomics and rapidly declining sequencing costs, it seems inevitable that there will be vastly increased demand for individuals understanding their own genome-scale data and its health implications. The personal and complex nature of personal genomic information and users' interaction with it raise important HCI questions.

Personal Genome Project

The Personal Genome Project [10] is a nonprofit organization that seeks to improve the scientific understanding of genetic and environmental contributions to human traits through the creation of a public genetic database of 100,000 volunteers [11-13]. Participants must be willing to share their genomic sequences, as well as health data, with the scientific community and the general public. The organization consists of sites spanning four countries. The longest running PGP site is based out of George Church's Lab at Harvard Medical School. The Harvard PGP was established in 2005. It began with a pilot study of 10 fully identified individuals, known as the PGP-10, and slowly scaled up. Today, more than 4000 US citizens are enrolled in the project through a process of "open consent" [14] to publicly share their genomic information. We established a design partnership with the Personal Genome Project and are collaborating closely with its researchers.

User Perspectives on Personal Genomics

Little empirical data exists about the attitudes and motivations of people who have their genome sequenced and interact with their data [15]. Only a few studies have recruited DTCGT consumers who had actually received their own personal genomic information. In these studies, curiosity was mentioned as the participants' primary motivation for undergoing genomic testing [15]. Most respondents wanted to learn more about themselves, were curious about their genetic makeup, or wanted to learn about individual genetic risk factors. Participants also stated that they would use information gained from the test to take personal responsibility for their future health [16]. Other themes included fascination with genealogy, contribution to research, and recreation [15]. Studies also identified several concerns among DTCGT users, including privacy, as well as the nature of the results and their future impact [17-20]. Only a small number of users around the world have had their entire genome sequenced and returned to them—500 as of June 2013 [3]—and to our knowledge, no studies have investigated the perspectives of such users. Further research is needed to understand personal genomic users' motivations and concerns, information needs and practices, and the factors that impact willingness to share information.

Related Work: Human-Computer Interaction for Genomics

To date, little HCI research has focused on direct user engagement with personal genomic information. Lachance et al [3] examined the features of websites in which consumers can directly purchase and receive genetic testing without the mandatory involvement of a health care provider. Their findings indicate that most users would struggle to find and understand the important information on the majority of sites. Other efforts have considered user engagement with genomic and biological information more broadly, focusing mainly on novel interaction techniques for large biological datasets. For example, Shaer et al [21] have discussed opportunities and challenges for applying tangible and embodied interaction for discovery and learning of genomics. Kuznetsov et al [22] described a possible role for HCI in supporting the growing community of do-it-yourself biology (DIYbio) citizen scientists. Schkolne et al [23] developed an immersive tangible interface for supporting scientists in the design of new DNA molecules. Also, several tabletop systems have also been developed to explore interactive visualization of large biological datasets—DeepTree [24] and PhyloGenie [25] allow users to explore and learn phylogenetic trees. Most closely related to our work is G-nome Surfer [26], a tabletop user interface for collaborative exploration and learning of genomic information. This tool was not, however, designed to support consumers as they explore their own personal genomic data.

Methods

Study 1: Understanding Users

To gain insight into the information needs and practices of consumers interested in directly engaging with their genomic information, we conducted an exploratory qualitative study [27]. We recruited 63 study participants (29 women, 46%), aged between 21 and 71 with an average age of 47 (SD 14) from the Personal Genome Project volunteer community. This population of early adopters consists of users of various genetic testing services, who already spent time working with different tools available to explore their data, thus allowing us to understand existing information practices and needs of consumers, who use a range of genetic testing services. The interactions between these early adopters and their data provide a strong basis for exploring future data visualizations that appeal to a more diverse population.

Participants completed an online questionnaire consisting of 10 open-ended questions (see Table 1) about their engagement with personal genomics services and data. Response length averaged 252 words per user. We analyzed the data using content analysis methods. First-level codes were developed from preliminary review of the data by two independent coders and were then collapsed into advanced categories based on frequency. Categories were analyzed for the identification of themes. From this, we reported results regarding users' information practices and needs.

Table 1. Open-ended questions from Study 1.

Question number	Questions from online questionnaire
1	What are the main reasons for your interest in exploring your personal genomic information?
2	What impact did your discoveries have on your life and attitude toward your health? Was there anything that you did, started doing, or stopped doing as a result of getting your personal genomic information?
3	What new or unexpected things did you learn as a result of genetic testing?
4	Did your discoveries lead you to social or formal interactions with other people and if so, who? For example, did you discuss your results with health professionals, family members, scientists, or support groups?
5	What websites and computational tools did you use for engaging with your personal (or your family's) genomic information? How did you use these tools to learn from your data?
6	What features or applications could help you manage and learn even more from your (or your family's) personal genomic data?
7	What are the main reasons for your decision to share your personal genomic information on PGP[a]?
8	What were valuable aspects of your experience exploring and sharing your personal genomic information?
9	What concerns do you have regarding exploring and sharing your personal genomic information?
10	Is there anything else you think we should ask you about your experience of engaging with your genomic data?

[a] Personal Genome Project (PGP).

Study 2: Informing Users

In order to gain further insight into how users engage with, and learn from, their annotated personal genomic reports we conducted a qualitative study of personal genomics users. Participants were once again recruited from the PGP volunteer community. This population was chosen specifically because of their deep understanding of the data and tools available, and because they are likely to be first adopters of any new tool for personal genomics. This study was held as a workshop, which took place during the Genomes, Environments, and Traits (GET) conference, organized by the Personal Genome Project in Cambridge, Massachusetts, in April 2014.

The study focused on interactions around a specific genome reporting tool, GET-Evidence [12], which is an interactive personal genomic report provided to all PGP volunteers. We chose to study this particular tool since it is one of the most comprehensive gene variant reports available for consumers. Other direct-to-consumer genetic testing providers return information to users related to their traits and ancestry, but not a health-related report. The service 23andMe provided risk assessment results for about 250 conditions up until December 2013, when they suspended reporting of health-related interpretations while it is undergoing FDA review, which examines whether test results are accurate and are adequately communicated to, and understood by, consumers [8]. All of the direct-to-consumer genetic testing services also return raw genotyping data to users, which can be used to engage with the data beyond the commercial provider's reports, for example, by seeking information about specific gene variants or conditions of interest.

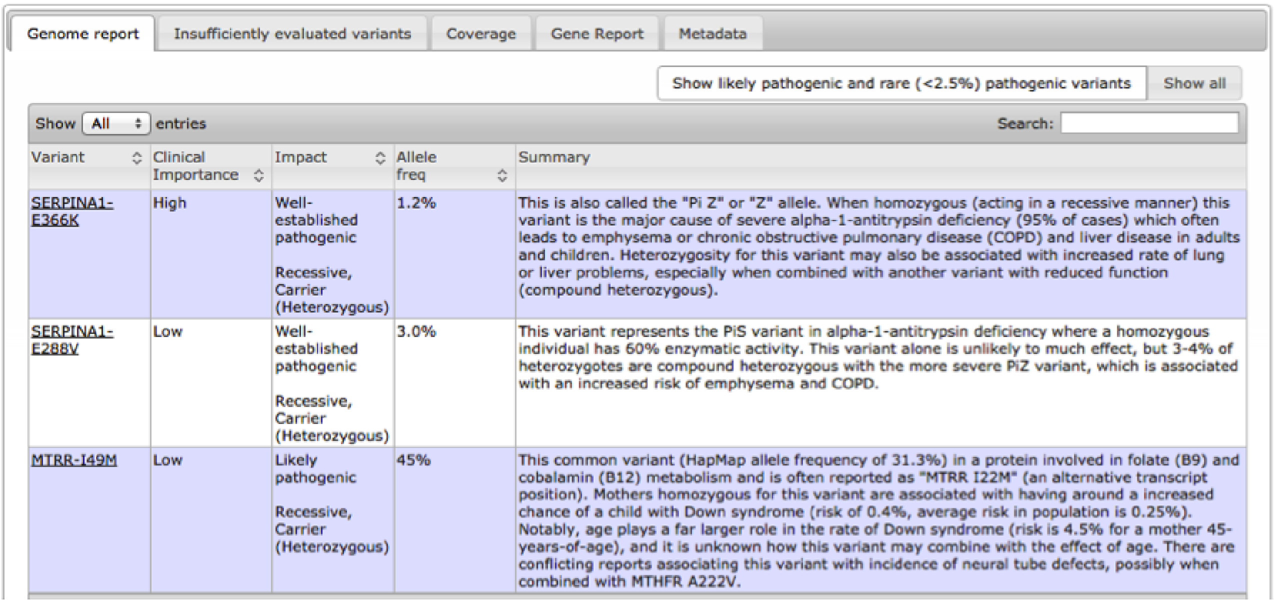
The GET-Evidence report presents detailed information in a tabular design, including a list of gene variants reported to cause particular conditions or traits, the frequency of each variant in the population, the potential impact of each variant and the certainty of that impact (eg, well-established pathogenic, likely protective, uncertain benign), the clinical importance of each variant (ie, low, medium, or high), and a summary describing the current knowledge about a variant. Commentary and links to additional articles and external resources are also available. The table is sorted by clinical importance, but users can further sort their report based on the characteristics above (eg, by potential impact). Figure 1 shows a screenshot of a GET-Evidence report.

Following a brief presentation that reviewed the goals of our research, 36 PGP volunteers—15 female (42%), aged 21 to 83 with an average age of 45 (SD 19)—were recruited to participate. We conducted in-depth, semistructured interviews with each participant. We asked users to explain their goals in engaging with personal genomic information, to share their information practices, and to show us how they use tools to learn from their data. We also asked participants to walk us through their workflow as they explore their personal GET-Evidence report (see Figure 1) [28]. Finally, to elicit ideas about new ways for visualizing and interacting with personal genomics, we presented users with a treemap visualization (see Figure 2) [29] of their own personal genomic data. Participants' personal genomic data were retrieved from the PGP public database. We chose treemaps as a starting point for a discussion about new ways for presenting personal genomic data because they have been successfully applied to the visualization of gene ontologies [30]. Their application to personal genomics for use by consumers, however, is new. We asked users to compare the tabular report with the new visual report and to suggest further ideas that could improve their engagement with the data.

The prototype treemap visualization of the GET-Evidence report (see Figure 2) was created using Google Charts application programming interface (API). It presents the same information and interpretation as the original tabular GET-Evidence report. The treemap groups genetic variants by their clinical importance: low, medium, or high. Each variant is represented by a rectangle with a size proportional to its importance. The color represents the impact of a particular gene variant: pathogenic, benign, or protective. The saturation of the color represents the certainty of the scientific findings determining the impact of a gene variant where highly saturated colors represent high certainty. A red-green color scheme (red—pathogenic, green—protective) was used because it is well accepted in biology and is typically used for visualizing gene expression. Additional information about the gene variant, including a summary, is presented when hovering above a particular variant's tile. Navigation between the two levels of the treemap is handled through selection.

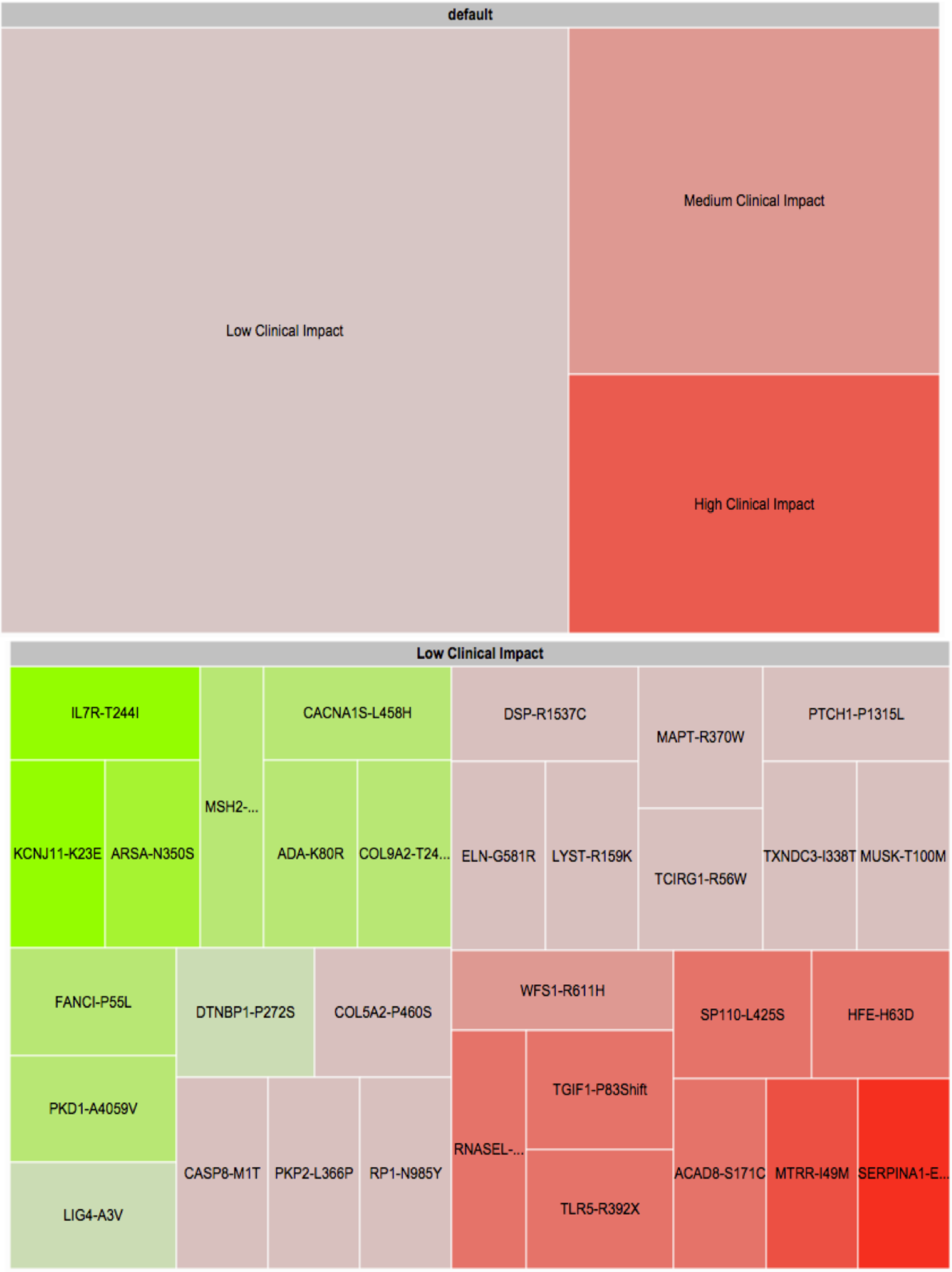
Data were collected, included recordings of participant interviews, detailed notes, logs of user actions as they explored their data, and responses to an online questionnaire. Recordings were later transcribed and data were analyzed using content analysis methods by two independent coders.

Figure 1. A screenshot of a GET-Evidence report, which utilizes tabular design.



Variant	Clinical Importance	Impact	Allele freq	Summary
SERPINA1-E366K	High	Well-established pathogenic Recessive, Carrier (Heterozygous)	1.2%	This is also called the "Pi Z" or "Z" allele. When homozygous (acting in a recessive manner) this variant is the major cause of severe alpha-1-antitrypsin deficiency (95% of cases) which often leads to emphysema or chronic obstructive pulmonary disease (COPD) and liver disease in adults and children. Heterozygosity for this variant may also be associated with increased rate of lung or liver problems, especially when combined with another variant with reduced function (compound heterozygous).
SERPINA1-E288V	Low	Well-established pathogenic Recessive, Carrier (Heterozygous)	3.0%	This variant represents the PiS variant in alpha-1-antitrypsin deficiency where a homozygous individual has 60% enzymatic activity. This variant alone is unlikely to much effect, but 3-4% of heterozygotes are compound heterozygous with the more severe PiZ variant, which is associated with an increased risk of emphysema and COPD.
MTRR-I49M	Low	Likely pathogenic Recessive, Carrier (Heterozygous)	45%	This common variant (HapMap allele frequency of 31.3%) in a protein involved in folate (B9) and cobalamin (B12) metabolism and is often reported as "MTRR I22M" (an alternative transcript position). Mothers homozygous for this variant are associated with having around a increased chance of a child with Down syndrome (risk of 0.4%, average risk in population is 0.25%). Notably, age plays a far larger role in the rate of Down syndrome (risk is 4.5% for a mother 45-years-of-age), and it is unknown how this variant may combine with the effect of age. There are conflicting reports associating this variant with incidence of neural tube defects, possibly when combined with MTHFR A222V.

Figure 2. Two-level treemap prototype visualization of genetic variants. The top screen is the landing page for the visualization, whereas the bottom screen shows what happens when a higher-level rectangle is clicked on. Red represents pathogenic impact. Green represents protective impact.



Study 3: Probing Users

Overview

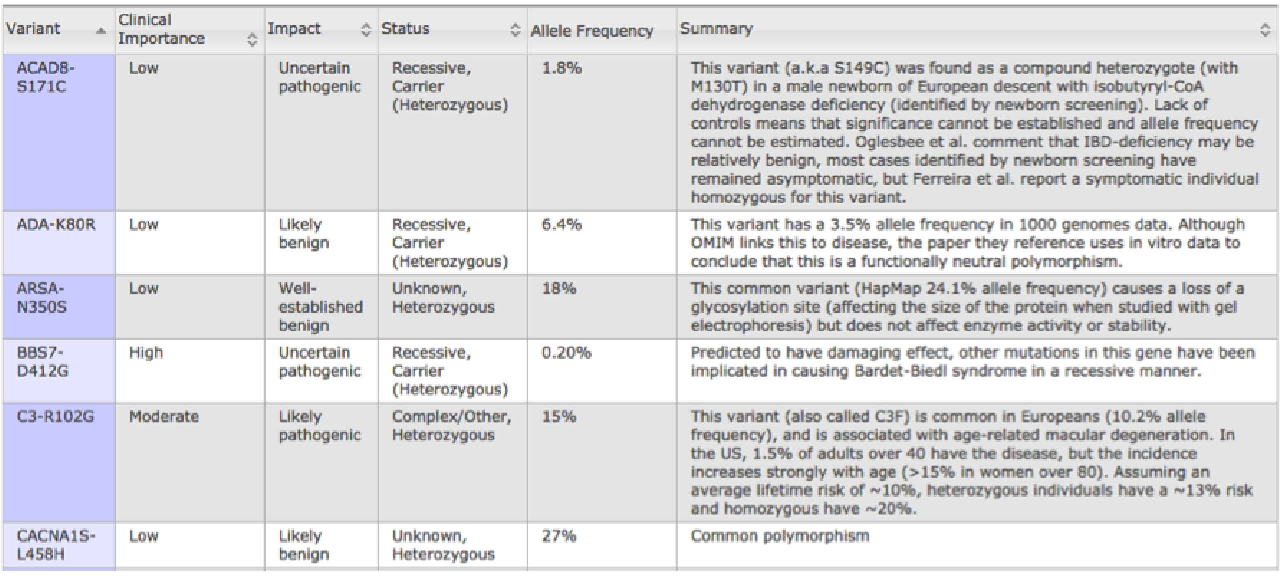
Informed by the findings of Study 2, which will be discussed later and indicated that a visual summary of the report could potentially help nonexpert users in navigating and understanding personal genomic data, we conducted a between-subjects experimental study comparing the effects of different interactive visual genomic reports on nonexperts' understanding of genomic data. These interactive reports were designed based on insights gained in Study 2.

An experimental website was developed specifically for this study, in which different versions of a personal genomics report using GET-Evidence interpretation (see Figures 3-9) were presented. The control condition for this study was a sortable table (see Figure 3), similar to the existing tabular GET-Evidence report. Genetic risk reports from other existing direct-to-consumer genetic testing services (eg, 23andMe) were not included in this evaluation, because they offer medical rather than genetic interpretation of the data.

We implemented the experimental Web platform using MySQL, PHP, JavaScript, Google Charts, and D3 libraries. Participants were recruited via Amazon Mechanical Turk and received US $1.00 for their time. Mechanical Turk is a crowdsourcing marketplace for online tasks that is widely used for HCI and medical informatics research [31-33]. We limited participation to US users with at least 100 prior Human Intelligence Tasks (HITs) at 99% or higher approval rate.

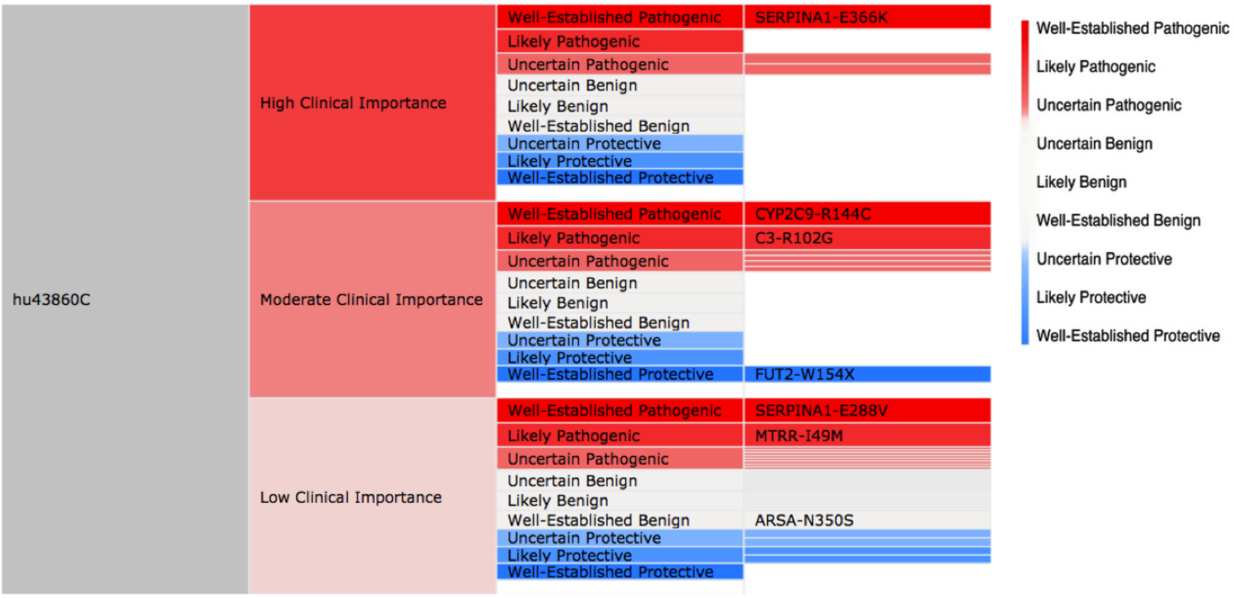
Participants first received a tutorial on the human genome and personal genomics using materials developed by the Personal Genetics Education Project [34]. Their understanding of the material was assessed through a short six-question quiz. If the participant was unable to answer at least three out of six questions correctly, their data were not used in the analysis. They were then presented with one of seven versions of the GET-Evidence report developed for this study.

Figure 3. Table (control condition): gene variants are sorted by name. Variants can be further sorted by clicking on the arrows in each of the columns. The table is similar to the existing GET-Evidence report.



Variant ▲	Clinical Importance ⇕	Impact ⇕	Status ⇕	Allele Frequency	Summary ⇕
ACAD8-S171C	Low	Uncertain pathogenic	Recessive, Carrier (Heterozygous)	1.8%	This variant (a.k.a S149C) was found as a compound heterozygote (with M130T) in a male newborn of European descent with isobutyryl-CoA dehydrogenase deficiency (identified by newborn screening). Lack of controls means that significance cannot be established and allele frequency cannot be estimated. Oglesbee et al. comment that IBD-deficiency may be relatively benign, most cases identified by newborn screening have remained asymptomatic, but Ferreira et al. report a symptomatic individual homozygous for this variant.
ADA-K80R	Low	Likely benign	Recessive, Carrier (Heterozygous)	6.4%	This variant has a 3.5% allele frequency in 1000 genomes data. Although OMIM links this to disease, the paper they reference uses in vitro data to conclude that this is a functionally neutral polymorphism.
ARSA-N350S	Low	Well-established benign	Unknown, Heterozygous	18%	This common variant (HapMap 24.1% allele frequency) causes a loss of a glycosylation site (affecting the size of the protein when studied with gel electrophoresis) but does not affect enzyme activity or stability.
BBS7-D412G	High	Uncertain pathogenic	Recessive, Carrier (Heterozygous)	0.20%	Predicted to have damaging effect, other mutations in this gene have been implicated in causing Bardet-Biedl syndrome in a recessive manner.
C3-R102G	Moderate	Likely pathogenic	Complex/Other, Heterozygous	15%	This variant (also called C3F) is common in Europeans (10.2% allele frequency), and is associated with age-related macular degeneration. In the US, 1.5% of adults over 40 have the disease, but the incidence increases strongly with age (>15% in women over 80). Assuming an average lifetime risk of ~10%, heterozygous individuals have a ~13% risk and homozygous have ~20%.
CACNA1S-L458H	Low	Likely benign	Unknown, Heterozygous	27%	Common polymorphism

Figure 9. Zoomable partition: organized first by clinical importance, and then by potential effect. Zoom by clicking on the rectangles. Detailed information about the variant appears upon hovering over it.



Experimental Conditions

We developed six alternative designs of interactive visual personal genomics reports based on the user requirements and feedback solicited in the previous two studies. In particular, we focused on three of the functional requirements identified in Study 1 (as discussed in the Results section): reviewing an annotated report (R1), integrating data resources (in particular, summary of, and links to, scientific literature) (R2), and making content accessible to nonexperts (R6).

The interventions (ie, interactive reports) developed varied in the visualization technique used for a visual summary, and in the interaction techniques provided for exploring the data. Figures 3-9 show the seven experimental conditions: a tabular control condition (Figure 3) and six visual interactive reports.

Based on our findings from Study 2 that are discussed in more detail later on, we used a new red-white-blue color-coding scheme across all visual conditions to represent the impact and certainty of a particular gene variant. Color represents the impact—pathogenic (red), benign (white), or protective (blue). Saturation represents the certainty of the scientific findings determining the impact of a gene variant, where highly saturated colors map to high certainty. We also added a clickable glossary to all of the reports.

Participants were randomly assigned to one of the following seven conditions: a tabular report modeled after the GET-Evidence report (the control condition) (Figure 3, n=105), a bar chart (Figure 4, n=103), a bubble graph (Figure 5, n=115), a treemap (Figure 6, n=102), a heat map (Figure 7, n=104), a zoomable treemap (Figure 8, n=96), or a zoomable partition (Figure 9, n=105).

We used the same personal genomics data across the different versions, allowing for direct comparison of the reports. This approach of using a fictional dataset to assess user comprehension is a common practice in studies of personal genomics, for example, as in Haga at al [35] and Kaufman et al [36]. We chose a fictional dataset in which sex and ethnicity do not have a specific effect. Actual personal genomic reports include information regarding sex and ethnicity as it may have an effect on particular variations. Once participants had viewed the mock genome reports, they were asked to answer two types of questions: (1) comprehension questions which measure the effectiveness of the interactive visualizations in conveying genomic information, and (2) subjective questions on the extent to which users perceived the report to be understandable. Participants also responded to open-ended questions soliciting their perspectives on useful features and areas for improvement.

Figure 4. Bar chart: bars represent gene variants. A bar’s height and color represent the potential effect of the variant. Variants are separated by clinical importance. Information about a variant appears upon hovering over a bar.

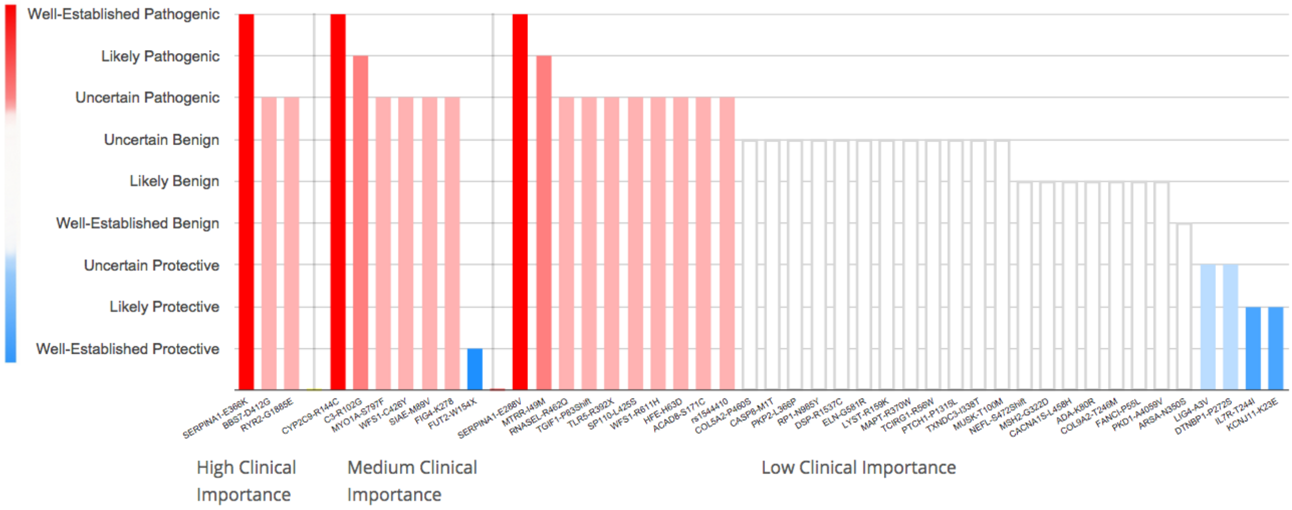


Figure 5. Bubble graph: each bubble represents a gene variant. A bubble's height and color represents the variant’s potential effect. Variants are separated by clinical importance. Detailed information about the variant appears upon hovering a bubble.

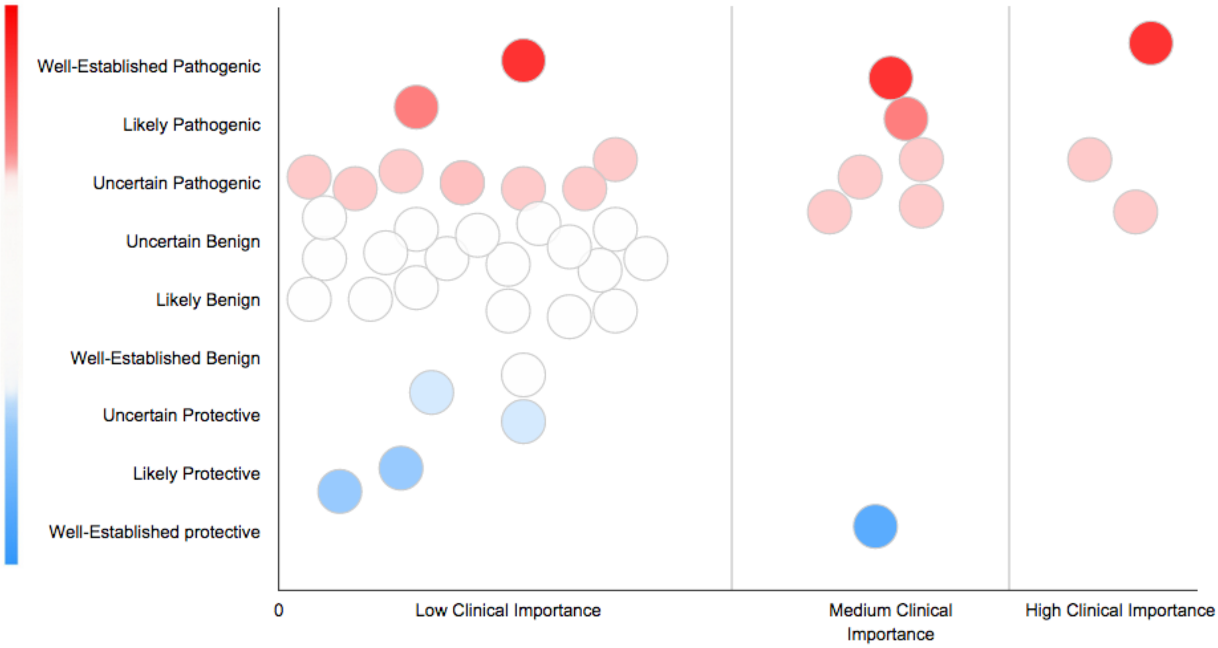


Figure 6. Treemap: gene variants are sorted by clinical importance. By clicking on the boxes or the buttons above the chart, variants within the clinical importance section selected appear and are color coded based on their potential effects. Detailed information about each variant appears upon hovering over it.

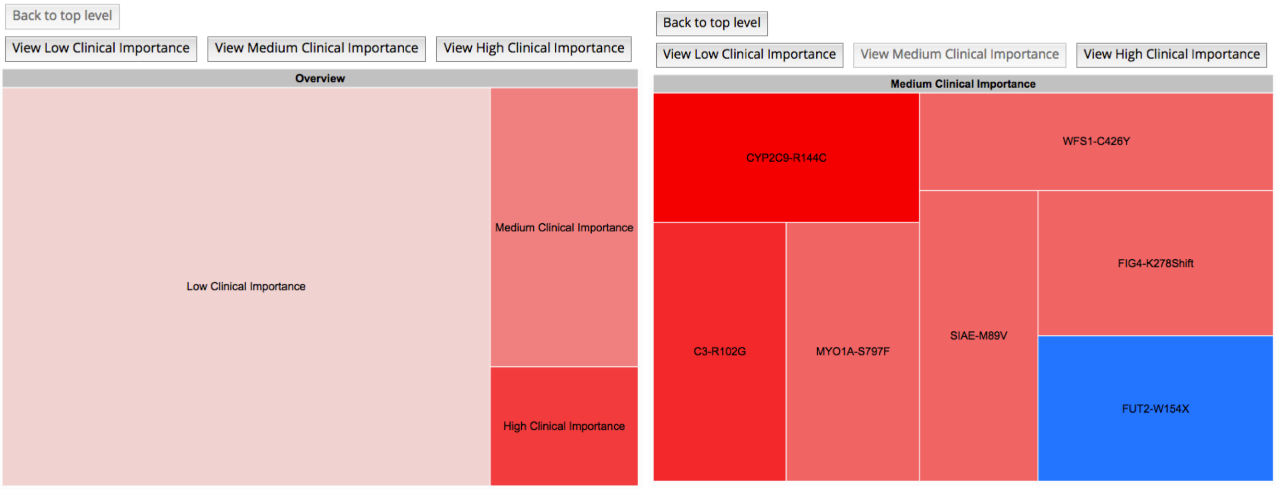


Figure 7. Heat map: each rectangle represents a gene variant. A rectangle's color represents the variant's potential effect. Variants are separated into the three gray boxes by clinical importance. Detailed information about the variant appears upon hovering over it.

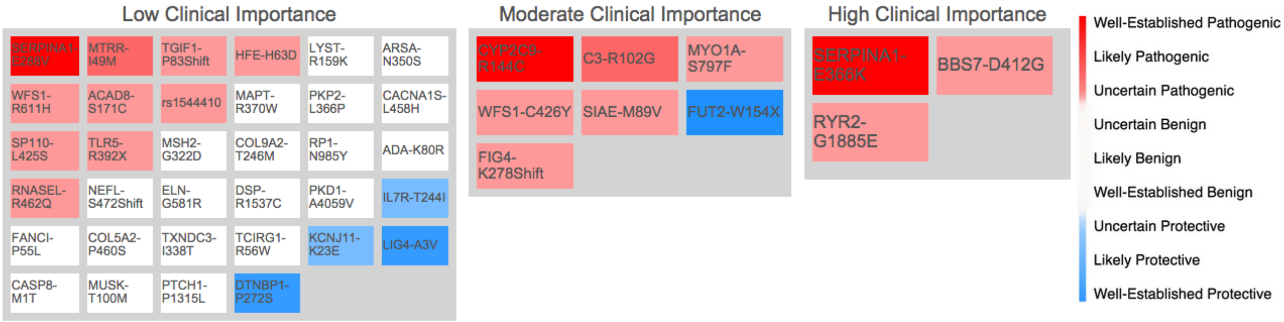
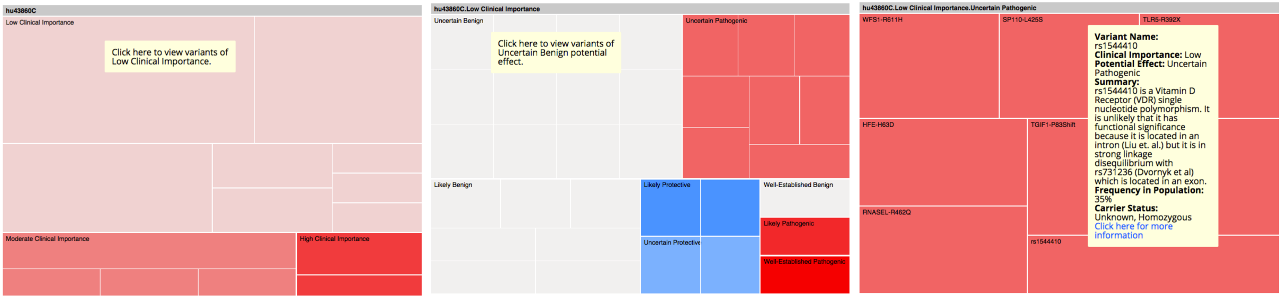


Figure 8. Zoomable treemap: the first level is sorted by clinical importance, the second level is sorted by potential effect, and the third level contains gene variants. The levels are navigable by clicking on the various rectangles. Detailed information about the variant appears upon hovering over it.



Variant Name: rs1544410
Clinical Importance: Low
Potential Effect: Uncertain Pathogenic
Summary: rs1544410 is a Vitamin D Receptor (VDR) single nucleotide polymorphism. It is unlikely that it has functional significance because it is located in an intron (Liu et. al.) but it is in strong linkage disequilibrium with rs731236 (Dvornyk et al) which is located in an exon.
Frequency in Population: 35%
Carrier Status: Unknown, Homozygous
Click here for more information

Visualization Questions

Participants were asked to answer a number of questions testing their understanding of the genomic information based on the visual report they were assigned (see Table 2, Q1 to Q9). The comprehension questions were written in collaboration with the Harvard Personal Genome Project's director of research. Participants also answered questions about their perception of the information (see Table 2, Q10 to Q18) and reported basic demographic information. Participants' responses were recorded in a database and their performance and opinions were compared across the different experimental conditions.

Table 2. Understanding and opinion questions from Study 3.

Question number	Question	Question type
Q1	The number of variants with high clinical importance: _______	Fill in the blank
Q2	The number of variants that are well-established pathogenic: _______	Fill in the blank
Q3	Based on the information above, the number of variants in Jamie's report with low clinical importance is ________ the number of variants with high clinical importance.	Multiple choice: Greater than, Equal to, Less than, I don’t know
Q4	Based on the information above, the number of uncertain pathogenic variants in Jamie's report is ________ the number of well-established pathogenic variants.	Multiple choice: Greater than, Equal to, Less than, I don’t know
Q5	Based on the information above, the number of potentially pathogenic variants in Jamie's report is ________ the number of potentially benign or protective variants.	Multiple choice: Greater than, Equal to, Less than, I don’t know
Q6	Which variants would Jamie be most likely to discuss with a health care provider?	Open response
Q7	Based on the information above, Jamie's risk of developing stomach flu is ________ the average person.	Multiple choice: Greater than, Equal to, Less than, I don’t know
Q8	Based on the information above, Jamie's risk of developing age-related macular degeneration is ________ the average person.	Multiple choice: Greater than, Equal to, Less than, I don’t know
Q9	If you were Jamie, knowing this information, which of the following conditions would you be interested in learning more about? Select all that apply.	Checkbox: Alzheimer's, Parkinson's, Liver Disease, Colon Cancer, Diabetes, Emphysema, Tuberculosis, Eye Disease
Q10	The information in the report was presented in an accessible manner.	Likert scale
Q11	The report is easy to understand.	Likert scale
Q12	Jamie's genes determine everything about them and their future.	Likert scale
Q13	If I were Jamie, I would need the help of a health care professional to understand the results in the report.	Likert scale
Q14	The scientific knowledge used to generate this report is well established.	Likert scale
Q15	If I were Jamie, I would show the results in the report to my doctor.	Likert scale
Q16	The report gives me a firm grasp of Jamie's health and biology.	Likert scale
Q17	Please use the space below to tell us which features were most helpful for understanding the report.	Open response
Q18	Please use the space below to tell us how we can improve the report to make it easier to understand.	Open response

Results

Study 1

Demographics

A total of 83% (52/63) of the participants held academic degrees, 32% (20/63) held doctoral degrees, and 30% (19/63) worked in life sciences-related fields. This demographic is consistent with the description of early adopters provided by Rogers’ theory of the diffusion of innovations [37]. Early adopters tend to have advanced education, expert knowledge—though not necessarily in the subject matter at hand—and a willingness to engage in trials of new technologies.

All participants had prior access to their personal genomic data. Some users received genomic data from more than one service. Table 3 describes the genetic testing services used by study participants.

Table 3. Personal genomic data sources (n=63).

Personal genomic data sources	Users, n (%)
23andMe	38 (60)
Complete Genomics	11 (17)
Family Tree DNA	11 (17)
AncestryDNA	5 (8)
NG Genographic	2 (3)
Microbiome	2 (3)
Sorenson	1 (2)
Medical tests	1 (2)
Exome	1 (2)

Existing Tools

Participants were asked about the websites and computational tools they use to engage with their personal genomic information, and how they use these tools to learn from their data. We found that about 11% (7/63) of participants have used tools beyond those offered by their genetic testing service. An additional 10% (6/63) of participants had tried to explore their genomes using the tools provided by their service provider but found the tools to be too complicated, confusing, or “not user friendly.” Table 4 lists the tools and websites used by our study participants and highlights the main features of each tool.

Table 4. Interactive tools for exploring personal genomics (n=63).

Tool	Users, n (%)	Main features
23andMe	22 (35)	Testing service and interpretative tool Health report on over 240 conditions with external links (until December 2013) Visualized ancestry information Download of raw data file
Promethease	8 (13)	Interpretative tool Annotated report with latest information from SNPedia Comparison of two genomes Family report
		Database Medical, genealogical, and phenotypic variant associations SNP[a] articles with links to publications and summary Facility for sharing data
SNPedia	5 (8)	Access to shared data
Personal Genome Project	7 (11)	Testing service and database Annotated report of variants sorted by clinical importance and impact Article for each variant includes summary and links to relevant publications Facility for sharing data Access to shared data
Gedmatch	7 (11)	Database and interpretative tool Comparison of user's data with public results Visualized information about selected matches Genetic distance calculator Relationship calculator
AncestryDNA	7 (11)	Testing service and interpretive tool Information on ancestry Updates of new matches Access to family tree Download of raw data file
PubMed	6 (10)	Database Access to e-books and journals about all aspects of medicine and life sciences

[a] Single nucleotide polymorphism (SNP).

Functional Requirements

While participants reported being motivated by a diverse set of questions ranging from learning about their traits, to identifying health risks, to learning about their ancestry, they used existing tools to perform six common information tasks: reviewing an annotated report, integrating resources, curating information, comparing genomes, sharing information, and making content accessible. Table 5 depicts these information tasks, and provides an example quote motivating each task.

These tasks constitute the functional requirements for new interactive systems designed for direct consumer engagement with personal genomic information:

1. Task R1: reviewing an annotated report. Participants described the difficulty of interpreting existing tabular and dense textual reports. They expressed a desire for visualizations that make the information easier to explore and understand.

2. Task R2: integrating resources. Participants expressed a need for integrating various data resources, including annotated genomes, scientific publications, various public databases, and health-related data, into a single tool.

3. Task R3: curating information. Participants articulated a need for collecting, relating, organizing, and storing diverse information artifacts (eg, scientific papers, popular articles, notes for doctor appointments, gene variants, and videos) found throughout their independent research of their genome.

4. Task R4: comparing genomes. Participants asked for the ability to triangulate data from several individuals in order to understand connections within families.

5. Task R5: facilitating sharing information. Participants highlighted a need for tools that facilitate information sharing with family, friends, and genetic researchers.

6. Task R6: making content accessible. Participants indicated a need for adapting the content and language of personal genomic reports toward nonexperts. Users also asked to integrate

educational materials within the reports and to point to actionable information.

Table 5. Information tasks and relevant quotes from users.

Task number	Task	Quote
R1	Reviewing an annotated report	“I'd be interested in seeing a graphic illustration of my chromosome sets.” “It would be great to show the SNPs[a] by chromosomal location and in relation to other genes.”
R2	Integrating resources	“Integrated databases of published research that allow the end user, through a seamless interface, to connect personal data with any possibly relevant literature and public data.”
R3	Curating information	“Features that show more clearly what reasonable actionable options there might be for dealing with or preventing various illnesses.”
R4	Comparing genomes	“...easy to use, at home programs, will be needed to compare one's data with those of friends.”
R5	Facilitating information sharing	“The thing that would help the most would be for people to be willing to share more information.”
R6	Making content accessible	“Every time I try to understand something, I have to educate myself via Google, instead of the interface that gives me my genetic data educating me. The research it takes holds me back from using my info more.”

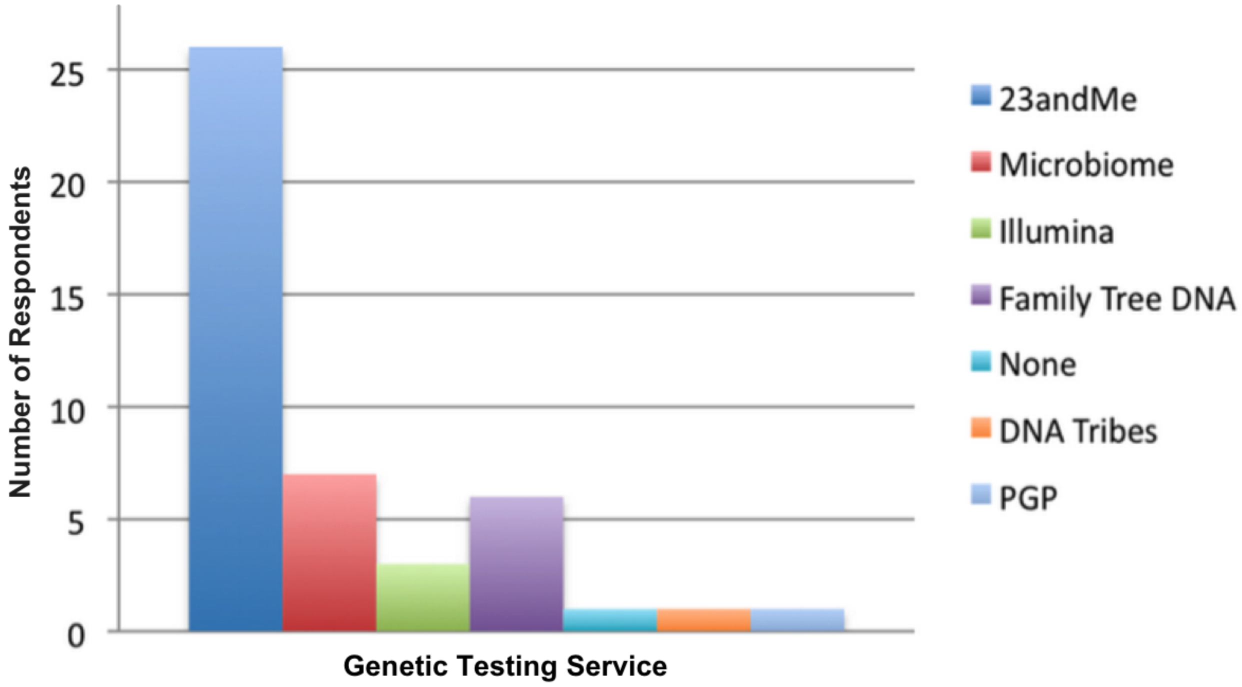
[a] Single nucleotide polymorphisms (SNPs).

Study 2

Demographics

A total of 88% (32/36) of participants held academic degrees, 31% (11/36) held doctoral degrees, and 47% (17/36) worked in life sciences-related fields. All users had previous access to their testing service’s personal genomic report (eg, 23andMe report). Figure 10 demonstrates the distribution of personal genetic testing services. A total of 11% (4/36) of users first viewed their results within the 2 months prior to the study, 6% (2/36) within 3 to 6 months, and 83% (30/36) received the results more than 6 months prior to the study. Approximately one-third of participants had previous access to the GET-Evidence report generated by the PGP. Two-thirds of the users reviewed their GET-Evidence report for the first time in the workshop.

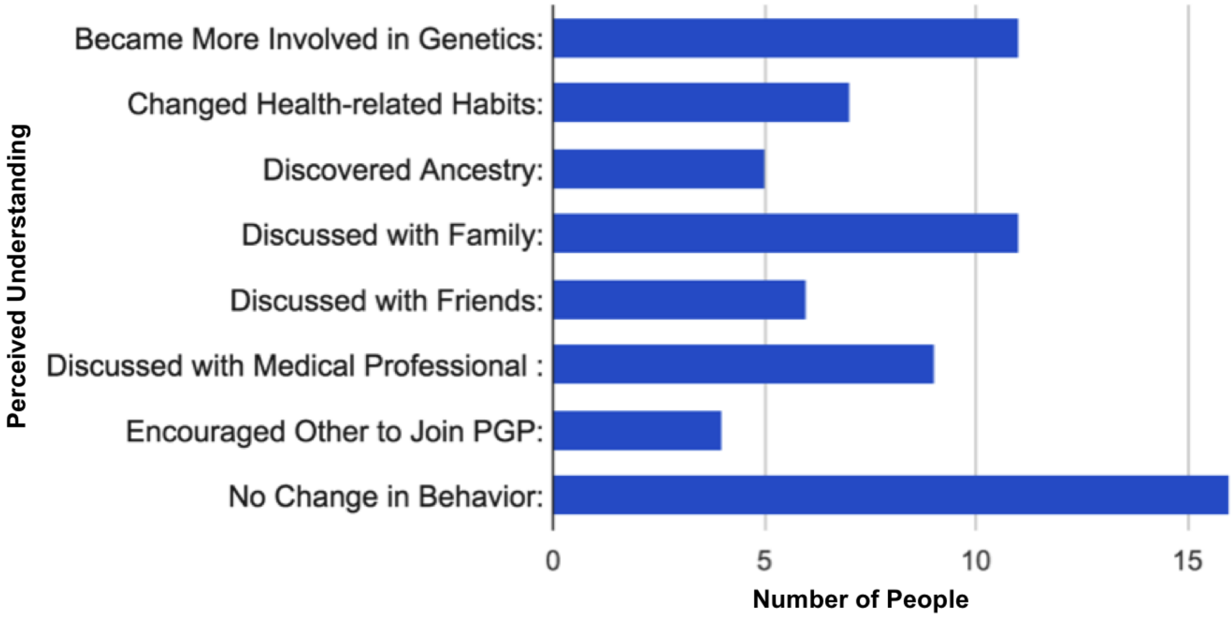
Figure 10. Genetic testing services used by study participants.



Motivation and Impact

Most participants provided more than one reason for exploring their genetic data, including understanding family and individual health risks, gaining insight into ancestry, satisfying curiosity, advancing science, and promoting open-source science. Specifically, 36% (13/36) listed understanding health risks as the primary reason for exploring personal genomics information, while over 66% (24/36) listed advancing genetic, scientific, and health research as their main reason. A total of 14% (5/36) noted that promoting open-source data was a motivation, and 19% (7/36) of users mentioned curiosity. Participants were also asked to describe how knowledge of their genetic traits and health risks impacted their lives. Figure 11 shows their responses. Multiple users listed more than one aspect of influence.

Figure 11. Reported impact of users' personal genomic information on their lives. Multiple users listed more than one influence.



Tabular Report

Users spent about 10 minutes exploring their tabular GET-Evidence report. Most users began by reviewing the table and attempting to sort it according to the impact of the various variants. Users prioritized variants that have well-established pathogenic impact with high clinical importance. One explained, “Most interesting for me is what is high clinical important—I zoom in on things that say high, pathogenic, well-established.”

Many users commented on the amount and nature of information presented: “To me, it seems clunky and more like raw data, and could use some more sorting,” and “Show me the most important vectors vs the entire shebang.” Another user commented, “This is overwhelming, I prefer it filtered by high importance.” Users requested more advanced sorting and filtering mechanisms: “Being able to sort the list, so that you can see all pathogenic mutations together, or the mutations ranked by how well-studied they are.” A total of 4 users did not realize that the table could be sorted at all and scanned the table entries individually in search for important variants: “It was difficult scanning rows.” Some users requested direct search functionality that includes the detailed summary entries.

Visual Report

In order to initiate a conversation about alternative ways for presenting personal genetic information, we asked users to explore their own data using a treemap report (see Figure 2).

Users spent about 10 minutes exploring their data using the treemap report followed by a semistructured interview about the strengths and weaknesses of interactive visual representations of genetic data.

Approximately one-third of the participants expressed a strong preference for the treemap visualization. In the words of one participant, “I like this better in every way. It provides quick visual summary and weights the low and moderate by size so I can quickly determine what to be concerned about if anything.” Many of these users found the color coding to be particularly helpful: “The color shading made it easy to tell which alleles were protective or pathogenic.” Others suggested the inclusion of a color key, the use of a color scheme accessible for users with red-green color deficiency, and better color distinction between benign and pathogenic variants.

On the other hand, many users commented on the treemap navigation either finding it confusing or preferring to see all the information at once: “I dislike navigating because [I] need to do additional actions to access desired browsing criteria.” While we attributed some confusion to the navigation mechanism implemented by the Google Visualization API, we also identified the importance of balancing the requirement to highlight important information with the need to present to users an overview of the entire dataset.

Accessibility for Nonexperts

Users requested “nonscientist-friendly” reports that provide access to glossaries and use nontechnical language. About half of the users commented on the technical jargon used in the reports, finding it difficult to understand: “As a person that doesn’t understand science, it’s overwhelming.” In particular, several users asked about the following terms: allele frequency, homozygous, pathogenic, and benign. Some users noted that variant names, which are used as labels, are too long and intimidating for nonscientists, and suggested adding information beyond scientific variant names.

Complexity and Uncertainty

We learned that one of the most challenging aspects of presenting personal genomic information to consumers is the complexity and uncertainty of the interpretation. Associations between gene variants and medical conditions or traits are established through scientific studies that vary in the strength of the evidence they present. The GET-Evidence report classifies evidence into three main categories: established, likely, and uncertain. Multiple users thus commented on the challenge of understanding the potential impact and clinical importance of some variants. For example, one user asked, "Are variants with low clinical impact just variants that have not been well characterized yet?" After using the treemap report another user noted, "So much inadequate evidence out there—successful that it communicates that."

In addition, many complex conditions such as diabetes or various cancers are associated with multiple genes rather than by a single gene variant. One participant requested, "A better understanding of how the factors combine to affect me." Another user suggested, "It would be helpful to see similar diseases grouped together. For instance I had one protective and one negative SNP for macular degeneration, and it would be hard to connect that only looking at this report." Furthermore, an individual may be a carrier for a certain trait, meaning that she or he is not impacted by particular gene variants but their children might be. As one user explains, "I am also interested in alleles that may not have an impact for me (as a carrier) but that could affect my future children if they end up getting two copies, so it would be nice to have a separate report that shows those."

Providing Evidence

Several users requested that the reports provide direct links to sources of scientific data while grading the rigor of the studies: "Include links to relevant studies—back it up."

Relating Genetic Variants to Medical Conditions

Finally, 6 participants out of 36 (17%) suggested focusing the results report on medical conditions rather than gene variants: "It would be helpful to have the health condition rather than the variant/trait displayed. [I] would rather see [medical] condition not variant at the tile label." Participants also requested information on how to mitigate the risk for particular conditions to which they are genetically predisposed. However, in December 2013, the US Food and Drug Administration ordered companies that offer personal genome testing, such as 23andMe, to cease providing such health reports to customers [8]. The FDA was concerned with the use of direct-to-consumer genetic tests for medical purposes because of the uncertainty and inaccuracy of these tests in predicting disease risk. As a result, new medical interpretation of an individual's genomic data, as requested by several participants, is beyond the scope of our investigation.

Study 3

Overview

A total of 745 participants were recruited for this study. Participants whose post-tutorial test included three or more incorrect responses (out of six) were excluded from the analysis. The sample used in the analyses thus came to 731 individuals whose average age was 36.6 years (SD 11.9). A total of 40.6 % (297/731) of the participants were women, 56.8% (415/731) held academic degrees, 2.2% (16/731) held a doctoral degree, and 7.8% (57/731) worked in life sciences-related fields. While our first study was with early adopters who were potentially experts in the field of biology or genetics, this study targeted a much more diverse distribution of people. Users spent just over 12 minutes on average (SD 8) exploring the reports.

Comprehension Scores

Comprehension questions assessed users' abilities to identify variants that indicated increased risk for a particular condition and to understand the certainty of the scientific evidence behind the interpretation. A comprehension score between 0 and 10 was assigned based on users' responses to nine questions in the form of multiple choice (Table 2; Q3, Q4, Q5, Q7, and Q8), fill in the blank (Table 2; Q1, Q2, and Q6), and select all that apply (Table 2, Q9). There were multiple answers for the "select all that apply" question, which were given separate points. Participants' abilities to find clinically important variants, both pathogenic and protective, were also assessed. A comparison of the participants' responses to the comprehension questions across the seven experimental conditions was made using analysis of variance (ANOVA) followed by a post hoc Tukey's test. Table 6 provides a summary of these findings.

Table 6. Summary of comprehension scores across interventions.

Intervention	Average score (SD)
Table (control)	5.65 (1.83)
Bar chart	6.21 (1.56)
Bubble chart	6.30 (1.44)
Treemap	5.74 (1.72)
Heat map	6.25 (1.40)
Zoomable treemap	4.63 (2.16)
Zoomable partitions	5.08 (1.90)

The analyses revealed significant differences between the report types (see Figure 12). The zoomable treemap and zoomable partition reports were found to be less effective in communicating genomic data than the visualizations in the other conditions—5.08 and 4.63 out of 10, respectively. Indeed, scores obtained using the zoomable treemap report resulted in significantly lower comprehension scores compared with all nonzoomable reports ($P<.001$), and the zoomable partition report led to significantly lower understanding compared with the bar chart (6.21/10), bubble graph (6.30/10), and heat map reports (6.25/10) (*P*<.001). In addition, the table report (5.65/10) was also found to be significantly less effective than the bubble graph report (*P*=.04).

Participants were asked to rate their perceived ease of understanding on a 5-point Likert scale (Table 2, Q11). A comparison of responses between the seven report types was made using ANOVA followed by a post hoc Tukey's test. Analysis demonstrated that perceived understanding was highest in the bubble graph (4.31 out of 5, SD 1.52) and lowest in the zoomable treemap (3.29, SD 1.87) (see Figure 13). The bubble graph was perceived as significantly easier to understand than both the zoomable treemap (3.29, SD 1.87, *P*=.01) and the tabular control condition (3.62, SD 1.74, *P*=.05).

We found significant differences in the results between participants who worked in life science-related fields and others who did not, in terms of both comprehension—where those in the life sciences scored higher—and perceived ease of understanding—where those in the life sciences found it easier to understand. However, when running the statistical analyses comparing the visualization types among people in the life sciences population, the differences between the visualization types were found to be similar to the differences between the visualization types among the entire population.

Figure 12. Users' comprehension of the reports across the report types. Error bars represent 95% CI.

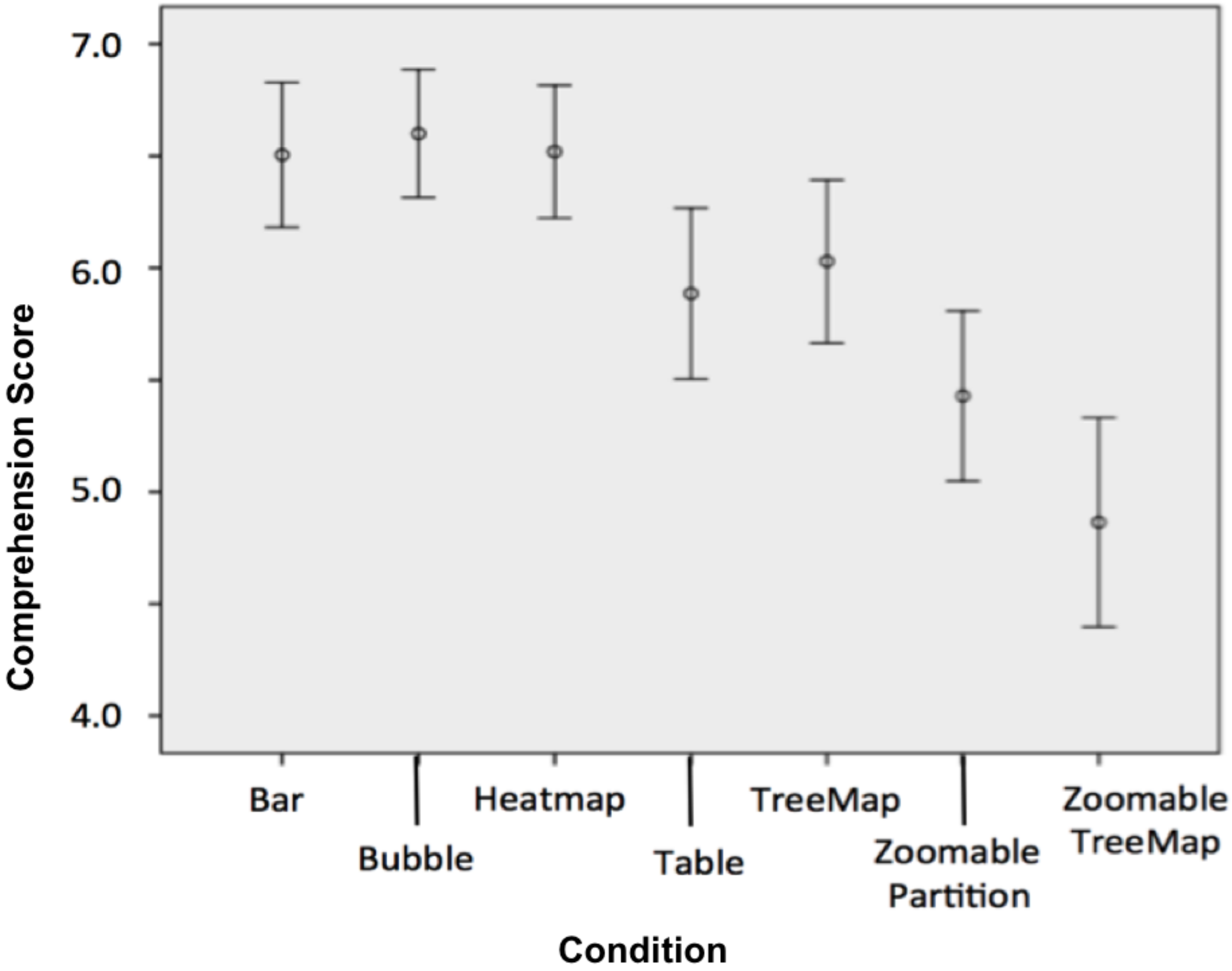
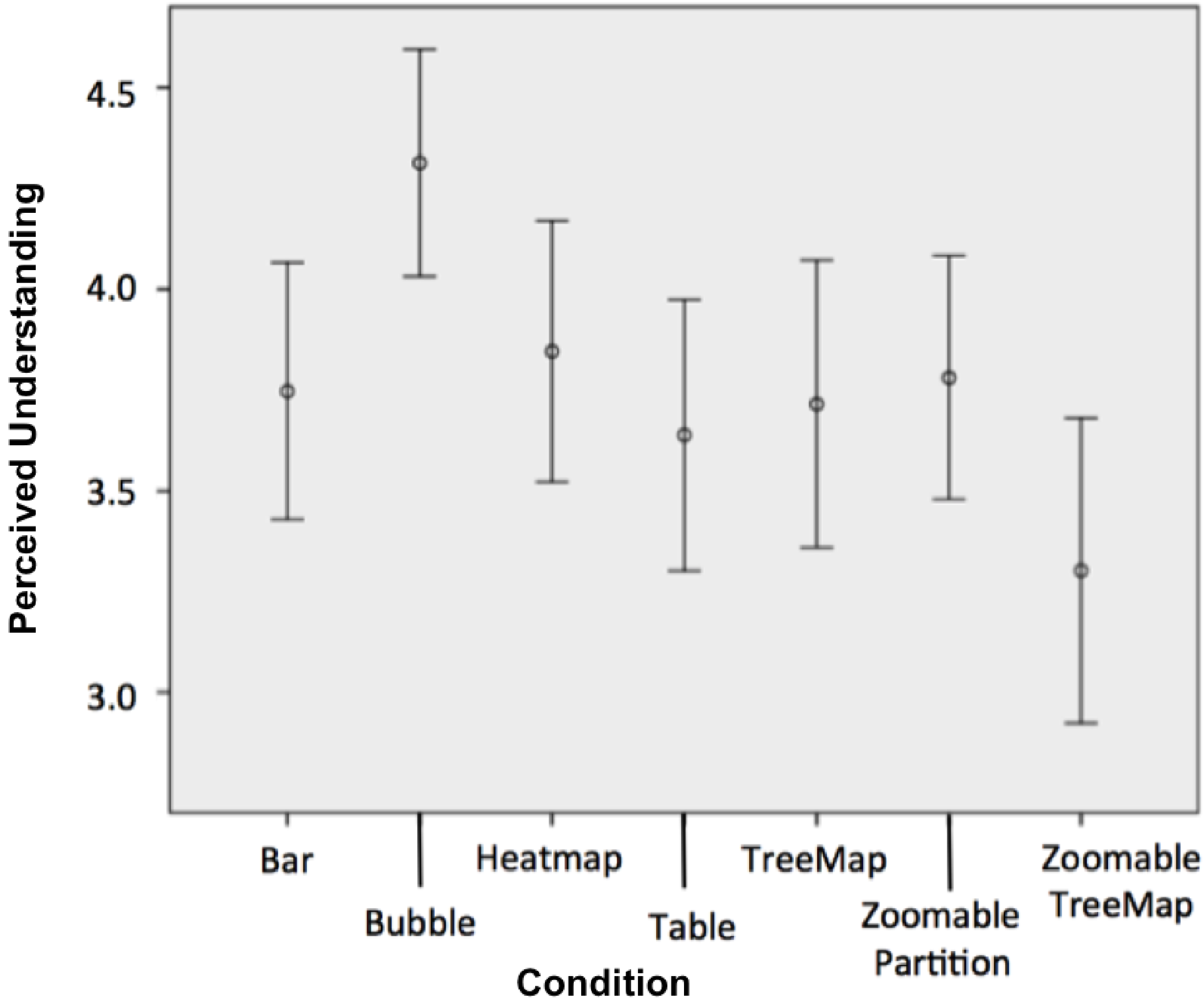


Figure 13. Users' subjective scores of perceived understanding. Error bars represent 95% CI.



Qualitative Results

Approximately 39.2% (240/612) of participants who responded to the open-ended questions and who were not assigned to the control condition, which did not use color coding, reflected positively on the color-coding scheme. For example, one participant noted that “The color-coded chart made it a lot easier to see the harmful and beneficial genes and how severe they potentially were.” Another participant mentioned, “The use of different colors made some information very easy to see right away.” Only 3.1% (19/612) of participants gave negative feedback on the color coding. About 33.9% (243/716) of all users still found the gene variant summaries too technical and difficult to understand, though 5.7% (41/716) noted that they found the glossary helpful. Approximately 9.8% (60/612) of participants, not including those in the control condition, indicated that presenting information in tooltips when hovering was not effective. Users pointed to trouble navigating and closing the tooltips, as well as maintaining context as the presented tooltip occluded part of the visualization.

Approximately 25.0% (48/192) of the participants in the zoomable visualization intervention (see Figures 8 and 9) commented that they found zooming confusing. For example, one participant described “...too many in and outs...once information is collapsed it's difficult to navigate." Another mentioned, “The information would be easier to understand in some sort of chart or perhaps more than one rather than the interactive diagram. One chart might have a brief overview and the next perhaps more details. I find the diagram frustrating and time-consuming because the information cannot be accessed all at once.” A total of 12.5% (13/104) of participants in the control condition described the ability to sort columns in the table as the most useful feature of the visualization. One user noted that “Being able to change how the information is sorted made it easier to determine what information was important.” The qualitative data did not, however, provide an explanation for why the bubble graph performed so well in the comprehension test.

Discussion

Study 1

Findings from Study 1 shed light on information practices and needs of early adopters of personal genomics. We identified preliminary functional requirements for new direct-to-consumer interactive tools for personal genomics including the following: (1) R1: reviewing an annotated report, (2) R2: integrating resources, (3) R3: curating information, (4) R4: comparing genomes, (5) R5: facilitating sharing information, and (6) R6: making content accessible. However, findings from this study also highlighted a need to further investigate detailed elements of interactive genomic reports that were not probed in this study. In particular, expanding our understanding of Task R1, which arguably represents the most basic functionality of existing direct-to-consumer personal genomic tools, in order to understand how users engage with interactive genetic reports to learn about their genomic data. Open-ended questions, which informed Study 2, included the following: How do users approach their personal genomic reports? What makes genomic reports difficult to understand? What features could make genomic information more accessible?

Study 2

To address these questions, we conducted a workshop with Personal Genome Project participants focusing on their interactions with annotated personal genomic reports. Our goals for this workshop were to gain a more nuanced understanding of how users currently engage with genomic data, how they learn from their own personal genomic data, and what factors impact their understanding and preferences.

Findings indicated that early adopters of personal genomics were motivated by various factors and were not necessarily approaching their personal genomic information to find an answer to a concrete question. Rather, they sought information about gene variants with well-established pathogenic impact and that were of high clinical importance. Results also highlighted factors that make direct-to-consumer genomic reports difficult to understand, including a large amount of textual information, scientific and technical jargon, and the complexity and uncertainty of the interpretation. Finally, several features were found to be potentially helpful in making personal genomic information more accessible and understandable, including the following: (1) presenting a visual summary (eg, a treemap) that highlights important variants based on their clinical importance and potential impact, (2) using nontechnical language and providing a glossary, and (3) allowing users to search and sort the report.

The user population that participated in this study was again consistent with early adopters as described by Rogers' theory of the diffusion of innovations [37]—users that tend to have advanced education, expert knowledge, and a willingness to engage in trials of new technologies. Thus, many open questions remain about how to make personal genomics accessible to a general nonexpert user population.

As direct-to-consumer genetic testing services become increasingly available to the general public it is important to also study nonexperts' interactions with genomic reports. Informed by the insights gained in this study with early adopters, we developed several new interactive visual personal genomic reports aimed at nonexperts. Study 3 focused on investigating how alternative visual designs for genomic reports impact nonexpert understanding of personal genomic information.

Study 3

Findings from this experiment indicate that HCI interventions can improve the understandability of interactive personal genomic reports for a diverse population of consumers compared to existing reports. Findings also highlight the following implications for the design of interactive, visual personal genomic reports:

1. Zoomability might compromise understandability. The findings show that while zoomable reports (see Figures 8 and 9) may provide additional layers of information, they seem to be less effective in conveying personal genomics information to nonexpert users. This may be because such interfaces are less familiar to most nonexpert users, or because a visual summary maintains better context. Offering additional explanation and tutoring may help users to benefit more from such interactive tools.

2. Overview and familiarity. The findings also suggest that nonzoomable report types, which offer an overview of the entire report through a visual summary (ie, bar chart, bubble graph, and heat map), may be better than tables at conveying personal genomic information. Comprehension scores obtained using the bubble graph interface were the only ones to reach statistical significance, but the findings call for more research comparing different report types and interactive features.

3. Comprehension and perceived understandability. Among the report types studied, the bubble-based report combined both high scores of objective understanding—using the comprehension test—and high scores of subjective perceived understandability (see Figures 12 and 13). This finding suggests that this report type is more likely than others to be useful for nonexpert users. Future work will explore the factors that make this report type more effective and preferable than others.

4. Communicating impact and certainty using color. Findings from Study 2 indicated that when exploring their report, users prioritized locating variants with well-established pathogenic impact. The use of color coding, which utilizes both hue and saturation, was found to be effective in helping users to identify high-priority gene variants. Based on the feedback received in Study 2, we chose a three-color, rather than two-color, coding scheme for Study 3—red (pathogenic), white (benign), blue (protective). This color scheme is accessible to users with color vision deficiencies and was found to be effective based on the qualitative results.

5. Hovering and tooltips. In order to simplify the text-heavy tabular design, all six interactive visualizations presented a summary of each gene variant in a tooltip when the user hovered over a gene variant. Findings identified several usability considerations and problems with hovering and tooltips, including (1) the action to trigger the presentation of a tooltip—while deliberate selection limits fluid exploration, hovering may trigger the presentation of tooltips without user intention, (2) visual elements should be large enough to allow the user to hover above a particular element, (3) when displayed, tooltips occluded parts of the visualization, hiding information that was important for maintaining context—several design techniques could be applied to resolve this problem, including semitransparent tooltips, expanding the visualization layer, and presenting information in an alternative area rather than in a tooltip, and (4) what interactive features should be supported

by tooltips—participants requested the ability to search within a tooltip, and to save and share the content of tooltips.

Conclusions and Future Work

We presented findings from three complementing studies that combined qualitative and quantitative methods to inform the design of personal genomic reports. Our findings offer useful insights for designers and researchers interested in the role HCI can play in making personal genomics understandable and useful for consumers. Study 1 explored the information practices and needs of early adopters and identified preliminary functional requirements for new direct-to-consumer personal genomics interactive tools. Extending Study 1 using face-to-face interviews and user demonstrations, Study 2 sought to understand why and how users engage with interactive genetic reports to learn about their personal genomic data. Building on the findings of the first two studies, Study 3 focused on designing and testing alternative interactive reports informed by the needs and practices identified earlier. The designs, using different interactive visualizations, were tested using online experiments with Amazon Mechanical Turk users to investigate how variations in interface design and data visualization affect users' understanding of, as well as preferences and attitudes toward, personal genomic reports.

To our knowledge, this paper presents the first study that focuses on information practices, requirements, and design considerations for nonexpert engagement with personal genomics. Future work may focus on the role demographic and other personal attributes may have on users' understanding of different report types. For example, emerging work shows that personality traits are important to how users perceive data visualizations [38]. Understanding how users' backgrounds and personalities affect their understanding of, and likelihood to act upon, personal genomic reports is important. We also plan to apply findings from the qualitative and experimental research to the design and development of new interactive tools that empower consumers to engage with, and make sense of, their personal genomic information.

Acknowledgments

This work was partially funded by Grants IIS-1017693 and IIS-1422706 from the National Science Foundation (NSF), Division of Information and Intelligent Systems (IIS).

Conflicts of Interest

None declared.

References

1. Davies K. The $1,000 Genome: The Revolution in DNA Sequencing and the New Era of Personalized Medicine. New York, NY: Free Press; Sep 07, 2010.
2. Dudley JT, Karczewski KJ. Exploring Personal Genomics. Oxford, UK: Oxford University Press; 2013.
3. Cadwalladr C. The Guardian. 2013 Jun 08. What happened when I had my genome sequenced? URL: http://www.theguardian.com/science/2013/jun/08/genome-sequenced [accessed 2015-03-06] [WebCite Cache ID 6WpTZa3qU]
4. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008 Mar;24(3):133-141. [doi: 10.1016/j.tig.2007.12.007] [Medline: 18262675]
5. AncestryDNA. URL: http://dna.ancestry.com/?s_kwcid=ancestry+dna&o_xid=58712&o_lid=58712&o_sch=Search [accessed 2015-03-04] [WebCite Cache ID 6WmfpSgtn]
6. FamilyTreeDNA. URL: https://www.familytreedna.com/ [accessed 2015-03-06] [WebCite Cache ID 6WpT7XReh]
7. 23andMe. URL: https://www.23andme.com/ [accessed 2015-03-06] [WebCite Cache ID 6WpSvqaP0]
8. Pollack A. The New York Times. 2013 Nov 25. F.D.A. orders genetic testing firm to stop selling DNA analysis service URL: http://www.nytimes.com/2013/11/26/business/fda-demands-a-halt-to-a-dna-test-kits-marketing.html?_r=1 [accessed 2015-03-06] [WebCite Cache ID 6WpTfBqL1]
9. Centers for Medicare & Medicaid Services (CMS), HHS, Centers for Disease Control and Prevention (CDC), HHS, Office for Civil Rights (OCR), HHS. CLIA program and HIPAA privacy rule; patients' access to test reports. Final rule. Fed Regist 2014 Feb 6;79(25):7289-7316 [FREE Full text] [Medline: 24605389]
10. Personal Genome Project. URL: http://www.personalgenomes.org/ [accessed 2015-03-06] [WebCite Cache ID 6WpTOwjx8]
11. Ball M, Bobe J, Chou M, Clegg T, Estep PW, Lunshof JE, et al. Harvard Personal Genome Project: lessons from participatory public research. Genome Med 2014;6(2):10 [FREE Full text] [doi: 10.1186/gm527] [Medline: 24713084]
12. Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, et al. A public resource facilitating clinical use of genomes. Proc Natl Acad Sci U S A 2012 Jul 24;109(30):11920-11927 [FREE Full text] [doi: 10.1073/pnas.1201904109] [Medline: 22797899]
13. Church GM. The personal genome project. Mol Syst Biol 2005 Dec 13;1:1-3 [FREE Full text] [doi: 10.1038/msb4100040] [Medline: 16729065]
14. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. Nat Rev Genet 2008 May;9(5):406-411. [doi: 10.1038/nrg2360] [Medline: 18379574]

15. Goldsmith L, Jackson L, O'Connor A, Skirton H. Direct-to-consumer genomic testing: systematic review of the literature on user perspectives. Eur J Hum Genet 2012 Aug;20(8):811-816 [FREE Full text] [doi: 10.1038/ejhg.2012.18] [Medline: 22333900]
16. McGowan ML, Fishman JR, Lambrix MA. Personal genomics and individual identities: motivations and moral imperatives of early users. New Genet Soc 2010 Sep 1;29(3):261-290 [FREE Full text] [doi: 10.1080/14636778.2010.507485] [Medline: 21076647]
17. Bloss CS, Ornowski L, Silver E, Cargill M, Vanier V, Schork NJ, et al. Consumer perceptions of direct-to-consumer personalized genomic risk assessments. Genet Med 2010 Sep;12(9):556-566. [doi: 10.1097/GIM.0b013e3181eb51c6] [Medline: 20717041]
18. Gordon ES, Griffin G, Wawak L, Pang H, Gollust SE, Bernhardt BA. "It's not like judgment day": public understanding of and reactions to personalized genomic risk information. J Genet Couns 2012 Jun;21(3):423-432 [FREE Full text] [doi: 10.1007/s10897-011-9476-4] [Medline: 22180182]
19. Leighton JW, Valverde K, Bernhardt BA. The general public's understanding and perception of direct-to-consumer genetic test results. Public Health Genomics 2012;15(1):11-21 [FREE Full text] [doi: 10.1159/000327159] [Medline: 21720148]
20. McGuire AL, Burke W. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. JAMA 2008 Dec 10;300(22):2669-2671 [FREE Full text] [doi: 10.1001/jama.2008.803] [Medline: 19066388]
21. Shaer O, Mazalek A, Ullmer B, Konkel M. From big data to insights: opportunities and challenges for TEI in genomics. In: Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction. 2013 Presented at: 7th International Conference on Tangible, Embedded and Embodied Interaction; February 10-13, 2013; Barcelona, Spain p. 109-116. [doi: 10.1145/2460625.2460642]
22. Kuznetsov S, Taylor AS, Regan T. At the seams: DIYbio and opportunities for HCI. In: Proceedings of the Designing Interactive Systems Conference. 2012 Presented at: Designing Interactive Systems (DIS) Conference; June 11-15, 2012; Newcastle, UK p. 258-267. [doi: 10.1145/2317956.2317997]
23. Schkolne S, Ishii H, Schroder P. Immersive design of DNA molecules with a tangible interface. In: Proceedings of IEEE Visualization 2004. 2004 Presented at: IEEE Visualization 2004; October 10-15, 2004; Austin, TX p. 227-234. [doi: 10.1109/VISUAL.2004.47]
24. Block F, Horn MS, Phillips BC, Diamond J, Evans E, Chia Shen. The DeepTree exhibit: visualizing the Tree of Life to facilitate informal learning. IEEE Trans Vis Comput Graph 2012 Dec;18(12):2789-2798. [doi: 10.1109/TVCG.2012.272]
25. Schneider B, Strait M, Muller L. Phylo-Genie: engaging students in collaborative 'tree-thinking' through tabletop techniques. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 3071-3080. [doi: 10.1145/2207676.2208720]
26. Shaer O, Strait M, Valdes C, Wang H, Feng T, Lintz M, et al. The design, development, and deployment of a tabletop interface for collaborative exploration of genomic data. Int J Hum Comput Stud 2012 Oct;70(10):746-764 [FREE Full text] [doi: 10.1016/j.ijhcs.2012.05.003]
27. Shaer O, Nov O, West A, Eastman D. Understanding information practices of interactive personal genomics users. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (Extended Abstracts). 2014 Presented at: The ACM CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, ON p. 2455-2460. [doi: 10.1145/2559206.2581145]
28. GET-Evidence. URL: http://evidence.pgp-hms.org/about [accessed 2015-03-06] [WebCite Cache ID 6WpTBTmoU]
29. Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. ACM Trans Graph 1992 Jan;11(1):92-99. [doi: 10.1145/102377.115768]
30. Baehrecke E, Dang N, Babaria K, Shneiderman B. Visualization and analysis of microarray and gene ontology data with treemaps. BMC Bioinformatics 2004 Jun 28;5:84 [FREE Full text] [doi: 10.1186/1471-2105-5-84] [Medline: 15222902]
31. Cui L, Carter R, Zhang GQ. Evaluation of a novel Conjunctive Exploratory Navigation Interface for consumer health information: a crowdsourced comparative study. J Med Internet Res 2014;16(2):e45 [FREE Full text] [doi: 10.2196/jmir.3111] [Medline: 24513593]
32. Muench F, van Stolk-Cooke K, Morgenstern J, Kuerbis AN, Markle K. Understanding messaging preferences to inform development of mobile goal-directed behavioral interventions. J Med Internet Res 2014;16(2):e14 [FREE Full text] [doi: 10.2196/jmir.2945] [Medline: 24500775]
33. Nov O, Arazy O, López C, Brusilovsky P. Exploring personality-targeted UI design in online social participation systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 361-370. [doi: 10.1145/2470654.2470707]
34. Personal Genetics Education Project. URL: http://www.pged.org/ [accessed 2015-03-06] [WebCite Cache ID 6WpTHnl1l]
35. Haga SB, Mills R, Pollak KI, Rehder C, Buchanan AH, Lipkus IM, et al. Developing patient-friendly genetic and genomic test reports: formats to promote patient engagement and understanding. Genome Med 2014;6(7):58 [FREE Full text] [doi: 10.1186/s13073-014-0058-6] [Medline: 25473429]
36. Kaufman DJ, Bollinger JM, Dvoskin RL, Scott JA. Risky business: risk perception and the use of medical services among customers of DTC personal genetic testing. J Genet Couns 2012 Jun;21(3):413-422. [doi: 10.1007/s10897-012-9483-0] [Medline: 22278220]

37. Rogers EM. Diffusion of Innovations. 5th edition. New York, NY: Free Press; 2003.
38. Ziemkiewicz C, Ottley A, Crouser RJ, Yauilla AR, Su SL, Ribarsky W, et al. How visualization layout relates to locus of control and other personality factors. IEEE Trans Vis Comput Graph 2013 Jul;19(7):1109-1121. [doi: 10.1109/TVCG.2012.180] [Medline: 23661008]

Abbreviations

ANOVA: analysis of variance
API: application programming interface
DIYbio: do-it-yourself biology
DTCGT: direct-to-consumer genetic testing
FDA: Food and Drug Administration
GET: Genomes, Environments, and Traits
HCI: human-computer interaction
HGP: Human Genome Project
HIT: Human Intelligence Task
IIS: Information and Intelligent Systems
NSF: National Science Foundation
PGP: Personal Genome Project
SNP: single nucleotide polymorphism

Edited by G Eysenbach; submitted 11.03.15; peer-reviewed by SH Yim, L Lovrecic; comments to author 02.04.15; revised version received 28.04.15; accepted 07.05.15; published 12.06.15.

Please cite as:
Shaer O, Nov O, Okerlund J, Balestra M, Stowell E, Ascher L, Bi J, Schlenker C, Ball M
Informing the Design of Direct-to-Consumer Interactive Personal Genomics Reports
J Med Internet Res 2015;17(6):e146
URL: http://www.jmir.org/2015/6/e146/
doi:10.2196/jmir.4415
PMID:26070951

Review

Patient Portals and Patient Engagement: A State of the Science Review

Taya Irizarry[1], BSN, MSN; Annette DeVito Dabbs[1], RN, FAAN, PhD; Christine R Curran[1], RN, PhD

[1] School of Nursing, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Taya Irizarry, BSN, MSN
School of Nursing
University of Pittsburgh
336 Victoria Building
3500 Victoria Street
Pittsburgh, PA, 15261
United States
Phone: 1 412 624 5314
Fax: 1 412 383 7227
Email: tai19@pitt.edu

Abstract

Background: Patient portals (ie, electronic personal health records tethered to institutional electronic health records) are recognized as a promising mechanism to support greater patient engagement, yet questions remain about how health care leaders, policy makers, and designers can encourage adoption of patient portals and what factors might contribute to sustained utilization.

Objective: The purposes of this state of the science review are to (1) present the definition, background, and how current literature addresses the encouragement and support of patient engagement through the patient portal, and (2) provide a summary of future directions for patient portal research and development to meaningfully impact patient engagement.

Methods: We reviewed literature from 2006 through 2014 in PubMed, Ovid Medline, and PsycInfo using the search terms “patient portal” OR “personal health record” OR “electronic personal health record”. Final inclusion criterion dictated that studies report on the patient experience and/or ways that patients may be supported to make competent health care decisions and act on those decisions using patient portal functionality.

Results: We found 120 studies that met the inclusion criteria. Based on the research questions, explicit and implicit aims of the studies, and related measures addressed, the studies were grouped into five major topics (patient adoption, provider endorsement, health literacy, usability, and utility). We discuss the findings and conclusions of studies that address the five topical areas.

Conclusions: Current research has demonstrated that patients’ interest and ability to use patient portals is strongly influenced by personal factors such age, ethnicity, education level, health literacy, health status, and role as a caregiver. Health care delivery factors, mainly provider endorsement and patient portal usability also contribute to patient’s ability to engage through and with the patient portal. Future directions of research should focus on identifying specific populations and contextual considerations that would benefit most from a greater degree of patient engagement through a patient portal. Ultimately, adoption by patients and endorsement by providers will come when existing patient portal features align with patients’ and providers’ information needs and functionality.

(J Med Internet Res 2015;17(6):e148) doi:10.2196/jmir.4255

KEYWORDS

electronic personal health record; patient portal; patient engagement; meaningful use

Introduction

Patient Engagement and Patient Portals

Patient engagement has been identified as an essential dimension of the multifaceted solution to the cost/quality crisis in US health care. The patient-centric definition of patient engagement by the Agency for Healthcare Research and Quality (AHRQ) is “the involvement in their own care by individuals (and others they designate to engage on their behalf), with the goal that they make competent, well-informed decisions about their health and health care and take action to support those decisions” [1].

AHRQ also defines patient engagement from a systems perspective as “a set of behaviors by patients, family members, and health professionals and a set of organizational policies and procedures that foster both the inclusion of patients and family members as active members of the health care team and collaborative partnerships with providers and provider organizations” [1].

Currently, there is an increasing awareness of health care system’s responsibility to provide easily accessible ways for patients to be engaged in their own care by creating effective partnerships that lead to the patient’s ability to make competent and well-informed decisions [2]. While an electronic personal health record (ePHR) tethered to an electronic health record (EHR), also known as a patient portal, is currently recognized as a promising mechanism to support greater patient engagement, questions remain about how health care leaders, policy makers, and designers can encourage adoption by both providers and patients and what factors might contribute to sustained utilization.

Definition and Background of Patient Portals

An ePHR that directly links, or is “tethered”, to an EHR is most commonly referred to as a patient portal. In general, patient information from the EHR such as the problem list, allergies, and lab test results populate the patient portal. In some instances, patients may enter data to populate the EHR. In contrast, an untethered ePHR is under the control of the patient. This means an individual manually enters all information or grants permission for the information to be transferred to the ePHR, from a specific source like a laboratory or pharmacy, and determines who will have access. Thus, the value of an untethered ePHR is determined by a person’s willingness to manage and maintain their ePHR information. Because there is little that health care organizations can do to initiate patient engagement using an untethered ePHR, this literature review is focused exclusively on the patient portal, directly linked to an EHR.

Patient portals were introduced and adopted by a few large health care organizations in the late 1990s (eg, MyChart at the Palo Alto Medical Foundation and Indivo at Boston Children’s Hospital) [3,4]. However, patient portals did not gain widespread use until 2006 when several initiatives coincided, including the launch of ePHRs by Microsoft and Google, the awarding of Centers for Medicare and Medicaid Services (CMS) contracts to private firms to conduct feasibility studies of ePHRs using existing claims data from Medicare programs, and Blue Cross and Blue Shield Association and America’s Health Insurance Plans’ announcement to develop data-sharing programs that would ultimately support ePHR development [5]. These initiatives also coincided with the broad social movement towards adoption and daily use of powerful information and communication sharing tools such as smartphones and social media, illustrating the readiness of the general population to embrace technology in a new socially interactive way.

The current principal driver of patient portal development is the meaningful use (MU) criteria of the CMS EHR incentive program [6]. Features mandated by MU that directly relate to patient portal functionality include providing (1) a clinical summary to the patient after each visit, (2) secure messaging (SM) between patient and provider, (3) ability to view, download, and transmit personal health record data, (4) patient specific education, (5) patient reminders for preventative services, and (6) medication reconciliation [7]. While these criteria clearly outline tasks and goals, they do little to reflect the value proposition to the end users (patients and providers) or the steps required to engage patients in a sustained and relevant way. Therefore, an aim of this review was to explore the current research addressing the encouragement and support of patient engagement through the patient portal.

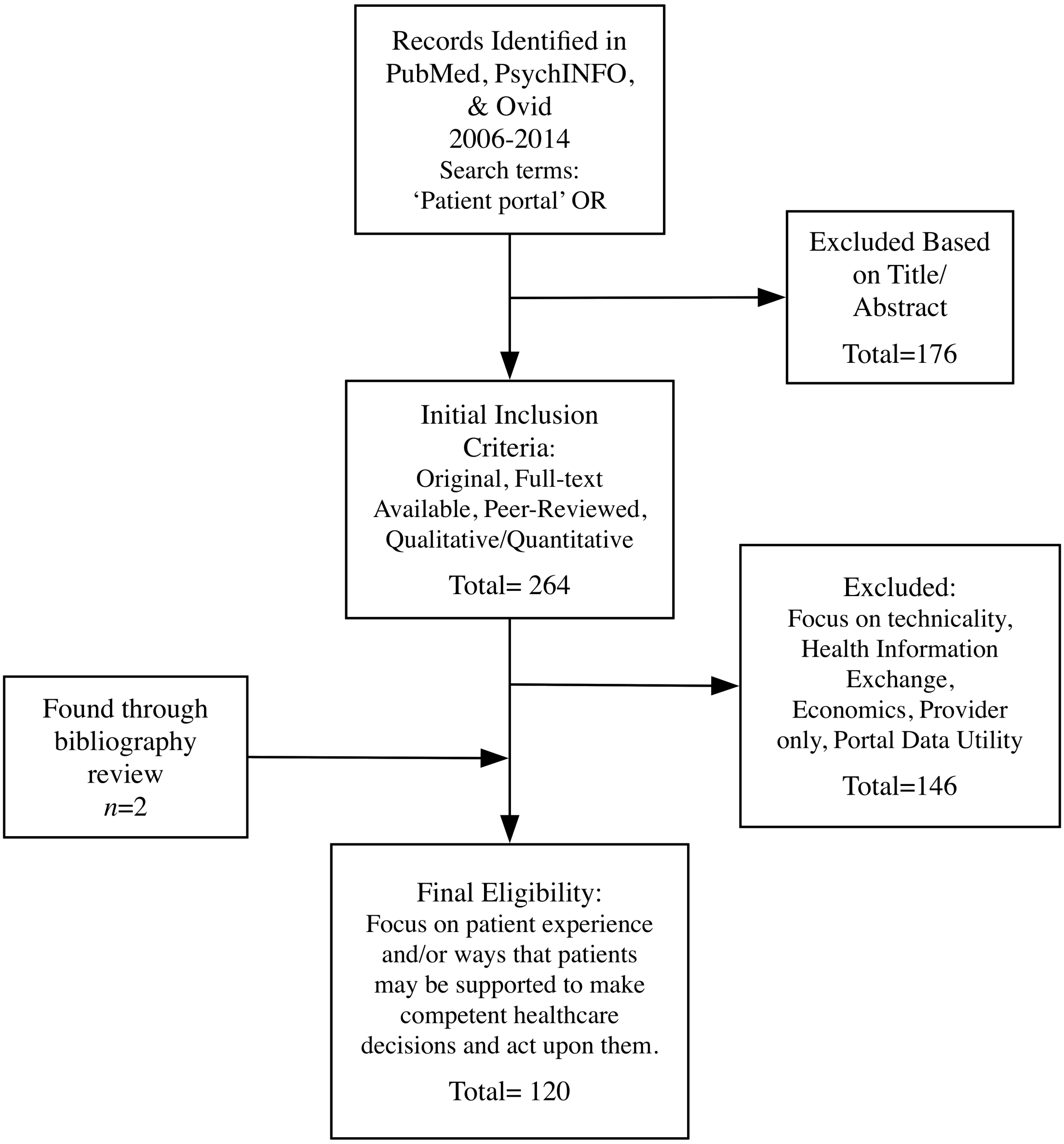
Methods

Search Strategy

Due to the advances in technology and consumer readiness in the mid-2000s, the review was limited to recent literature to better reflect current trends in design, functionality, and perceived user readiness of patient portals. We reviewed literature from 2006 through 2014 in PubMed, Ovid Medline, and PsycInfo using the search terms “patient portal” OR “personal health record” OR “electronic personal health record”. Bibliographies and the literature reviews from these sources were used to identify additional studies [8,9]. Initial inclusion criteria were (1) original, peer-reviewed, qualitative, and quantitative research of tethered ePHRs or patient portals, (2) English language, and (3) available in full text. The final inclusion criterion was that the studies reported on the patient experience and/or ways that patients may be supported to make competent health care decisions and act on those decisions using patient portal functionality. Studies were not targeted to any particular patient subgroup, disease, or clinical setting.

Of the 440 articles identified by the search, 176 were excluded based on title and abstract. Further review based on the final inclusion criterion resulted in 120 articles, which were reviewed in depth (see Multimedia Appendix 1 for summaries of each). Excluded articles focused on the provider perspective only, technicalities of patient portal implementation (eg, policy issues, safety, security), implications for Health Information Exchange, economics impacts, or the utility of patient portal data for research purposes (see Figure 1).

Figure 1. Literature review flow chart.



Results

Overview

We grouped the studies into five major topics based on the research questions, explicit and implicit aims of the studies, and related measures addressed. The topics identified included patient adoption, provider endorsement, health literacy, usability, and utility (Table 1). Of the 120 articles that were reviewed, 66 (55.0%) were non-experimental descriptive, 26 (21.7%) were qualitative or mixed-methods, 14 (11.7%) were randomized controlled trials, 10 were pilot studies or case reports (8.3%), and 4 were cohort studies (3.3%) (Table 2). Only 11 articles explicitly identified a guiding theoretical framework, with the Chronic Care Model being the most common among them. The year 2011 was a turning point in the number of published articles, which coincides with the initiation of CMS EHR incentives program. The topical areas that showed the greatest increase in volume were patient adoption and utility (Table 1). See Multimedia Appendix 1 for a brief description of each article and the topical areas addressed. The following section describes each topical area and discusses relevant implications for research, development, and implementation of patient portals.

Table 1. Summary of articles on categories of patient portals for patient engagement.

Year	Provider endorsement	Health literacy	Usability	Patient adoption	Utility	Total # of articles
2006	0	0	0	1	2	3
2007	1	1	1	2	3	4
2008	0	1	1	4	6	8
2009	0	0	1	3	4	7
2010	1	2	2	7	7	11
2011	3	3	2	11	8	17
2012	1	3	3	11	10	16
2013	2	3	5	12	17	27
2014	0	3	5	11	19	27
Total	8	16	20	62	76	120

Table 2. Levels of evidence adapted from Melnyk & Fineout-Overholt, 2005.

Type of study	Level of evidence	# of studies
RCT	2	14
Cohort/Quasi-experimental	3	4
Descriptive	4	
Non-experimental (survey, correlational, etc)		66
Qualitative/Mixed method		26
Pilot study/case report	5	10

Patient Adoption

Before a patient portal can serve as a tool for individuals to become more engaged and involved in their own care, patients must first adopt it. CMS 2014 stage 2 MU regulations define adoption in terms of institutional reporting for reimbursement and require that 5% of the institutions' patient population (1) download or view electronic health information and (2) use secure electronic messages (eg, email) [6]. However, in our review, various operational definitions of adoption were used. For example, many observational studies used usage data of the initial login to the patient portal site to represent adoption; others used data from surveys about patients' intention to use the portal. Several randomized controlled trials (RCTs) used rates of patient portal intervention adherence to study protocol to define adoption, and for some of these trials, those who completed the studies were considered adopters; in others, adoption was defined as the frequency of intervention use.

Of the 62 articles [5,10-70] that focused on or described patient portal adoption as part of the report, six RCTs included detailed descriptions of intervention group participants who completed the study (and therefore were considered adopters) in comparison to those who did not. We found 12 qualitative or mixed-method studies that collected data about adoption from patients through focus groups or semistructured interviews; 21 studies focused on interest and barriers to adoption for specific populations or patient portal functions (eg, elderly, safety-net, human immunodeficiency populations, secure messaging, prescription refills).

The term "digital divide" is often used to describe major potential barriers to access of electronic tools such as a patient portal and refers to disparities among subgroups based on access to the Internet and computer literacy. However, this term does not encompass the many other factors that may contribute to adoption such as language barriers, age, race and ethnicity, social economic status, and level of patient activation [32,50,54,71]. Several studies examining adoption have shown that ethnic minorities (African American, Latino, Asian) and patients who are younger (under 35 years), healthier, and less educated were less likely to adopt patient portals [15,55,72]; however, results are mixed regarding gender differences [50,63]. People with disabilities and chronic conditions, frequent users of health care services, and caregivers of elderly parents or children tend to have the most interest in patient portals [28,50,62,73]. Other important factors of patient portal adoption include provider acceptance and promotion, and usability of the patient portal interface including ease of registration, navigation, and perceived privacy and security [18-20,74].

Provider Endorsement

Provider endorsement and continued engagement with the patient portal have been identified as important factors in a patient's decision to adopt and continue to use the patient portal functions to achieve and sustain anticipated positive outcomes [19,75]. Of the 8 articles that addressed physician endorsement [12,19,34,76-80], 5 studies were qualitative or mixed-method studies, and one RCT included a retrospective survey of physicians' use and satisfaction.

Four of the studies sought to capture attitudes of clinicians towards patient portals prior to having firsthand experience interacting with them. Prior to actual use of patient portals, clinicians expressed concerns related to patient engagement including: the potential for inducing patient anxiety regarding test results; the accuracy of patient entered data; the potential liability for tracking and acting on critical clinical information, such as blood glucose levels and blood pressure readings; implications for changes in the patient-provider relationship; and the anticipated increased workload [34,77,78,81].

Retrospective studies showed that the pre-portal concerns regarding patient anxiety about test results were not justified as demonstrated by numerous patients who found the test result feature one of the most useful [82]. In addition, while perceived increases in workload and duration of clinic visits varied among studies, clinicians believed patients were more interested in participating in their care and found that verifying the additional information in the patient portal provided during face-to-face visits was helpful, thus eliminating the accuracy concern [19]. Overall, the workflow of individual providers and the health care team as a whole, including nurses, pharmacists, support staff, and physicians, must be adapted in order to incorporate patient portal functionality, and the patient engagement it allows, into the delivery of preventative services and illness management processes [45].

Health Literacy

The definition of health literacy developed for the National Library of Medicine and used by the Healthy People 2010 initiative is "the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions" [83]. Of the 16 studies that specifically addressed health literacy [11,14,30,40,42,64,65,74,84-91], the majority included self-reported health literacy measures via survey questions or open-ended questions; only Noblin et al (42) and Taha et al [91] included validated health literacy measures. Four studies [64,85,88,91] identified conceptual knowledge, numeracy, and computer skills as particularly important literacy factors that contributed to successful patient engagement via a patient portal.

Noblin et al [42] found that 65% of participants who intended to adopt the outpatient clinic's patient portal had a higher eHealth literacy score than those who were not interested in patient portal adoption. Taha et al [91] results indicated that if health texts involved numeric concepts, users encountered problems, even if they were considered to have "adequate" health literacy. These studies underscore the importance of evaluating health literacy and health numeracy separately in order to identify specific risk factors and design flaws that could impact patient comprehension and ultimately jeopardize the accuracy of patient input and interpretation of results.

Four studies directly addressed the impact of health literacy of intended users on the successful completion of specific tasks [64,84,88,92]. Results showed that patients responded better when medical jargon and abbreviations were translated into "patient friendly" language. These results echo Haggstrom et al [85] and Monkman & Kushniruk's [88] findings of the dangers of low health and computer literacy to safe and effective use of patient portals.

Schnipper et al [92] and Sox et al [84] revealed that, despite patient involvement in early design and testing of patient portals, subsequent scenario-based usability testing uncovered navigation difficulties primarily due to the unfamiliarity with complex medical language and confusion of how and when to correct identified errors. Monkman & Kushniruk [88] suggest that including health literacy assessments in usability testing of consumer health information systems, such as patient portals, would inform the design of systems for better navigation, data input, and conceptual understanding of health information included throughout the patient portal.

Monkman & Kushniruk [88] also proposed a specific heuristic for health literacy whose purpose is to identify and categorize when clinical information within the patient portal would most likely be misunderstood by a layperson who does not possess a health care background. This study, along with several other qualitative studies showed that specific health topics (eg, medications, lab results, and allergies) required extra attention to designing with health literacy considerations in mind [45,89,93]. Proposed navigation and aiding tools that increased patients' ability to understand their personal health information more fully include integrating links to definitions of terms and detailed explanations, using movies and illustrations, substituting lay language for medical terminology and using graphs to track trending data, such as blood pressure and blood glucose levels [84,85,94].

Usability

Usability testing is the term used to describe the assessment of how easy a user interface is to operate. The word "usability" also refers to methods for improving ease of use during the design process [95]. One such method is heuristic evaluation, a method of testing a preliminary prototype by examining the interface and judging its compliance with recognized usability principles (ie, "heuristics"). Further iterative usability testing is accomplished using a series of prototypes and participatory scenario-based and "think-aloud" sessions with intended users in order to redesign the interface and workflows to better match user needs and preferences. Early usability testing, and its role in patient portal design, is important because it directly impacts whether or not a patient can easily adopt a patient portal. It also impacts the ability of the user to successfully navigate portal functions, accurately input information, and comprehend the information presented, ultimately contributing to its usefulness as a tool for patient engagement.

Of the 20 studies that addressed usability of patient portals, 6 performed some form of heuristic and usability testing with objective observation and various forms of "think aloud" sessions [25,84,85,92,94,96]. Only Schnipper et al [92] included usability testing of both the clinician and patient interfaces. The remaining 14 studies assessed users' subjective satisfaction and ease of use with questionnaires and/or interviews to evaluate overall adoption and utilization [11,38,45,47,48,64,65,73,82,88,89,91,97,98].

Schnipper et al [92] addressed the needs of both end users (ie, clinicians and patients) in the usability testing of a medication management module embedded within the patient portal. The study highlighted the need for end user-specific interfaces and functionality in order to make the user experience easier and more efficient, thus demonstrating its value and promoting sustained use. For patients, this meant striking a balance between free-text, structured, and coded data fields in order to leverage the usefulness of patient-entered data without confusing or overwhelming patients. For example, drop-down menus and scrolls bars were found to be less confusing and more efficient than dynamic text boxes that would react to the word being typed when inputting data, such as medications and allergies. In the case of clinicians, this meant integrating the clinician side of the application with their workflow so that clinicians could verify and correct patient-entered data while simultaneously facilitating the flow of that data into the EHR.

Much of the literature surrounding usability confirms that adoption and sustained use of technology are directly related to ease of navigation and the perceived usefulness of the available information [99]. While nearly all the patient portal usability studies that used subjective assessments showed positive results for ease of use and satisfaction, the in-depth objective usability studies were more effective at uncovering a variety of barriers to safe and effective use.

Utility

Utility refers to the availability of needed features. Utility and usability are equally important and together determine whether something is useful [99]; 76 studies focused in some way on patient portal utility [5,12,13,15,19,22,23,25-27,30,34,37,41,44,47,52,53,56,57,59,60,64,65,69,70,79,82,84-87,89,90,92,96,98,100-137]. The majority of descriptive, qualitative, and mixed-method studies focused on eliciting patient preferences for specific functions. Patients preferred functions that offered convenience, such as an easy way to contact and communicate with providers, order prescription refills, and access multiple family medical records. Easy-to-read, printer-friendly summaries were also viewed as helpful for sharing information with family members and providers who did not have patient portal access. The top two patient portal qualities that were deemed most utilitarian for patients were personalization and collaborative communication between patients and providers [67,138].

Personalization

While numerous descriptive and qualitative studies attest to the desire for personalized patient portal functionality, there is little research about what kind of personalization would lead to greater patient engagement. Currently, the greatest research focus is on chronic disease medication management and preventative services. Only 3 RCTs specifically tested the efficacy of patient-tailored interventions [13,30,90]. Grant et al [13] provided patient-tailored decision support and enabled the patient to author a “Diabetes Care Plan” for electronic submission to the physician prior to upcoming appointments. This intervention led to increases in pre-visit use of the patient portal and increased rates of diabetes-related medication adjustment at 12 months. Krist et al [62] provided a personally tailored list of prevention recommendations and found that at 16 months, 1 in 4 users were up-to-date on all preventive services—nearly double that of non-users. Sequist et al [30] sent personalized electronic messages that included (1) alerts for overdue health screenings and information on screening options, (2) a mechanism for patients to submit requests to schedule screening examinations, and (3) a link to a Web-based tool for patients to assess their personal risk of colorectal cancer. Findings showed that screening rates were significantly higher at 1 month for patients who received electronic messages than for those who did not, but the difference was no longer significant at 4 months.

Collaborative Communication

Collaborative communication refers to the ability for patients and providers to share timely and pertinent information, enabling patients to participate as active members of the care team beyond the hospital or clinic setting. SM and medication reconciliation are the two most common patient portal functions that offer the opportunity for such communication. Both functions also pose the greatest potential changes to provider workflow and overall impact on the patient-provider relationship.

For example, the difficulty aligning information management tools with current provider workflow and care delivery priorities was highlighted in a study of an interactive medication reconciliation module that emailed primary care physicians when a patient added or changed information [106]. Results showed that patients were willing and able to annotate their medication list, offering the most up-to-date and complete information, but email notifications were ineffective at prompting providers to update the EHR medication list outside of a clinic visit [106]. Thus, while the notion of designing patient portals to support patient involvement in their care, such as opportunities for their participation in medication reconciliation, shows promise, their effectiveness will depend on the ability to better incorporate these functions into provider workflow and delivery of care.

Other implications of electronic forms of communication via a patient portal are the potential to improve efficiency by way of substituting SM for face-to-face encounters and using SM reminders to decrease missed appointments and promote timely preventative care. However, research shows mixed results leading researchers to believe that the relationship between SM and utilization is more complex than the simple substitution of online for in-person care suggests. For example, while an earlier study at Kaiser Permanente showed a decrease in face-to-face encounters after the initiation of SM [22], a subsequent study in a different Kaiser region showed the opposite effect [115]. A study done at the Mayo Clinic, aimed at clarifying this discrepancy, focused on frequency of messages, long-term use, and importance of SM among certain subgroups [121], which showed neither an increase nor decrease in face-to-face provider visits with the use of SM.

SM is also being used as a one-way communication tool to deliver reminders for preventative care and appointments. A 2011 study at seven Duke medical clinics showed that email reminders, in combination with scheduling functionality within the patient portal, demonstrated significant declines in “no-shows” [27]. A meta-analysis and systematic review by

Guy et al [139] demonstrated a substantial increase in the likelihood of attending clinic appointments when patients received SM reminders. Perhaps the most encouraging results with SM were the large reduction in missed appointments among historically disadvantaged groups, such as Medicaid recipients, the uninsured, and black patients [27].

SM reminders via email have also been shown to be generally successful at encouraging higher rates of preventative services use. For example, a multi-practice randomized controlled trial showed improvement in the rates of certain preventive screenings and vaccinations, but preventative services as a whole were not impacted [113]. Findings suggest that SM reminders are most effective when they are tailored to the population and context, thus targeting specific goals such as herpes zoster vaccinations for older adults, or pediatric preventative care visit reminders for parents [119,129].

Discussion

Principal Findings

The current principal driver of patient portal development is CMS and Medicaid EHR incentive program meaningful use (MU) criteria [6]. While MU criteria clearly outline requirements of basic functionality and targeted adoption rates, they do not delineate the steps or features required to engage patients in a sustained and relevant way. Presently there is no clear definition of patient portal adoption beyond the minimum use requirements outlined in the MU criteria. However, in order for health care institutions to track the success of patient portals in terms of patient engagement, a multi-dimensional definition of portal adoption should include both motivating factors for initiation and use over time A definition of this kind would inform a set of universal quality and efficiency reporting measures beyond the current minimal MU criteria to include more relevant patient engagement data.

Current research has demonstrated that patients' interest and ability to use patient portals is strongly influenced by personal factors such age, ethnicity, education level, health literacy, health status, and role as a caregiver. Health care delivery factors, mainly provider endorsement and patient portal usability, also contribute to patients' ability to engage through and with the patient portal.

While health literacy has been identified as an important factor in the successful use of patient portals, few studies have used validated health literacy measures, making it difficult for future research to build on the findings. Research demonstrates that specific aspects of health literacy, mainly numeracy and familiarity with medical terminology, greatly impact the ability of patients to accurately input data and interpret the information provided in the patient portal. The direct relationship between health literacy and effective use of the patient portal supports the argument for the use of specific health literacy heuristics as part of overall usability testing.

Research also demonstrates that objective testing (as opposed to solely subjective) should also be a part of patient portal usability testing. Although objective usability testing is expensive and time consuming, studies demonstrate the need for continued work in this area in order to ensure patient portal interfaces promote patient comprehension and data entry accuracy. The promotion of content accuracy and patient comprehension impacts the overall usefulness of the information for both patients and providers.

The perceived usefulness of patient portals from the providers' perspectives cannot be underestimated. Provider endorsement is one of the most influential factors impacting patients' initial adoption, as well as its continued use as a tool for collaborative communication [20]. Yet, current research demonstrates the difficulty in aligning information management tools, such as the patient portal, with current provider workflow and care delivery priorities.

While current development and research is focused on demonstrating feasibility and efficiency of medication reconciliation and SM reminders, the research has revealed roadblocks to successful implementation rooted in the lack of provider workflow adaptations A greater understanding of the essential adjustments in provider workflow, including potential changes in the roles and responsibilities of the care team overall, is necessary in order to translate findings into practice. Few studies have focused on exploring how patient portal use should unfold within the context of the patient-provider interaction, or how it might impact the overall organization and workflow of the health care team including potential liability concerns, reimbursement, and relationships with patients.

Ultimately, successful implementation requires health care institutions to invest time and resources to systematically assess the health needs of their specific patient and caregiver populations, their individual stages of readiness to adopt a patient portal, and the types of assistance needed to do so [140]. Ideally, interactive sites would collect information on individuals' health, health behaviors and personal goals, and assess health literacy and functional ability, which would then inform the adaptation of the patient portal to accommodate the needs of the individual and/or what additional or alternative resources may be useful [2]. Such adaptations include personalized content and tailored data presentations specifically designed to enhance interpretation and comprehension of key personal health concerns and timely and pertinent action steps.

In addition, external environmental and contextual factors, such as distance between patient and clinic, and complexity and trajectory of health concerns, may impact which form of access is preferred for a specific person, provider, location, and situation. Future directions of research should focus on identifying specific populations and contextual considerations that would benefit most from a greater degree of patient engagement through a patient portal. This information could then lead to the creation of health care service policies that promote the use of a patient portal by both providers and patients within the most appropriate settings.

Conclusions

If institutions are to engage patients via the patient portal in a way that encourages them to become active members of the care team, support their competence in making health-related decisions, and help them to act on those decisions, institutional leaders must consider the contributing factors that impact efficacy and sustained use of patient portals. According to this review, these factors include attention to the topical areas of patient adoption, provider endorsement, health literacy, usability, and utility. Ultimately, adoption by patients and endorsement by providers will come when existing patient portal features align with patients' and providers' information needs and functionality. Conceptualizing patient portals as a dynamic component of the patient-provider relationship and health care delivery system as a synergetic whole, rather than an isolated repository of information or a set of disconnected functions meant to collect patient data for provider use, may help to inform future research, improve patient portal design, and efforts to promote adoption and effectiveness.

Acknowledgments

TI was funded on T32: Technology Research in Chronic and Critical Illness (2NR008857, Erlen & DeVito Dabbs, Directors). The funding agency had no role in the preparation, review, or approval of the manuscript. The opinions, results, and conclusions reported in this paper are those of the authors and are independent of the funding source.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A brief description of each article and the topical areas it addresses.

[PDF File (Adobe PDF File), 306KB - jmir_v17i6e148_app1.pdf]

References

1. Rockville MD, Maurer M, Dardess P, Carman KL, Frazier K, Smeeding L. Guide to Patient and Family Engagement: Environmental Scan Report. 2012. URL: http://www.ahrq.gov/research/findings/final-reports/ptfamilyscan/ptfamilyscan.pdf [accessed 2015-01-17] [WebCite Cache ID 6Vel4kPos]
2. Kupchunas WR. Personal health record: new opportunity for patient education. Orthop Nurs 2007;26(3):185-91; quiz 192. [doi: 10.1097/01.NOR.0000276971.86937.c4] [Medline: 17538475]
3. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
4. Mandl KD, Simons WW, Crawford WCR, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC Med Inform Decis Mak 2007;7:25 [FREE Full text] [doi: 10.1186/1472-6947-7-25] [Medline: 17850667]
5. Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. J Med Internet Res 2009;11(2):e14 [FREE Full text] [doi: 10.2196/jmir.1187] [Medline: 19403467]
6. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]
7. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
8. Kim J, Jung H, Bates DW. History and Trends of "Personal Health Record" Research in PubMed. Healthc Inform Res 2011 Mar;17(1):3-17 [FREE Full text] [doi: 10.4258/hir.2011.17.1.3] [Medline: 21818452]
9. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
10. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]
11. Keselman A, Slaughter L, Smith CA, Kim H, Divita G, Browne A, et al. Towards consumer-friendly PHRs: patients' experience with reviewing their health records. AMIA Annu Symp Proc 2007:399-403 [FREE Full text] [Medline: 18693866]
12. Wald JS, Burk K, Gardner K, Feygin R, Nelson E, Epstein M, et al. Sharing electronic laboratory results in a patient portal--a feasibility pilot. Stud Health Technol Inform 2007;129(Pt 1):18-22. [Medline: 17911670]
13. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 8;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]

14. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. BMJ 2008 Jun 7;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
15. Lafky DB, Horan TA. Prospective personal health record use among different user groups: results of a multi-wave study. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences.: IEEE; 2008 Presented at: 41st Annual International Conference on System Sciences; Jan. 7-10, 2008; Hawaii. [doi: 10.1109/HICSS.2008.363]
16. Zickmund SL, Hess R, Bryce CL, McTigue K, Olshansky E, Fitzgerald K, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med 2008 Jan;23 Suppl 1:20-26 [FREE Full text] [doi: 10.1007/s11606-007-0273-6] [Medline: 18095039]
17. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009 Apr;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
18. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
19. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]
20. Wald JS. Variations in patient portal adoption in four primary care practices. AMIA Annu Symp Proc 2010;2010:837-841 [FREE Full text] [Medline: 21347096]
21. Wen KY, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
22. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375 [FREE Full text] [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
23. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
24. Kahn JS, Hilton JF, Van NT, Leasure S, Bryant KM, Hare CB, et al. Personal health records in a public hospital: experience at the HIV/AIDS clinic at San Francisco General Hospital. J Am Med Inform Assoc 2010;17(2):224-228 [FREE Full text] [doi: 10.1136/jamia.2009.000315] [Medline: 20190069]
25. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010;341:c5814 [FREE Full text] [Medline: 21081595]
26. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
27. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011;13(2):e41 [FREE Full text] [doi: 10.2196/jmir.1702] [Medline: 21616784]
28. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
29. 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]
30. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med 2011 Apr 11;171(7):636-641 [FREE Full text] [doi: 10.1001/archinternmed.2010.467] [Medline: 21149743]
31. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
32. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
33. North F, Hanna BK, Crane SJ, Smith SA, Tulledge-Scheitel SM, Stroebel RJ. Patient portal doldrums: does an exam room promotional video during an office visit increase patient portal registrations and portal use? J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i24-i27 [FREE Full text] [doi: 10.1136/amiajnl-2011-000381] [Medline: 21976028]

34. McInnes DK, Solomon JL, Bokhour BG, Asch SM, Ross D, Nazi KM, et al. Use of electronic personal health record systems to encourage HIV screening: an exploratory study of patient and provider perspectives. BMC Res Notes 2011;4:295 [FREE Full text] [doi: 10.1186/1756-0500-4-295] [Medline: 21843313]
35. Levy MA, Giuse DA, Eck C, Holder G, Lippard G, Cartwright J, et al. Integrated information systems for electronic chemotherapy medication administration. J Oncol Pract 2011 Jul;7(4):226-230 [FREE Full text] [doi: 10.1200/JOP.2011.000259] [Medline: 22043185]
36. Jung C, Padman R, Shevchik G, Paone S. Who are portal users vs. early e-Visit adopters? A preliminary analysis. AMIA Annu Symp Proc 2011;2011:1070-1079 [FREE Full text] [Medline: 22195168]
37. Nagykaldi Z, Aspy CB, Chou A, Mold JW. Impact of a Wellness Portal on the delivery of patient-centered preventive care. J Am Board Fam Med 2012 Apr;25(2):158-167 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110130] [Medline: 22403196]
38. Nielsen AS, Halamka JD, Kinkel RP. Internet portal use in an academic multiple sclerosis center. J Am Med Inform Assoc 2012 Feb;19(1):128-133 [FREE Full text] [doi: 10.1136/amiajnl-2011-000177] [Medline: 21571744]
39. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
40. Johnson AJ, Easterling D, Nelson R, Chen MY, Frankel RM. Access to radiologic reports via a patient portal: clinical simulations to investigate patient preferences. J Am Coll Radiol 2012 Apr;9(4):256-263. [doi: 10.1016/j.jacr.2011.12.023] [Medline: 22469376]
41. Rosenbloom ST, Daniels TL, Talbot TR, McClain T, Hennes R, Stenner S, et al. Triaging patients at risk of influenza using a patient portal. J Am Med Inform Assoc 2012;19(4):549-554 [FREE Full text] [doi: 10.1136/amiajnl-2011-000382] [Medline: 22140208]
42. Noblin AM, Wan TTH, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
43. Wakefield DS, Kruse RL, Wakefield BJ, Koopman RJ, Keplinger LE, Canfield SM, et al. Consistency of patient preferences about a secure Internet-based patient communications portal: contemplating, enrolling, and using. Am J Med Qual 2012;27(6):494-502. [doi: 10.1177/1062860611436246] [Medline: 22517909]
44. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol YH, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012 Aug;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
45. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res 2012;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
46. Tsai J, Rosenheck RA. Use of the internet and an online personal health record system by US veterans: comparison of Veterans Affairs mental health service users and other veterans nationally. J Am Med Inform Assoc 2012;19(6):1089-1094 [FREE Full text] [doi: 10.1136/amiajnl-2012-000971] [Medline: 22847305]
47. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
48. Ahlers-Schmidt CR, Nguyen M. Parent intention to use a patient portal as related to their children following a facilitated demonstration. Telemed J E Health 2013 Dec;19(12):979-981. [doi: 10.1089/tmj.2013.0041] [Medline: 24050608]
49. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
50. Sanders MR, Winters P, Fortuna RJ, Mendoza M, Berliant M, Clark L, et al. Internet access and patient portal readiness among patients in a group of inner-city safety-net practices. J Ambul Care Manage 2013;36(3):251-259. [doi: 10.1097/JAC.0b013e31829702f9] [Medline: 23748274]
51. Ronda MCM, Dijkhorst-Oei LT, Gorter KJ, Beulens JWJ, Rutten GEHM. Differences between diabetes patients who are interested or not in the use of a patient Web portal. Diabetes Technol Ther 2013 Jul;15(7):556-563 [FREE Full text] [doi: 10.1089/dia.2013.0023] [Medline: 23777369]
52. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
53. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to Electronic Health Records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
54. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. J Am Med Inform Assoc 2013;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
55. 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]

56. Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. BMJ Open 2013;3(7):1-10 [FREE Full text] [doi: 10.1136/bmjopen-2013-002931] [Medline: 23901027]
57. Hanberger L, Ludvigsson J, Nordfeldt S. Use of a web 2.0 portal to improve education and communication in young patients with families: randomized controlled trial. J Med Internet Res 2013;15(8):e175 [FREE Full text] [doi: 10.2196/jmir.2425] [Medline: 23973555]
58. Butler JM, Carter M, Hayden C, Gibson B, Weir C, Snow L, et al. Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users. AMIA Annu Symp Proc 2013;2013:152-161 [FREE Full text] [Medline: 24551328]
59. Gu Y, Day K. Propensity of people with long-term conditions to use personal health records. Stud Health Technol Inform 2013;188:46-51. [Medline: 23823287]
60. Gerber DE, Laccetti AL, Chen B, Yan J, Cai J, Gates S, et al. Predictors and intensity of online access to electronic medical records among patients with cancer. J Oncol Pract 2014 Sep;10(5):e307-e312 [FREE Full text] [doi: 10.1200/JOP.2013.001347] [Medline: 25006222]
61. Kondylakis H, Kazantzaki E, Koumakis L, Genitsaridi I, Marias K, Gorini A, et al. Development of interactive empowerment services in support of personalised medicine. Ecancermedicalscience 2014;8:400 [FREE Full text] [doi: 10.3332/ecancer.2014.400] [Medline: 24567757]
62. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
63. Jung C, Padman R. Virtualized healthcare delivery: understanding users and their usage patterns of online medical consultations. Int J Med Inform 2014 Dec;83(12):901-914. [doi: 10.1016/j.ijmedinf.2014.08.004] [Medline: 25193501]
64. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
65. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
66. Ancker JS, Silver M, Kaushal R. Rapid growth in use of personal health records in New York, 2012-2013. J Gen Intern Med 2014 Jun;29(6):850-854. [doi: 10.1007/s11606-014-2792-2] [Medline: 24519102]
67. Abramson EL, Patel V, Edwards A, Kaushal R. Consumer perspectives on personal health records: a 4-community study. Am J Manag Care 2014 Apr;20(4):287-296 [FREE Full text] [Medline: 24884860]
68. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
69. Neuner J, Fedders M, Caravella M, Bradford L, Schapira M. Meaningful use and the patient portal: patient enrollment, use, and satisfaction with patient portals at a later-adopting center. Am J Med Qual 2015;30(2):105-113. [doi: 10.1177/1062860614523488] [Medline: 24563085]
70. Riippa I, Linna M, Rönkkö I. The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study. J Med Internet Res 2014;16(11):e257 [FREE Full text] [doi: 10.2196/jmir.3462] [Medline: 25413368]
71. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2014 Apr 30:1-13. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
72. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
73. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 2013 Dec;28(4):1124-1139. [doi: 10.1037/a0033911] [Medline: 24364414]
74. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
75. Logue MD, Effken JA. Modeling factors that influence personal health records adoption. Comput Inform Nurs 2012 Jul;30(7):354-362. [doi: 10.1097/NXN.0b013e3182510717] [Medline: 22525046]
76. Wynia MK, Torres GW, Lemieux J. Many physicians are willing to use patients' electronic personal health records, but doctors differ by location, gender, and practice. Health Aff (Millwood) 2011 Feb;30(2):266-273 [FREE Full text] [doi: 10.1377/hlthaff.2010.0342] [Medline: 21289348]
77. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]

78. Yau GL, Williams AS, Brown JB. Family physicians' perspectives on personal health records: qualitative study. Can Fam Physician 2011 May;57(5):e178-e184 [FREE Full text] [Medline: 21642732]
79. van der Vaart R, Drossaert CHC, Taal E, van de Laar MAFJ. Giving rheumatology patients online home access to their electronic medical record (EMR): advantages, drawbacks and preconditions according to care providers. Rheumatol Int 2013 Sep;33(9):2405-2410 [FREE Full text] [doi: 10.1007/s00296-012-2408-2] [Medline: 22453527]
80. Urowitz S, Wiljer D, Apatu E, Eysenbach G, Delenardo C, Harth T, et al. Is Canada ready for patient accessible electronic health records? A national scan. BMC Med Inform Decis Mak 2008;8:33 [FREE Full text] [doi: 10.1186/1472-6947-8-33] [Medline: 18652695]
81. Witry MJ, Doucette WR, Daly JM, Levy BT, Chrischilles EA. Family physician perceptions of personal health records. Perspect Health Inf Manag 2010;7:1d [FREE Full text] [Medline: 20697465]
82. Pai HH, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec;20(6):e561-e569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]
83. Selden CR, Zorn M, Ratzan SC, Parker RM. National Library of Medicine Current Bibliographies in Medicine: Health Literacy. Washington, DC: U.S. Government Printing Office; 2000. US Department of Health and Human Services. Healthy People 2010. Introduction URL: http://chirr.nlm.nih.gov/health-literacy.php [accessed 2015-06-02] [WebCite Cache ID 6YzJ53nyH]
84. Sox CM, Gribbons WM, Loring BA, Mandl KD, Batista R, Porter SC. Patient-centered design of an information management module for a personally controlled health record. J Med Internet Res 2010;12(3):e36 [FREE Full text] [doi: 10.2196/jmir.1269] [Medline: 20805091]
85. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]
86. Krist AH, Peele E, Woolf SH, Rothemich SF, Loomis JF, Longo DR, et al. Designing a patient-centered personal health record to promote preventive care. BMC Med Inform Decis Mak 2011;11:73 [FREE Full text] [doi: 10.1186/1472-6947-11-73] [Medline: 22115059]
87. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
88. Monkman H, Kushniruk A. Applying usability methods to identify health literacy issues: an example using a Personal Health Record. Stud Health Technol Inform 2013;183:179-185. [Medline: 23388278]
89. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
90. Krist AH, Woolf SH, Rothemich SF, Johnson RE, Peele JE, Cunningham TD, et al. Interactive preventive health record to enhance delivery of recommended care: a randomized trial. Ann Fam Med 2012;10(4):312-319 [FREE Full text] [doi: 10.1370/afm.1383] [Medline: 22778119]
91. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. J Appl Gerontol 2014 Jun;33(4):416-436. [doi: 10.1177/0733464812447283] [Medline: 24781964]
92. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Design and implementation of a web-based patient portal linked to an electronic health record designed to improve medication safety: the Patient Gateway medications module. Inform Prim Care 2008;16(2):147-155. [Medline: 18713531]
93. Fuji KT, Abbott AA, Galt KA. Personal health record design: qualitative exploration of issues inhibiting optimal use. Diabetes Care 2014;37(1):e13-e14 [FREE Full text] [doi: 10.2337/dc13-1630] [Medline: 24356602]
94. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc 2009;16(5):660-669 [FREE Full text] [doi: 10.1197/jamia.M3154] [Medline: 19567793]
95. Kuziemsky CE, Kushniruk A. Context mediated usability testing. Stud Health Technol Inform 2014;205:905-909. [Medline: 25160319]
96. Do NV, Barnhill R, Heermann-Do KA, Salzman KL, Gimbel RW. The military health system's personal health record pilot with Microsoft HealthVault and Google Health. J Am Med Inform Assoc 2011;18(2):118-124 [FREE Full text] [doi: 10.1136/jamia.2010.004671] [Medline: 21292705]
97. Gaskin GL, Longhurst CA, Slayton R, Das AK. Sociotechnical Challenges of Developing an Interoperable Personal Health Record: Lessons Learned. Appl Clin Inform 2011;2(4):406-419 [FREE Full text] [doi: 10.4338/ACI-2011-06-RA-0035] [Medline: 22003373]
98. Ennis L, Robotham D, Denis M, Pandit N, Newton D, Rose D, et al. Collaborative development of an electronic Personal Health Record for people with severe and enduring mental health problems. BMC Psychiatry 2014;14(1):305 [FREE Full text] [doi: 10.1186/s12888-014-0305-9] [Medline: 25403285]
99. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]

100. Tuil WS, ten Hoopen AJ, Braat DDM, de Vries RPF, Kremer JAM. Patient-centred care: using online personal medical records in IVF practice. Hum Reprod 2006 Nov;21(11):2955-2959 [FREE Full text] [doi: 10.1093/humrep/del214] [Medline: 16982658]
101. Grant RW, Wald JS, Poon EG, Schnipper JL, Gandhi TK, Volk LA, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for diabetes collaborative care. Diabetes Technol Ther 2006 Oct;8(5):576-586 [FREE Full text] [doi: 10.1089/dia.2006.8.576] [Medline: 17037972]
102. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]
103. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Usage patterns of a personal health record by elderly and disabled users. AMIA Annu Symp Proc 2007:409-413 [FREE Full text] [Medline: 18693868]
104. Bergman DA, Brown NL, Wilson S. Teen use of a patient portal: a qualitative study of parent and teen attitudes. Perspect Health Inf Manag 2008;5:13 [FREE Full text] [Medline: 18923702]
105. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
106. Staroselsky M, Volk LA, Tsurikova R, Newmark LP, Lippincott M, Litvak I, et al. An effort to improve electronic health record medication list accuracy between visits: patients' and physicians' response. Int J Med Inform 2008 Mar;77(3):153-160. [doi: 10.1016/j.ijmedinf.2007.03.001] [Medline: 17434337]
107. Wald JS, Grant RW, Schnipper JL, Gandhi TK, Poon EG, Businger AC, et al. Survey analysis of patient experience using a practice-linked PHR for type 2 diabetes mellitus. AMIA Annu Symp Proc 2009;2009:678-682 [FREE Full text] [Medline: 20351940]
108. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
109. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
110. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, Quartey NK, et al. The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak 2010;10:46 [FREE Full text] [doi: 10.1186/1472-6947-10-46] [Medline: 20809950]
111. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]
112. Shaw RJ, Ferranti J. Patient-provider internet portals--patient outcomes and use. Comput Inform Nurs 2011 Dec;29(12):714-8; quiz 719. [doi: 10.1097/NCN.0b013e318224b597] [Medline: 21697705]
113. Wright A, Poon EG, Wald J, Feblowitz J, Pang JE, Schnipper JL, et al. Randomized controlled trial of health maintenance reminders provided directly to patients through an electronic PHR. J Gen Intern Med 2012 Jan;27(1):85-92 [FREE Full text] [doi: 10.1007/s11606-011-1859-6] [Medline: 21904945]
114. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
115. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. JAMA 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]
116. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
117. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
118. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
119. Otsuka SH, Tayal NH, Porter K, Embi PJ, Beatty SJ. Improving herpes zoster vaccination rates through use of a clinical pharmacist and a personal health record. Am J Med 2013 Sep;126(9):832.e1-832.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2013.02.018] [Medline: 23830534]
120. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]

121. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
122. Murray MF, Giovanni MA, Klinger E, George E, Marinacci L, Getty G, et al. Comparing electronic health record portals to obtain patient-entered family health history in primary care. J Gen Intern Med 2013 Dec;28(12):1558-1564 [FREE Full text] [doi: 10.1007/s11606-013-2442-0] [Medline: 23588670]
123. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, Martin T, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014 Feb;21(e1):e157-e162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
124. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013 Jul;163(1):280-1.e1. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
125. Aberger EW, Migliozzi D, Follick MJ, Malick T, Ahern DK. Enhancing patient engagement and blood pressure management for renal transplant recipients via home electronic monitoring and web-enabled collaborative care. Telemed J E Health 2014 Sep;20(9):850-854. [doi: 10.1089/tmj.2013.0317] [Medline: 25046403]
126. Barron J, Bedra M, Wood J, Finkelstein J. Exploring three perspectives on feasibility of a patient portal for older adults. Stud Health Technol Inform 2014;202:181-184. [Medline: 25000046]
127. Borbolla D, Del FG, Taliercio V, Otero C, Campos F, Martinez M, et al. Integrating personalized health information from MedlinePlus in a patient portal. Stud Health Technol Inform 2014;205:348-352. [Medline: 25160204]
128. Wright A, Feblowitz J, Maloney FL, Henkin S, Ramelson H, Feltman J, et al. Increasing patient engagement: patients' responses to viewing problem lists online. Appl Clin Inform 2014;5(4):930-942. [doi: 10.4338/ACI-2014-07-RA-0057] [Medline: 25589908]
129. Tom JO, Chen C, Zhou YY. Personal health record use and association with immunizations and well-child care visits recommendations. J Pediatr 2014 Jan;164(1):112-117. [doi: 10.1016/j.jpeds.2013.08.046] [Medline: 24120019]
130. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
131. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
132. Hess R, Fischer GS, Sullivan SM, Dong X, Weimer M, Zeith C, et al. Patterns of response to patient-centered decision support through a personal health record. Telemed J E Health 2014 Nov;20(11):984-989. [doi: 10.1089/tmj.2013.0332] [Medline: 25243350]
133. Hall MJ, Herda MM, Handorf EA, Rybak CC, Keleher CA, Siemon M, et al. Direct-to-patient disclosure of results of mismatch repair screening for Lynch syndrome via electronic personal health record: a feasibility study. Genet Med 2014 Nov;16(11):854-861. [doi: 10.1038/gim.2014.42] [Medline: 24784156]
134. Druss BG, Ji X, Glick G, von Esenwein SA. Randomized trial of an electronic personal health record for patients with serious mental illnesses. Am J Psychiatry 2014 Mar;171(3):360-368. [doi: 10.1176/appi.ajp.2013.13070913] [Medline: 24435025]
135. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff (Millwood) 2014 Oct;33(10):1817-1822. [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
136. Chrischilles EA, Hourcade JP, Doucette W, Eichmann D, Gryzlak B, Lorentzen R, et al. Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 2014;21(4):679-686. [doi: 10.1136/amiajnl-2013-002284] [Medline: 24326536]
137. Cabitza F, Simone C, De MG. User-driven prioritization of features for a prospective InterPersonal Health Record: perceptions from the Italian context. Comput Biol Med 2015 Apr;59:202-210. [doi: 10.1016/j.compbiomed.2014.03.009] [Medline: 24768267]
138. Hargreaves JS. Will electronic personal health records benefit providers and patients in rural America? Telemed J E Health 2010 Mar;16(2):167-176. [doi: 10.1089/tmj.2009.0063] [Medline: 20082592]
139. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
140. Simmons LA, Wolever RQ, Bechard EM, Snyderman R. Patient engagement as a risk factor in personalized health care: a systematic review of the literature on chronic disease. Genome Med 2014;6(2):16 [FREE Full text] [doi: 10.1186/gm533] [Medline: 24571651]

Abbreviations

AHRQ: Agency for Health care Research and Quality
ePHR: electronic personal health record

EHR: electronic health record
CMS: Centers for Medicare and Medicaid Services
MU: meaningful use
SM: secure messaging

Edited by G Eysenbach; submitted 20.01.15; peer-reviewed by D Amante, A Benis; comments to author 05.02.15; revised version received 10.03.15; accepted 03.04.15; published 23.06.15.

Please cite as:
Irizarry T, DeVito Dabbs A, Curran CR
Patient Portals and Patient Engagement: A State of the Science Review
J Med Internet Res 2015;17(6):e148
URL: http://www.jmir.org/2015/6/e148/
doi:10.2196/jmir.4255
PMID:26104044

Review

“Nothing About Me Without Me”: An Interpretative Review of Patient Accessible Electronic Health Records

Sagar Ramesh Jilka[1], BSc, MRes, PhD; Ryan Callahan[1], BSFS; Nick Sevdalis[2], PhD; Erik K Mayer[1], PhD; Ara Darzi[1], FRS

[1]Centre for Health Policy, Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Centre for Implementation Science, Health Service & Population Research Department, King’s College London, London, United Kingdom

Corresponding Author:
Sagar Ramesh Jilka, BSc, MRes, PhD
Centre for Health Policy
Department of Surgery and Cancer
Imperial College London
1029, 10th Floor, QEQM Building
St. Mary’s Hospital
London, W2 1NY
United Kingdom
Phone: 44 7708715627
Fax: 44 2033126950
Email: sagarjilka@gmail.com

Abstract

Background: Patient accessible electronic health records (PAEHRs) enable patients to access and manage personal clinical information that is made available to them by their health care providers (HCPs). It is thought that the shared management nature of medical record access improves patient outcomes and improves patient satisfaction. However, recent reviews have found that this is not the case. Furthermore, little research has focused on PAEHRs from the HCP viewpoint. HCPs include physicians, nurses, and service providers.

Objective: We provide a systematic review of reviews of the impact of giving patients record access from both a patient and HCP point of view. The review covers a broad range of outcome measures, including patient safety, patient satisfaction, privacy and security, self-efficacy, and health outcome.

Methods: A systematic search was conducted using Web of Science to identify review articles on the impact of PAEHRs. Our search was limited to English-language reviews published between January 2002 and November 2014. A total of 73 citations were retrieved from a series of Boolean search terms including “review*” with “patient access to records”. These reviews went through a novel scoring system analysis whereby we calculated how many positive outcomes were reported per every outcome measure investigated. This provided a way to quantify the impact of PAEHRs.

Results: Ten reviews covering chronic patients (eg, diabetes and hypertension) and primary care patients, as well as HCPs were found but eight were included for the analysis of outcome measures. We found mixed outcomes across both patient and HCP groups, with approximately half of the reviews showing positive changes with record access. Patients believe that record access increases their perception of control; however, outcome measures thought to create psychological concerns (such as patient anxiety as a result of seeing their medical record) are still unanswered. Nurses are more likely than physicians to gain time efficiencies by using a PAEHR system with the main concern from physicians being the security of the PAEHRs.

Conclusions: This review implements a novel scoring system, which shows there is a lack of rigorous empirical testing that separates the effect of record access from other existing disease management programs. Current research is too targeted within certain clinical groups’ needs, and although there are positive signs for the adoption of PAEHRs, there is currently insufficient evidence about the effect of PAEHRs on health outcomes for patients or HCPs.

(J Med Internet Res 2015;17(6):e161) doi:10.2196/jmir.4446

KEYWORDS

patient accessible electronic health records; online record access

Introduction

Modern technology is changing the role of the passive patient to a more informed and engaging stakeholder in their own care [1]. Technology is making personal health-related data and documents digitally accessible and shareable between patients and physicians, with the aim of improving the safety, quality, and effectiveness of care [2]. According to the Council of Europe, patients should be in a position to access their medical records at their request and also be able to control who else can see their records [3]. Despite such calls, it is still not common practice for patients to access their medical records [4].

The use of patient accessible electronic health records (PAEHRs) has been considered by health organizations since the early 1990s [5]; however, PAEHRs have only recently received attention for their use in improving access to patient data [6,7]. In their early days, PAEHRs failed to gain approval for adoption because of prohibitive financial cost and the difficulty of transitioning from paper-based records [8]. With the advancements of modern technology, PAEHR systems should be technologically easier to implement and administer, yet the question still remains: Why has modern medicine not yet seen more widespread application and implementation of PAEHR in patient care?

One potential reason is that research has still not resolved whether patients want to access their medical records. Assuming patients would like access to their records, it is not yet known how helpful their medical record (in its typical current form) will be to them and whether patients will understand its content [1,4]. Furthermore, we currently have no knowledge of the impact that patient access to their PAEHRs would have on health care providers (HCPs) [9].

To date, research on the impact of PAEHRs has been focused on a particular clinical group, or on a limited number of outcome measures, from either the perspective of patients or doctors. Furthermore, no data have been published regarding the impact of changes in information supply—whether qualitative or quantitative—on patients' psychological status, for example, their anxiety about their health [10]. To address the above issues, we provide a review of existing reviews that aims to critically evaluate the current state of the evidence regarding PAEHRs. The main objective of this paper was to synthesize relevant research to provide a quantitative insight into the impact of PAEHRs across a range of outcome measures in a number of clinical populations and investigate differences between patients and HCPs.

Methods

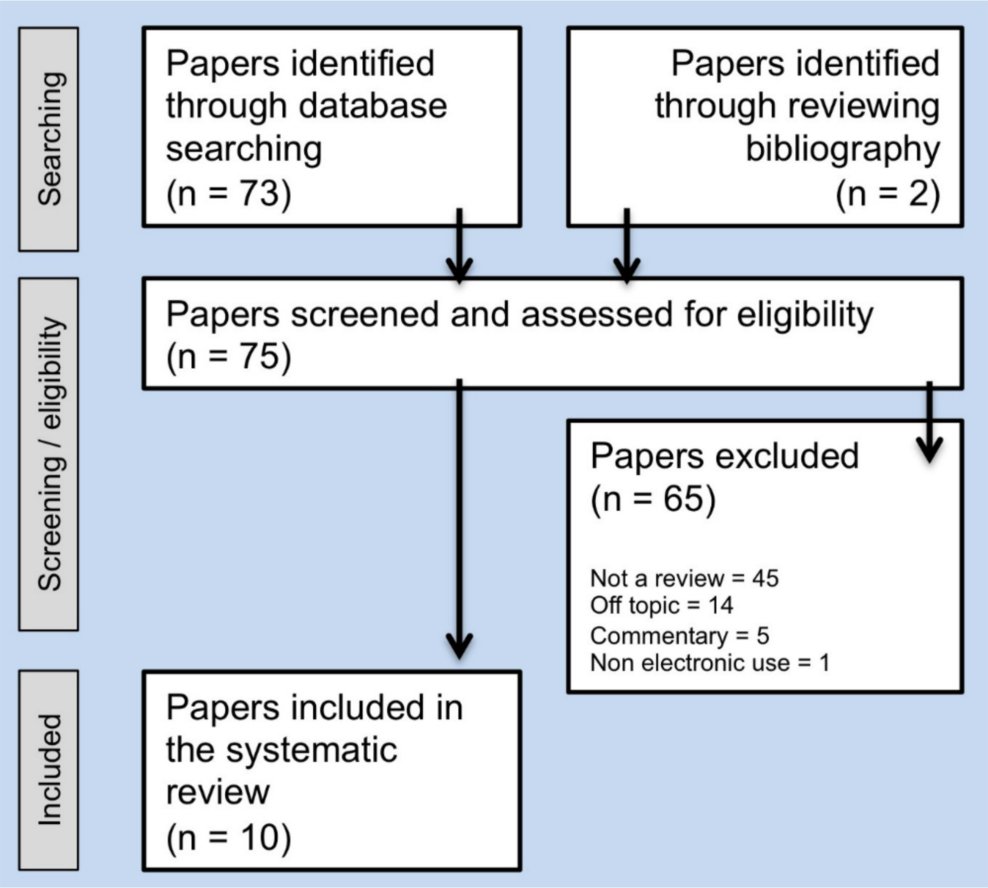
Study Search and Selection

We searched English-language articles indexed in any databases in Web of Science with a publication date between January 2002 and November 2014. Potentially relevant review articles were identified using a combination of medical subject headings, free text phrases, and Boolean searches. These included "review*" with (1) "patient access to records" (n=49 citations), (2) "patient portal" (n=18 citations), and (3) "patient accessible" (n=6 citations) across all Web of Science databases, including Web of Science core collection, MEDLINE, and BIOSIS Citation Index. This allowed us to focus the current paper as a review of reviews within the existing literature resulting in 73 citations. The references of selected reviews were also examined to search for additional articles satisfying inclusion criteria (n=1).

Eligibility Criteria

We defined PAEHRs as patient accessible information held by the physician and/or health care system. We included systematic reviews that assessed the effect of PAEHRs on a variety of quality and clinically related outcome measures in adult populations. The reviews investigated patients suffering chronic disease such as diabetes and hypertension as well as patients seen in primary care. Inclusion criteria included suitable research questions, description of methods supporting the paper as a review, and reported a narrative on the impact of PAEHRs. Exclusion criteria were non-English, non-peer-reviewed, duplicates, non-empirical, and papers with a non-electronic use of record access or if the focus of the paper was on the design of a patient portal system. The majority of citations were excluded because they did not provide a review of the existing literature on patient/HCP outcome measures based on a review of the abstract and/or study title (Figure 1).

Figure 1. Flow diagram of study methodology.



Scoring System

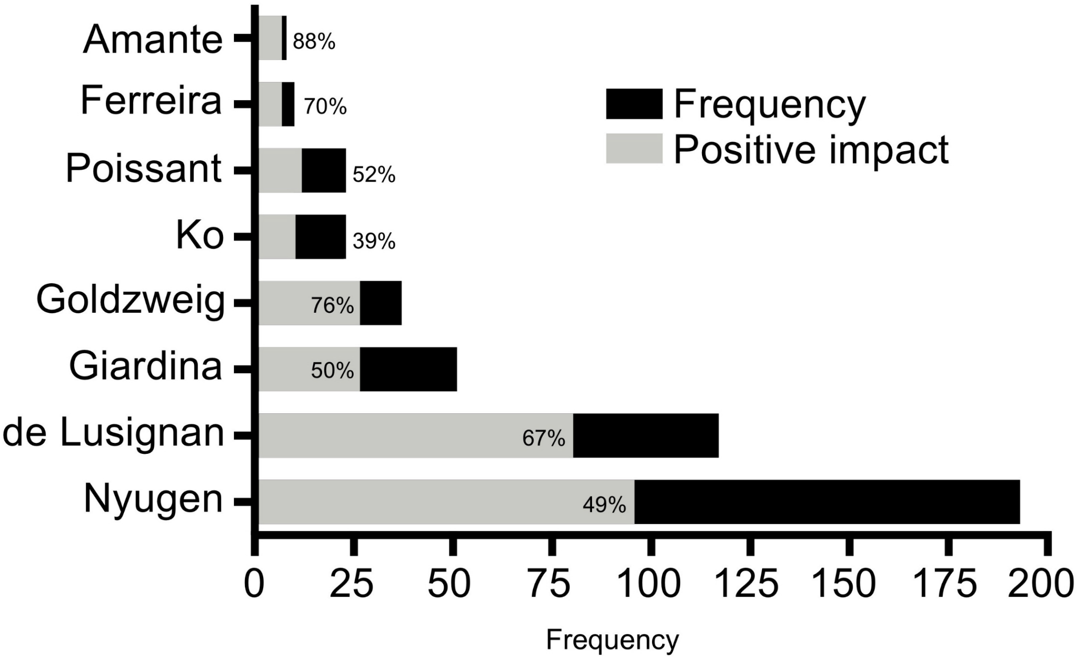
A scoring system was developed to weight the impact of an outcome measure quantitatively and thereby investigate the impact of PAEHRs by individually assessing their impact on each outcome measure described in the reviews. These outcome measures were subsequently categorized forming 16 measurable domains (Multimedia Appendix 1). The definitions of these outcome measures are either (1) derived directly from one of the original review sources (eg, “effectiveness of record access” and “usefulness and usability” have been concatenated to make the definition “usefulness/effectiveness of record access”), or (2) a logical definition has been applied based on the original definitions (eg, “glycemic control, change in gyrated hemoglobin and blood pressure control” have been concatenated to make the definition “clinical outcome”). The citations of each included review were assessed to determine which outcome measures were investigated (frequency) as well as the result of that outcome measure, that is, if the investigated outcome measure was found to improve as a result of PAEHR access (positive impact). For example, in a review by Giardina et al [9], a study was included carried out by McCarrier et al, which evaluated the effectiveness of electronic patient portals in a group of diabetic patients [11]. McCarrier found that there was no improvement in glycemic control in patients with PAEHR access (clinical measure), but patients became more involved in their clinical care through the use of PAEHRs (self-efficacy - patient involvement), therefore providing a positive outcome score of 1 in the “self-efficacy - patient involvement” outcome measure and a no improvement score of 0 in the “clinical outcome” outcome measure.

Results

Overview

The systematic search provided ten review articles reporting on PAEHR implementations across different health care contexts and clinical groups (Multimedia Appendix 2; [12]). Eight review articles were used in the final analysis. One review was excluded because of duplicate citations [1], and another study [7] contained 32 citations that were not referenced directly within the outcome measures described in their paper. Figure 2 summarizes the total number of times an outcome measure was reported in each review against the number of times these outcome measures were reported to have had a positive impact across each individual study.

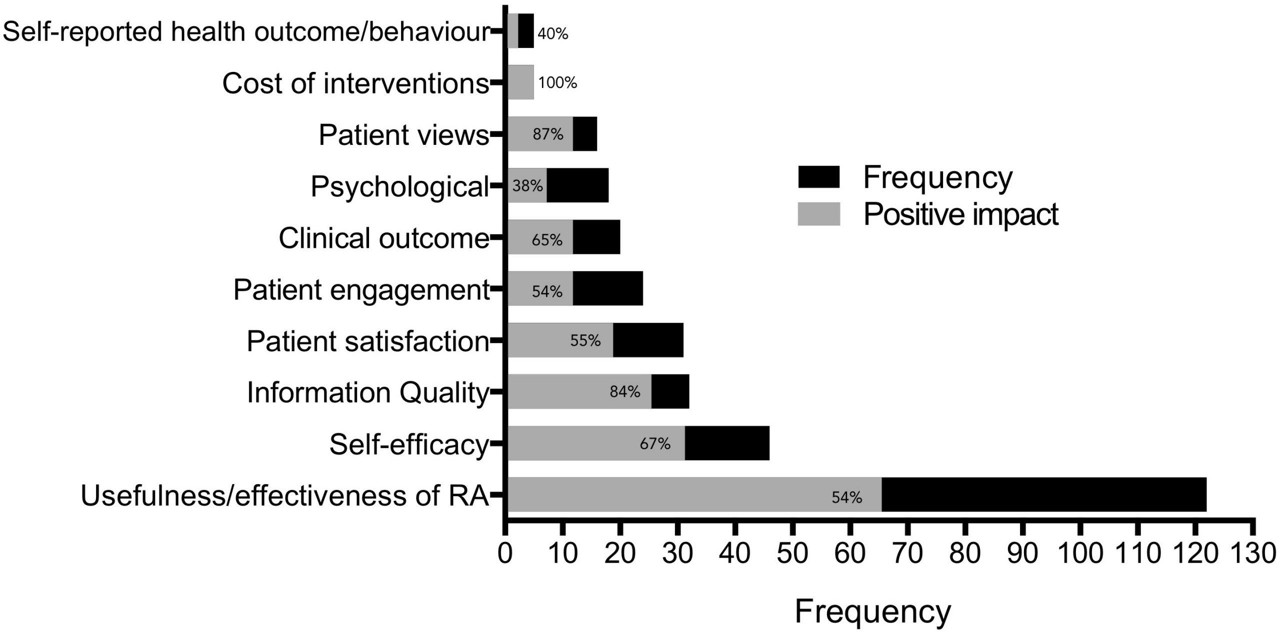
Figure 2. Frequency of all outcome measures across the 8 reviews analyzed in this study. The number of times a review (y-axis) reported on any outcome measure (black bar) against the number of times these outcome measures were found to have a positive impact (gray bar).



Patients' Perspective

Across reviews, we found some uncertainty regarding whether access to PAEHRs makes a difference and whether patients actually want access to their PAEHRs (Figure 3).

Figure 3. The frequency of patient outcomes (black bars) against frequency of positive change (gray bars); eg, usefulness/effective of record access (RA) has been investigated 122 times as an outcome measure with only 66 of those investigations reporting a positive impact (gray bar). We therefore infer that the proportion of black on the horizontal bars illustrates that there are studies that have found RA to have a negative impact or at least no impact on the outcome factors.



Usefulness/Effectiveness of Patient Access to Electronic Health Records

The usefulness/effectiveness of PAEHRs included outcomes such as the usefulness, interoperability, and adoption rate. A more detailed list can be found in Multimedia Appendix 1. It is unclear from the current evidence whether PAEHRs are useful or effective for patients. Giardina found 40% (2/5) of studies showing positive outcomes of PAEHR usefulness [9], Nyugen found 50% (43/86), and Poissant found 53% (10/19). Two reviews found PAEHRs showed an overall positive impact:

100% for both de Lusignan (9/9) and Goldzweig (2/2). Whereas Ferreira found the opposite effect (0/1).

Nyugen et al reported that patients questioned the usefulness of PAEHRs because they were not well designed and did not integrate well with other existing clinical systems, for example, the National Health Service (NHS) HealthSpace [13]. Four themes emerged from the current review that act as a framework for usefulness: (1) promotion of a sense of illness ownership, (2) patient driven communication, (3) personalized support, and (4) mutual trust between patient and provider.

Patient Satisfaction

Patient satisfaction was investigated with outcomes such as mood states and satisfaction with care [14] and is further defined in Multimedia Appendix 1. We found six reviews that reported on patient satisfaction. Of those, two reviews [4,14] found no change in patient satisfaction (0/1 in both reviews) and one review reported 14% (1/7) that included showing a positive impact on patient satisfaction after PAEHR use [9]. Nyugen found 40% (2/5) of studies [13] and Goldzweig reported 57% (4/7) of studies showing a positive impact on patient satisfaction [15]. De Lusignan reported 100% (10/10) of studies showing a positive change in patient satisfaction [16].

A barrier to PAEHR uptake is poor patient satisfaction with a PAEHR system. Satisfaction can be a result of various aspects of patient experience, such as the (perceived) quality of care, consultation, or information provided [9]. Giardina et al found 11 studies that reported on patient satisfaction with eight of them showing no significant differences in satisfaction as a result of PAEHR access [9]. Similarly, Ferreira et al found that use of PAEHRs produces only modest benefits in satisfaction [4].

Patients’ Self-Efficacy

Self-efficacy involved various aspects that encompass a patient’s beliefs about how they feel, including patient involvement, communication, and patient empowerment as a result of PAEHR access. Overall, we found 67% (31/46) of positive changes as a result of PAEHR use across all self-efficacy domains, as made up by patient involvement (67%, 10/15), patient empowerment (78%, 18/23), and patient communication (38%, 3/8).

The most common reasons that patients wanted to look at their medical records were to see what their physician said about them (74%), to be more involved in their health care (74%), and to understand their condition better (72%) [4]. Ko et al report patient empowerment outcomes in 3 clinical groups, namely oncology (n=2), and palliative care (n=1) demonstrating positive change after PAEHR use, and a negative change in a group of rheumatoid arthritis patients (n=1), and two studies in oncology with patient communication as an outcome (both showing no change in communication with PAEHR access) [14].

Psychological Outcomes

Psychological outcomes examined across reviews included measures of anxiety, depression, contentment, and quality of life, using behavioral measures such as the Spielberger State-Trait Anxiety Inventory and the European Organization for Research and Cancer quality-of-life questionnaire (EORTC QLQ-C30) [17]. We found a typical pattern of mixed outcomes with 11 studies showing no change in psychological outcomes from a total of 18. For example, a study reported in Goldzweig et al randomly assigned couples having in vitro fertilization in the Netherlands to usual care versus PAEHR access and found no change in anxiety or change in depression between the 2 groups as a result of PAEHR access [18]. Poissant et al also report that PAEHR access was not found stressful by patients [19]. Ferreira et al found no consistent pattern in the impact of PAEHRs on psychological outcomes and suggested it is worthwhile to carry out a larger study on the effects of PAEHR use on such outcomes [4].

Health Outcomes/Behaviors and Clinical Outcomes

Health outcomes/behavior include diet, alcohol intake, medication changes, and smoking or exercise habits and are different to “clinical outcomes”, which refer to outcomes that can be empirically tested such as hemoglobin A1c levels. Giardina et al’s review shows a typical pattern of PAEHR impact, whereby they found a mix of results relating to specific clinical measures (such as blood pressure and various diabetes measures) with 50% (2/4) of studies reporting a positive change in clinical measure [9]. Goldzweig et al found most positive changes with 75% (6/8) of studies in their review reporting a positive change as a result of PAEHR access [15].

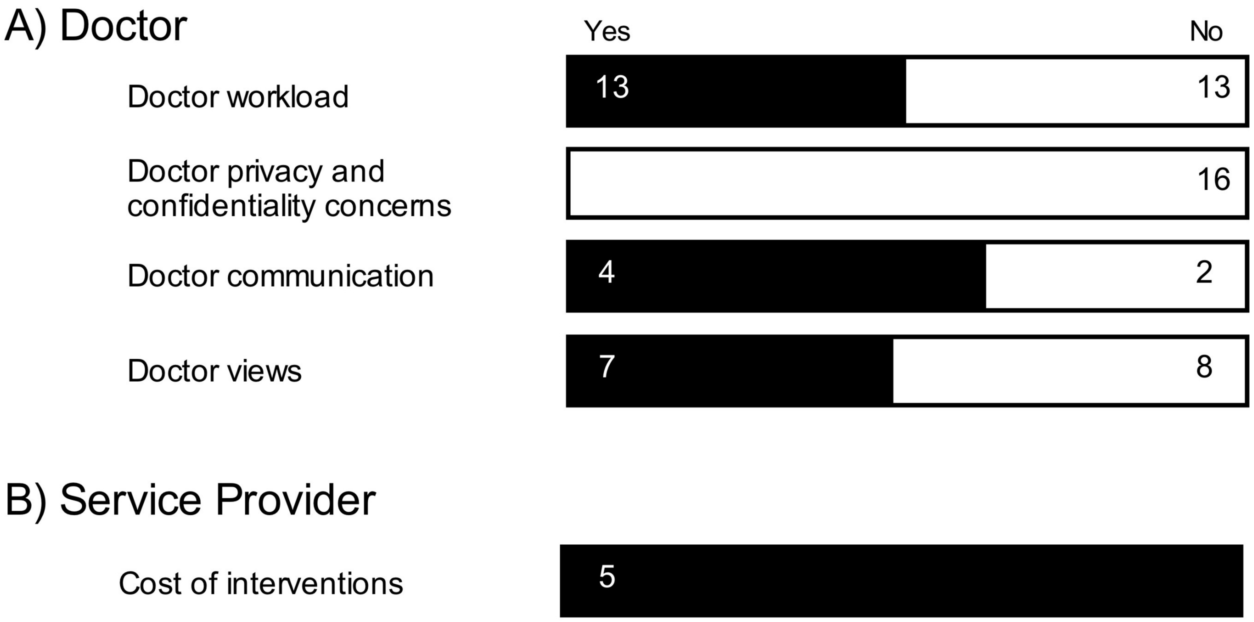
Ammenwerth et al found that the impact of PAEHR access on health outcomes is limited with respect to impact on clinical outcome, health resource consumption, patient adherence, and patient-physician communication. They report that the parameters studied did not show a statistically significant difference between intervention and control groups and in particular, no statistically significant changes could be observed for parameters related to clinical outcome. Ammenwerth’s findings suggest that the available evidence does not support the assumption that PAEHRs improve patient care [1].

Health Care Professionals’ Perspective

There were a number of articles that evaluated the benefit of PAEHRs from the HCPs’ perspective (Multimedia Appendix 1), although relatively fewer studies focused on the HCPs’ perspective of PAEHRs when compared with patient perspective [20] (Figure 4). The types of HCP evaluated were mainly doctors and nurses [21-30].

Several of these, in addition to stating a qualitative benefit, described the measurable impact of any benefit as outcome measures including workload, privacy and confidentiality concerns, cost, and communication. These are described in more detail below.

Figure 4. Frequency of studies showing a positive change (black bars) and negative/no change (white bars) after patients were given record access from the point of view of doctors and service providers.



Doctors' Views, Privacy, and Confidentiality Concerns

The poor uptake of electronic health records (EHRs) may be driven by HCPs who are wary of patient access to medical records, fearing it may cause patient anxiety. De Lusignan et al found eight studies where physicians feared that PAEHR access without a physician available to interpret the information might cause patients to worry [16]. Although these risks are low [31], doctors have concerns about shared medical records and see less potential for benefit than patients [32]. These concerns included doctors finding a computer system "stressful", having spent twice as much time using the computer than they had previously using their hand-written notes [19].

These concerns are also extended to the security of the electronic records, with HCPs reporting professional concerns about privacy and confidentiality in 16 studies of de Lusignan's review [16]. The security and confidentiality of patient data must be put at the forefront of EHR services in order to achieve widespread consumer acceptance and adoption [9], and patients should have the right to decide who can access and edit their medical records [33], which was found to be a common barrier for PAEHR uptake [15].

Workload

HCPs do not want changes to the current medical record system to negatively impact their time [34]. Research has shown an interesting mix of findings on the impact of PAEHRs on workload. The most striking finding is a study that recently investigated changes in HCP workload [16]. De Lusignan et al found that half of the studies in their review (13/26) showed PAEHRs have a positive impact on changes to workload or workflow (ie, a decrease in workload).

Poissant et al's review focused on the effects of PAEHR access on HCPs documentation time. They found that that decreased documentation time in a PAEHR project is not likely to be realized, especially for physicians. From a total of 23 studies included in their review, they found that 11 studies examined the impact of PAEHRs on time efficiencies of nurses, of which six studies found that nurses are more likely than physicians to gain time efficiencies by using a computer system to document patient information. Two studies found that bedside PAEHR increased documentation time, and one study reported different results depending on the specific content of the information being documented [19].

With respect to physicians, ten studies examined the impact of PAEHR on time efficiencies of physicians. Poissant et al found that using a PAEHR system increased physician documentation time by 17%. Of their studies, 60% (6/10) reported significant results in the direction of unfavorable impact on initial visit time, and 10% (1/10) lacked sufficient information to identify whether the results were significant. In the remaining three studies, there were no significant differences between computer and paper documentation time [19]. Ferreira et al report that physicians found no change in their workload or no adverse consequences as a result of PAEHRs, and all the physicians supported the use of PAEHRs [4].

Doctor Communication

Improving doctor-patient relationship is one of the few outcomes that can be investigated from a physician point of view, yet studies still report how doctor-patient relationships improve from the patient point of view. Ferreira et al report only one study that investigated doctor-patient relationships. They found that the majority of doctors (and patients) were unanimous in their belief that the paehr access was positive for both physicians and patients and improved the level of communication between them [4]. Furthermore, Ferreira et al report three randomized clinical trial studies whereby hcps found access to paehrs via the internet easy to use, useful, and considered that it could improve their communication with other HCPs [4].

Cost for Patient Accessible Electronic Health Records

Fewer studies across the reviews examined PAEHRs from the perspective of the service provider (eg, a hospital providing PAEHR access). In one study, Poissant et al found that using PAEHRs for writing all inpatient orders significantly lowered patient charges and hospital costs [19]. Nyugen et al reviewed three studies that demonstrated how PAEHRs in the United States could provide a positive return on investment providing evidence of major financial benefit [13].

Data Quality

Apart from patients, HCPs, and service provider factors, we considered study design, which informs the quality of the evidence analyzed in our review. Poissant et al reviewed 23 papers of which only 5 were randomized controlled trials (RCT), with other studies being posttest control studies (n=6), and one-group pretest-posttest designs (n=12) [19]. Ferreira et al outlined the number of articles implementing an RCT (n=18), a transversal study (n=39), a longitudinal study (n=5), and a letter (n=20) [4].

Not all studies highlighted the historically small proportion of randomized studies. In Giardina et al’s more recent review, however, 20 studies were RCTs with only seven studies being uncontrolled observational studies [9], suggesting that the quality of evidence is continuing to improve.

Discussion

Principal Findings

The systematic reviews included in our synthesis aimed to investigate the effect of record access on various outcomes. We found that these reviews showed mixed outcomes in aspects of patient safety, usefulness, satisfaction, and self-efficacy across patients and HCPs. This is typically represented by Giardina et al’s review, who found an absence of positive evidence on these outcome measures, with only 50% of studies showing positive changes with record access [9]. Positively, the little work carried out on the cost of PAEHRs has shown that implementing PAEHR systems would lower hospital costs.

We next highlight some of the issues that surround the study of patient access to their medical records in terms of both technical and scientific rigor, which leads to the root of the problem: for such a large problem, there is very little data-driven evidence coming from a large population. We believe a large factor contributing to the lack of success in PAEHR access has been a lack of data-driven evidence about the opinions, wants, and needs of large clinical consumer groups. This setback comes to the heart of the issue in this field: PAEHR developers are still not clear whether providing patients with record access makes a difference to either the patients themselves or their physician.

Lack of Empirical and Rigorous Testing

Current research is targeted to certain clinical groups and their needs, which makes the findings difficult to implement across a large non-disease-specific population. More than half of the patient portal evaluations reviewed by Otte-Trojel were targeted at chronic disease patients, such as the management of diabetes, hypertension, and depression [7]. The problem with disease-specific studies is that they are more vulnerable to a “ceiling effect” due to the breadth and quality of the well-established existing disease management programs. This problem is also highlighted in Goldzweig’s review, which identified examples where record access was associated with improved outcomes for patients with chronic diseases, such as diabetes, hypertension, and depression, but these studies generally used the PAEHR in conjunction with case management [15]. As a result, the effects of PAEHRs are small and could provide an explanation into why PAEHR effects are often inconsistent. Future work could consider investigating the effects of PAEHRs on various mechanisms (such as patient empowerment) outside the remit of disease-specific groups to avoid issues surrounding care coordination [7].

A large proportion of studies that investigate the impact of PAEHRs on various outcome measures follow a quasi-experimental design implementing interviews and/or surveys to measure the impact of each intervention. There is the potential to implement better quality study designs and use more objective and rigorous measuring techniques to determine whether a cause and effect relationship exists between PAEHRs and outcome measures. Future research should examine the processes of PAEHR and their direct effects by implementing a pretest and posttest design where participants are tested on a specific set of outcome measures before and after exposure to a PAEHR system.

Research should also aim to address our understanding of how PAEHRs can bridge the gap between patient and doctor with a focus on using up-to-date technologies. Over the last 20 years, there have been large technological improvements, both in terms of hardware and software. As a result, research carried out in the last century may not be comparable with modern day technologies. We found that a large proportion of studies that investigated the effects of PAEHRs were published between 1996 and 2005. The implementation of PAEHRs should no longer be a technological problem as the technology has been available for some years now [13], therefore, it is important for research to reflect these advances.

Limitations

Our study focused only on English language reviews, which neglects PAEHR advancements from other parts of the world. Furthermore, our review of reviews covers a small overall evidence base compared to a systematic review focused on one group (eg, patients) and a lack of quantitative synthesis is arguable, as the reviews presented heterogeneous datasets/studies. However, we believe that the reviews analyzed here cover a large number of primary studies across a variety of outcome measures and our scoring system provides a quantifiable way of synthesizing the literature. PAEHR systems conceptually vary, and our review brings together results across a variety of PAEHR systems, as do the reviews that make up our work, which could be contributing to the nature of the results. The small number of RCTs investigating patient access to their medical records was further limited by the small sample sizes in the studies used in this review, therefore compromising the quality of a scientific study. However, there is currently little solid evidence from RCTs of proven effectiveness in improved patient health outcomes through the use of PAEHRs [35].

Conclusions

Our synthesis of available systematic reviews examined the impact of patient access to electronic medical records and revealed few overarching results. There was minimal evidence to support the universal use of PAEHRs both from a patient or HCP point of view; however, PAEHRs appear to have a positive impact on patient empowerment. Patients appear to have positive views after using PAEHRs and the information quality in PAEHRs is positive, although major drawbacks include security, privacy, and confidentiality concerns. HCPs also appear to be divided in terms of whether using a PAEHR reduces their workload. The topic of PAEHRs appears to be one that divides both patients and HCPs and is certainly a field where more rigorous research is needed to evaluate practice and improve system design and implementation.

Acknowledgments

This work was supported by the Sowerby Foundation (UK). Author NS's research was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Authors' Contributions

SRJ designed and performed the review and analyzed data. SRJ, RC, NS, EM, and AD wrote the manuscript.

Conflicts of Interest

Author NS delivers regular patient safety and human factors training on a consultancy basis to hospitals in the United Kingdom and internationally.

Multimedia Appendix 1

Definitions of outcome measures across synthesized reviews.

[PDF File (Adobe PDF File), 70KB - jmir_v17i6e161_app1.pdf]

Multimedia Appendix 2

Evidence synthesis overview.

[PDF File (Adobe PDF File), 88KB - jmir_v17i6e161_app2.pdf]

References

1. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
2. Kay M. Building Foundations for EHealth. 2006. World Health Organization URL: http://www.who.int/goe/publications/bf_FINAL.pdf [accessed 2015-06-22] [WebCite Cache ID 6ZU0Mk3Yr]
3. Council of Europe. Recommendation No. R (97) 5 of the Committee of Ministers to member states on the protection of medical data. Int Dig Health Legis 1998;49(3):502-508. [Medline: 11657540]
4. Ferreira A, Correia A, Silva A, Corte A, Pinto A, Saavedra A, et al. Why facilitate patient access to medical records. Stud Health Technol Inform 2007;127:77-90. [Medline: 17901601]
5. McLaren P. The right to know. BMJ 1991 Oct 19;303(6808):937-938 [FREE Full text] [Medline: 1954414]
6. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
7. Otte-Trojel T, de BA, Rundall T, van de Klundert Joris. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
8. Haste F. Personal health record evaluation. Project report. London, UK: Health Education Authority; 1992.
9. Davis GT, Menon S, Parrish D, Sittig D, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014;21(4):737-741. [doi: 10.1136/amiajnl-2013-002239] [Medline: 24154835]
10. Gravis G, Protière C, Eisinger F, Boher JM, Tarpin C, Coso D, et al. Full access to medical records does not modify anxiety in cancer patients: results of a randomized study. Cancer 2011 Oct 15;117(20):4796-4804 [FREE Full text] [doi: 10.1002/cncr.26083] [Medline: 21607939]
11. McCarrier K, Ralston J, Hirsch I, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009 Apr;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]

12. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
13. Nguyen L, Bellucci E, Nguyen L. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform 2014 Nov;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011] [Medline: 25085286]
14. Ko H, Turner T, Jones C, Hill C. Patient-held medical records for patients with chronic disease: a systematic review. Qual Saf Health Care 2010 Oct;19(5):e41. [doi: 10.1136/qshc.2009.037531] [Medline: 20511601]
15. Goldzweig C, Orshansky G, Paige N, Towfigh A, Haggstrom D, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
16. de LS, Mold F, Sheikh A, Majeed A, Wyatt J, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
17. Gravis G, Protière C, Eisinger F, Boher JM, Tarpin C, Coso D, et al. Full access to medical records does not modify anxiety in cancer patients: results of a randomized study. Cancer 2011 Oct 15;117(20):4796-4804 [FREE Full text] [doi: 10.1002/cncr.26083] [Medline: 21607939]
18. Tuil WS, Verhaak CM, Braat DDM, de Vries RPF, Kremer JAM. Empowering patients undergoing in vitro fertilization by providing Internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [doi: 10.1016/j.fertnstert.2006.11.197] [Medline: 17416366]
19. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. J Am Med Inform Assoc 2005;12(5):505-516 [FREE Full text] [doi: 10.1197/jamia.M1700] [Medline: 15905487]
20. Fitton C, Fitton R, Hannan A, Fisher B, Morgan L, Halsall D. The impact of patient record access on appointments and telephone calls in two English general practices: a population-based study. London Journal of Primary Care 2014;6:8-15.
21. Bosman RJ, Rood E, Oudemans-van Straaten HM, Van der Spoel JI, Wester JPJ, Zandstra DF. Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery. Intensive Care Med 2003 Jan;29(1):83-90. [doi: 10.1007/s00134-002-1542-9] [Medline: 12528027]
22. Marasovic C, Kenney C, Elliott D, Sindhusake D. A comparison of nursing activities associated with manual and automated documentation in an Australian intensive care unit. Comput Nurs 1997;15(4):205-211. [Medline: 9260381]
23. Minda S, Brundage DJ. Time differences in handwritten and computer documentation of nursing assessment. Comput Nurs 1994;12(6):277-279. [Medline: 7812894]
24. Wong DH, Gallegos Y, Weinger MB, Clack S, Slagle J, Anderson CT. Changes in intensive care unit nurse task activity after installation of a third-generation intensive care unit information system. Crit Care Med 2003 Oct;31(10):2488-2494. [doi: 10.1097/01.CCM.0000089637.53301.EF] [Medline: 14530756]
25. Menke J, Broner C, Campbell D, McKissick M, Edwards-Beckett J. Computerized clinical documentation system in the pediatric intensive care unit. BMC Med Inform Decis Mak 2001;1(1):3-3. [doi: 10.1186/1472-6947-1-3]
26. Pabst MK, Scherubel JC, Minnick AF. The impact of computerized documentation on nurses' use of time. Comput Nurs 1996;14(1):25-30. [Medline: 8605657]
27. Pierpont GL, Thilgen D. Effect of computerized charting on nursing activity in intensive care. Crit Care Med 1995 Jun;23(6):1067-1073. [Medline: 7774218]
28. Bradshaw KE, Sittig DF, Gardner RM, Pryor TA, Budd M. Computer-based data entry for nurses in the ICU. MD Comput 1989;6(5):274-280. [Medline: 2486506]
29. Shu K, Boyle D, Spurr C, Horsky J, Heiman H, O'Connor P, et al. Comparison of time spent writing orders on paper with computerized physician order entry. Stud Health Technol Inform 2001;84(Pt 2):1207-1211. [Medline: 11604922]
30. Kovner C, Schuchman L, Mallard C. The application of pen-based computer technology to home health care. Comput Nurs 1997;15(5):237-244. [Medline: 9329225]
31. Ross S, Lin C. The Effects of Promoting Patient Access to Medical Records: A Review. Journal of the American Medical Informatics Association 2003 Mar 01;10(2):129-138. [doi: 10.1197/jamia.M1147]
32. Ross S, Todd J, Moore L, Beaty B, Wittevrongel L, Lin CT. Expectations of patients and physicians regarding patient-accessible medical records. J Med Internet Res 2005;7(2):e13 [FREE Full text] [doi: 10.2196/jmir.7.2.e13] [Medline: 15914460]
33. Markwell DC, Fogarty L, Hinchley A. Validation of a European message standard for electronic health records. Stud Health Technol Inform 1999;68:818-823. [Medline: 10725010]
34. Rimmer A. GP Online January. 2012. Patient access to medical records will add to GP workload URL: http://www.gponline.com/patient-access-medical-records-will-add-gp-workload/article/1113203 [accessed 2015-06-22] [WebCite Cache ID 6ZU02jNdw]
35. Archer N, Fevrier-Thomas U, Lokker C, McKibbon K, Straus S. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]

Abbreviations

EHR: electronic health record
HCP: health care provider
PAEHR: patient accessible electronic health records
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.03.15; peer-reviewed by B Fisher, R Fitton; comments to author 09.04.15; revised version received 28.04.15; accepted 24.05.15; published 29.06.15.

Please cite as:
Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A
"Nothing About Me Without Me": An Interpretative Review of Patient Accessible Electronic Health Records
J Med Internet Res 2015;17(6):e161
URL: http://www.jmir.org/2015/6/e161/
doi:10.2196/jmir.4446
PMID:26123476

Original Paper

Medical Student and Tutor Perceptions of Video Versus Text in an Interactive Online Virtual Patient for Problem-Based Learning: A Pilot Study

Luke A Woodham[1,2], BSc (Hons), MSc; Rachel H Ellaway[3], PhD; Jonathan Round[1,4]; Sophie Vaughan[1,4]; Terry Poulton[1], PhD.; Nabil Zary[2], PhD

[1]Institute of Medical and Biomedical Education, St. George's, University of London, London, United Kingdom
[2]Department of Learning Informatics Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[3]Human Sciences, Northern Ontario School of Medicine, Sudbury, ON, Canada
[4]St George's University Hospitals NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Luke A Woodham, BSc (Hons), MSc
Institute of Medical and Biomedical Education
St. George's, University of London
6th Floor, Hunter Wing
Cranmer Terrace
London, SW17 0RE
United Kingdom
Phone: 44 20 8725 ext 5559
Fax: 44 20 8725 0806
Email: lwoodham@sgul.ac.uk

Abstract

Background: The impact of the use of video resources in primarily paper-based problem-based learning (PBL) settings has been widely explored. Although it can provide many benefits, the use of video can also hamper the critical thinking of learners in contexts where learners are developing clinical reasoning. However, the use of video has not been explored in the context of interactive virtual patients for PBL.

Objective: A pilot study was conducted to explore how undergraduate medical students interpreted and evaluated information from video- and text-based materials presented in the context of a branched interactive online virtual patient designed for PBL. The goal was to inform the development and use of virtual patients for PBL and to inform future research in this area.

Methods: An existing virtual patient for PBL was adapted for use in video and provided as an intervention to students in the transition year of the undergraduate medicine course at St George's, University of London. Survey instruments were used to capture student and PBL tutor experiences and perceptions of the intervention, and a formative review meeting was run with PBL tutors. Descriptive statistics were generated for the structured responses and a thematic analysis was used to identify emergent themes in the unstructured responses.

Results: Analysis of student responses (n=119) and tutor comments (n=18) yielded 8 distinct themes relating to the perceived educational efficacy of information presented in video and text formats in a PBL context. Although some students found some characteristics of the videos beneficial, when asked to express a preference for video or text the majority of those that responded to the question (65%, 65/100) expressed a preference for text. Student responses indicated that the use of video slowed the pace of PBL and impeded students' ability to review and critically appraise the presented information.

Conclusions: Our findings suggest that text was perceived to be a better source of information than video in virtual patients for PBL. More specifically, the use of video was perceived as beneficial for providing details, visual information, and context where text was unable to do so. However, learner acceptance of text was higher in the context of PBL, particularly when targeting clinical reasoning skills. This pilot study has provided the foundation for further research into the effectiveness of different virtual patient designs for PBL.

(J Med Internet Res 2015;17(6):e151) doi:10.2196/jmir.3922

KEYWORDS

problem-based learning; educational technology; multimedia; Internet; audiovisual aids

Introduction

Virtual patients are interactive online tools that present learners with simulated patient encounters [1]. They are used in a range of contemporary medical educational settings, including small-group learning, lectures, self-directed learning, and assessment [2,3], as well as in other disciplines such as nursing [4] and primary care [5]. Virtual patients are generally Web-based which allows for a wide range of resources, such as multimedia or multiple-choice questions, to be included in their design.

The use of virtual patients has been linked to the development of learners' clinical reasoning skills [6] by allowing them to be active participants in a clinical situation, interpreting the available information, and making decisions based on what they know. The design of virtual patients generally follows 1 of 2 models: linear or branching [7]. Branching virtual patients are based on a decision tree that allows learners to make decisions at selected option points, thereby changing their path through the case. In contrast, learner interactions with linear virtual patients do not change the narrative of the scenario. Different paths can have different consequences, which can help learners to develop their clinical reasoning skills in ways that are safe, structured, and rich in feedback and instruction [8].

Research into the effective use of simulation for learning has identified the benefits of feedback and repetitive practice [9-11]. Virtual patients are a form of simulation and many of these factors have been used to guide their design [12]. Low-fidelity simulations, such as virtual patients, have a number of advantages over high-fidelity mannequin-based simulations. For instance, they are cheaper to produce and deploy and can (by being Web-based) be scaled to larger numbers of concurrent users. Indeed, Norman et al [13] argue that there is little educational advantage in using high-fidelity simulations over lower fidelity solutions, whereas Maran and Glavin [14] make a distinction between "engineering fidelity" and "psychological fidelity." Low-fidelity simulations, such as virtual patients, arguably have a low level of engineering fidelity (ie, the degree to which the physical characteristics of the task are represented) that can reduce their cost without reducing their psychological fidelity (ie, the degree to which skills of the task are captured by the simulation).

St George's, University of London in the United Kingdom developed a range of virtual patients for use in problem-based learning (PBL) [15]. This was done by adapting existing "paper" cases to include branches where learners could move through a case by selecting different patient-management options [16]. Each option took the learners down a different path, each of which was set up with different consequences for the development of the case. These case designs were rendered in an online virtual patient system so that multiple groups of students could use the same case simultaneously while tracking the different routes they took through the case.

In 2010, St George's, University of London transformed their PBL curriculum, replacing the paper cases with interactive online virtual patients and delivering this throughout the academic year [17,18]. This initiative was well-received by learners, the majority of whom preferred the virtual patient cases to paper-based cases [16]. This model of PBL has continued to be used to the present day. However, for the technology to be effective and sustainable, it required that the project team take a wider view of how to integrate technology into the PBL learning environment, establishing new procedures and guidelines beyond simply switching the paper resource to a branching virtual patient [17].

The educational community is taking an increasingly holistic view of the role of technology in education, acknowledging that an effective learning exercise depends greatly on the way that is implemented and the context within which it is implemented [19,20]. Ellaway [21] has proposed that virtual patients be considered from an activity-theoretical viewpoint; that learning is not intrinsic within the technological artifacts themselves, and that research should instead focus on the educational activities that virtual patients can be used to mediate. From this perspective, the virtual patient is a part of the scaffolding on which an activity is built along with factors such as the environment in which it takes place and the role of the tutor or facilitator. Educational technologies can be used in a variety of ways by learners with varying degrees of effectiveness [22], and likewise a single virtual patient can be used as a part of many different activities [23].

The need to situate a virtual patient resource within the activities that make use of them is not unique to virtual patients; a similar approach is required to guide the use of any technical artifact used in a learning activity. The primarily Web-based nature of virtual patient resources allows additional forms of media, such as images and video, to be easily incorporated within the virtual patient and their effectiveness should also be considered in the context of the activities in which they are used.

Although there have been several studies that have explored the impact of using video within traditional PBL curricula [24-27], there is little published evidence regarding the impact of using video in PBL that uses branching virtual patients targeting clinical reasoning, particularly for undergraduate learners. Bowen [28] identifies key elements of the clinical reasoning process to be that of data acquisition and the subsequent generation and identification of the problem to be addressed, leading ultimately to the generation of a hypothesis and diagnosis. Studies involving undergraduate medical students demonstrated a reduction in this type of critical thinking in nonbranching PBL following the introduction of video-based cases [25,26]. Kamin et al [29] identified a particular decrease in critical thinking at the point of identifying problems when using video-based PBL (compared to the same information provided as text), attributing this to the learners' need to perceive and articulate information from video. In contrast, this study also identified beneficial effects attributable to the use of video in other, later, stages of critical thinking. Studies focusing on postgraduate learners have identified similar benefits to the use of video [24]. De Leng et al [27] explored learner perceptions of video in a traditional PBL setting and proposed a series of guidelines for its effective use. They identified 4 key areas in which videos could enhance and add value to PBL cases: they were more authentic and illustrative, they provided a more comprehensive view of a scenario, they were more challenging for the learner, and they were more memorable. However, these guidelines did not take into account the particular characteristics that branching virtual patient resources brought to these activities, nor was there a specific focus on the development of clinical reasoning skills. Therefore, there is a need for evidence-based guidance on how to use video when developing Web-based virtual patients for use in PBL activities.

Our starting hypothesis was that virtual patients, and particularly branching virtual patients, are better suited to developing clinical reasoning skills than traditional PBL [15,30,31] and that (based on evidence from previous studies) the introduction of video elements to PBL can reduce the ability of undergraduate students to engage in critical thinking [25,26,29]. More specifically, we wanted to explore whether the use of video within a branching virtual patient for PBL could reduce the development of students' clinical reasoning skills due to the difficulty of critically evaluating the information provided in a video format compared to a text format. Therefore, we designed the study to address the question of how undergraduate medical students interpret and evaluate information provided by video, when compared with text, presented in the context of a branched interactive online virtual patient designed for PBL?

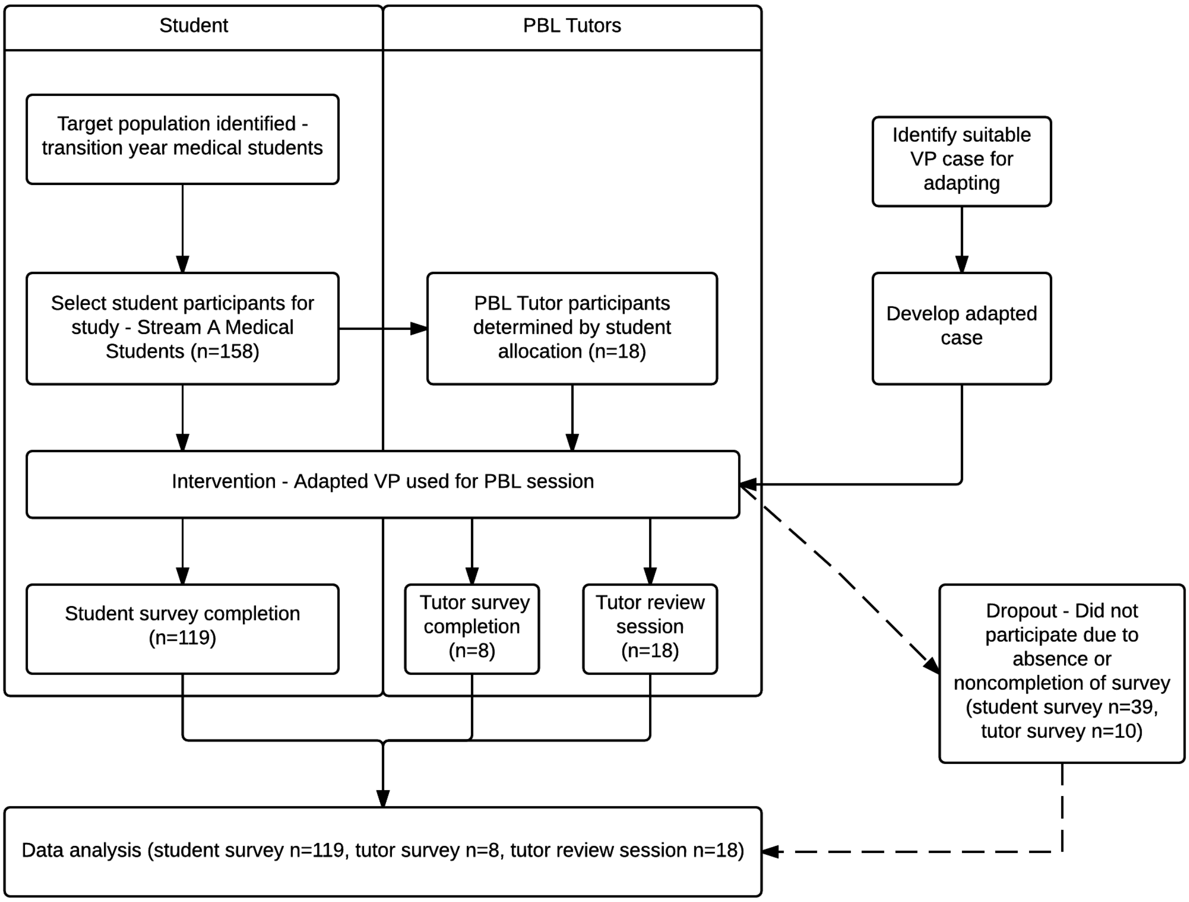
Methods

Overview

We created an educational intervention in the form of an adapted virtual patient case, in which the early stages of the case replaced text content with video. This virtual patient case was introduced into one week of the PBL curriculum for undergraduate medical students at St George's, University of London.

We used a mixed-methods approach [32,33] to capture the experiences of these students and the PBL tutors that facilitated the sessions. All participants had previous experience of PBL tutorials that used text-based virtual patients, so text was used as a baseline for comparison with video. Utilizing a convergent parallel study design [33] as described in Figure 1, we surveyed students and tutors through written questionnaires and ran a review session with tutors to capture their verbal feedback and merged the results in the analysis phase.

Figure 1. Flow diagram describing study design for using virtual patients (VPs) in problem-based learning (PBL).



Participants and Setting

The primary participants for the study were undergraduate medical students enrolled in the transition year at St George's, University of London. This was the second year of the program for graduate entry students and the third year for undergraduate entry students and was the point at which these 2 program streams joined.

Students undertook a program of PBL tutorials during the transition year. Running twice a week, there were 18 themed tutorials, each of which took place over 3 separate timetabled 3-hour sessions. For practical reasons, a student cohort was divided into 2 streams, with the groups completing the PBL tutorials at different stages of the year. Students from one stream (stream A) received the intervention in October 2013. Because PBL sessions were a mandatory component of the course, all students in the selected stream participated in the learning activity. Participation in the feedback activity was encouraged, but was not mandatory. All survey responses were anonymous.

Each PBL group consisted of 8 or 9 students, a mix of graduate entry and undergraduate entry learners. The PBL sessions were facilitated by tutors who also participated in the study. Their role was noninterventional; they did not teach but guided the session to ensure that the groups stay focused and covered the requisite learning objectives. Each PBL group was facilitated by a tutor who worked with the same learner groups throughout the year. The role of a tutor in PBL requires a specific approach and a particular set of guidelines must be followed; all PBL tutors received training in their roles and PBL techniques, but were not selected on the basis of any subject knowledge. The tutorial in which the intervention was introduced was conducted in the same manner as any other PBL session; experienced PBL tutors did not require or receive any specific training related to the intervention. However, they were informed before the date that the intervention would be introduced and given the opportunity to raise questions about the study with the research team.

The study was reviewed and approved by the chair of the St George's, University of London Ethics Review Board and approved by the undergraduate program course director. To provide assurance that students received no advantage through either participation or nonparticipation, all students were provided with access to both the video and text-based versions of the tutorial after the PBL session had taken place using the institutional learning management system. The video virtual patient was reviewed by the module leads before the intervention was delivered to confirm that the content was suitable for use, accurate, and that all the required information for the case was still delivered. The module leads agreed that the content was suitable and gave their consent to its use.

Intervention

One section of a preexisting virtual patient PBL case was adapted by replacing the textual information with video clips. We used the following predetermined criteria to select a suitable virtual patient case for adaptation from the existing St George's, University of London PBL curriculum:

1. The case had to have a narrative that could be staged using the equipment and performers available in the St George's Advanced Patient Simulator simulation and skills training center. This eliminated cases that took place primarily in nonclinical environments.
2. The part of the case to be filmed also had to require learners to interpret only information that could be effectively shown and visualized through video, meaning sections of the case which required analysis of detailed test results, scans, or other similar information were not considered suitable for video representation.
3. The timing of the case in the curriculum was also a critical factor; it was necessary that it took place early in the year so that students were not so familiar with text-based cases that it would prejudice their perceptions against a change to video.

Having reviewed all the available virtual patient cases against these criteria to determine their suitability for adaptation, we selected the second part of a 3-part virtual patient case regarding a patient suffering an abdominal aortic aneurysm. Considering the established typology for virtual patients [7], the selected virtual patient had a number of defining characteristics. The case was presented in the English language using a branching path model to target undergraduate medical students in the context of a PBL tutorial. It was designed to be used in a PBL tutorial with standard-sized groups of 8 or 9 participants and a duration of 3 hours. In the virtual patient scenario, students were asked to assume the role of a Foundation Year 2 (second year of postgraduate training) doctor, with the focus being decisions to be about patient treatment [17].

We storyboarded the first 9 stages of this for filming, taking in the first 2 decision points that learners were required to negotiate during the PBL session. This was the first case that was scheduled to run during the academic year, helping to ensure that the learners were not already too familiar with the text-based tutorials used elsewhere in the curriculum in advance. Each clip was designed to provide both the scenario narrative and the relevant information that learners needed to make effective patient-management decisions.

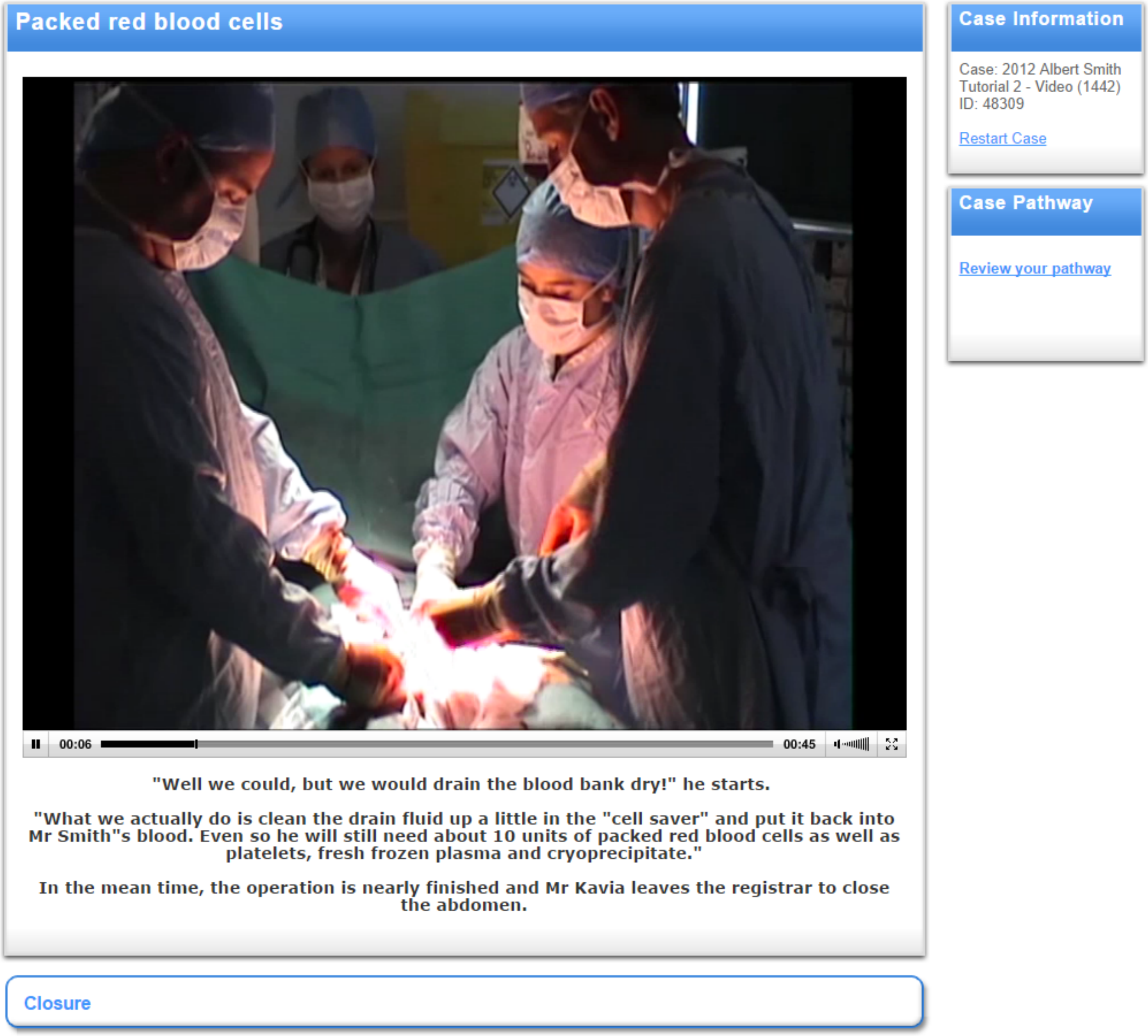
The video material was created in partnership with the St George's Advanced Patient Simulator center because they had the facilities to create an approximation of the required settings. The videos were recorded in simulation rooms from 4 angles, with sound captured from room-mounted and individual wireless microphones worn by performers.

The filming was completed over 2 half-day sessions using volunteer actors and 5 cameras. Four cameras were fixed-viewpoint cameras available in the simulation center, whereas an additional portable camera was used to capture close-up shots and other viewpoints. The simulation center was used to stage scenes representing a recovery room, an operating theater, and a ward environment. Approximately 4 hours of footage was captured from each camera and this was edited down into 9 video clips varying from between 45 seconds to 4 minutes in length. These video clips were then embedded into the virtual patient case, replacing the text in this part of the tutorial. Examples of the video clips are provided in Multimedia

Appendices 1 and 2. The video content was reviewed for suitability and accuracy by the academic leads for the relevant module of the course and approved for use. A screenshot of a video in a virtual patient case is shown in Figure 2.

Teaching sessions took place in dedicated small-group teaching rooms; students and tutors were arranged around a table with an Internet-connected computer workstation attached to an interactive SMART Board [34] and projector situated at one end of the room, which was used to display the virtual patient case to the group.

Figure 2. Screenshot showing a video clip embedded in the online virtual patient activity used in this study.



Packed red blood cells

"Well we could, but we would drain the blood bank dry!" he starts.

"What we actually do is clean the drain fluid up a little in the "cell saver" and put it back into Mr Smith"s blood. Even so he will still need about 10 units of packed red blood cells as well as platelets, fresh frozen plasma and cryoprecipitate."

In the mean time, the operation is nearly finished and Mr Kavia leaves the registrar to close the abdomen.

Closure

Case Information

Case: 2012 Albert Smith
Tutorial 2 - Video (1442)
ID: 48309

Restart Case

Case Pathway

Review your pathway

Instrument Development

We conducted this mixed-methods study using structured and unstructured data gathered from 3 sources: a student survey, a survey of PBL tutors, and a discussion and review session with PBL tutors.

We developed the survey instrument from an established instrument for exploring student experiences using virtual patients [35-37]. Questions were added regarding students' perceived ability to evaluate the information available in the scenario, their understanding of the context provided by the scenario, and their sense of engagement with the case. There were 19 questions in total. The first 2 questions (room number and course type—undergraduate entry or graduate entry) were used to categorize the data for analysis and the third question asked for a perception of how many times their group watched each video on average. The remaining questions provided for structured responses in the form of a multiple-choice or Likert scale answer followed by an unstructured response to provide further detail or explanation. The full survey instrument is provided in Multimedia Appendix 3. Given the established basis of the parent instrument and time constraints in executing the study, we did not pilot the survey instrument further. We provided the survey instrument to student participants in paper form as part of the packs that accompanied the PBL session to encourage on-the-spot completion. We entered the paper-based responses into a Web-based system [38] to allow for combined reporting and analysis.

The experiences and thoughts of the PBL tutors were captured through a review session and a distinct survey instrument tailored to the tutor experience. The survey instrument was based on those developed in a previous study relating to tutor perceptions of virtual patients in a virtual world [39] and adapted to the context of this study. The survey instrument (see Multimedia Appendix 4) was provided to the tutor participants in paper form for on-the-spot completion at the time of the PBL session in which the intervention was delivered. One of the researchers (LW) conducted the review session in November 2013 and a semistructured approach was identified as being appropriate for the review session due to the research team’s existing knowledge of the domain [40] and a question script was developed to guide the discussion (see Multimedia Appendix 5). The session was audiorecorded and later transcribed for analysis. The participants in the review session were PBL tutors that had facilitated the tutorial in which the video-based virtual patient intervention was introduced.

Analytic Approach

We analyzed the unstructured free-text data from the survey responses and reviewed session transcripts using a theoretical thematic analysis approach [41]. The data were manually coded by one of the authors (LW) using ATLAS.ti software [42] and the codes generated were developed through iterative readings of the datasets. Individual sentences in transcripts and free-text responses were identified as the units of analysis for coding [43] to ensure that all themes expressed could be identified with sufficient granularity. We used an open-coding approach for the first reading of the data, in which themes grounded in the data were noted. A second reading continued with this approach, identifying information overlooked during the first iteration. In accordance with a theoretical thematic analysis model and in contrast to an inductive analysis of the data, the coding process was conducted with a view to the specific area of research examined in this study and did not attempt to codify the responses beyond this context. Subsequent iterations moved toward an axial coding model [44]. With each reading the generated codes were further refined, thematic linkages between codes were noted, and codes with common meaning were merged and grouped. After the sixth reading of the data, all the thematic groupings were clearly distinct and no new codes emerged, and the researchers were satisfied that the coding process had allowed a number of broad, descriptive themes to be identified in the data.

We analyzed the structured responses by converting Likert scale matrix values to ordinal form (strongly disagree=1; strongly agree=5) and by generating descriptive statistics. We categorized data by course of study (undergraduate entry or graduate entry) to control for any bias caused by differing levels of experience among student participants and we used 2-tailed Mann-Whitney *U* tests to identify if there were any statistically significant differences between the 2 groups.

Results

Overview

Out of 158 students registered to attend the PBL session, 119 responded to the student survey giving a response rate of 75.3%, although some students did not answer every question. We considered this response rate to be satisfactory and attributed noncompletion of the survey to a combination of absence from the session or students electing not to provide feedback. In addition, responses were not received from 2 of the 18 student groups, suggesting that the tutors responsible for collecting the student responses in those rooms had either not distributed or simply did not return the survey instruments. In total, students provided 274 open-text comments in their survey responses.

The tutor survey received 8 responses, a response rate of 44% (8/18), with 21 open-text comments. In both datasets, responses that were null, “n/a,” or simply “no” when asked for further details were excluded. Due to the small sample size (n=8) for the PBL tutor survey response, we determined that generating descriptive statistics for this would be of little value and unreliable, although the unstructured tutor responses were included in the qualitative analysis. The low response rate for the PBL tutor survey was attributed to the timing of the survey data collection because it took place at the end of the session when tutors had a number of other tasks to complete, such as gathering materials and feedback from students, meaning that time for them to complete the survey was scarce. It was also noted that the tutor survey had been provided in the information pack provided to tutors, which also included the student survey forms. The 2 instruments were both printed on white paper and were, therefore, not immediately distinguishable. Although tutors had been previously briefed, a possible explanation for the low response rate was that they were unaware the tutor survey had been included.

Of the 16 groups that returned responses, 7 reached a consensus on the number of times each video was watched, 6 stated that they had watched each video once on average and 1 group reported having watched each video twice. No consensus was reached for the other 9 groups, with responses ranging between 1 and 2 viewings.

The descriptive statistics for the Likert scale items from the student participants are shown in Table 1.

Table 1. Descriptive statistics for Likert scale student survey responses (N=119).

Statement and medium encountered	Responses, n	Mean score (SD)
While working on this case, I felt I had to make the same decisions a doctor would in real life.		
Text	119	3.87 (0.75)
Video	116	3.82 (0.90)
While working on this case, I felt I were the doctor caring for this patient.		
Text	119	3.53 (0.85)
Video	114	3.41 (1.04)
Watching the scenario take place in the videos made me feel more emotionally involved with the case than when playing the role of an F2 doctor in the text.	118	2.99 (1.06)
Playing the role of an F2 doctor in the text-based parts of the tutorials increased my engagement with the scenario compared with watching the videos.	119	3.30 (0.88)
The use of video brought the scenario to life.	119	3.49 (1.02)
The use of video made the scenario more memorable.	118	3.62 (1.07)
The use of video influenced the option choices that my group made.	119	2.94 (0.87)
The use of video helped me to relate the scenario to real-life experience.	119	3.49 (0.93)
I was able to obtain all the information from the videos that I needed in order to make informed patient-management decisions.	119	2.99 (1.03)
I felt that it was easier to identify relevant information from text than the videos.	119	3.75 (1.00)
The use of video had a positive impact on the group discussion.	114	3.29 (0.89)

To control the impact of PBL groups with students with different levels of prior experience we categorized data by group; graduate entry students and undergraduate entry students. This yielded 2 independent ordinal datasets for each Likert item. We ran nonparametric 2-tailed Mann-Whitney *U* tests on the datasets to test whether the distributions for graduate entry and undergraduate entry students were significantly different at the 5% level ($P<.05$) (see Table 2). In each case, the null hypothesis (ie, that the 2 groups had the same distribution) could not be rejected indicating that the graduate entry or undergraduate entry status of students, and thus the different nature of the prior experience of these 2 groups, did not significantly impact on their experiences of the intervention and did not have a material impact in skewing the data. We concluded that responses for the different groups did not have to be separated in the analysis and that conclusions drawn would be applicable and generalizable across both groups.

Table 2. Descriptive statistics for Likert survey items categorized by learner stream (undergraduate entry or graduate entry).

Survey question	Graduate entry (n=46)			Undergraduate entry (n=73)			*U*	*P*
	Responses, n (%)	Mean (SD)	SEM	Responses, n (%)	Mean (SD)	SEM		
While working on this case, I felt I had to make the same decisions a doctor would in real life.								
Text	46 (100)	3.87 (0.72)	0.11	73 (100)	3.86 (0.77)	0.09	1651.50	.87
Video	46 (100)	3.96 (0.89)	0.13	70 (96)	3.73 (0.90)	0.11	1341.50	.10
While working on this case, I felt I were the doctor caring for this patient.								
Text	46 (100)	3.35 (0.97)	0.14	73 (100)	3.64 (0.75)	0.09	1451.50	.18
Video	45 (98)	3.38 (1.09)	0.16	69 (95)	3.43 (1.01)	0.12	1533.50	.91
Watching the scenario take place in the videos made me feel more emotionally involved with the case than when playing the role of an F2 doctor in the text.	46 (100)	3.07 (1.08)	0.16	72 (99)	2.94 (1.04)	0.12	1545.50	.52
Playing the role of an F2 doctor in the text-based parts of the tutorials increased my engagement with the scenario compared with watching the videos.	46 (100)	3.24 (0.87)	0.13	73 (100)	3.34 (0.88)	0.10	1581.00	.57
The use of video brought the scenario to life.	46 (100)	3.48 (0.91)	0.13	73 (100)	3.49 (1.09)	0.13	1601.50	.65
The use of video made the scenario more memorable.	46 (100)	3.63 (1.08)	0.16	72 (99)	3.61 (1.07)	0.13	1633.00	.89
The use of video influenced the option choices that my group made.	46 (100)	2.91 (0.81)	0.12	73 (100)	2.96 (0.90)	0.11	1670.00	.96
The use of video helped me to relate the scenario to real-life experience.	46 (100)	3.52 (0.94)	0.14	73 (100)	3.47 (0.93)	0.11	1589.50	.59
I was able to obtain all the information from the videos that I needed in order to make informed patient-management decisions.	46 (100)	3.17 (0.90)	0.13	73 (100)	2.88 (1.09)	0.13	1424.00	.15
I felt that it was easier to identify relevant information from text than the videos.	46 (100)	3.67 (0.94)	0.14	73 (100)	3.79 (1.04)	0.12	1514.00	.33
The use of video had a positive impact on the group discussion.	43 (100)	3.40 (0.88)	0.13	71 (97)	3.23 (0.90)	0.11	1394.00	.41

Analysis

Our thematic coding process drew on the unstructured responses of both tutors and students, and led to 67 distinct codes grounded in the data during the open-coding process. Five codes were automatically generated using ATLAS.ti [42] for the purpose of categorizing the open-text survey responses and did not relate directly to the research question. Therefore, we disregarded these for the purpose of the thematic mapping exercise.

Table 3 shows the high-level themes we identified in the analysis. We identified thematic links between the remaining codes by reviewing the codes for similarities in meaning and relevance. As a result of this exercise, 8 clear thematic groupings emerged (the level of engagement appeared twice with some identifying text as more engaging and others feeling that video was more engaging).

Table 3. Summary of high-level themes identified and the number of quotations coded against each theme.

High-level theme	Code-quotation count
Video made the scenarios more real	70
Hard to identify key information in video	68
Video more engaging	55
Poor sound quality	44
Text can be reviewed	34
Would favor a text script to complement video	31
Text more engaging	21
Video slows the pace of PBL	15
Video well-suited to showing procedures	10

Level of Engagement

We found quite varied perspectives on the effectiveness and desirability of using video as part of the PBL process. For instance, 55 comments indicated that the video had had a positive impact, whereas 20 statements described a negative effect. Those students who described a positive engagement with video noted that it provided the scenario with more “immediacy” and “involvement.” Several students raised the countervailing idea that text “allows more room for imagination.”

The structured data reflected a similar lack of consensus when considering the merits of using video to heighten engagement. We noted before the intervention that the role of the student was altered by the introduction of video; when using text the student was addressed as if they were the doctor, whereas in the video the student was at most observing the doctor. We anticipated that this may have had a negative impact on student engagement with the scenario, although there were other more positive factors, such as providing richer visual information and context. To explore this effect in more detail, individual students were asked whether they felt they had to make the same decisions as a doctor and whether they felt they were the doctor caring for the patient for both the video and text components of the virtual patient. We used sign tests (*Z*) because of the ordinal and dependent datasets to test whether the median values were different for video and text (Table 4). In each case, the null hypothesis (ie, that the distributions were the same) could not be rejected at the 5% level (*P*<.05), indicating that the use of video did not significantly impact on student responses to this statement when compared to their response for text.

Table 4. Results from 2-tailed sign test (*Z*) for individual student responses to Likert items comparing text and video.

Likert item	Negative differences, n	Positive differences, n	Ties, n	Total, n	*Z*	*P*
While working on this case, I felt I had to make the same decisions a doctor would in real life.	23	23	70	116	<0.001	>.99
While working on this case, I felt I were the doctor caring for this patient.	27	25	62	114	–0.139	.89

It Can Be Harder to Identify Relevant Information From Video

Students found that it was harder to identify relevant information in the video compared to text. Many of the students identified that they missed key bits of information in the video and that this confused the group. One participant commented that the video was “unclear and lacked direction and confused us more as a group.” This perceived information deficit was felt by some participants to have reduced the quality of discussion and that they “learned more and had more information available to discuss with text in front of me.”

Text Can Be Reviewed More Easily Than Video

We identified that a key advantage of text was that it can be reviewed and revisited more easily and on an individual basis by students during the PBL session. Many students wanted to be able to refer back to the text and identified that the nature of video meant that during discussions they did not have the ability to refer back to the source material. Although the video could be replayed, and many groups confirmed that they played the video multiple times, it was not possible to view the video and discuss simultaneously. One student commented that “it’s easier to check facts when debating” using text.

Video Slows the Pacing of Problem-Based Learning Activities

Many of the students and tutors identified that the use of video had a significant impact on the pace of a PBL session. Some students identified that this caused an increase in the overall length of the session by approximately 30 minutes. One tutor in the review session identified that the pace and responsiveness of the discussion was slowed by the introduction of video:

> *I think we took longer because after they watched the video we still had to stop and talk about things that the video raised, but we couldn’t talk about them at*

> *the time as we were watching the video, so for us the case actually took longer than it might have done if we were to raise learning objectives and discussing.*

The student survey indicated that 38.5% (45/117) of the students felt that they on average watched each video twice or more than twice, which also had the effect of slowing the pace of the session. Comments also considered this in relation to the greater difficulty in reviewing video compared to text, stating that they had to “repeatedly watch” the video. It was suggested that providing a text transcript would alleviate this issue by eliminating the need for repeated viewings.

Video Made the Scenario Seem More Real

Many students commented on ways in which they felt the use of video made the scenario more real. The comments reflected a broad range of reasons, but with a common thread that the use of video “brought it to life.” Some comments related to the students being able to identify themselves and their role in the scenario: “it made it more real as if I was present whilst the whole situation was happening.” Another felt that video “portrayed the urgency of the decision that needed to be made.”

A particularly common observation among the students was that the video provided additional visual cues that were not present in the text. The nature of these visual cues varied greatly, some students identified the impact of seeing the social interactions taking place or of seeing reactions and facial expressions in the video. Others mentioned the importance of observing the environment in which the patient encounter occurred, saying that it “contextualizes the scenario” and provided “clues about what is in environment, IV drips, blood, etc.” Additionally, some students described the impact that these visual clues had on the decision-making process and stated that “visualizing blood loss on the video altered our initial opinion on what to do next.”

Video is Well-Suited to Displaying Procedures

A number of the participants (both tutor and student) suggested that video could provide particular advantages over text in representing clinical procedures. A significant component of the video we used showed a simulated surgical procedure to treat an abdominal aortic aneurysm. However, we should be clear that the clip was not intended to teach the procedure, but to further the narrative of the scenario. Nevertheless, its inclusion triggered responses indicating that video was perceived by students to be a superior way of learning about such procedures compared to text, comments including “some of the practical procedures (ie surgery) are better explained and understood if it was demonstrated ‘in action’ in the form of a video.” Other comments recognized the benefits of seeing video of procedures, but questioned the value of embedding them within patient scenarios, stating that it would be just as valuable to link to YouTube videos showing the same procedures.

Students Favor a Combination of Text and Video

We asked student participants whether they felt that the video was effective and whether they preferred video or text. The responses to these questions showed similar patterns regardless of year of study, with a majority of students feeling that the use of video was effective (Figure 3). However, when asked to state a preference for video- or text-based scenarios, the majority of students expressed a preference for text (Figure 4).

Several students suggested that an optimum arrangement would be to have a combination of video and text; one student suggested “a mix so memorable but also easy to understand.” Another agreed, commenting that a combined approach “would allow us to see the key case easily whilst seeing a more realistic scenario in the video.” Others asked for a text transcript of the content in the video. Other views indicated that the combination approach would best be achieved by providing the video first then the text equivalent.

Figure 3. Bar graph of student responses to question “Do you feel that the use of video in the tutorial was effective?”.

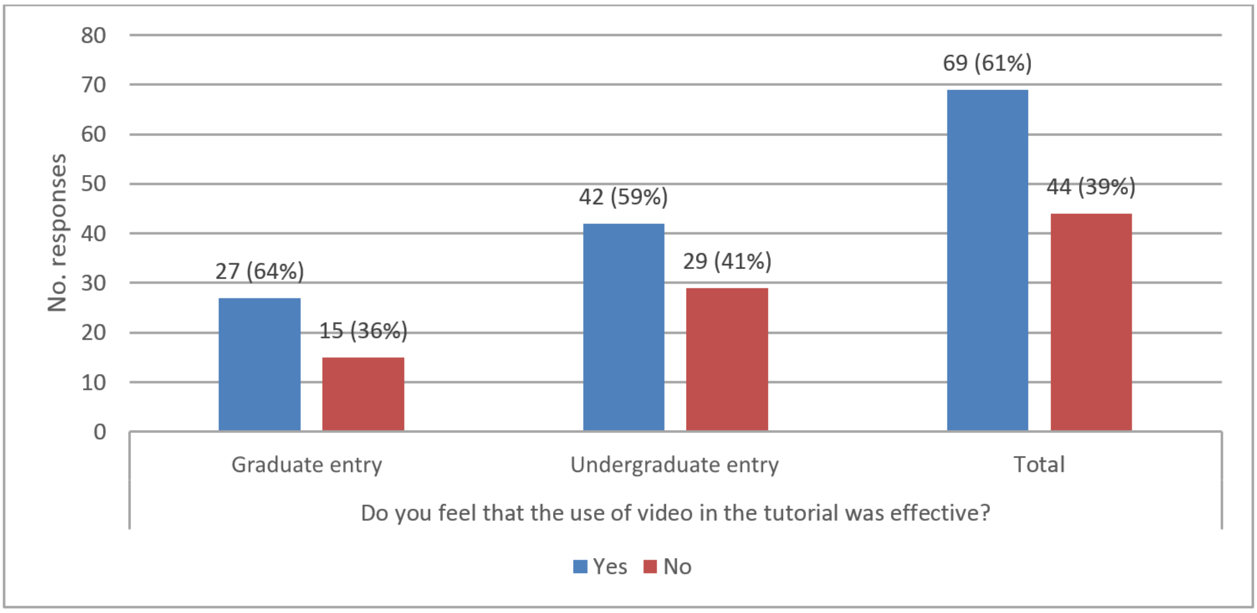
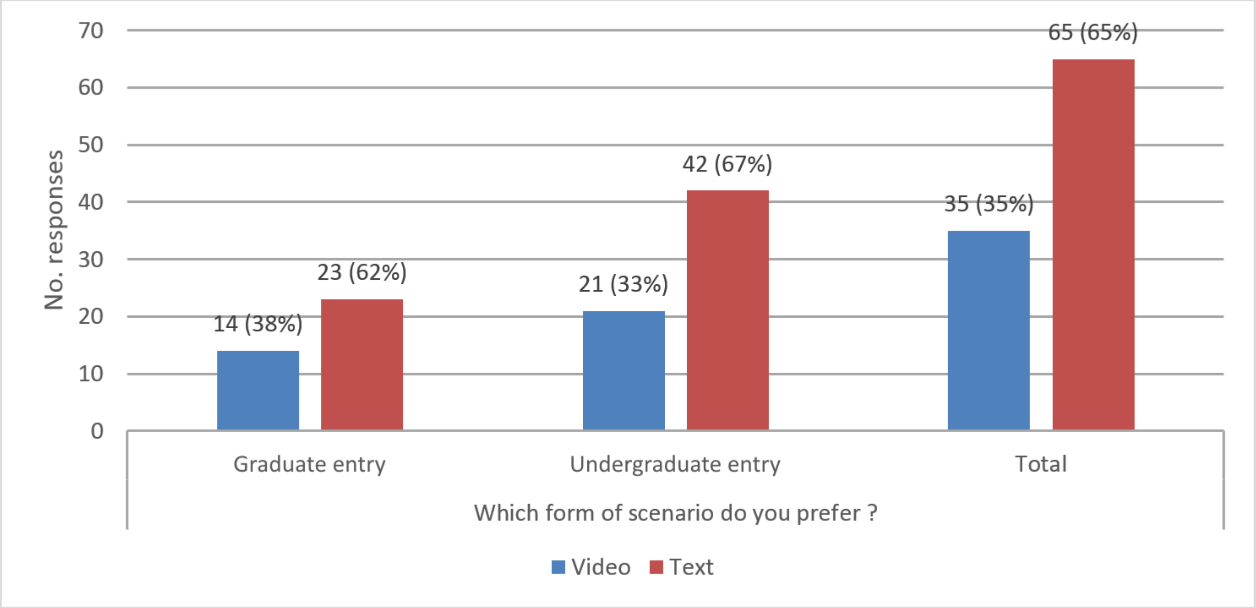


Figure 4. Bar graph of student responses to question “Which form of scenario do you prefer?”.



The Quality of the Video Resource

Many tutors and students commented on the quality of the video resource and the impact that it had on the session. A particular concern of many was the perceived poor sound quality, with participants stating that the sound was “hard to hear,” “muffled,” and “occasionally distorted.” Examination of the video clips revealed that there was audible distortion present in the original recording at 2 points in the videos. Inspection of the rooms in which the sessions took place also revealed that the character of the frequency response of the audio systems in those rooms had the effect of muffling the speech in the clips and accentuating the background noise making the speech less clear than had been apparent when preparing the clips using headphones.

Additional comments were made about the quality of the video relating to the editing of the clips (eg, cuts were too frequent) and, in particular, the quality of the acting. The actors in the clips were volunteers from the project team and as a consequence were perceived to lack proficiency. It was felt by some that the acting quality distracted from the learning task and that “random clips of poor acting doesn’t translate information effectively.”

Discussion

Principal Results

Our study investigated the impact of video clips replacing related text content on a PBL session run using an interactive virtual patient. In particular, we considered the ability of students to process and evaluate the information provided in the video. What became clear on analysis of the results was, although a number of themes emerged, the findings showed a diversity of opinions among students. When asked the question explicitly, a majority of the student participants indicated that they preferred text-based virtual patients for PBL, while acknowledging the benefits of using video in certain circumstances. The contrast between the majority of students’ stated preference for text, yet their widespread identification of the benefits of video, may potentially be explained by understanding the motivation and challenges faced by today’s medical students, who are required to assimilate a significantly increased volume of knowledge and to continue to do so throughout their professional development without a comparable increase in time [45]. Sobral [46] described the motivation of medical students as depending on both extrinsic and controlled factors (ie, the course structure and need to pass certain assessment targets) as well as intrinsic and autonomous ones (ie, the enjoyment of learning). Our results indicate that extrinsic factors were of primary importance to our students, who were necessarily focused on acquiring the ever-widening pool of knowledge needed to qualify and pass their exams. The introduction of video required learners to employ a greater level of critical analysis to extract and evaluate the available information; hence, the students expressed a widespread preference for text.

Participants recognized the value that video provided in terms of engagement and the provision of visual information, commenting that it brought the scenario to life for them. However, the additional challenges that video brought were also demonstrated clearly in our results; the widespread belief among participants was that information was harder to identify in a video clip. Existing studies identify that the use of video makes the initial stages of critical thinking more challenging for undergraduate students [25,26,29], particularly at the problem-identification stage that requires the information provided to be evaluated and synthesized by students. This is made more difficult using video because students have to filter and evaluate a larger volume of information, including visual and auditory information, to extract the key points. Cognitive load theory describes the nature of this increased challenge; learning is impaired when the cognitive load of a task is greater than an individual’s working memory [47]. The introduction of video imposes cognitive load that learners perceive as extraneous and unnecessary for them to achieve their immediate learning objective, which is to acquire the knowledge to pass exams. It is arguable that because the goal of the branching virtual patient is to develop clinical reasoning skills for real-life practice, the ability to critically filter the available information is far from extraneous. However, as a consequence of this potential misalignment in the perceived learning objectives for the virtual patient between educators and students, the effect was that video was perceived to be less efficient than text as a method of acquiring information. Further research is required to investigate this contention and to more fully understand learner motivation when participating in PBL activities.

The time taken during the learning activity was shown to be a priority issue among participants, with many responses noting that video slowed the pace of the PBL session. Groups were generally unable to agree on the extent of this effect, however, with only a minority of groups able to reach a consensus on how many viewings of each video were required. Moreover, video was considered harder to review multiple times to identify the information; repeat viewings had to always be watched by the whole group because an individual could not rewatch the video in isolation. Text, on the other hand, provided scrolling was not necessary, allowed all the information to be visible on screen at one time and could be reviewed independently by individuals without requiring the whole group's attention. This loss of individual agency over the resource when wanting to review the information compounded the difficulty that learners faced in effectively evaluating the information provided and was a key factor in student perception of its effectiveness. This supports the conclusions reached in previous studies [48], which identified the importance of student control when using multimedia resources.

Student perceptions of the technical qualities of the video also impacted upon the perceived effectiveness of the resource. In videos where the students perceived the audio quality or the proficiency of the actors in the clip to be deficient, many commented that it had a negative effect on their learning. The key point from this is that video is not intrinsically educationally useful (or not) because production quality, editing, and a number of other contextual factors can separately impact its utility.

Existing work has proposed specific design principles for virtual patients [12], which include the appropriate use of media and the authenticity of the interface. Although the proposed principles do not address virtual patients for PBL specifically (which have very specific requirements and warrant consideration in isolation from virtual patients intended for other purposes), the use of video as media within a virtual patient for PBL has a profound impact on the authenticity of the virtual patient interface as a means for engaging the student with the scenario. The principles identify that the use of media should be preferred when it provides a superior means for explaining or providing information to learners. Comparing our results with the factors identified by De Leng et al [27], learners acknowledged that video can provide a comprehensive and illustrative representation of a scenario, it can convey a large amount of visual information, and it can help to make the scenario feel more real. Our study did not attempt to establish whether the information provided in video is more memorable and further research would be required to address that particular question.

However, the claim that video is more authentic is potentially a contentious one. Differing perspectives on what is represented by authenticity make it difficult to validate claims of video being more authentic; the concept of "thick" authenticity suggested by Shaffer and Resnick [49] identifies different types of authenticity in a learning experience, each of which is interdependent with other types. Similarly, the literature on simulation distinguishes between engineering fidelity and psychological fidelity [13,14]. Several of the participants in this study felt that the use of the video made the scenario seem more real, suggesting increased authenticity, and it is clear that the means in which video represents a scenario mirrors real life more closely than text (ie, video provides a higher level of engineering fidelity). Yet when viewing the intervention as a learning activity in the context of the PBL session itself as proposed by Ellaway [21], we must also consider other factors: the video clips were paused and watched multiple times and decisions were reached by consensus without the pressure and time restrictions that would be present in real life. This indicates that the use of video did not serve to increase the psychological fidelity of the learning activity. Given this wider context, the suggestion that the learning activity is more authentic when using video than it is when using text begins to break down and raises the further question of whether authenticity should necessarily be an aspirational characteristic for a learning activity. We have established that the increased challenge provided by video during this intervention reduced the efficiency in which learners were able to achieve their learning goals. Norman et al [13] report that the fidelity of a simulation has little bearing on the effective transfer of learning and our results suggest that indiscriminate use of video aiming to simply increase the authenticity of virtual patient resources may show a similar pattern, particularly if little thought is given to the authenticity of the learning activity mediated by the virtual patient.

However, if effectively targeted to information where it is well-suited, our results indicate that the use of video can be an effective complement to text in PBL activities. It was widely commented in our study that video was well-suited to demonstrating procedures in a way that text cannot. A combined approach, favored by many of the participants, in which a combination of text and video is used, would provide a means to focus the use of video on areas of the virtual patient where it was beneficial to the intended learning activity, while using text to efficiently deliver learning in the areas where it is most effective. The provision of a text transcript that accompanies the video would run counter to the principles proposed by Mayer's cognitive theory of multimedia learning [50], who points to a redundancy effect achieved when utilizing multiple modalities of delivering information in multimedia resources [51]. However, it is noted that this study is primarily focused on individual and self-directed learning materials. Our results indicate that the group dynamic in PBL sessions based around branched virtual patients may warrant an alternative approach.

Limitations

This study represents a first step toward investigating the effects of such interventions, piloting the use of video within an interactive virtual patient designed for PBL. However, there are a number of limitations to the study:

1. The findings from this study are based on a single intervention and setting. Further research with a greater number of similar interventions will be required to validate the findings and to establish the broader generalizability.
2. Although the student survey instrument was based on an existing validated instrument [35,36], both this and the tutor survey instrument could have been more robustly validated. Our decision not to engage in substantial validation reflects the pilot nature of the study.
3. Our focus on experience and perception rather than quantifiable outcome measures (ie, exam or clinical skill performance) reflects the broad and relatively unstructured nature of PBL outcomes [52] and the challenges of assessing PBL as a whole [53]. For a pilot study, we were more interested in understanding how learners responded to these different stimuli and how they understood and rationalized these responses rather than quantifying their responses.
4. Although the use of self-reported data would be a potential limitation for a more quantitative study (because of recall and response bias), this was not a significant concern given the proximity of survey completion to the educational event, the use of multiple sources of data, and the pilot nature of the study.
5. Although several students identified that videos of procedures would be beneficial and provide value, the piloted intervention did not include instructional material targeting knowledge of procedures, instead including a simulated procedure to serve the narrative of the scenario. Therefore, the suggestion that video may be well-suited to targeting this area has not been tested in this study and should be the focus of future inquiry.
6. There was widespread dissatisfaction with the quality of the sound in the video clips that were developed for the intervention and this seemed to have had a substantial impact on learner perceptions of the utility of the resource. There was some suggestion that the sound on the clips was not clearly audible using the equipment in the PBL rooms and that this may have exacerbated some of the effects we observed. Learners played clips multiple times to fully understand the provided information may have slowed the PBL process more than it otherwise would have done. Any future studies should include procedures to test the media in the PBL rooms as well as on individual workstations to ensure that it plays with sufficient clarity in that environment.
7. We felt that learner perceptions of the number of times they had played video clips was a relatively crude and unreliable measure and that in future work we would seek to use log data to track the number of times a clip had been played.
8. Finally, this study examined the effect of the intervention at a specific institution, where all the learner participants in the study were previously familiar with the PBL process and had experience of text-based scenarios as a baseline experience. More research is required before the conclusions reached here can be safely generalized to other educational settings. We also urge caution in generalizing our findings to other types of learning activity in which learning may be self-directed or lecture-based. Indeed, our finding that the circumstances of the PBL activity reconfigured the utility of the video and text resources would suggest that the utility and efficacy of video material should be explicitly tested for different educational activities and settings.

Conclusions

This pilot study introduced video clips into a virtual patient resource, replacing existing text content, and tested it with undergraduate learners as part of an interactive online PBL session. We explored the impact that this intervention had on participants' ability to access and evaluate the information provided in the resource and their perceptions of how effective the approach was. The results identified both positive and negative effects from the introduction of the intervention. Students identified value in the video resource, but when asked to state a preference, the majority chose the text-based resource. Course of study (graduate entry or undergraduate entry), and accordingly varying levels of experience, did not impact learners' stated preference.

Our results lead us to conclude that a combination approach may be a superior one within the context of undergraduate PBL. Using video only for elements to which it is particularly suited (ie, displaying procedures) may reduce any negative impact on the pace of the learning activity and would reduce any extraneous cognitive load introduced by video that might reduce the efficacy of the learning resource. Further research is necessary, in particular larger scale studies using a greater number of virtual patient interventions and contexts. However, despite the provisional nature of our findings, we have illustrated the context dependency of the perceived value of different multimedia components in a small-group PBL setting and in doing so we have developed a richer understanding of the role of educational multimedia in health professional education.

Acknowledgments

We wish to acknowledge the learner and tutor participants in this study, the staff at the St George's Advanced Patient Simulation (GAPS) center, the volunteer actors in the video clips, and the PBL team and e-Learning Unit at St George's, University of London.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video clip entitled “Surgical Progress”.

[MP4 File (MP4 Video), 13MB - jmir_v17i6e151_app1.mp4]

Multimedia Appendix 2

Video clip entitled “Packed red blood cells”.

[MP4 File (MP4 Video), 13MB - jmir_v17i6e151_app2.mp4]

Multimedia Appendix 3

Survey instrument for student participants.

[PDF File (Adobe PDF File), 133KB - jmir_v17i6e151_app3.pdf]

Multimedia Appendix 4

Survey instrument for tutor participants.

[PDF File (Adobe PDF File), 75KB - jmir_v17i6e151_app4.pdf]

Multimedia Appendix 5

Questions used to structure the tutor review meeting.

[PDF File (Adobe PDF File), 79KB - jmir_v17i6e151_app5.pdf]

References

1. Ellaway R, Candler C, Greene P, Smothers V. An Architectural Model for MedBiquitous Virtual Patients. 2006. URL: http://groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture [accessed 2014-10-06] [WebCite Cache ID 6T87kiIw9]
2. Round J, Conradi E, Poulton T. Improving assessment with virtual patients. Med Teach 2009 Aug;31(8):759-763. [Medline: 19811215]
3. Poulton T, Balasubramaniam C. Virtual patients: a year of change. Med Teach 2011;33(11):933-937. [doi: 10.3109/0142159X.2011.613501] [Medline: 22022903]
4. Georg C, Zary N. Web-based virtual patients in nursing education: development and validation of theory-anchored design and activity models. J Med Internet Res 2014;16(4):e105 [FREE Full text] [doi: 10.2196/jmir.2556] [Medline: 24727709]
5. Salminen H, Zary N, Björklund K, Toth-Pal E, Leanderson C. Virtual patients in primary care: developing a reusable model that fosters reflective practice and clinical reasoning. J Med Internet Res 2014;16(1):e3 [FREE Full text] [doi: 10.2196/jmir.2616] [Medline: 24394603]
6. 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]
7. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach 2009 Aug;31(8):743-748. [Medline: 19811212]
8. Ellaway RH, Poulton T, Smothers V, Greene P. Virtual patients come of age. Med Teach 2009 Aug;31(8):683-684. [Medline: 19811203]
9. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010 Jan;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]
10. Issenberg SB. The scope of simulation-based healthcare education. Simul Healthc 2006;1(4):203-208. [doi: 10.1097/01.SIH.0000246607.36504.5a] [Medline: 19088590]
11. Issenberg SB, McGaghie WC, Petrusa ER, Lee GD, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005 Jan;27(1):10-28. [doi: 10.1080/01421590500046924] [Medline: 16147767]
12. Huwendiek S, Reichert F, Bosse H, de Leng BA, van der Vleuten CPM, Haag M, et al. Design principles for virtual patients: a focus group study among students. Med Educ 2009 Jun;43(6):580-588. [doi: 10.1111/j.1365-2923.2009.03369.x] [Medline: 19493183]
13. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. Med Educ 2012 Jul;46(7):636-647. [doi: 10.1111/j.1365-2923.2012.04243.x] [Medline: 22616789]
14. Maran NJ, Glavin RJ. Low- to high-fidelity simulation - a continuum of medical education? Med Educ 2003 Nov;37 Suppl 1:22-28. [Medline: 14641635]

15. Poulton T, Ellaway RH, Round J, Jivram T, Kavia S, Hilton S. Exploring the efficacy of replacing linear paper-based patient cases in problem-based learning with dynamic Web-based virtual patients: randomized controlled trial. J Med Internet Res 2014;16(11):e240 [FREE Full text] [doi: 10.2196/jmir.3748] [Medline: 25373314]
16. Poulton T, Conradi E, Kavia S, Round J, Hilton S. The replacement of 'paper' cases by interactive online virtual patients in problem-based learning. Med Teach 2009 Aug;31(8):752-758. [Medline: 19811214]
17. Ellaway RH, Poulton T, Jivram T. Decision PBL: A 4-year retrospective case study of the use of virtual patients in problem-based learning. Med Teach 2014 Oct 14:1-9. [doi: 10.3109/0142159X.2014.970627] [Medline: 25313934]
18. Bakrania T, Poulton T, Beaumont C. G4-JISC Final Report. 2010. URL: http://www.jisc.ac.uk/media/documents/programmes/curriculumdelivery/G4_Final_Reportv2.pdf [accessed 2014-10-06] [WebCite Cache ID 6T88lKXia]
19. Dolmans DHJM, De Grave W, Wolfhagen IHAP, van der Vleuten CPM. Problem-based learning: future challenges for educational practice and research. Med Educ 2005 Jul;39(7):732-741. [doi: 10.1111/j.1365-2929.2005.02205.x] [Medline: 15960794]
20. Phillips RA, McNaught C, Kennedy G. Towards a generalised conceptual framework for learning: the Learning Environment, Learning Processes and Learning Outcomes (LEPO) framework. In: ED-MEDIA. 2010 Presented at: 22nd annual World Conference on Educational Multimedia, Hypermedia & Telecommunications; June 29, 2010; Toronto, Canada p. 2495-2504.
21. Ellaway RH. Virtual patients as activities: exploring the research implications of an activity theoretical stance. Perspect Med Educ 2014 Sep;3(4):266-277 [FREE Full text] [doi: 10.1007/s40037-014-0134-z] [Medline: 25082311]
22. Sandars J. It appeared to be a good idea at the time but … A few steps closer to understanding how technology can enhance teaching and learning in medical education. Med Teach 2011;33(4):265-267. [doi: 10.3109/0142159X.2011.550972] [Medline: 21456982]
23. Ellaway RH, Davies D. Design for learning: deconstructing virtual patient activities. Med Teach 2011;33(4):303-310. [doi: 10.3109/0142159X.2011.550969] [Medline: 21456988]
24. Balslev T, de Grave WS, Muijtjens AMM, Scherpbier AJJA. Comparison of text and video cases in a postgraduate problem-based learning format. Med Educ 2005 Nov;39(11):1086-1092. [doi: 10.1111/j.1365-2929.2005.02314.x] [Medline: 16262803]
25. Basu RR, McMahon GT. Video-based cases disrupt deep critical thinking in problem-based learning. Med Educ 2012 Apr;46(4):426-435. [doi: 10.1111/j.1365-2923.2011.04197.x] [Medline: 22429179]
26. Ghanchi NK, Khan S, Afridi A, Sajid S, Afzal S, Ahmed I, et al. Video or paper for delivery of problem-based learning cases? Med Educ 2013 Nov;47(11):1131. [doi: 10.1111/medu.12353] [Medline: 24117575]
27. de Leng B, Dolmans D, van de Wiel M, Muijtjens A, van der Vleuten CPM. How video cases should be used as authentic stimuli in problem-based medical education. Med Educ 2007 Feb;41(2):181-188. [doi: 10.1111/j.1365-2929.2006.02671.x] [Medline: 17269952]
28. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med 2006 Nov 23;355(21):2217-2225. [doi: 10.1056/NEJMra054782] [Medline: 17124019]
29. Kamin C, O'Sullivan P, Deterding R, Younger M. A comparison of critical thinking in groups of third-year medical students in text, video, and virtual PBL case modalities. Acad Med 2003 Feb;78(2):204-211. [Medline: 12584102]
30. Bateman J, Hariman C, Nassrally M. Virtual patients can be used to teach clinical reasoning. Clin Teach 2012 Apr;9(2):133-134. [doi: 10.1111/j.1743-498X.2012.00546.x] [Medline: 22405376]
31. Posel N, Mcgee JB, Fleiszer DM. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. Med Teach 2014 Dec 19:1-6. [doi: 10.3109/0142159X.2014.993951] [Medline: 25523009]
32. Creswell J. Qualitative, Quantitative, and Mixed Methods Approaches. 2nd edition. Thousand Oaks, CA: Sage Publications, Inc; 2002.
33. Creswell J, Clark V. Designing and Conducting Mixed Methods Research. 2nd edition. Thousand Oaks, CA: Sage Publications, Inc; 2010.
34. SMART Technologies. SMART Interactive Solutions for Education, Business and Government. URL: http://www.smarttech.com/Home%20Page/Landing%20Page [accessed 2014-10-06] [WebCite Cache ID 6T89brb4g]
35. De Leng B, Huwendiek S, Donkers J. eViP Deliverable 5.2: Published evaluation instrument for Virtual Patients and for learning and teaching activities with Virtual Patients. 2009. URL: http://www.virtualpatients.eu/wp-content/uploads/2009/06/evip_deliverable_5.2_final-3.pdf [accessed 2014-10-06] [WebCite Cache ID 6T89mam0W]
36. Huwendiek S, De Leng BA, Kononowicz AA, Kunzmann R, Muijtjens AMM, Van Der Vleuten CPM, et al. Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. Med Teach 2014 Oct 14:1-8. [doi: 10.3109/0142159X.2014.970622] [Medline: 25313931]
37. eViP Electronic Virtual Patients. URL: http://www.virtualpatients.eu/ [accessed 2014-07-15] [WebCite Cache ID 6R5DlYHBd]
38. SurveyMonkey: Free online survey software & questionnaire tool. URL: https://www.surveymonkey.com/ [accessed 2014-07-15] [WebCite Cache ID 6R5DbsDxv]
39. Conradi E, Kavia S, Burden D, Rice A, Woodham L, Beaumont C, et al. Virtual patients in a virtual world: Training paramedic students for practice. Med Teach 2009 Aug;31(8):713-720. [Medline: 19811207]

40. Morse JM. The implications of interview type and structure in mixed-method designs. In: Gubrium JF, Holstein J, Marvasti AB, McKinney KD, editors. The SAGE Handbook of Interview Research: The Complexity of the Craft. Thousand Oaks, CA: Sage Publications, Inc; 2012:193-204.
41. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
42. ATLAS.ti: The Qualitative Data Analysis & Research Software. URL: http://www.atlasti.com/index.html [accessed 2014-07-15] [WebCite Cache ID 6R5DE2nd3]
43. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. Med Teach 2012;34(10):850-861. [doi: 10.3109/0142159X.2012.704439] [Medline: 22913519]
44. Kennedy TJT, Lingard LA. Making sense of grounded theory in medical education. Med Educ 2006 Feb;40(2):101-108. [doi: 10.1111/j.1365-2929.2005.02378.x] [Medline: 16451236]
45. Prober CG, Heath C. Lecture halls without lectures--a proposal for medical education. N Engl J Med 2012 May 3;366(18):1657-1659. [doi: 10.1056/NEJMp1202451] [Medline: 22551125]
46. Sobral DT. What kind of motivation drives medical students' learning quests? Med Educ 2004 Sep;38(9):950-957. [doi: 10.1111/j.1365-2929.2004.01913.x] [Medline: 15327676]
47. Young JQ, Van MJ, Durning S, Ten CO. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach 2014 May;36(5):371-384. [doi: 10.3109/0142159X.2014.889290] [Medline: 24593808]
48. Ellaway RH, Pusic M, Yavner S, Kalet AL. Context matters: emergent variability in an effectiveness trial of online teaching modules. Med Educ 2014 Apr;48(4):386-396. [doi: 10.1111/medu.12389] [Medline: 24606622]
49. Shaffer DW, Resnick M. “Thick” authenticity: new media and authentic learning. Journal of Interactive Learning Research 1999;10(2):195-215.
50. Mayer R. Multimedia Learning. Cambridge: Cambridge University Press; 2001.
51. Mayer RE, Heiser J, Lonn S. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. Journal of Educational Psychology 2001;93(1):187-198. [doi: 10.1037/0022-0663.93.1.187]
52. Norman GR, Schmidt HG. The psychological basis of problem-based learning: a review of the evidence. Acad Med 1992 Sep;67(9):557-565. [Medline: 1520409]
53. Norman G. Problem-based learning makes a difference. But why? CMAJ 2008 Jan 1;178(1):61-62 [FREE Full text] [doi: 10.1503/cmaj.071590] [Medline: 18166733]

Abbreviations

PBL: problem-based learning

Edited by G Eysenbach, P Bamidis; submitted 07.10.14; peer-reviewed by S Huwendiek, I Nicolaidou, P Staccini; comments to author 22.03.15; revised version received 19.04.15; accepted 16.05.15; published 18.06.15.

Please cite as:
Woodham LA, Ellaway RH, Round J, Vaughan S, Poulton T, Zary N
Medical Student and Tutor Perceptions of Video Versus Text in an Interactive Online Virtual Patient for Problem-Based Learning: A Pilot Study
J Med Internet Res 2015;17(6):e151
URL: http://www.jmir.org/2015/6/e151/
doi:10.2196/jmir.3922
PMID:26088435

Original Paper

A Virtual Emergency Telemedicine Serious Game in Medical Training: A Quantitative, Professional Feedback-Informed Evaluation Study

Iolie Nicolaidou[1*], BA (Hons), MSc, PhD; Athos Antoniades[2,3*], BSc, PhD; Riana Constantinou[4], BSc, MSc, PhD (c); Charis Marangos[5], BSC, MSc; Efthyvoulos Kyriacou[6], PhD; Panagiotis Bamidis[7], MSc, PhD; Eleni Dafli[7], MD, MSc, PhD; Constantinos S Pattichis[5], PhD

[1]Department of Communication and Internet Studies, Cyprus University of Technology, Limassol, Cyprus
[2]STREMBLE Ventures Ltd, Limassol, Cyprus
[3]University of Cyprus, Department of Computer Science, Nicosia, Cyprus
[4]Cyprus Emergency and Pre-hospital care Association, KSEPA, Nicosia, Cyprus
[5]Department of Computer Science, University of Cyprus, Nicosia, Cyprus
[6]Department of Computer Science & Engineering, Frederick University, Limassol, Cyprus
[7]Lab of Medical Physics, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece
[*]these authors contributed equally

Corresponding Author:
Iolie Nicolaidou, BA (Hons), MSc, PhD
Department of Communication and Internet Studies
Cyprus University of Technology
P.O. Box 50329
Limassol, 3603
Cyprus
Phone: 357 99329897
Fax: 357 25 312914
Email: iolie.nicolaidou@cut.ac.cy

Abstract

Background: Serious games involving virtual patients in medical education can provide a controlled setting within which players can learn in an engaging way, while avoiding the risks associated with real patients. Moreover, serious games align with medical students' preferred learning styles. The Virtual Emergency TeleMedicine (VETM) game is a simulation-based game that was developed in collaboration with the mEducator Best Practice network in response to calls to integrate serious games in medical education and training. The VETM game makes use of data from an electrocardiogram to train practicing doctors, nurses, or medical students for problem-solving in real-life clinical scenarios through a telemedicine system and virtual patients. The study responds to two gaps: the limited number of games in emergency cardiology and the lack of evaluations by professionals.

Objective: The objective of this study is a quantitative, professional feedback-informed evaluation of one scenario of VETM, involving cardiovascular complications. The study has the following research question: "What are professionals' perceptions of the potential of the Virtual Emergency Telemedicine game for training people involved in the assessment and management of emergency cases?"

Methods: The evaluation of the VETM game was conducted with 90 professional ambulance crew nursing personnel specializing in the assessment and management of emergency cases. After collaboratively trying out one VETM scenario, participants individually completed an evaluation of the game (36 questions on a 5-point Likert scale) and provided written and verbal comments. The instrument assessed six dimensions of the game: (1) user interface, (2) difficulty level, (3) feedback, (4) educational value, (5) user engagement, and (6) terminology. Data sources of the study were 90 questionnaires, including written comments from 51 participants, 24 interviews with 55 participants, and 379 log files of their interaction with the game.

Results: Overall, the results were positive in all dimensions of the game that were assessed as means ranged from 3.2 to 3.99 out of 5, with user engagement receiving the highest score (mean 3.99, SD 0.87). Users' perceived difficulty level received the lowest score (mean 3.20, SD 0.65), a finding which agrees with the analysis of log files that showed a rather low success rate

(20.6%). Even though professionals saw the educational value and usefulness of the tool for pre-hospital emergency training (mean 3.83, SD 1.05), they identified confusing features and provided input for improving them.

Conclusions: Overall, the results of the professional feedback-informed evaluation of the game provide a strong indication of its potential as an educational tool for emergency training. Professionals' input will serve to improve the game. Further research will aim to validate VETM, in a randomized pre-test, post-test control group study to examine possible learning gains in participants' problem-solving skills in treating a patient's symptoms in an emergency situation.

(J Med Internet Res 2015;17(6):e150) doi:10.2196/jmir.3667

KEYWORDS

telemedicine; emergency telemedicine; serious games; virtual patients; medical education; professional feedback-informed evaluation; emergency assessment and management

Introduction

Background

New media technologies such as serious games involving virtual patients align with medical students' preferred learning styles and are seen positively by them [1-3] as an innovative way for learning skills that are necessary in the medical profession. This is mostly because of the affordances of serious games and virtual patients in engaging learners while at the same time avoiding the risks associated with real patients. Most studies involving virtual patients in serious games in medical education focus on the description of the design and development of innovative ways to teach doctors and medical students [4-8]. However, there is evidence from controlled trial designs or similar rigorous methodologies [9-15] that indicates the effectiveness of proposed serious games and virtual patient interventions in medical education. Not many serious games exist that focus on supporting medical students' skills in responding to medical emergency situations to manage the symptoms of patients facing cardiovascular complications through telemedicine. Moreover, there is a lack of evaluation/impact measurements with actual professional groups, as most studies that aimed to evaluate serious games and virtual patients had students as participants [1,2,5]. The present study attempted to fill these two gaps with the design and development, and attempt to quantify the potential of a virtual emergency telemedicine serious game in medical training through a professional feedback-informed evaluation study, with 90 professional ambulance crew nursing personnel specializing in the assessment and management of emergency cases.

Medical Students' Attitudes Toward Games

Serious games align with digital natives' [16] preferred learning styles and this seems to apply in medical education as well [17,18]. In the domain of medical education, several studies showed that medical students' attitudes toward games in medical education, and toward learning from new media, such as immersive 3D virtual environments or virtual mentoring systems, are favorable [1-3]. In a recent study with 217 medical students from two US universities [1], the vast majority (98%) of students liked the idea of using technology to enhance health care education, felt that education should make better use of new media technologies (96%), and believed that games can have an educational value (80%). The attitudes of doctors in training toward virtual mentoring systems were also reported to be favorable in the study of Jaffer et al [3], which examined the effect of a short introduction on how to use virtual systems on 57 junior doctors in the United Kingdom. Evidence for medical students' preferences in learning from new media, such as immersive 3D virtual environments also comes from a study with 90 second-year Master's students in pharmacy [2]. The attitudes of 62 medical students were also favorable on the use of serious games' interactive algorithms involving electronic virtual patients in medical education, as they found the interactive algorithms were effective learning tools, facilitating enhanced knowledge in the field of acute medicine [19].

Ample evidence indicates that serious games align with medical students' preferred learning styles; therefore, a next step is to evaluate whether empirical evidence, documenting learning gains of such games, also supports their integration in medical education, as it was suggested by [20] in the area of emergency medicine in particular.

Serious Games Involving Virtual Patients in Medical Education

Serious games development and implementation for medical education is a growing domain. According to Graafland et al [21], who conducted a systematic review of serious games for medical education and surgical skills training, which included 25 research studies and covered 30 serious games published between 1995 and 2012, serious games form an innovative approach toward the education of medical professionals. Serious games attempt to deliver affordable, accessible, and usable interactive virtual worlds, supporting applications in training and education [9]. Following these trends, traditional instructor-centered teaching is yielding to a learner-centered model that puts learners in control of their own learning in medical education [22].

Some of the attributes of games involving virtual patients in medical education that make them attractive and useful include the fact that game environments provide a safe and controlled setting within which players can learn in an engaging way, while avoiding the risks associated with real patients [4]. Virtual patient simulations have the significant advantages of requiring fewer personnel and resources, being accessible at any time, and being highly standardized [23], and may support learning processes and be a valuable complement in teaching communication skills, patient-centeredness, clinical reasoning, and reflective thinking [24]. According to the systematic review by Ghanbarzadeh et al [25], virtual patients can be used by trainees such as nurses, surgeons, students, and other medical staff, and their performance can be assessed and benchmarked in different ways. However, virtual patients are notoriously difficult and costly to author, adapt, and exchange. Historically, this has limited their uptake and utility, despite their being able to provide high-quality learning opportunities [26] and despite enthusiasm about the educational potential of three-dimensional virtual worlds and virtual patients for medical educators [27].

There are many recent examples of studies that involved virtual patients in serious games for medical education in areas such as cardiology for practicing doctors [4], rehearsing professional behaviors, such as taking a patient's history for medical students [5], improving the efficiency of junior doctor training through a junior doctor medical simulator [6], teaching insulin therapy for primary care physicians [7], surgical training and developing clinical skills to respond to injuries sustained during catastrophic incidents [8], teaching about medical ethics, medical law, and medical professionalism [28], and repurposing Web-based virtual patients to multi-user virtual environments for undergraduate dental education [29]. Of special interest for this study are virtual patient games on emergency treatment in cardiology, which are discussed next.

In the area of cardiology but not focusing on emergency treatment, Dafli, Bamidis, and Dombros [4] described the design and implementation of a pilot scenario with a simulated virtual patient for potential use in Greek medical education. In their scenario, the practicing doctor could interact with a virtual patient to examine him, select diagnostic tests and inquiry methods, select different approaches, and reach a decision with respect to the treatment of a cardiological incident. The content was realistic and it was enhanced with the addition of audiovisual material to simulate medical reality and allow users to develop clinical skills through the system. Even though the application of the virtual patient has not yet been evaluated by a large number of users, it appears to be a promising application that can extend the users' experience with real patients and allow them to practice their clinical skills in a systematic, safe, and protected way, adjusted to their own needs and level of experience [30].

In the area of emergency treatment but not focusing on cardiology, the University of Auckland's Second Life simulation island, Medical Centre, and Emergency Room simulations represent a case where more experiential and immersive multi-user virtual environments have been used for designing and hosting virtual patients in serious games and experiential learning tools in emergency medicine and care [31]. Another case is the use of the Second Life virtual simulation environment for mock oral emergency medicine examinations targeting emergency medicine residents, who have the requirement for board certification in order to become emergency physicians [32].

A common theme identified in the literature of studies that involved virtual patients in serious games for medical education is that even though they provide indications that virtual world medical simulations have the potential to enable students to practice professional behaviors in a risk-free environment, providing opportunities for skills practice prior to real-world patient encounters, their work is not always validated through empirical research. This is true for studies such as the one by Danforth et al [5]. The same applies for Diehl et al [7], who plan to evaluate their game, InsuOnline, using a randomized controlled trial design in future studies and Guise et al [33] who only performed initial usability testing for two narrative virtual patients that they developed for vocational mental health nurse training. McEvoy et al [34] examined virtual patients as an educational intervention to improve pediatric basic specialist trainee education in the management of suspected child abuse. Their evaluation methodology focused on the use of a questionnaire, developed to determine trainees' perception of the value of the virtual patient as an educational tool, and it has not yet been evaluated by professionals.

Research Evidence From Controlled Trials Supporting the Effectiveness of Serious Games in Medical Education

As shown in the previous section of this paper, there is an abundance of studies that involve virtual patients in serious games in medical education that focus on the description of the design and development of innovative ways to teach doctors and medical students. However, only a few of them refer to virtual patient games for emergency treatment in cardiology.

Studies that advanced to controlled trial designs or similar rigorous methodologies, involving random assignments to experimental and control groups, to research the effectiveness of proposed serious games interventions in medical education have not been numerous. Research evidence from controlled trials supports that serious games [9] and the deployment of virtual patients [11,13] offer the potential to enhance learning and improve subsequent performance when compared to traditional educational methods in areas such as in basic life support skills [11], knowledge acquisition about pediatric respiratory diseases [13], hematology and cardiology topics [14], and cardiac examination competency in medical students [15]. Research also showed that small duration (eg, 1 hour) interventions of virtual patient-based e-learning programs are not necessarily more effective compared to traditional training in the area of improving physicians' substance abuse management skills [12]. There are also some studies in the literature that, even though they used a robust methodology, did not measure learning gains but focused on the effect of virtual patient training on students' confidence instead, in areas such as history-taking and clinical breast examination [35,36].

Of interest for this study are controlled trial studies that focused on emergency treatment in cardiology through virtual patients and these studies are discussed next. In an area that partly relates to emergency cardiology incidents' treatment, as it involves basic life support with the use of a defibrillator, Kononowicz et al [11] introduced a voluntary virtual patients' module into a basic life support with an automated external defibrillator (BLS-AED) course to examine whether this addition would improve the knowledge and skills of students taking the course. Half of the students were randomly assigned to an experimental group and given voluntary access to a virtual patient module consisting of six cases presenting BLS-AED knowledge and skills. The study was conducted over 6 weeks and involved 226 first-year medical students. The voluntary module was used by 61 of the 114 entitled study participants. The group that used virtual patients demonstrated better results in knowledge acquisition and in some key BLS-AED action skills than the group without access, or those students from the experimental group deliberately not using virtual patients.

In a study that involves cardiology but not emergency treatment, positive results that indicate better retention with virtual patients than with traditional learning methods have also been reported by Botezatu et al [14], who conducted a randomized controlled study on early and delayed assessment results of 49 students using virtual patients for learning and examination of hematology and cardiology topics in an internal medicine course.

In the area of emergency treatment but not necessarily in cardiology, Dev et al [37] tested the architecture of a virtual emergency department patient for scenarios in emergency medicine in a multi-person learning environment based on online gaming technology. The efficacy of the model and the virtual emergency department learning environment was evaluated in a study where 12 advanced medical students and first-year residents managed six trauma cases, in groups of four. Their pre- and post-test performance results showed significant learning, with results comparable to those obtained in human mannequin simulators.

Other researchers did not focus on the comparison between virtual patients and traditional modules but rather tried to maximize the benefits of both approaches by combining them. These attempts demonstrated that virtual patients in virtual worlds or as part of serious games offer significant learning potential when used as a supplement to the traditional teaching techniques of medical education. Evidence for this comes from studies with virtual patients as a supplemental teaching tool for pediatric dentistry from Papadopoulos et al [10] and for clinical skills training for practicing health care workers Triola et al [23]. Similar studies in the areas of cardiology and emergency treatment were not found in the literature.

Trends in the Literature on Serious Games Involving Virtual Patients

The evidence outlined above demonstrates not only medical students' potential uptake of serious games for education and training but also the potential effectiveness and capability of serious games employing virtual patients to increase students' learning in various domains of medical education, including a few studies in emergency treatment in cardiology, which is the focus of this study. However, the literature review of recent research studies in the area of virtual patient implementation in medical education shows two gaps. The first gap is that there is a lack of evaluation/impact measurements with actual professional groups; as in most of the existing studies, participants were typically medical students. Two exceptions to the studies described above are the work of Heinrichs et al [38] (2010) and the work of Salminen et al [24]. Heinrichs et al [38] attempted to determine whether a virtual emergency department, designed after Stanford University Medical Center's Emergency Department was an effective clinical environment for training emergency department physicians and nurses for mass-casualty incidents. The participants of the study were professionals, more specifically 10 physicians with an average of 4 years of post-training experience, and 12 nurses with an average of 9.5 years of post-graduate experience. Similarly, in the work of Salminen et al [24], the virtual patient model that was developed to facilitate medical students' reflective practice and clinical reasoning as well as the case created using the developed model were validated by a group of 10 experienced primary care physicians and then further improved by a work group of faculty involved in that medical program.

The second gap identified in the relevant literature is that not many serious games exist that focus on supporting medical students' skills in responding to medical emergency situations to treat the symptoms of patients facing cardiovascular complications through telemedicine. Some of the scenario-based games that exist, for example the ones that have been developed and repurposed as part of the eViP virtual patients' project in the area of "Cardiology, Emergency Medicine, ECG", are based on mainly textual information presented to the user, even though some incorporate audio and video. The user is typically asked to make a diagnosis choice among given options in multiple choice form (eg, eViP Mr Horcek, developed by DecisionSim). Even though some of these games provide instant or enquiry-based feedback to the user, others do not and are linear in nature. An important limitation of these games is that they are not real-time games simulating realistic conditions of emergency care, in the sense that the user does not have any time pressure and is not typically required to solve the problem in a limited amount of time, as is the case in reality, where if medical professionals do not react promptly they may lose the patient. Another limitation is that the user is not necessarily in contact with a virtual patient all the time; he or she may only be interacting with on-screen textual instructions or choices rather than observing the virtual patient or the signals of an electrocardiogram (ECG) at any given time.

Another category of existing games are commercial games, such as MicroSim [39], which can provide resource-efficient, self-directed learning through the simulation of realistic patient scenarios to help learners develop decision-making and critical-thinking skills. In the MicroSim pre-hospital version, patient cases are set in a pre-hospital environment as well as in an ambulance. The learner has access to most of the tools and drugs that are available in an ambulance as well as the initial scene of the pre-hospital environment [39]. Another example of a serious game implementing an immersive virtual learning space, which was developed for training health care professionals in clinical skills is Pulse!! - The Virtual Clinical Learning Lab. In this game, graphics recreate a lifelike, interactive, virtual training environment in which civilian and military health care professionals practice clinical skills in order to better respond to injuries sustained during catastrophic incidents, such as combat or bioterrorism. The game is designed to support a range of the training needs nurses and medical professionals require [8]. However, commercial simulations, such as Microsim and Pulse!! typically involve a high cost to be made available to hospitals for medical education, and are not always research-validated.

This study attempts to address the two identified gaps by not only designing and developing the Virtual Emergency

TeleMedicine (VETM) serious game, but also involving a number of professionals, who are at the same time stakeholders, specifically professional ambulance crew nursing personnel specializing in the treatment of emergency cases, in the evaluation of the educational potential of this game. The design and development phase of the VETM game is described in the next section, which is followed by a quantitative, professional feedback-informed evaluation of the game.

An Example of a Serious Game for Medical Emergency Situations: Virtual Emergency Telemedicine (VETM)

The Rationale Behind VETM

The VETM game is a simulation-based serious game that was developed in response to calls to integrate serious games in medical education and training. The game was also developed to provide a learning environment that can be used as a supplement to traditional training in emergency situations and that is, at the same time, compatible with today's medical students' preferences toward new media and new learning technologies. The game makes use of data from an ECG and is designed and developed to train practicing doctors, medical students, or other health care professionals, such as nurses and paramedics, for problem-solving in real-life emergency clinical scenarios through a telemedicine system. Users of the game learn how to respond to medical emergency situations to assess and manage the symptoms of a virtual patient, who is located in an ambulance, through different scenarios involving cardiovascular complications. The game allows users to practice their skills while receiving immediate feedback by the system and virtual patient. The VETM game is based on principles of adult learning and problem-based learning, such as self-pacing, contextualization, and a hands-on approach in which the learner is an active participant [7] and employs popular gamification tactics to engage the users. More specifically, the game is based on scenarios, clear goals are provided to the user who is committed to achieve them, points are allocated for correct performance actions for reinforcement purposes, progress of the user in the game is visible to enable progress monitoring, and constant feedback is provided to the user through the experience, as suggested by [40].

The concept of the VETM game is based on a real system developed by the eHealth labs of the University of Cyprus and Frederick University. This work initially aimed at the development of an integrated portable medical device for emergency telemedicine. The system enables the transmission of critical biosignals (ECG, blood pressure, heart rate, oxygen saturation, temperature) and still images of the patient, from the emergency site to an emergency call center; thus enabling physicians to direct pre-hospital care in a more efficient way, improving patient outcome and reducing mortality. The system was designed in order to operate over several communication links such as mobile, satellite, ADSL [41,42]. The system has been continuously updated on the technological issues related to telecommunications and operating system environment. In addition to the ongoing work on the emergency telemedicine systems, the eHealth lab group is working on the transmission of real-time video as presented in [43-46].

Technical Specifications of VETM

With regard to the technical specifications, the VETM game implements a scripting language called Scribulance, which is a blend of C and Pascal. Scribulance is a user-friendly high-level scripting language developed specifically for the purpose of the game. It enables the creation of custom scenarios, which are currently text-based. The game was created using the programming language C# with XNA for the game. In order to support a high-level scripting language for developing the game's scenarios, a scenario compiler was also created. For testing and evaluation purposes, several scripts that used the features of the scripting language were implemented to examine whether they worked properly and a detailed command-by-command debugging was conducted to examine that the flow of the code worked as intended [47,48].

Instructional Design of VETM

As Figure 1 shows, the interface of the game includes the "electrocardiogram" display with leads in the upper part, the "Actions" and "Drugs" available to the user in the left part, the virtual "Patient" in the right part, and additional information with regard to the physical state of the patient (heart rate, respiration, temperature, etc) in the further right part of the screen. The "Actions" menu provides several options for the user to choose from. In this particular example, the user can "speak to the patient" to see if he is responding, "measure his temperature", "ensure that his airway is open", etc. The order of these options changes randomly every time the game is played and changes according to whether there are complications to the patient in the scenario. If the user chooses to provide drugs to the patient, then a drop-down menu is available under "Drugs" with several options of medication that are typically available in an ambulance (such as atropine sulphate, adrenaline, acetylsalicylic acid, paracetamol, salbutamol, hydrocortisone, etc) and the option to choose the dosage of each one.

Several "Tools" in graphical format are also provided. These are located around the patient. Tools include an "oxygen mask", "providing IV fluids", a "stethoscope", the option to examine the patient with "palpations" or "perform Cardiopulmonary Resuscitation (CPR)", a "flashlight" to examine the patient's eye pupils, and the option to "enable the ECG" to monitor the patient's heart (Figure 1).

After every one of the user's actions (which can take place in one of three possible formats, either in the form of using a tool, or requesting an action, or providing medication), instant feedback is provided in the bottom part of the interface. In the example of Figure 1, the user attempted to measure the patient's temperature and the feedback received was the following: "Temperature is 36.2 degrees Celsius". The user receives a positive or negative score after each action. Users can check their score at any given time while playing the game. If they do, they get a detailed report of correct and incorrect actions up to that point. Alternatively, users can get a detailed report of actions that were correct (these are indicated in green color) and actions that were incorrect (these are indicated in red color) at the end of the game. The final report also includes the time needed to save the patient and the total score of the user.

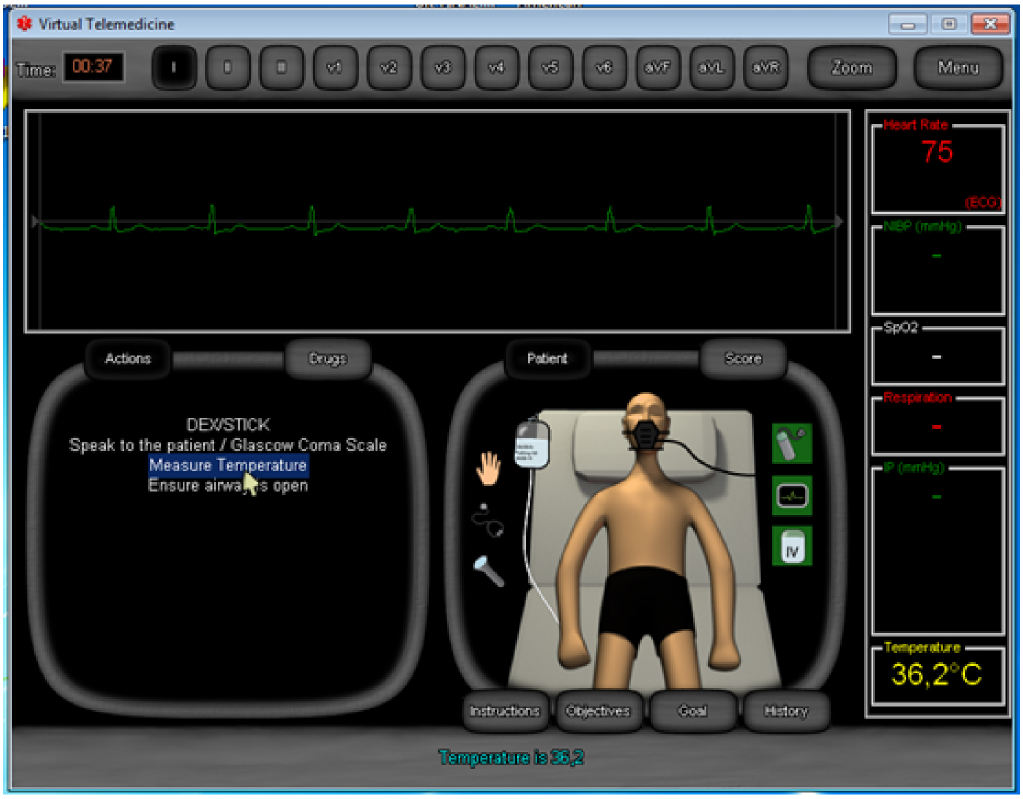
As part of the instructional design of the game, the user can also access the specific “Objectives” for that scenario, the “Goals” of the game and the scenario’s “Instructions” that put the scenario into context. As shown in Figure 1, these are located below the virtual patient. In this scenario, named “Dyspnea and chest discomfort”, the “Instructions” were the following:

> *The Emergency Dispatch Centre received an emergency call and sent you to treat the following incident: “Husband feels bad, came home from work and is getting worse, dyspneic and with precordial pain”. You are the ambulance crew. You have to examine and treat the patient, define working diagnosis and differential diagnosis, administer the therapy, define direction according to local situation and possible following steps. Conditions on the scene: September 2, 2013, 13:30pm, clear sky, calm, temperature 20°C (68°F). Call to hospital is 10 minutes.*

The user can also access and should consult the “History” of the patient prior to attempting to save him. In this scenario, the history of the patient that was available to the user was the following:

> *A 65-year-old man presented with dyspnea and chest discomfort. A week ago he had a similar episode, but symptoms have worsened now. Medical history: Lung cancer being treated with chemotherapy.*

Figure 1. Screenshot of the Virtual Emergency TeleMedicine game that shows the interface with the virtual patient.



Temperature is 36,2

The Scenario-Editor of VETM

The game has two types of target populations: students and trainers. On the one hand, it can be used for training and practicing purposes in the context of given scenarios when the target population is users such as medical students, nurses, paramedics, and health professionals. On the other hand, it can be used for teaching or training purposes, when the target population is advanced users such as medical professors or people involved in training medical professionals, who can write scenarios of their own, with very limited prior knowledge of programming required, by using the scenario-editor of the game.

The Scenario Editor in VETM imports scenarios written in Scribulance, a very simple state-based scripting language similar to C. This language can describe a scenario in such a way that the choices and their effects are well-defined by the author. The scenario can then be added into the game easily. If there are no errors in the scenario, it will be added successfully and made available to play.

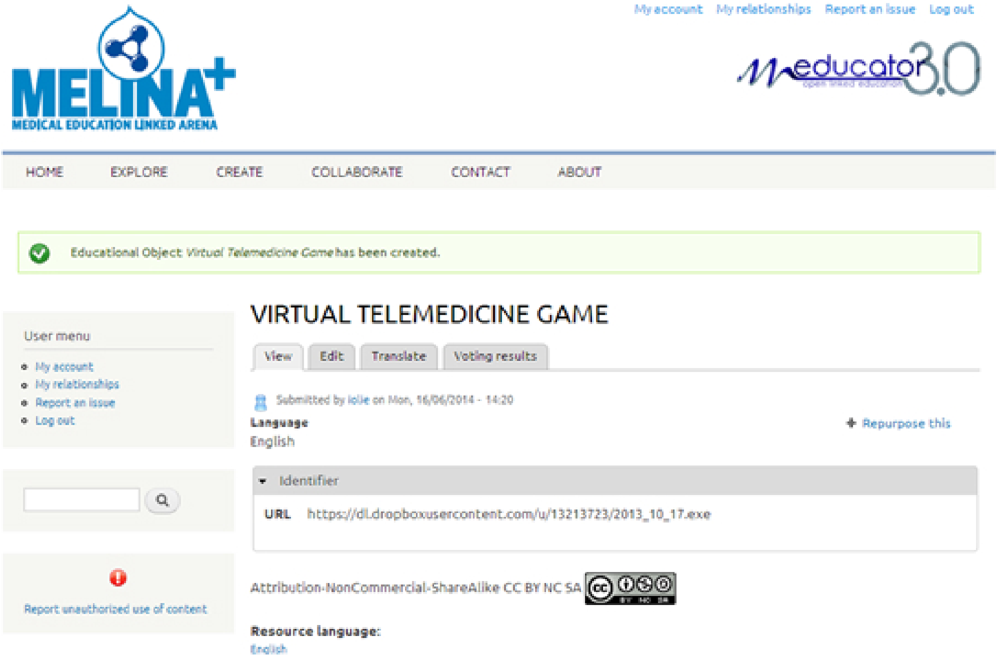
A medical professor or a trainer of medical professionals who is interested in adding a scenario to the game needs to know the basic structure of a scenario at a technical level, and this is described next. A Scenario is made up of States, Functions, and Variables.

A State may have any of the following Events: (1) a collection of Options that the player will see (an Option has a set of commands to be followed whenever that Option is selected), (2) an Enter Event (a set of commands to be followed whenever that State is activated), (3) an Exit Event (a set of commands to be followed whenever that State is deactivated), and (4) a collection of Time Events. For each Time Event, the user must specify a time in minutes and seconds. A Time Event is a set of commands to be followed when the specified time passes after the State has been activated.

Variables hold temporary values needed for a scenario. A variable can be, for example, the amount of times a patient was given a certain medication. A Function is a set of commands not associated to a State. It can have any number of Parameters (also known as Arguments) and may return a single value. Functions are typically used for code that is repeated multiple times in a Scenario. An Annotation is information, in text form, for one of the following: "History", "Objectives", and "Instructions", which can be customized by the user in each scenario. A Scenario must have at least one State, while everything else is optional.

The VETM game is currently working as a standalone application on personal computers (PCs) with the Windows operating system [47,48]. The game can be shared and repurposed (through changing its scenarios, or creating new ones and compiling them with the included virtual emergency telemedicine compiler). It is discoverable through different instantiations of mEducator (Multi-type Content Sharing and Repurposing in Medical Education) [49], such as mEducator 3. 0 Melina+, an extended version of Drupal 7, which is offered as an installation profile and enables website administrators to install a learning management system, focused on medical education. As Figure 2 shows, the VETM game is listed there as a shared Internet resource, described with appropriate metadata.

Figure 2. Interface of Melina+ (Medical Education Linked Arena), showing the Virtual Emergency TeleMedicine Game educational resource and metadata.



Educational Object *Virtual Telemedicine Game* has been created.

VIRTUAL TELEMEDICINE GAME

Submitted by iolie on Mon, 16/06/2014 - 14:20

Language
English

Identifier

URL https://dl.dropboxusercontent.com/u/13213723/2013_10_17.exe

Attribution-NonCommercial-ShareAlike CC BY NC SA

Resource language:
English

Research Question of the Study

The present study refers to the evaluation of one scenario ("Dyspnea and chest discomfort") of the VETM game by professionals and attempts to provide an answer to the following research question: What are professionals' perceptions of the potential of the Virtual Emergency TeleMedicine game for training people involved in treating emergency cases?

The "professionals" in the context of this study were professional ambulance crew ambulance crew nursing personnel specializing in assessing and managing pre-hospital emergency cases. The professionals' "perceptions" referred to participants' perceived advantages and limitations of the game. These were operationalized in the evaluation questionnaire that was developed under six dimensions of the game: (1) user interface, (2) difficulty level, (3) feedback, (4) educational value, (5) engagement, and (6) terminology.

Methods

Context of the Evaluation

The evaluation of the VETM game was conducted as part of a training seminar targeting professional ambulance crew nursing personnel specializing in assessing and managing pre-hospital emergency cases. The training was organized by the Cyprus Ministry of Health; it has a total duration of 250 hours spread over 2 years in about 50 five-hour meetings, and it is compulsory.

A total of 90 participants organized in four groups participated in the evaluation. The four training meetings took place on Oct. 6-7 and Oct. 13-14, 2014 in the computer lab of a pedagogical institute with 12 desktop computers. The duration of each meeting was approximately 1.5 hours. The evaluation sessions consisted of: (1) demo, (2) hands-on experience of the game, and (3) evaluation. The first part consisted of a brief 5-minute presentation of the game by the first author, to demonstrate the user interface, the tools that are available to the user, the ECG part, etc, and the signing of informed consent forms by participants that allowed the voluntary and anonymous use of their demographic data and questionnaire evaluation data, the audiotaping of their comments, and the examination of the log files that were automatically created when they interacted with the game. The participants were informed that they would be asked to complete a questionnaire to evaluate the game and to provide their comments verbally on a voluntary basis.

In the second part, the participants were asked to work in groups of two on each desktop computer to explore the game and to try to solve the scenario “Dyspnea and Chest Discomfort”. The scenario involved the examination, treatment, and diagnosis of a virtual patient who experienced dyspnea and precordial pain. The participants’ goal was dual: to solve the scenario correctly and also have a hands-on experience of the functionality of the game so that they could evaluate it. A total of 81 participants worked in groups and 9 participants worked as individuals. In 83% (75/90) of cases, participants worked in groups of two and in 7% (6/90) of cases, participants worked in groups of three.

In the third part, participants were asked to individually complete a questionnaire and also verbally express their thoughts via brief interviews on the improvement of the game, its value, if any, and points of confusion, etc, to the first author, who was audiotaping the participants’ responses.

Participants

The participants of the study were 90 professional ambulance crew nursing personnel specializing in the assessment and management of pre-hospital emergency cases. With regard to ethical considerations, participation in the study was anonymous and voluntary, participants signed an informed consent form and the study protocol was previously approved by a review board at the University of Cyprus.

Data Sources

Overview

The study included three data sources: a game evaluation questionnaire, the game’s automatically created log files, and participants’ comments in the form of short interviews, conducted either while they interacted with the game or immediately after. These data sources are described next.

Game Evaluation Questionnaire

As far as the description of the instrument for the evaluation of the game is concerned, this consisted of three parts. The first part consisted of demographic data such as: gender, age, previous experience with a telemedicine system, years of experience, domain of expertise, previous experience with educational games, and previous experience with medical education games.

The second part consisted of 36 questions on a 5-point Likert scale (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree). The instrument assessed six dimensions of the game: (1) user interface (8 questions, Q1-Q8); (2) difficulty level (5 questions, Q9-Q13); (3) feedback (5 questions, Q14-18); (4) educational value (9 questions, Q19-27); (5) user engagement (6 questions, Q28-Q33); and (6) terminology and language (3 questions, Q34-36).

An internal consistency reliability analysis was performed on these 36 items, which showed that the instrument appeared to have good internal consistency (Cronbach alpha=.9).

The third part of the questionnaire included 5 open-ended questions related to things that they liked or disliked in the game, things that were difficult or confusing, suggestions for improvement of the game, and whether they would be interested to use the game for training purposes. Users’ hand-written answers in the 5 open-ended questions of the questionnaire were typed and analyzed.

Game Log Files

Each time the game was played, a log file was automatically created with information on the time it took for the user to solve the scenario, the types of mistakes that he or she made, etc. The total number of log files that were analyzed was 379.

Users’ Audiotaped Comments From Short Interviews

Users’ audiotaped comments from short interviews were transcribed verbatim. A total of 24 short interviews were conducted with a total of 55 participants (40 male, 15 female) with a total duration of 106 minutes. Each brief interview lasted for an average of 4.4 minutes (minimum=1 minute, maximum=18 minutes).

Analysis

With regard to the analysis of the responses of the questionnaire, data were coded in a statistical package (IBM/SPSS Statistics 20), using the numbers 1 to 5 for coding the answers (Completely Disagree to Completely Agree, respectively), and methods of descriptive analyses (main statistics of valid frequencies, means, and standard deviations) were calculated for all 36 questions. Four questions (Q10, Q12, Q17, Q20) were negatively phrased (eg, Q10 - The game is not challenging for me, Q12 - I needed more time to be able to solve the problem, Q17 - The feedback I receive when I make a choice is confusing, and Q20 - It was not clear what I could learn from the game) and their score was reversed prior to the calculation of compound scores. The calculation of compound scores was achieved as follows: the participants’ scores in the respective items under each one of the six dimensions that were examined were summed, and then recoded to reflect the 5-point Likert scale of the original questions. For example, for the second dimension, “difficulty level”, the sum of the five relevant questions (Q9-Q13) resulted in a possible minimum score of 5 and a maximum possible score of 25. This was recoded as follows: 5-8 were coded as value 1, 9-12 were coded as value 2, 13-16 were coded as value 3, 17-20 were coded as value 4, and 21-25 were coded as value 5. The same process was followed for all six dimensions, for consistency in reporting results.

The information contained in log files was transferred in SPSS for analysis. Overall, 23 different types of mistakes were coded and analyzed. The overall time for which all users in all four sessions engaged with playing the game was 13.5 hours.

The participants’ responses in the 5 open-ended questions of the questionnaire were coded for each question individually and frequencies of each code were calculated to give a sense of participants' reactions to the game.

Users' audiotaped comments from short interviews were analyzed qualitatively to complement themes from the analysis of the open-ended questions and to identify emerging themes of users' perceptions on the VETM game. The users' comments in the short interviews were initially classified in three broad categories corresponding to the areas that four out of five questions of the questionnaire examined: positive aspects of the game, negative or confusing aspects of the game, and suggestions for improvement. An additional theme that emerged examined the educational value of the game.

Results

Participants' Demographic Information

The participants of the study were 90 professional ambulance crew nurses (67%, 60/90 male and 33%, 30/90 female). As can be seen from Table 1, participants' average age was 32.23 years (SD 5.25, n=89) and it ranged from 25 to 49 years old. They had an average of 8.55 years of professional experience (mean 8.55, SD 5.26, n=85), which ranged from 1 to 26 years. The vast majority of participants 81% (71/90) had more than 5 years of experience, with 30% (27/90) of them having between 10 and 26 years of experience.

The vast majority of participants 87% (78/90) did not have any previous experience with games, either educational games or medical domain games. Of the people who had some experience with either type of game (13%, 12/89), their average time of playing games per week was approximately half an hour (0.57 hours per week). The majority of participants 70% (63/90) did not have any previous experience with a telemedicine system either.

Table 1. Participants' demographic information (n=90).

	n	Mean	SD
Age (years)	89	32.23	5.25
Professional experience (years)	85	8.55	5.27
Game experience (hours per week)	70	0.57	2.17

Users' Behavior in Solving the Dyspnea and Chest Discomfort Problem

The analysis of log files showed that the game was played for a total of 379 times during the four evaluation sessions. Users attempted to solve the problem for an average of 8 times (mean 8.39, SD 3.6) with a minimum of 2 times and a maximum of 21. The result was "success", which means that the patient was saved, in 20.6% (78/379) of these cases and the result was "failure", which means that the virtual patient was lost, in 79.4% (301/379) of these attempts. The average time for which the users interacted with the game was 2.17 minutes at a time (SD 52.3 seconds) and it ranged from 15 seconds to 5 minutes for each time the scenario was played. Table 2 shows the frequencies of the 23 most common mistakes professionals made when attempting to solve the scenario Dyspnea and Chest Discomfort of the VETM game.

Table 2. Most common mistakes in solving the Dyspnea and Chest Discomfort scenario of the VETM (Virtual Emergency TeleMedicine) game (n=915).

Mistake number	Mistake description	Frequency n (%)
M4	Your patient is dying	301 (32.9)
M1	Eye pupils not checked	97 (10.6)
M3	Providing Adrenaline is inappropriate at this moment	92 (10.1)
M14	You have provided Oxygen before ensuring the airway was open	91 (9.9)
M12	The patient’s state has gotten worse!	59 (6.4)
M5	Providing Salbutamol is inappropriate at this moment	42 (4.6)
M8	CPR at this point was inappropriate	32 (3.5)
M6	Providing Hydrocortisone is inappropriate at this moment	30 (3.3)
M9	Temperature not checked	26 (2.8)
M7	Providing Atropine Sulphate is inappropriate at this moment	22 (2.4)
M10	Did not use DEX/STICK	20 (2.2)
M13	You turned the Oxygen off	20 (2.2)
M16	Providing Acetylsalicylic Acid is inappropriate at this moment	17 (1.9)
M2	Providing Furosemide is inappropriate at this moment	15 (1.6)
M19	Providing Nitroglycerin is inappropriate at this moment	11 (1.2)
M11	Did not attempt to interact with the patient	10 (1.1)
M17	Providing Glucagon is inappropriate at this moment	10 (1.1)
M18	Providing Nitrous Oxide is inappropriate at this moment	5 (0.6)
M15	Did not palpate the patient’s chest	4 (0.4)
M20	Providing Narcan is inappropriate at this moment	4 (0.4)
M23	ECG monitor not activated	4 (0.4)
M21	Providing Paracetamol is inappropriate at this moment	2 (0.2)
M22	Providing Morphine is inappropriate at this moment	1 (0.1)

Users’ Evaluation of the VETM Game

Results of the evaluation are presented in Table 3 in two formats: (1) using frequencies (percentages of participants who fell into each one of the four categories ranging from “strongly disagree” to “strongly agree”), and (2) using descriptive statistics (means and standard deviations). Descriptive statistics are used to identify general trends in participants’ perceptions of the game and frequencies are used to examine the results of the participants’ response to each question in more detail (Table 3).

Table 4 summarizes the descriptive statistics of each one of the six dimensions evaluated in the VETM game (interface, feedback, difficulty level, educational value, engagement, and terminology). As can be seen from Table 4, overall results are relatively high with regard to users’ engagement (mean 3.99, SD 0.87, n=84), the game’s interface (mean 3.83, SD 1.0, n=83), and educational value (mean 3.83, SD 1.05, n=84), and a bit lower with regard to the feedback provided in the game (mean 3.4, SD 0.79, n=81), the difficulty level (mean 3.2, SD 0.65, n=84), and the terminology used in the game (mean 3.32, SD 1.0, n=84). The principal results from participants’ evaluation of the VETM game are further discussed in the next section of the paper.

Table 3. Results of the usability evaluation of the VETM (Virtual Emergency TeleMedicine) game through frequencies and descriptive statistics.

	Domain examined	n	Frequencies[a]					Descriptive statistics	
			S (%)	D (%)	N (%)	A (%)	SA (%)	mean	SD
User interface (8 questions)									
1.	Accessing the game objectives was easy.	87	1.1	8.9	24.1	33.3	32.2	3.86	1.01
2.	Accessing the game instructions was easy.	88	2.3	9.1	27.3	37.5	23.9	3.72	1.01
3.	Accessing the patient history was easy.	88	3.4	3.4	15.9	37.5	39.8	4.07	1.00
4.	The game is user-friendly.	87	3.4	6.9	13.8	37.9	37.9	4.00	1.06
5.	I like the interface of the game.	87	4.6	5.7	19.5	41.4	28.7	3.84	1.06
6.	The game graphics are adequate.	87	4.6	11.5	33.3	33.3	17.2	3.47	1.05
7.	The response time of the game is as expected.	85	4.7	10.6	27.1	40.0	17.6	3.55	1.05
8.	The game is easy to navigate.	86	1.2	19.8	15.1	37.2	26.7	3.69	1.11
Difficulty level (5 questions)									
9.	Solving the problem was easy.	88	9.1	18.2	34.1	30.7	8.0	3.10	1.08
10.	The game is not challenging for me.	87	17.2	25.3	25.3	16.1	16.1	2.89	1.32
11.	The time allowed by the game for the doctor to save the patient is sufficient.	88	6.8	13.6	25.0	37.5	17.0	3.44	1.13
12.	I needed more time to be able to solve the problem in Scenario Dyspnea and Chest Discomfort.	86	9.3	24.4	34.9	25.6	5.8	2.94	1.06
13.	The game is complicated.	86	17.4	23.3	33.7	19.8	5.8	2.73	1.14
Feedback (5 questions)									
14.	Keeping track of my score while playing the game was easy.	86	4.7	12.8	22.1	34.9	25.6	3.64	1.14
15.	The game provides ways to recover after making a mistake.	85	29.4	14.1	29.4	21.2	5.9	2.6	1.27
16.	The feedback I receive when I make a choice is adequate.	85	4.7	15.3	41.2	25.9	12.9	3.27	1.03
17.	The feedback I receive when I make a choice is confusing.	86	8.1	12.9	46.5	22.1	10.5	3.14	1.04
18.	I can learn from my mistakes when I play the game.	85	5.9	10.6	20.0	31.8	31.8	3.73	1.19
Educational value (9 questions)									
19.	I learned how to diagnose and treat complications of … in the game.	87	9.2	9.2	26.4	39.1	16.1	3.44	1.15
20.	It was not clear what I could learn from the game.	86	9.3	23.3	40.7	22.1	4.7	2.9	1.0
21.	I found the game educational.	85	4.7	7.1	24.7	36.5	27.1	3.74	1.08
22.	The game will be interesting for medical students.	87	3.4	6.9	17.2	39.1	33.3	3.92	1.05
23.	The game will be useful for medical student training.	87	5.7	8.0	19.5	36.8	29.9	3.77	1.14
24.	This game is a useful learning aid.	87	3.4	3.4	24.1	40.2	28.7	3.87	0.99
25.	Learning objectives are clearly identified.	87	4.6	3.4	34.5	32.2	25.3	3.7	1.04
26.	I would recommend the game to my colleagues.	87	5.7	4.6	25.3	37.9	26.4	3.75	1.08
27.	If I were an instructor I would like to use the game in a classroom setting with my students.	87	6.9	8.0	16.1	37.9	31.0	3.78	1.18
Engagement (6 questions)									
28.	I was motivated to undertake the challenge of the game.	86	2.3	9.3	23.3	46.5	18.6	3.7	0.96

	Domain examined	n	Frequencies[a]					Descriptive statistics	
			S (%)	D (%)	N (%)	A (%)	SA (%)	mean	SD
29.	I am interested in learning about how to react in cardiovascular emergency situations.	87	0.0	3.4	16.1	39.1	41.4	4.18	0.83
30.	I feel “in control” when I play the game.	87	6.9	11.5	26.4	41.4	13.8	3.44	1.09
31.	I was absorbed in the activity of the game.	87	2.3	3.4	27.6	44.8	21.8	3.8	0.9
32.	I felt that time passed quickly.	86	2.3	3.5	22.1	43.0	29.0	3.93	0.93
33.	The game was worthwhile.	85	2.4	2.4	22.4	41.2	31.8	3.98	0.93
Terminology (2 questions)									
34.	The terminology used is correct.	85	4.7	12.9	31.8	36.5	14.1	3.42	1.04
35.	The terminology used is consistent.	85	4.7	11.8	31.8	42.4	9.4	3.4	0.98
36.	I would prefer the VETM game in Greek rather than English.	86	2.3	1.2	16.3	16.3	64.0	4.38	0.96

[a] SD=Strongly disagree, D=Disagree, N=Neither agree nor disagree, A=Agree, SA=Strongly agree.

Table 4. Descriptive statistics of the compound scores of the six dimensions of the game.

Dimensions of the game	n	Mean	SD
Interface	83	3.83	1.01
Feedback	81	3.40	0.79
Difficulty level	84	3.20	0.65
Educational value	84	3.83	1.05
User engagement	84	3.99	0.87
Terminology	84	3.32	1.00

Users’ Attitudes Toward the Game

Users’ attitudes toward the game were examined using two qualitative data sources: the users’ written comments in the five questions of the questionnaire and the users’ verbal comments in brief interviews. Users’ audiotaped comments complemented data that came from the open-ended questions of the questionnaire and provided more extensive explanations for users’ reactions and attitudes toward the game. As negative comments and game features that users found confusing or difficult spontaneously led to suggestions for improvement, these are grouped together in the following section, which starts with users’ negative comments and naturally leads to their suggestions for improving the game.

Users’ Identification of Confusing or Difficult Features of the Game and Suggestions for Improvement of the Game

Users identified features of the game that created confusion or hindered their problem-solving process while interacting with the game in Questions 2 and 3 and provided suggestions for improving the game in Question 4 of the questionnaire.

More specifically, Question 2 (What did you not like in the VETM game?) was answered by 57 people. One in five people (21%, 15/57) noted a few bugs that need to be improved, 8% (6/57) of people thought that the time allowed to save the patient especially when in state of complication was not adequate, and the same number of people disliked the fact that the game was written in the English language instead of the Greek language (8%, 6/57). Other negative features that were reported by only three people referred to the need to have a larger number of choices available (3/57), and the graphics (3/57) and colors of the game (3/57).

Question 3 (What things were difficult or caused confusion in the game?) was answered by 51 people; 28% (13/51) of people again referred to some bugs in the game, while 11% (5/51) of people said that it was not difficult so they didn’t have anything to report. Three people referred to other difficulties within the game such as providing the exact dosage of medication (3/51), the use of the English language (3/51), and difficulty in evaluating the state of the patient (3/51).

Question 4 (Do you have any suggestions for improving the game?) was answered by 57 people; 13% of people suggested collaboration with experienced nursing personnel to improve the game and correcting the bugs that were identified (8/57) and 8% suggested translating the game into Greek (5/57), adding options for possible interventions that nurses can take (3/57), and improving graphics (3/57).

With respect to negative or confusing aspects of the game as these were coded from the analysis of the users’ interviews, the following features were identified, at least once, by users who provided suggestions to alleviate difficulties: more closely following the “ABCDE” protocols with which nurses are familiar (M14, Interview 7; M26, Interview 13), using the generic names of drugs (Male Trainer, Interview 6), adding all possible drugs that could be administered, have them visible in all scenarios that are developed (F6, Interview 10; F7, Interview 11; M37, Interview 22; M38, Interview 23) and include the standardized dosage for each one (M21, Interview 10; M35, Interview 21), adding the heart, lungs, and ECG sounds to the game (M14, F15 Interview 7), making the time that passes when performing CPR more visible by increasing the font and changing the color of the written feedback (M15, Interview 8; M37, Interview 22; M39, Interview 23), provide the ability to give adrenaline while performing CPR (M22, Interview 10; M32, Interview 19; M37, Interview 22), having the defibrillator present at all times rather than making it appear only in scenarios where it is needed (M28, Interview 14), having the user chose current voltage and joules for defibrillation (M29, Interview 15), and last, translate the game into the Greek language as users are more familiar with the terminology in Greek rather than English (M28, Interview 14; M38, Interview 23).

An interesting aspect that emerged from the analysis of the interviews had to do with the desire of experienced users to have less guidance and less scaffolding in the game. Users commented that, if the target group is professionals, then the game should provide more advanced options (M7, Interview 4; M15, Interview 8; M23, Interview 11; M27, Interview 14; M39, Interview 23; F15, Interview 24) and not make these options immediately available to the user (M23, Interview 11). They also suggested additions of more advanced options, such as the addition of a greater number of types of oxygen masks (M23, Interview 11; M9, Interview 4; F2, Interview 5; F8, Interview 13) so that the user can learn how to identify the correct mask to be applied depending on each medical incident that is presented in the game and the addition of more advanced questions, such as “how much oxygen to provide” instead of simply providing oxygen (M9, Interview 4; F2, Interview 5), or “identify which type of IV should be induced” rather than simply providing IV for experienced users (M15, Interview 8; M14, Interview 7), or “getting a different sound or input when examining a different part of the lungs” (M16, Interview 8; M20, Interview 10). Experienced participants could, however, understand that the game in its current state would be useful for medical students and first-year nurses who are inexperienced. Another suggestion included changing the animated character of the virtual patient into a human character to make the game more realistic (F5, Interview 8).

Discussion

Principal Results

This study aimed to perform a professional feedback-informed evaluation of the VETM game to identify professionals’ perceptions of the game’s potential for training people involved in the assessment and management of emergency cases in cardiology. Overall, the results were positive in all six dimensions that were assessed: game interface, feedback, difficulty level, educational value, user engagement, and terminology used in the game (with means ranging from 3.2 to 3.99 out of 5 in the six dimensions). This finding indicates that professionals can see the potential of the VETM game for training, practicing, or evaluating users’ problem-solving skills in real-life clinical scenarios through a telemedicine system and a virtual patient and is in agreement with previous studies that documented positive attitudes of medical students on games [1], favorable attitudes of junior doctors on new media, and favorable attitudes of medical students on the use of serious games involving virtual patients in medical education [19].

Evaluation results for each dimension are discussed in more detail. The first dimension of the evaluation referred to the interface of the game and it was assessed based on the cumulative results of 8 relevant questions. As can be seen from Table 4, the compound mean score of the game’s interface was 3.83 (SD 1.01), which provides an indication that the interface of the game was satisfactory. As can be shown from more detailed results reported in Table 3, three out of four participants (76%, 66/87) thought that the game was user friendly (Q4), 70% (61/87) of them liked the interface of the game (Q5), and 64% (55/86) thought that the game was easy to navigate (Q8). More than 60% of participants agreed or strongly agreed that accessing the game objectives (Q1, 66%, 57/87) and instructions (Q2,61%, 54/88), and the patient’s history (Q3, 77%, 68/88) was easy. However, relatively lower scores were received for the response time of the game (Q7, mean 3.55, SD 1.05, n=85) and for the evaluation of the graphics of the game (Q6, mean 3.47, SD 1.05, n=87). The last point may relate to the suggestion made by three participants to improve the graphics of the game and the comment about changing the animated character to a human character in one of the interviews.

The second dimension of the evaluation referred to whether the difficulty level of the game was appropriate (mean 3.2, SD 0.65, n=84) and it was assessed based on the cumulative results of 5 relevant questions, two of which that were negatively phrased (Q10 and Q12) were reversed. More than half of participants (55%, 48/88) thought that the time allowed to solve the problem was sufficient (Q11, mean 3.44, SD 1.13, n=88). Almost 40% (34/88) of participants thought that solving the problem was easy, while 27% (24/88) of participants thought that solving the problem was difficult. It is important to note that around 34% (30/88) of participants were not sure whether the scenario they tried was easy or difficult. This is partly reflected by the success rate that was calculated from the analysis of 379 log files of the game that showed that participants tried to solve the problem for an average of 8 times and their success rate was generally low (20.6%). It is possible that participants’ personal characteristics such as previous professional experience, experience with a telemedicine system and experience with playing educational games may have influenced their perceptions with regard to the difficulty level of the game.

The third dimension of the evaluation referred to the adequacy of feedback provided by the game to support users’ learning (mean 3.40, SD 0.79, n=81) and it was assessed based on the cumulative results of 5 relevant questions. A total of 63% (54/85) thought that they could learn from their mistakes while playing the game (Q18). Keeping track of their score while playing the game was characterized as easy by 61% (52/86) of participants (Q14). The feedback received was characterized as adequate by 39% (33/85) of participants (Q16) and 27% (23/85) agreed that the game provides ways to recover after making a mistake (Q15). The last point relates to what seemed to be an area of misconception among participants who thought that the game should provide ways to recover after making any type of mistake as opposed to the design goal of the game to additionally simulate mistakes that are irreversible in real life and thus should be irreversible in the game environment as well.

The fourth dimension of the evaluation referred to the value of the game and it was assessed based on the cumulative results of 9 relevant questions. Results showed that participants greatly valued the VETM game (mean 3.83, SD 1.05, n=84). The majority of participants thought that the game is educational (64%, 54/85), it will be interesting for medical students (72%, 63/87), useful for medical student training (67%, 58/87), and useful as a learning aid (69%, 60/87). Even though the hands-on experience with playing the game was short, more than half of participants (55%, 48/87) thought that they learned how to diagnose and treat complications of cardiovascular type through the game and 57% (50/87) thought that the learning objectives were clearly identified. Last, 64% (54/85) of participants would recommend the game to their colleagues and 69% (60/87) would like to use it for teaching purposes. The last finding agrees with what participants also noted in their written comments in question 5 of the questionnaire and with what they reported in interviews, in which they referred to the originality of the game, they expressed an intention to download it to practice with more scenarios, and they referred to advantages such as the elimination of cost for training, the elimination of danger associated with dealing with real patients, and providing the ability for users to learn from their mistakes in a safe environment.

These findings are in accordance with what was reported in the literature with regard to the affordability, accessibility, and usability of serious games, reported by [9,23], and with regard to avoiding risks associated with real patients reported by [4,30]. Moreover, participants' positive reaction toward the educational value of the game agrees with findings of previous studies that have been reported in the literature that documented medical students' positive attitudes toward games [1-3,19].

The fifth dimension of the evaluation referred to engagement and it was assessed with 6 questions. It has the highest mean score of the six dimensions that were examined (mean 3.99, SD 0.87, n=84). The majority of participants felt that the game was worthwhile (73%, 62/85), that time passed quickly (72%, 62/86), they were absorbed in the activity (67%, 58/87), motivated to undertake the challenge of the game (65%, 56/86), and felt "in control" (55%, 48/87). Again, this finding also agrees with the generally positive attitudes of medical students and professionals toward games that were reported in the literature [1-3,19].

The sixth and last dimension of the evaluation referred to whether the terminology used in that particular scenario was correct and consistent (mean 3.32, SD 1.0, n=84) and it was assessed based on the cumulative results of two relevant questions.

As documented in the literature review, there is a lack of evaluation/impact measurements with actual professional groups, as most studies that aimed to evaluate serious games and virtual patients had students as participants. From a methodological standpoint, this study builds on the work of Heinrichs et al [38] who validated their work on a virtual emergency department with a small sample of professionals, specifically 22 participants.

Not many open source serious games exist that focus on supporting medical students' skills in responding to medical emergency situations to manage the symptoms of patients facing cardiovascular complications through telemedicine. From a game design and development standpoint, this study overcomes some of the limitations of games developed as part of the eViP virtual patients' project in the area of "cardiology, emergency medicine, ECG". More specifically, as opposed to other non-commercial games that already exist in this area, the VETM game is not linear in nature, the order of choices provided to the user changes randomly each time the game is played, it is a real-time game simulating realistic conditions of emergency care where the user is required to solve the problem under time pressure, and it allows immediate contact with a virtual patient all the time as both the patient and signals of the ECG can be observed at any given time through the telemedicine system.

Limitations

Even though the VETM game was validated by professionals with regard to its potential educational value in this study and results were positive, it has not yet been validated using a controlled trial design, which is part of the directions for future research in this area.

Directions for Future Research

Most games for health care have not been validated as tools for education [7]. As Graafland et al [21] pointed out, games developed or used to train medical professionals need to be validated before they are integrated into teaching methods. Some weaknesses that were pointed out by participants will serve as input for the re-design of the VETM game. Further research should define valid performance parameters and formally validate any serious game before it can be seen as a fully-fledged teaching instrument for medical professionals. A direction of further research, to address the need for employing rigorous methodologies, such as controlled trials for the evaluation of virtual patients, will therefore be to determine the educational efficacy of VETM, in a controlled setting involving an experimental and a control group. The experimental group participants will interact with VETM as a way to practice their skills in emergency treatment while the control group participants will follow traditional teaching methods. Participants of both groups will be pre- and post-tested with respect to their problem-solving skills in treating a patient's symptoms in an emergency situation.

Furthermore, the real emergency telemedicine system will be expanded with (1) the addition of emergency scenarios for patient handling [50]), (2) health care services and patient location management during major disasters [51], and (3) ultrasound emergency video transition from the ambulance to the hospital [44-46]. It is foreseen that the above functionality will be integrated in the VETM serious game for the support of advanced emergency telemedicine services.

Conclusions

In response to calls to integrate serious games in medical education and training and in light of research findings that report the effectiveness of virtual patient implementation in medical education, this paper attempted to describe the design and evaluation of the Virtual Emergency TeleMedicine (VETM) game, a simulation-based virtual patient game that was developed in collaboration with the mEducator Best Practice network [49]. The VETM game makes use of data from an electrocardiogram to train practicing doctors or medical students for problem solving in real-life emergency clinical scenarios through a telemedicine system and virtual patients.

What makes the VETM game innovative and different from previous attempts to use telemedicine for training in emergency situations is that the game not only allows users to practice in the context of given scenarios but also allows them to write scenarios of their own with very limited knowledge of programming required, as the scenario editor uses a very simple state-based scripting language similar to C. Even though this functionality was not evaluated in the present study, it may provide a partial solution to the concern raised in the literature that refers to the difficulty and added costs for virtual patients to be authored, adapted, and exchanged. Another element that makes the VETM game different is the fact that it is made freely available through the Melina+ (Medical Education Linked Arena) platform for repurposing in different contexts, with different scenarios, in different languages, etc [49].

In response to the gap identified in the literature of a lack of evaluation/impact measurements with actual professional groups, this study followed a quantitative, professional feedback-informed evaluation of the educational potential of the VETM game. Professionals, who were also stakeholders in this case, included professional nurses, whose expertise is the assessment and management of emergency cases.

The results of the evaluation are promising with regard to the value of the game and provide a strong indication of the potential of this educational game in telemedicine. According to Cugelman [40], “users are the ultimate judges of intervention efficacy, so any gamified interventions will require user testing, to determine if they can work or not”(p. 4). Following Cugelman’s suggestion [40], the first users who tested the game were professionals, to provide their input and feedback as part of a quantitative evaluation of the game. This study showed that the inclusion of game elements in the domain of emergency treatment involving cardiovascular complications has the potential to enhance medical students’ learning experience and may increase their intrinsic motivation to practice, making the learning experience more enjoyable and potentially more effective, even though the latter remains to be seen. VETM can potentially become an attractive option for large-scale continuous medical education to help improve the knowledge of medical students, nurses, paramedics, rescuers, etc, on emergency treatment and potentially improve the diagnosis and treatment of patients who face cardiovascular or other problems.

Acknowledgments

This work was partially supported by the project mEducator (Multi-type Content Sharing and Repurposing in Medical Education), funded by the eContentplus Programme, a multiannual community program to make digital content in Europe more accessible, usable, and exploitable (Grant ECP 2008 EDU 418006). The completion of this work would not have been possible without the support and guidance of the Cyprus Emergency and Pre-Hospital Care Association (KSEPA). We would also like to acknowledge and thank the participants for voluntarily agreeing to participate to this study.

Conflicts of Interest

None declared.

References

1. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ 2010;10:50 [FREE Full text] [doi: 10.1186/1472-6920-10-50] [Medline: 20576125]
2. Richardson A, Bracegirdle L, McLachlan SIH, Chapman SR. Use of a three-dimensional virtual environment to teach drug-receptor interactions. Am J Pharm Educ 2013 Feb 12;77(1):11 [FREE Full text] [doi: 10.5688/ajpe77111] [Medline: 23459131]
3. Jaffer U, John N, Standfield N. Surgical Trainee Opinions in the United Kingdom Regarding a Three-Dimensional Virtual Mentoring Environment (MentorSL) in Second Life: Pilot Study. JMIR Serious Games 2013;1(1):e2 [FREE Full text] [doi: 10.2196/games.2822] [Medline: 25658652]
4. Dafli EL, Bamidis PD, Dombros NV. Development of a pilot application of virtual patients in Greek medical education. Archives of Hellenic Medicine 2010;27(5):803-810.
5. Danforth D, Procter M, Chen R, Johnson M. Development of Virtual Patient Simulations for Medical Education. Journal For Virtual Worlds Research 2009;2(2):3-11. [doi: 10.4101/jvwr.v2i2.707]
6. Sliney A, Murphy D. JDoc: A Serious Game for Medical Learning. 2008 Feb 10 Presented at: First International Conference on Advances in Computer-Human Interaction; 2008; Sainte Luce, Martinique p. 131-136. [doi: 10.1109/ACHI.2008.50]

7. Diehl LA, Souza RM, Alves JB, Gordan PA, Esteves RZ, Jorge MLSG, et al. InsuOnline, a Serious Game to Teach Insulin Therapy to Primary Care Physicians: Design of the Game and a Randomized Controlled Trial for Educational Validation. JMIR Res Protoc 2013;2(1):e5 [FREE Full text] [doi: 10.2196/resprot.2431] [Medline: 23612462]
8. De Freitas S. A review of game-based learning. Learning in immersive worlds URL: http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf [accessed 2014-07-03] [WebCite Cache ID 6QmndrvJ4]
9. Knight JF, Carley S, Tregunna B, Jarvis S, Smithies R, de Freitas S, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. Resuscitation 2010 Sep;81(9):1175-1179. [doi: 10.1016/j.resuscitation.2010.03.042] [Medline: 20732609]
10. Papadopoulos L, Pentzou A, Louloudiadis K, Tsiatsos T. Design and evaluation of a simulation for pediatric dentistry in virtual worlds. J Med Internet Res 2013 Oct;15(11):e240 [FREE Full text] [doi: 10.2196/jmir.2651] [Medline: 24168820]
11. Kononowicz AA, Krawczyk P, Cebula G, Dembkowska M, Drab E, Frączek B, et al. Effects of introducing a voluntary virtual patient module to a basic life support with an automated external defibrillator course: a randomised trial. BMC Med Educ 2012;12:41 [FREE Full text] [doi: 10.1186/1472-6920-12-41] [Medline: 22709278]
12. Harris JM, Sun H. A randomized trial of two e-learning strategies for teaching substance abuse management skills to physicians. Acad Med 2013 Sep;88(9):1357-1362 [FREE Full text] [doi: 10.1097/ACM.0b013e31829e7ec6] [Medline: 23887001]
13. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. Simul Healthc 2012 Feb;7(1):10-17. [doi: 10.1097/SIH.0b013e31823652de] [Medline: 22228285]
14. Botezatu M, Hult H, Tessma MK, Fors U. Virtual patient simulation: knowledge gain or knowledge loss? Med Teach 2010 Jul;32(7):562-568. [doi: 10.3109/01421590903514630] [Medline: 20653378]
15. Vukanovic-Criley JM, Boker JR, Criley SR, Rajagopalan S, Criley JM. Using virtual patients to improve cardiac examination competency in medical students. Clin Cardiol 2008 Jul;31(7):334-339. [doi: 10.1002/clc.20213] [Medline: 18636531]
16. Thompson P. The digital natives as learners: Technology use patterns and approaches to learning. Computers & Education 2013 Jul;65:12-33. [doi: 10.1016/j.compedu.2012.12.022]
17. Kato PM. Video games in health care: Closing the gap. Review of General Psychology 2010;14(2):113-121. [doi: 10.1037/a0019441]
18. Rosenberg BH, Landsittel D, Averch TD. Can video games be used to predict or improve laparoscopic skills? J Endourol 2005 Apr;19(3):372-376. [doi: 10.1089/end.2005.19.372] [Medline: 15865530]
19. Schwarz D, Štourač P, Komenda M, Harazim H, Kosinová M, Gregor J, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res 2013 Jul;15(7):e135 [FREE Full text] [doi: 10.2196/jmir.2590] [Medline: 23835586]
20. Štourač P, Harazim H, Schwarz D. AKUTNE.CZ algorithms and SEPSIS-Q scenarios as interactive tools for problem based learning sessions in medical education. MEFANET Journal 2013;1(2):61-73.
21. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
22. Ruiz J, Mintzer M, Leipzig RM. The impact of E-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [Medline: 16501260]
23. Triola M, Feldman H, Kalet AL, Zabar S, Kachur EK, Gillespie C, et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. J Gen Intern Med 2006 May;21(5):424-429 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00421.x] [Medline: 16704382]
24. Salminen H, Zary N, Björklund K, Toth-Pal E, Leanderson C. Virtual patients in primary care: developing a reusable model that fosters reflective practice and clinical reasoning. J Med Internet Res 2014;16(1):e3 [FREE Full text] [doi: 10.2196/jmir.2616] [Medline: 24394603]
25. Ghanbarzadeh R, Ghapanchi A, Blumenstein M, Talaei-Khoei A. A decade of research on the use of three-dimensional virtual worlds in health care: a systematic literature review. J Med Internet Res 2014;16(2):e47 [FREE Full text] [doi: 10.2196/jmir.3097] [Medline: 24550130]
26. MedBiquitous Consortium. URL: http://www.medbiq.org/ [accessed 2014-07-03] [WebCite Cache ID 6Qmnka0KR]
27. Boulos MNK, Hetherington L, Wheeler S. Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Info Libr J 2007 Dec;24(4):233-245. [doi: 10.1111/j.1471-1842.2007.00733.x] [Medline: 18005298]
28. 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 Aug;34(8):674-675. [doi: 10.3109/0142159X.2012.689450] [Medline: 22830327]
29. Antoniou P, Athanasopoulou C, Dafli E, Bamidis P. Exploring design requirements for repurposing dental virtual patients from the web to second life: a focus group study. J Med Internet Res 2014;16(6):e151 [FREE Full text] [doi: 10.2196/jmir.3343] [Medline: 24927470]
30. Dafli E, Antoniou P, Ioannidis L, Dombros N, Topps D, Bamidis PD. Virtual patients on the semantic Web: a proof-of-application study. J Med Internet Res 2015 Jan;17(1):e16 [FREE Full text] [doi: 10.2196/jmir.3933] [Medline: 25616272]

31. Diener S, Windsor J, Bodily D. Educause Australasia. 2009. Design and Development of Medical Simulations in Second Life and OpenSim URL: http://www.caudit.edu.au/educauseaustralasia09/assets/papers/monday/Scott-Diener.pdf [accessed 2014-07-03] [WebCite Cache ID 6QmnqCrCV]
32. Schwaab J, Kman N, Nagel R, Bahner D, Martin DR, Khandelwal S, et al. Using second life virtual simulation environment for mock oral emergency medicine examination. Acad Emerg Med 2011 May;18(5):559-562. [doi: 10.1111/j.1553-2712.2011.01064.x] [Medline: 21521404]
33. Guise V, Chambers M, Conradi E, Kavia S, Välimäki M. Development, implementation and initial evaluation of narrative virtual patients for use in vocational mental health nurse training. Nurse Educ Today 2012 Aug;32(6):683-689. [doi: 10.1016/j.nedt.2011.09.004] [Medline: 22056146]
34. McEvoy MM, Butler B, MacCarrick G, Nicholson AJ. Virtual patients: an effective educational intervention to improve paediatric basic specialist trainee education in the management of suspected child abuse? Ir Med J 2011 Sep;104(8):250-252. [Medline: 22125882]
35. Deladisma AM, Johnsen K, Raij A, Rossen B, Kotranza A, Kalapurakal M, et al. Medical student satisfaction using a virtual patient system to learn history-taking communication skills. Stud Health Technol Inform 2008;132:101-105. [Medline: 18391266]
36. Deladisma AM, Gupta M, Kotranza A, Bittner JG, Imam T, Swinson D, et al. A pilot study to integrate an immersive virtual patient with a breast complaint and breast examination simulator into a surgery clerkship. Am J Surg 2009 Jan;197(1):102-106. [doi: 10.1016/j.amjsurg.2008.08.012] [Medline: 19101251]
37. Dev P, Heinrichs WL, Youngblood P. Virtual Patient Model for Multi-Person Virtual Medical Environments. In: AMIA Annu Symp Proc. 2007 Presented at: AMIA Annual Symposium; 2007; Chicago, Illinois p. 181-185.
38. Heinrichs WL, Youngblood P, Harter P, Kusumoto L, Dev P. Training healthcare personnel for mass-casualty incidents in a virtual emergency department: VED II. Prehosp Disaster Med 2010;25(5):424-432. [Medline: 21053190]
39. MicroSim. URL: http://www.laerdal.com/docid/5899175/microsim [accessed 2014-07-03] [WebCite Cache ID 6QmnytRCOCugelman]
40. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
41. Mougiakakou S, Kyriacou E, Perakis K, Papadopoulos H, Androulidakis A, Konnis G, et al. BioMedical Engineering OnLine. 2011. Towards the Integrated Provision of eHealth Tools and Services: INTERMED and the Experience in the Region of South-East Mediterranean URL: http://www.biomedical-engineering-online.com/content/10/1/49 [accessed 2014-07-03] [WebCite Cache ID 6Qmo3vj1N]
42. Kyriacou E, Pattichis C, Karaolis M, Loizou C, Christodoulou C, Pattichis M, et al. An Integrated System for Assessing Stroke Risk. IEEE Eng Med Biol Mag 2007 Sep;26(5):43-50. [doi: 10.1109/EMB.2007.901794]
43. Pattichis C, Kyriacou E, Voskarides S, Pattichis M, Istepanian R, Schizas C. Wireless telemedicine systems: an overview. IEEE Antennas Propag Mag 2002 Apr;44(2):143-153. [doi: 10.1109/MAP.2002.1003651]
44. Panayides A, Eleftheriou I. Open-Source Telemedicine Platform for Wireless Medical Video Communication. International Journal of Telemedicine and Applications 2013:1-12. [doi: 10.1155/2013/457491]
45. Panayides A, Antoniou ZC, Mylonas Y, Pattichis MS, Pitsillides A, Pattichis CS. High-resolution, low-delay, and error-resilient medical ultrasound video communication using H.264/AVC over mobile WiMAX networks. IEEE J Biomed Health Inform 2013 May;17(3):619-628. [doi: 10.1109/TITB.2012.2232675] [Medline: 23232416]
46. Panayides AS, Pattichis MS, Pattichis CS. Mobile-Health Systems Use Diagnostically Driven Medical Video Technologies [Life Sciences]. IEEE Signal Process Mag 2013 Nov;30(6):163-172. [doi: 10.1109/MSP.2013.2276512]
47. Nicolaidou I, Antoniades A, Marangos C, Dafli E, Kyriacou E, Bamidis P, et al. The instructional design of the Virtual Telemedicine: A simulation-based serious game in health care. In: 5th International Conference of Education, Research and Innovation. 2012 Presented at: ICERI2012; Nov. 2012; Madrid, Spain p. 5746-5755.
48. Antoniades A, Marangos C, Kyriacou E, Nicolaidou I, Dafli E, Bamidis P, et al. Virtual Telemedicine: A serious game to develop problem solving skills in medical education. In: 4th International Conference on Education and New Learning Technologies. 2012 Presented at: EDULEARN2012; July 2012; Barcelona, Spain p. 3612-3619.
49. mEducator (Multi-type Content Sharing and Repurposing in Medical Education). URL: http://www.meducator.net/ [accessed 2015-06-12] [WebCite Cache ID 6ZFGjSHBC]
50. Kyriacou E, Constantinides P, Pattichis C, Pattichis M, Panayides A. eEmergency Health Care Information Systems. In: Proceedings of the 33rd IEEE EMBS Annual International Conference. 2011 Presented at: IEEE EMBS Annual International Conference; 2011; Boston, MA.
51. Kyriacou E, Kyprianou S, Christou S, Constantinou R, Panayides A, Neofytou M, et al. Avaris Net- CrisisEmergency Management of Health Services. 2014 Presented at: IEEE EMBC; 2014; Chicago, Illinois.

Abbreviations

BLS-AED: Basic Life Support Automated External Defibrillation
ECG: electrocardiogram

VETM: Virtual Emergency TeleMedicine

Edited by G Eysenbach, D Giordano; submitted 03.07.14; peer-reviewed by D Davies, P Stourac, L Diehl; comments to author 09.08.14; revised version received 12.01.15; accepted 15.02.15; published 17.06.15.

Please cite as:
Nicolaidou I, Antoniades A, Constantinou R, Marangos C, Kyriacou E, Bamidis P, Dafli E, Pattichis CS
A Virtual Emergency Telemedicine Serious Game in Medical Training: A Quantitative, Professional Feedback-Informed Evaluation Study
J Med Internet Res 2015;17(6):e150
URL: http://www.jmir.org/2015/6/e150/
doi:10.2196/jmir.3667
PMID:26084866

Original Paper

Mapping Power Law Distributions in Digital Health Social Networks: Methods, Interpretations, and Practical Implications

Trevor van Mierlo[1,2*], BA (Hons), MScCH, MBA, GEMBA, MScBMR; Douglas Hyatt[3*], PhD; Andrew T Ching[3*], PhD

[1]Research Associate, Henley Business School, University of Reading, Oxfordshire, United Kingdom
[2]Evolution Health Systems Inc, Toronto, ON, Canada
[3]Rotman School of Managment, University of Toronto, Toronto, ON, Canada
[*]all authors contributed equally

Corresponding Author:
Trevor van Mierlo, BA (Hons), MScCH, MBA, GEMBA, MScBMR
Research Associate, Henley Business School, University of Reading
Greenlands, Henley-on-Thames
Oxfordshire, RG93AU
United Kingdom
Phone: 44 1491 571454
Fax: 44 1491 571454
Email: tvanmierlo@evolutionhs.com

Abstract

Background: Social networks are common in digital health. A new stream of research is beginning to investigate the mechanisms of digital health social networks (DHSNs), how they are structured, how they function, and how their growth can be nurtured and managed. DHSNs increase in value when additional content is added, and the structure of networks may resemble the characteristics of power laws. Power laws are contrary to traditional Gaussian averages in that they demonstrate correlated phenomena.

Objectives: The objective of this study is to investigate whether the distribution frequency in four DHSNs can be characterized as following a power law. A second objective is to describe the method used to determine the comparison.

Methods: Data from four DHSNs—Alcohol Help Center (AHC), Depression Center (DC), Panic Center (PC), and Stop Smoking Center (SSC)—were compared to power law distributions. To assist future researchers and managers, the 5-step methodology used to analyze and compare datasets is described.

Results: All four DHSNs were found to have right-skewed distributions, indicating the data were not normally distributed. When power trend lines were added to each frequency distribution, R^2 values indicated that, to a very high degree, the variance in post frequencies can be explained by actor rank (AHC .962, DC .975, PC .969, SSC .95). Spearman correlations provided further indication of the strength and statistical significance of the relationship (AHC .987. DC .967, PC .983, SSC .993, $P<.001$).

Conclusions: This is the first study to investigate power distributions across multiple DHSNs, each addressing a unique condition. Results indicate that despite vast differences in theme, content, and length of existence, DHSNs follow properties of power laws. The structure of DHSNs is important as it gives insight to researchers and managers into the nature and mechanisms of network functionality. The 5-step process undertaken to compare actor contribution patterns can be replicated in networks that are managed by other organizations, and we conjecture that patterns observed in this study could be found in other DHSNs. Future research should analyze network growth over time and examine the characteristics and survival rates of superusers.

(J Med Internet Res 2015;17(6):e160) doi:10.2196/jmir.4297

KEYWORDS

social networks; eHealth; 1% rule; Pareto Principal; power law; 90-9-1 principle; moderated support; peer-to-peer support

Introduction

Background

Empirical examination of digital health social networks (DHSNs) began in the mid-1980s. In 1986, Schneider examined abstinence rates from smoking among 28 actors of an online system named the Electronic Information Exchange System (EIES) [1]. Actors logged on to EIES by typing the word “smoker” and could read and post messages to a bulletin board. During the same year, Robinson and Walters outlined Health-Net, an interactive computer network linking personal computers in student residences, libraries, academic buildings, and the Student Health Center at Stanford University [2]. Like EIES, Health-Net contained a bulletin board. These researchers all noted the potential impact of these networks on personal health, especially in regards to information access and knowledge sharing.

Decades later, DHSNs, otherwise known as bulletin boards, peer-to-peer support groups, online forums, or computer-mediated communication now proliferate the digital health landscape. As of December 2014, over 40,000 health-related communities exist on Yahoo! Groups. PatientsLikeMe, a for-profit health care company focusing on peer-to-peer support, has communities for over 2300 conditions. In 2013, Bender et al identified and examined 111 DHSNs dedicated to breast cancer survivors, with extensive archives of personal experiences [3].

The Internet also continues to evolve as an important health resource. A 2013 Pew Research Center report found that within the past year 59% of US adults used the Internet to search for health information, and 26% of Internet users read or watched someone else’s experience about a health or medical issue [4].

Although the research community is in the process of establishing the efficacy of DHSNs [5], peer-to-peer support groups remain an important component of the digital health ecosystem. A separate stream of research is evolving, which seeks to understand the mechanisms of DHSNs, how they are structured, how they function, and how their growth can be nurtured and managed. Other disciplines have analyzed complex networks, and measured specific interactions within [6], yet their theories and models have yet to be rigorously applied to digital health.

Network Effects

For decades, the fields of economics and marketing have sought to understand the structure, stagnation, growth, and distribution patterns of networks. The study of networks in demand-side economics has found that the value of a product or service is directly related to the number of others who use it [7-10]. This increase in value, otherwise known as a *positive network externality*, occurs with each sale of an additional unit.

This increase in value can be illustrated in many consumer goods. An example is FaceTime, a popular feature of Apple products. FaceTime is a videotelephony service (or video call app) that allows consumers to talk with each other via Voice over Internet Protocol (VoIP). FaceTime is available only on Apple products, so consumers must purchase an Apple product in order to join the FaceTime network. There were an estimated 19 million FaceTime-equipped devices in October 2010, growing to over 300 million by the end of 2012 [11]. For consumers (and presumably Apple), the value of the FaceTime network continues to increase with the sale of each additional Apple device.

Power Laws and Power Curves

A power law is an exponential relationship between two values that is scaled and is proportional. A power curve is the graphical representation of this phenomenon.

If plotted on a graph, the distinguishing feature of a power curve is a straight line with a slope of b or an equation of y=-x+b; the closer the data fit the straight line, the greater chance of the graphed relationship being defined as a power curve.

An example of a well-known power law is the Pareto Principal, colloquially known as the 80-20 rule. In the late 19th century, the Italian scholar Vilfredo Pareto noted that 80% of the land in Italy was owned by 20% of the population [12]. Likewise, it is common for those in business to note that 80% of their sales are generated from 20% of their customers or that 80% of absences can be attributed to 20% of staff.

One specific type of power law is a Zipfian distribution, otherwise known as Zipf’s law, eponymously named after George Kingsley Zipf, an American linguist and philologist who was a university lecturer at Harvard University [13]. Zipf first noted a statistical relationship in the frequency of word use but extended his method to other subjects, such as the size of cities and concentration of economic power [14].

Examples of power laws are ubiquitous. One resource lists over 80 types of natural and social power law phenomena in fields such as physics (eg, brush-fire damage, water levels in the Nile, earthquakes, size of asteroid hits), biology (eg, genetic circuitry, tumor growth, death from heart attack, predicting premature births, mass extinctions), social science (eg, word use, structure of World Wide Web, publications and citations, global terrorism events, traffic jams), and management research (eg, cotton prices, distribution of wealth, intra-firm decision events, alliance networks among biotech firms) [15]. In a separate study, power law distributions were consistent in 17 of 24 datasets ranging from linguistics (count of word use), biology (protein interaction degree), ornithology (bird species sightings), meteorology (solar flare intensity), and political science (intensity of wars) [16].

Power laws have been studied extensively, with Paul Kruman, a Nobel Prize winning economist, describing the phenomenon as “disturbing” or “baffling” [17]. However, defining power relationships are important as the models can help us with intuition and to begin to understand relationships between two distinct variables.

Distribution Patterns and Digital Health Social Networks

All networks have the potential to increase in value when an additional user or actor is added. Generally, if a network contains n users, potential connections between users is $n(n-1)$. However, value creation differs among various network types.

In our previous FaceTime example, network connections are ephemeral; a conversation between two actors terminates when a party ends the conversation. DHSNs differ from traditional networks as actor contributions are permanent. An actor's post remains on the network and can be accessed or read numerous times (Figure 1).

As actor posts are permanent in DHSNs, positive network externalities occur in two instances. The first is when a new actor joins the network and creates one or more posts. The second is when an existing actor authors a new post. In both instances, the network increases in size and value is added. Not surprisingly, actors post in varying frequencies, and some actors create more posts than others. The mathematical relationship between these two quantities (number of actors and number of posts) often constitutes a power law. Power laws are in contrast with traditional Gaussian averages in that they demonstrate correlated phenomena [15].

Recent research has started to investigate the power law phenomenon in DHSNs. A 2014 study found that the 1% rule, a marketing "rule of thumb", was consistent across four separate DHSNs [18]. Shortly afterwards, the 1% study was replicated within an Australian DHSN for depression [19]. This study confirmed the 1% rule and found that the ranked distribution of actor contributions fit a specific power law known as a Zipfian distribution.

As outlined previously, DHSNs have the potential to positively impact patients and may play a key role in normalizing disease and influencing medication and treatment adherence [4,5]. If they follow properties of power laws, managers and researchers may be able to account for, and anticipate, fluctuations in growth.

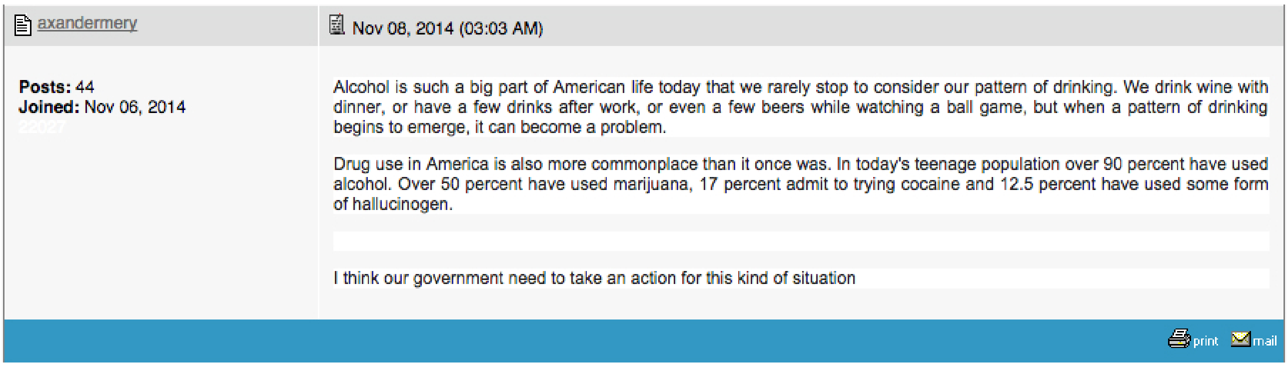


Figure 1. Post by actor axandermery on the social network Alcohol Help Center.

axandermery

Nov 08, 2014 (03:03 AM)

Posts: 44
Joined: Nov 06, 2014
22027

Alcohol is such a big part of American life today that we rarely stop to consider our pattern of drinking. We drink wine with dinner, or have a few drinks after work, or even a few beers while watching a ball game, but when a pattern of drinking begins to emerge, it can become a problem.

Drug use in America is also more commonplace than it once was. In today's teenage population over 90 percent have used alcohol. Over 50 percent have used marijuana, 17 percent admit to trying cocaine and 12.5 percent have used some form of hallucinogen.

I think our government need to take an action for this kind of situation

Objective

The objective of this study was to investigate whether the distribution frequency of four DHSNs, each addressing a unique condition, could be described as power curves. To assist future researchers with assessing the distribution frequencies of other DHSNs, a second objective was to describe the method used to determine the comparison.

Methods

Overview

The four DHSNs used in this study are Alcohol Help Center (AHC), Depression Center (DC), Panic Center (PC), and Stop Smoking Center (SSC). All DHSNs are moderated, are free to participants, do not offer any advertising or product promotion, and are components of each website's behavior-change program.

The DHSNs have been in existence for a considerable amount of time, ranging from 4.0 to 10.9 years (see Table 1).

Table 1. Digital health social networks.

	Problem drinking	Depression	Panic disorder	Smoking cessation
Date of first post	July 25, 2008	April 5, 2003	January 7, 2002	September 17, 2001
Date of last post	August 7, 2012	August 5, 2012	August 7, 2012	August 7, 2012
Days, n	1474	3411	3866	3978
Years, n	4.0	9.3	10.6	10.9
Posts, n	7148	12,583	45,032	513,586
Registrations, n	2584	5151	11,372	44,870
Actors who made at least one post, n (%)	449 (17.7%)	1230 (23.9%)	2767 (24.3%)	7963 (17.7%)

Each program has been extensively studied in the literature [20-32], and program features and functionality have been described elsewhere [18]. The major theoretical underpinnings used to develop the interventions are described in Table 2 [33-42].

Table 2. Theoretical underpinnings of behavior-change programs.

	Problem drinking	Depression	Panic disorder	Smoking cessation
Brief Intervention [33]	✔	✔	✔	✔
Cognitive Behavioral Therapy [34]		✔	✔	
Gamification [35]	✔	✔	✔	✔
Health Belief Model [36]	✔	✔	✔	✔
Motivational Interviewing [37]	✔	✔	✔	✔
Social Cognitive Theory [38,39]	✔	✔	✔	✔
Structured Relapse Prevention [40]	✔			✔
Targeting and Tailoring [41]	✔	✔	✔	✔
Transtheoretical Model / Stages of Change [42]	✔			✔

The four DHSNs are funded and managed by Evolution Health System Inc (EHS) and are part of the firm's social business model. EHS is a private, research-based organization that builds evidence-based digital programs designed to increase medication and treatment adherence.

All data collection procedures adhered to international privacy guidelines [43-45] and were in accordance with the Helsinki Declaration of 1975, as revised in 2008 [46]. The study was consistent with the University Research Ethics Committee procedures at Henley Business School, University of Reading, and was exempt from full review.

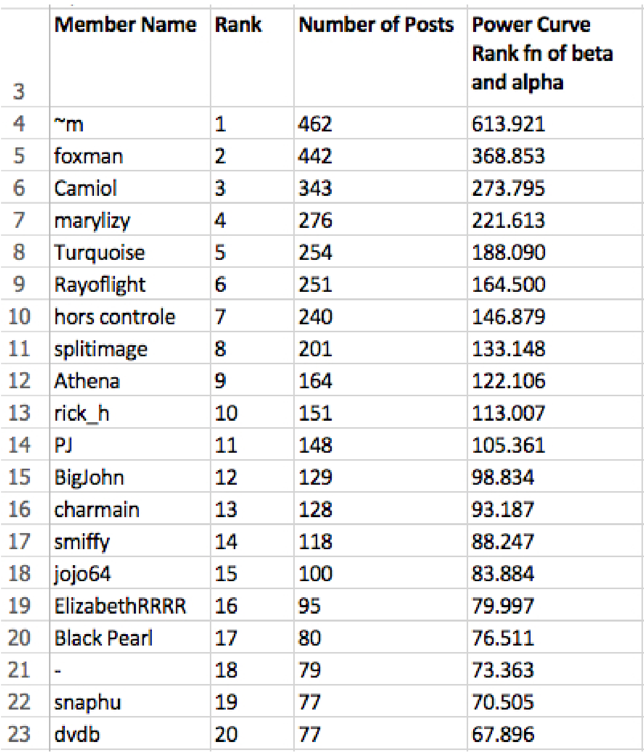
A 5-step process was undertaken to compare actor contributory patterns of the four DHSNs to power curves, as follows.

Step One

Data on all actors who posted one or more posts were imported from each DHSN's structured query language (SQL) server database to Microsoft Excel. Actors were then ranked, with the actor creating the greatest number of posts assuming the first position, the actor creating the second greatest number of posts in the second position, and so on.

Figure 2 illustrates this ranking process with actors from the AHC DHSN. The actor *~m* created the greatest number of posts in the network (n=462), assuming the rank of one. This is followed by the actor *foxman* who assumed the rank of two (n=442), and the actor *Camiol*, who assumed the third rank (n=343).

Figure 2. Ranking of top 20 actors contributing to Alcohol Help Center.

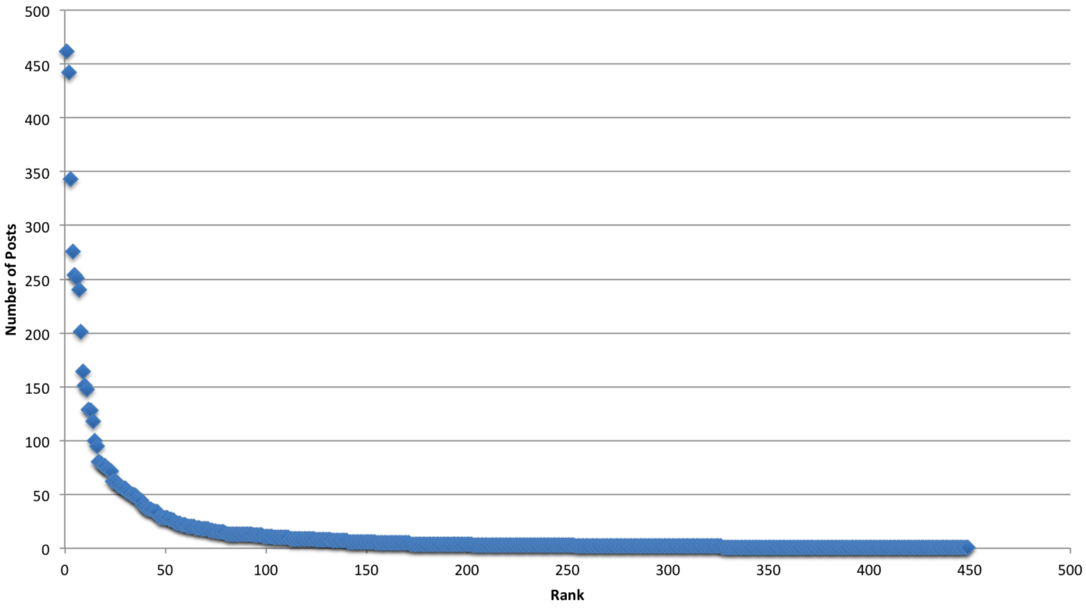


3	Member Name	Rank	Number of Posts	Power Curve Rank fn of beta and alpha
4	~m	1	462	613.921
5	foxman	2	442	368.853
6	Camiol	3	343	273.795
7	marylizy	4	276	221.613
8	Turquoise	5	254	188.090
9	Rayoflight	6	251	164.500
10	hors controle	7	240	146.879
11	splitimage	8	201	133.148
12	Athena	9	164	122.106
13	rick_h	10	151	113.007
14	PJ	11	148	105.361
15	BigJohn	12	129	98.834
16	charmain	13	128	93.187
17	smiffy	14	118	88.247
18	jojo64	15	100	83.884
19	ElizabethRRRR	16	95	79.997
20	Black Pearl	17	80	76.511
21	-	18	79	73.363
22	snaphu	19	77	70.505
23	dvdb	20	77	67.896

Step Two

Power distributions in ranked data have skewed distributions [14]. To visually determine the skewness, or asymmetry of the DHSNs rank distribution, X-Y scatter plots were created in Microsoft Excel (Figure 3).

Figure 3. Cumulative posting trends in Alcohol Help Center.



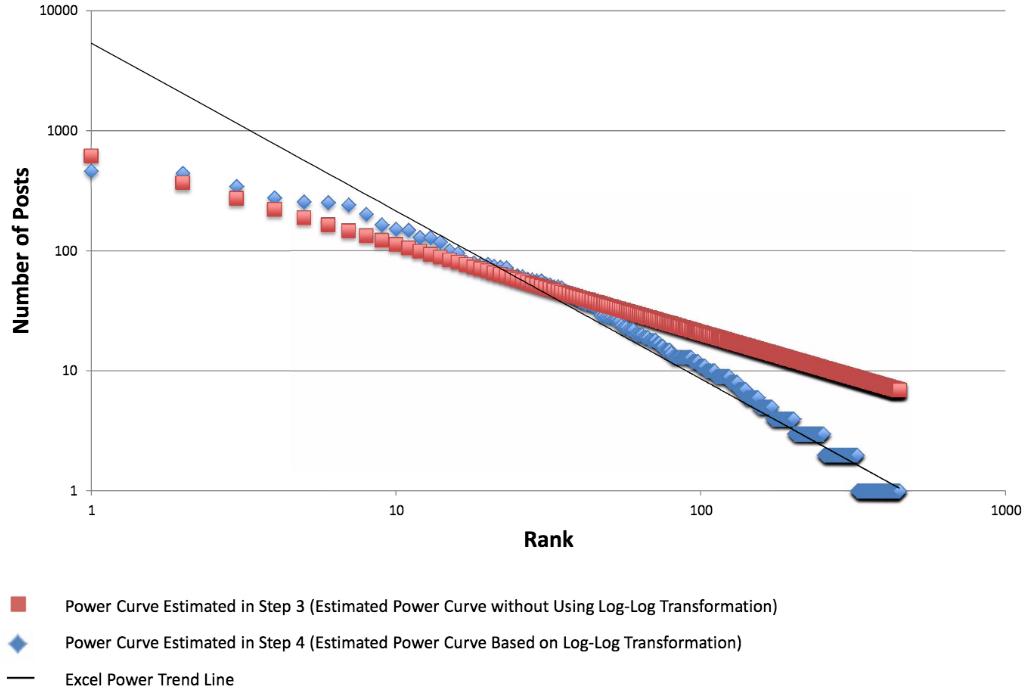
Step Three

In Excel, each actor's ranking was mapped to an equal position on a power curve with a slope of beta defined as y=10^(alpha+beta*Log^{10}x) (see Figure 2, Column D). We estimated alpha and beta for each of the networks by minimizing the sum of squared residuals based on the observed y and the predicted y.

Step Four

To visually compare each DHSN posts with its corresponding power curve rank function of alpha and beta, X-Y scatter plots were generated in Microsoft Excel, with both axes transformed to logarithmic scales. For DHSN posts, an Excel power trend line was added with the R^2 option selected (Figure 4). In this Excel built-in option, Excel applies ordinary least squares (OLS) to estimate Log(y)=alpha + beta*Log(x) + ε. In other words, Excel estimates alpha and beta by minimizing the sum of square residuals based on the observed Log(y) and Log(predicted y). However, in Step 3, we use y and predicted y to compute sum of squared residuals. This is why the predicted line generated by Excel power trend line option in Step 4 differs from the one we generate in Step 3.

Figure 4. Alcohol Help Center actor ranking and power curve raking with trendline and R2 value.



Step Five

In SPSS, Spearman correlations were used to compare DHSN posts to the power curve rank function of alpha and beta (Columns C and D in Figure 2). Spearman correlation was employed as the comparison method because it is commonly used in non-linear distributed data and does not make assumptions about the frequency distribution of variables [47].

Results

All four DHSNs were found to have right skewed distributions, indicating that the data were not normally distributed. This also confirmed that a small number of actors created the vast majority of content (Figure 5).

When logged, each of the DHSN's rank and post frequency data closely resembled power distributions. When Excel power trend lines were added, R^2 values indicated that to a very high degree, the variance in post frequencies is explained by actor rank (Figure 6).

To assess strength of the linear relationship between actor rank and number of contributions, and power curve rank, Spearmen correlations were calculated (Table 3).

Table 3. Comparison of log-log scatter plots to power curves.

Social network	R^2	Spearman correlation (sig)
Problem drinking	.96207	.987 (*P*<.001)
Anxiety	.96875	.972 (*P*<.001)
Depression	.97508	.967 (*P*<.001)
Smoking cessation	.94979	.993 (*P*<.001)

Figure 5. Right skewed distributions in four DHSNs.

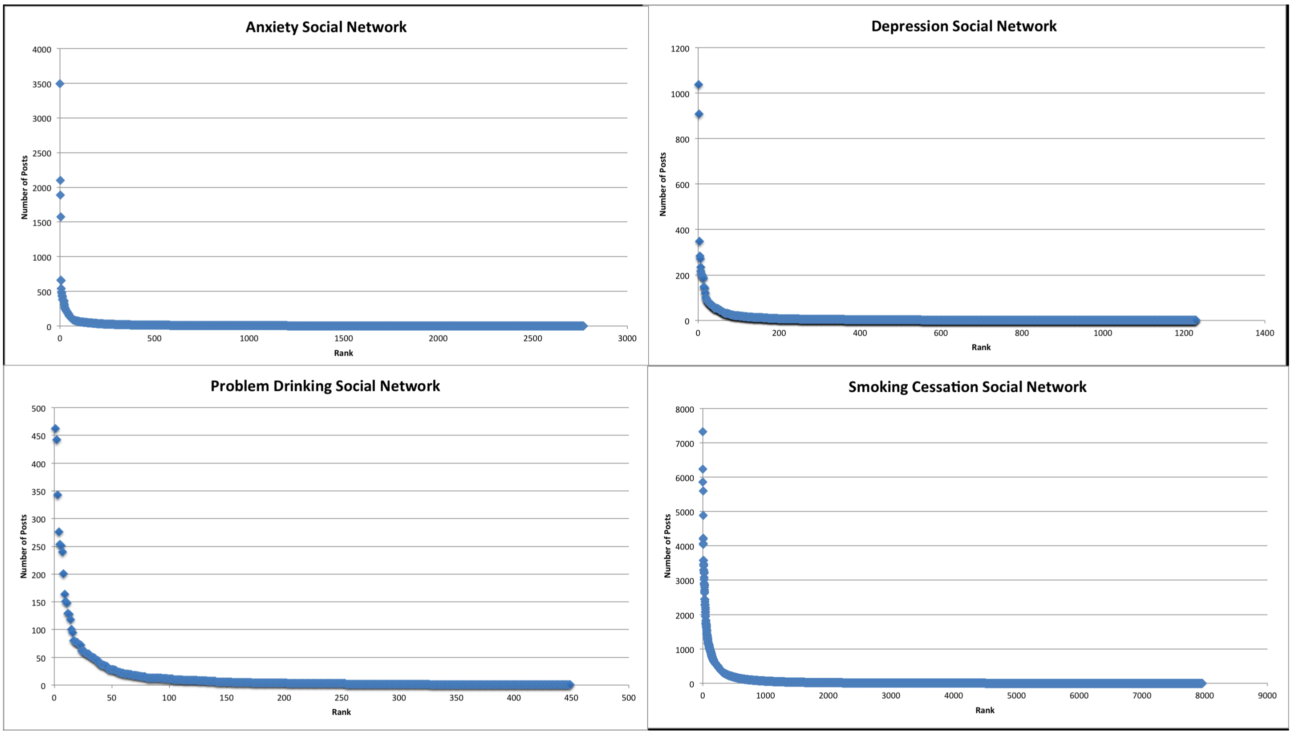
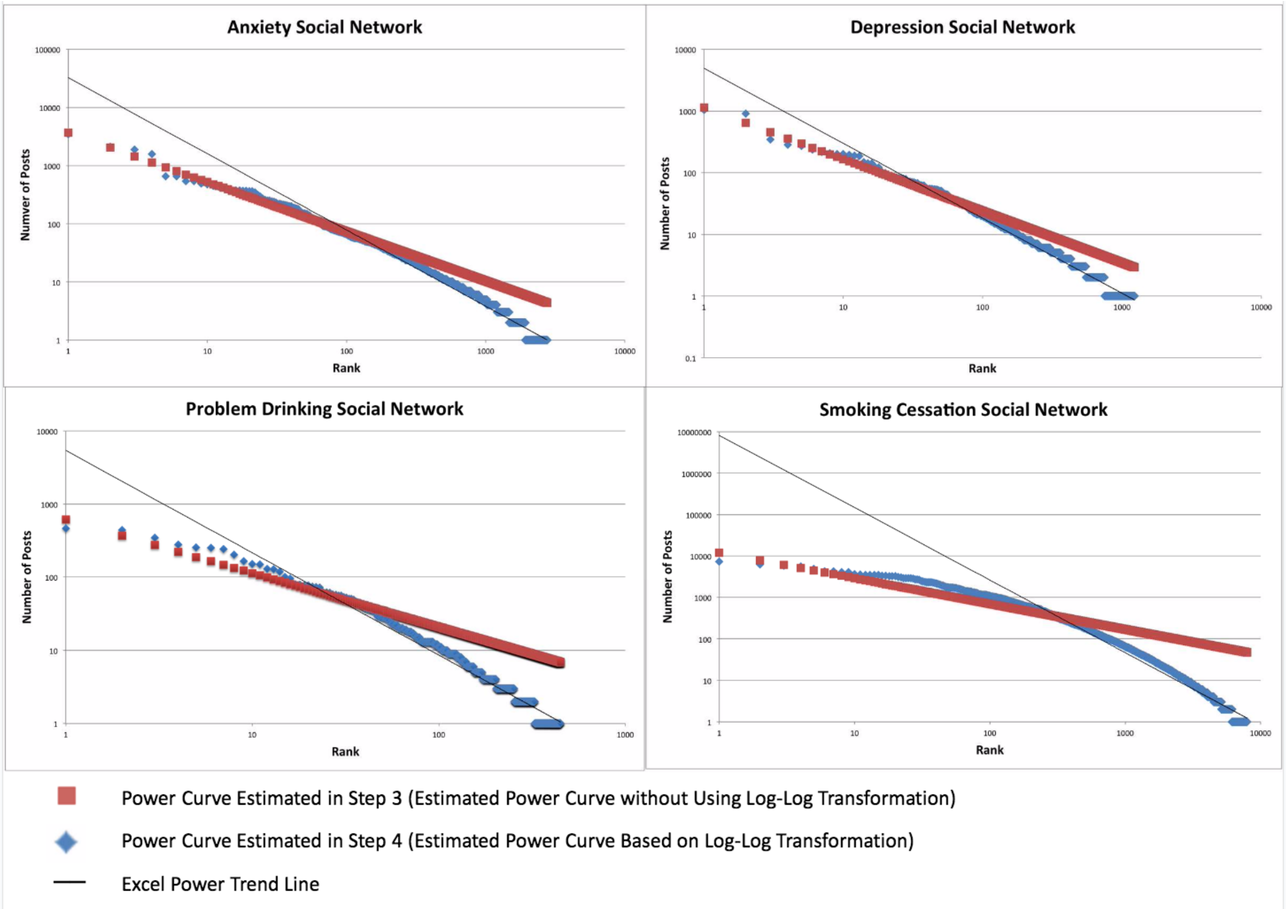


Figure 6. DHSN actor ranking and power curve ranking with trendline and R2 value.



Discussion

Principal Findings

The four DHSNs analyzed in this experiment differed in several areas. First, they addressed unique topics (two mental health, two addictions). Second, all four groups existed for different periods of time (minimum=4.0 years, maximum=10.9), had varying actor populations (minimum=449, maximum=7963), and total number of posts (minimum=7148, maximum=513,586).

Despite differences in condition addressed, program duration and data collection period, populations, and number of post sizes, results indicate that to a high degree, the distribution patterns of DHSNs resemble that of a power law. As power laws demonstrate correlated phenomena, they can help describe the topology of DHSNs.

Practical Implications

The results of this study have several practical implications for DHSN owners and managers. Unlike the earlier FaceTime example where actor conversations are ephemeral, all DHSN posts remain on the network, and each additional post increases network value. By following the 5-step process outlined in this manuscript, managers can map the growth of their networks and graphically isolate specific types of actors.

As positive network externalities occur with the addition of each post, it is imperative for managers to develop methods designed to retain actors who frequently create content. Past research has identified these actors as superusers [48,49]; however, very little is known about superuser demographic or psychographic characteristics.

Many DHSNs are managed by trained moderators. In health care settings, moderators are often required to read and approve posts, answer usability questions, and manage disputes. Support group moderation is a relatively new but growing profession [50], and to date there are few best practices designed to estimate labor costs. As staffing is often dependent on network size, observing power law distributions may help managers establish budgets and expenditures, such as employee recruitment or training.

Strengths and Limitations

A strength of this study is the use of four separate DHSNs with varying topics, population sizes, and periods of existence. A second strength is that the programs are not actively advertised or promoted, and there is no cost to join or participate. This has resulted in a dataset that contains a naturalistic, self-seeking population with limited participatory barriers.

However, this same strength may also be a weakness. Typically, networks have barriers to entry such as registration fees or membership requirements. A further weakness is that all four networks are managed and maintained by the same organization, and the information architecture of the programs is similar.

Another factor to be considered is that other phenomenon may be better suited to explain network patterns. For example, many smokers make an attempt to quit at the start of a new year [51], and seasonality may be better suited to explain both short- and long-term DHSN growth.

Also deserving of consideration is that the definition of network value in this manuscript is derived from the economics literature, where the addition of each post creates a positive network externality. In practice, all posts are not of equal value. Certain posts will be frequently visited and commented on more than others, and the value of these posts are arguably greater than posts that are less popular. Trained moderators also viewed, approved, and in some cases edited all posts in this study. Some posts were also deleted due to inappropriate content. In this context, future research may refine the definition of network value.

Finally, the efficaciousness of DHSNs has yet to be firmly established in the literature. Research continues to focus on possible relationships between social network use and increased treatment adherence and measurable health outcomes.

Future Research

To further validate our results, the method used in this study should be replicated in networks that are managed by other organizations, and it would be helpful to focus on a variety of conditions.

Future research should also analyze network growth over time through analyzing longitudinal or panel data. The 5-step method outlined in this paper could be applied to an investigation observing the strength and consistency of power distributions throughout the life span of a single DHSN.

The results of this study indicate that superusers may be important for network growth. Future research should investigate the direction of the causal relationship between superusers and network size. Future research should also seek to gain a better understanding of superuser characteristics, demographics, psychographics, and their survival rates.

Due to the availability of big data, other disciplines are now investigating the importance of the small number of consumers who account for a large percentage of profits [52,53], and health care should follow suit. Leveraging the expertise, wisdom, and experience of patients who are dedicated to sharing their knowledge and experience could possibly translate to increased treatment adherence and efficacy.

Conclusions

This is the first study to investigate power curves across multiple DHSN. To a high degree, the rank and post frequencies of the four DHSNs hold properties of power laws. The implications of the results are important as they give insight to both researchers and managers into the nature and inner mechanisms of DHSNs. Future research examining the characteristics, survival rates, and role of superusers is required.

Acknowledgments

The authors would like to thank Dr Harry Krashinksy, Associate Professor of Industrial Relations at the University of Toronto, for his assistance in the study design.

Conflicts of Interest

Author TvM is the CEO & Founder of Evolution Health Systems Inc, the owner of Alchol Help Center, Depression Center, Panic Centre, and Stop Smoking Center as well as other eHealth and mHealth platforms.

References

1. Schneider SJ. Trial of an on-line behavioral smoking cessation program. Computers in Human Behavior 1986 Jan;2(4):277-286. [doi: 10.1016/0747-5632(86)90008-7]
2. Robinson T, Walters PA. Health-net: an interactive computer network for campus health promotion. J Am Coll Health 1986;34(6):284-285. [doi: 10.1080/07448481.1986.9938952]
3. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013 May;21(5):1253-1263. [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
4. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Research Center's Internet & American Life Project; 2013. URL: http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf [accessed 2015-06-22] [WebCite Cache ID 6ZTfj6tNF]
5. 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]
6. Albert R, Barabási A. Statistical mechanics of complex networks. Rev Mod Phys 2002 Jan 30;74(1):47-97. [doi: 10.1103/RevModPhys.74.47]
7. Grabher G, Stark D. Organizing Diversity: Evolutionary Theory, Network Analysis and Postsocialism. Regional Studies 1997 Jul;31(5):533-544. [doi: 10.1080/00343409750132315]
8. Katz M, Shapiro C. Network Externalities, Competition, and Compatibility. American Economic Review 1985;75(3):424-440.
9. Leibenstein H. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. The Quarterly Journal of Economics 1950;64(2):183-207.

10. Uzzi B. The Sources and consequences of embeddedness for the economic performance of organizations: The network effect. American Sociological Review 1996;61(4):674-698.
11. Elmer-DeWitt P. FaceTime's network effect.: Fortune; 2010 Oct 21. URL: http://fortune.com/2010/10/21/facetimes-network-effects/ [accessed 2015-06-22] [WebCite Cache ID 6ZTgnmvfJ]
12. Sanders R. The Pareto Principle: its Use and Abuse. Jnl of Product & Brand Mgt 1992 Jun;1(2):37-40. [doi: 10.1108/10610429210036762]
13. The Harvard Crimson. 1950 Sep 27. Zipf Dies After 3-Month Illness URL: http://www.thecrimson.com/article/1950/9/27/zipf-dies-after-3-month/ [accessed 2015-06-22] [WebCite Cache ID 6ZTh4209q]
14. Newman M. Power laws, Pareto distributions and Zipf's law. Contemporary Physics 2005 Sep;46(5):323-351. [doi: 10.1080/00107510500052444]
15. Andriani P, McKelvey B. Beyond Gaussian averages: redirecting international business and management research toward extreme events and power laws. J Int Bus Stud 2007 Oct 18;38(7):1212-1230. [doi: 10.1057/palgrave.jibs.8400324]
16. Clauset A, Shalizi CR, Newman M. Power-Law Distributions in Empirical Data. SIAM Rev 2009 Nov 04;51(4):661-703. [doi: 10.1137/070710111]
17. Krugman P. Confronting the mystery of urban hierarchy. Journal of the Japanese and International Economies 1996 Dec;10(4):399-418. [doi: 10.1006/jjie.1996.0023]
18. van Mierlo T. The 1% Rule in Four Digital Health Social Networks: An Observational Study. J Med Internet Res 2014;16(1):e33. [doi: 10.2196/jmir.2966]
19. Carron-Arthur B, Cunningham JA, Griffiths KM. Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. Internet Interventions 2014 Oct;1(4):165-168. [doi: 10.1016/j.invent.2014.09.003]
20. Cunningham JA. Comparison of two internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]
21. Cunningham JA, Humphreys K, Kypri K, van Mierlo T. Formative evaluation and three-month follow-up of an online personalized assessment feedback intervention for problem drinkers. J Med Internet Res 2006;8(2):e5 [FREE Full text] [doi: 10.2196/jmir.8.2.e5] [Medline: 16867968]
22. Cunningham JA, Selby P, van Mierlo T. Integrated online services for smokers and drinkers? Use of the check your drinking assessment screener by participants of the Stop Smoking Center. Nicotine Tob Res 2006 Dec;8 Suppl 1:S21-S25. [Medline: 17491167]
23. Cunningham JA, van Mierlo T. Methodological issues in the evaluation of Internet-based interventions for problem drinking. Drug Alcohol Rev 2009 Jan;28(1):12-17. [doi: 10.1111/j.1465-3362.2008.00001.x] [Medline: 19320670]
24. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
25. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. Twelve-month follow-up results from a randomized controlled trial of a brief personalized feedback intervention for problem drinkers. Alcohol Alcohol 2010;45(3):258-262 [FREE Full text] [doi: 10.1093/alcalc/agq009] [Medline: 20150170]
26. Davis J. The Panic Center. Child Adolesc Ment Health 2007 Feb;12(1):49-50. [doi: 10.1111/j.1475-3588.2007.00437_3.x]
27. Farvolden P, Cunningham J, Selby P. Using E-Health Programs to Overcome Barriers to the Effective Treatment of Mental Health and Addiction Problems. Journal of Technology in Human Services 2009 Feb 03;27(1):5-22. [doi: 10.1080/15228830802458889]
28. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
29. Farvolden P, McBride C, Bagby RM, Ravitz P. A Web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res 2003;5(3):e23. [doi: 10.2196/jmir.5.3.e23]
30. McDonnell DD, Kazinets G, Lee H, Moskowitz JM. An internet-based smoking cessation program for Korean Americans: results from a randomized controlled trial. Nicotine Tob Res 2011 May;13(5):336-343. [doi: 10.1093/ntr/ntq260] [Medline: 21330285]
31. Pike KJ, Rabius V, McAlister A, Geiger A. American Cancer Society's QuitLink: randomized trial of Internet assistance. Nicotine Tob Res 2007 Mar;9(3):415-420. [doi: 10.1080/14622200701188877] [Medline: 17365773]
32. Rabius V, Pike KJ, Wiatrek D, McAlister AL. Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 2008;10(5):e45 [FREE Full text] [doi: 10.2196/jmir.1008] [Medline: 19033146]
33. O'Donnell A, Wallace P, Kaner E. From efficacy to effectiveness and beyond: what next for brief interventions in primary care? Front Psychiatry 2014;5:113. [doi: 10.3389/fpsyt.2014.00113]
34. Herbert JD, Forman EM. The Evolution of Cognitive Behavioral Therapy: the rise of psychological acceptance and mindfullness. In: Acceptance and Mindfulness in Cognitive Behavior Therapy: Understanding and applying the new therapies. New York: Wiley; 2011:1-26.
35. Deterding S. Gamification. interactions 2012 Jul 01;19(4):14-17. [doi: 10.1145/2212877.2212883]

36. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]
37. Miller W, Rollnick S. Motivational Interviewing: Preparing People to Change. New York: The Guilford Press; 2002.
38. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
39. McAlister L, Perry CL, Parcel GS. How Individuals, Environments, and Health Behaviors Interact. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Reserach and Practice. San Francisco: Jossey-Bass; 2008:169-188.
40. Sanchez-Craig M. Drink Wise: How to quit drinking or cut down. Toronto: Addiction Research Foundation; 1995.
41. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000;15(1):1-4. [Medline: 10788196]
42. Prochaska JO, DiClemente CC. Transtheoretical therapy: Toward a more integrative model of change. Psychotherapy: Theory, Research & Practice 1982;19(3):276-288. [doi: 10.1037/h0088437]
43. Directive/58/EC of the European Parliament and of the Council of 12 July Concerning the Processing of Personal Data and the Protection of in the Electronic Communications Sector (Directive on Privacy and Electronic Communications).: European Parliament and the Council of the European Union; 2002. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML [accessed 2015-06-22] [WebCite Cache ID 6ZTkmkoeC]
44. Personal Information Protection and Electronic Documents Act.: Office of the Privacy Commissioner of Canada; 2000. URL: http://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html [accessed 2015-06-22] [WebCite Cache ID 6ZTlA2ntM]
45. Summary of the HIPAA Privacy Rule.: US Department of Health & Human Services; 2003. URL: http://www.hhs.gov/ocr/privacy/ [accessed 2015-06-22] [WebCite Cache ID 6ZTlHrMvu]
46. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.: World Medical Association; 2005. URL: http://www.wma.net/en/30publications/10policies/b3/17c.pdf [accessed 2015-06-22] [WebCite Cache ID 6ZTlRwibQ]
47. Hauke J, Kossowski T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quaestiones Geographicae 2011;30(2):87-93. [doi: 10.2478/v10117-011-0021-1]
48. O'Neill B, Ziebland S, Valderas J, Lupianez-Villanueva F. User-generated online health content: a survey of Internet users in the United Kingdom. J Med Internet Res 2014;16(4):e118. [doi: 10.2196/jmir.3187]
49. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
50. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013;15(6):e119. [doi: 10.2196/jmir.2501]
51. Walsh RA, Paul CL. New Year quit smoking resolutions. Aust N Z J Public Health 2002 Apr;26(2):181. [Medline: 12054341]
52. Yoon E. Tap into your Super-Consumers.: Harvard Business Review; 2009 Nov 25. URL: https://hbr.org/2009/11/surprising-insights-from-super [accessed 2015-06-22] [WebCite Cache ID 6ZTmaLkTC]
53. Avery J, Norton M. Pay Attention to Your 'Extreme Customers'.: Harvard Business School; 2014 Jul 14. URL: http://hbswk.hbs.edu/item/7570.html [accessed 2015-06-22] [WebCite Cache ID 6ZTmtrJca]

Abbreviations

AHC: Alcohol Help Center
DC: Depression Center
DHSN: digital health social networks
EIES: Electronic Information Exchange System
EHS: Evolution Health Systems Inc
PC: Panic Center
SCC: Stop Smoking Center
VoIP: Voice over Internet protocol

Edited by G Eysenbach; submitted 28.01.15; peer-reviewed by B Carron-Arthur, B O'Neill, S McIntosh, D de Beurs, M Chary; comments to author 03.03.15; revised version received 10.04.15; accepted 24.05.15; published 25.06.15.

Please cite as:
van Mierlo T, Hyatt D, Ching AT
Mapping Power Law Distributions in Digital Health Social Networks: Methods, Interpretations, and Practical Implications
J Med Internet Res 2015;17(6):e160
URL: http://www.jmir.org/2015/6/e160/
doi:10.2196/jmir.4297
PMID:26111790

Original Paper

Automated Detection of HONcode Website Conformity Compared to Manual Detection: An Evaluation

Célia Boyer[1*], MSc; Ljiljana Dolamic[1*], PhD

Health On the Net Foundation, Chêne-Bourg, Switzerland

[*]all authors contributed equally

Corresponding Author:
Célia Boyer, MSc
Health On the Net Foundation
2 Chemin du Petit Bel-Air
Chêne-Bourg, 1225
Switzerland
Phone: 41 793092960
Fax: 41 3728885
Email: Celia.Boyer@HealthOnNet.org

Abstract

Background: To earn HONcode certification, a website must conform to the 8 principles of the HONcode of Conduct In the current manual process of certification, a HONcode expert assesses the candidate website using precise guidelines for each principle. In the scope of the European project KHRESMOI, the Health on the Net (HON) Foundation has developed an automated system to assist in detecting a website's HONcode conformity. Automated assistance in conducting HONcode reviews can expedite the current time-consuming tasks of HONcode certification and ongoing surveillance. Additionally, an automated tool used as a plugin to a general search engine might help to detect health websites that respect HONcode principles but have not yet been certified.

Objective: The goal of this study was to determine whether the automated system is capable of performing as good as human experts for the task of identifying HONcode principles on health websites.

Methods: Using manual evaluation by HONcode senior experts as a baseline, this study compared the capability of the automated HONcode detection system to that of the HONcode senior experts. A set of 27 health-related websites were manually assessed for compliance to each of the 8 HONcode principles by senior HONcode experts. The same set of websites were processed by the automated system for HONcode compliance detection based on supervised machine learning. The results obtained by these two methods were then compared.

Results: For the privacy criterion, the automated system obtained the same results as the human expert for 17 of 27 sites (14 true positives and 3 true negatives) without noise (0 false positives). The remaining 10 false negative instances for the privacy criterion represented tolerable behavior because it is important that all automatically detected principle conformities are accurate (ie, specificity [100%] is preferred over sensitivity [58%] for the privacy criterion). In addition, the automated system had precision of at least 75%, with a recall of more than 50% for contact details (100% precision, 69% recall), authority (85% precision, 52% recall), and reference (75% precision, 56% recall). The results also revealed issues for some criteria such as date. Changing the "document" definition (ie, using the sentence instead of whole document as a unit of classification) within the automated system resolved some but not all of them.

Conclusions: Study results indicate concordance between automated and expert manual compliance detection for authority, privacy, reference, and contact details. Results also indicate that using the same general parameters for automated detection of each criterion produces suboptimal results. Future work to configure optimal system parameters for each HONcode principle would improve results. The potential utility of integrating automated detection of HONcode conformity into future search engines is also discussed.

(J Med Internet Res 2015;17(6):e135) doi:10.2196/jmir.3831

KEYWORDS

HONcode; classification; artificial intelligence; natural language processing; quality standards

Introduction

The Internet has brought about immense change in the way individuals obtain and access health information [1]. It transformed health information distribution from occurring only in the doctor's office during patient visits (top-down information flow) to a multilateral, asynchronous form of communication. Patients feel empowered to gather and share their own information and to make more informed decisions regarding their own health care [2,3].

A recent study showed that 35% of US adults had used the Internet at one time or another to gather health information about a medical condition that they or someone else had [4]. Of these Internet users, 46% had also sought the advice of a health professional. Conversely, 38% of persons accessing the Internet for health information stated that they managed the health condition at home. Given that more than 30% of US adults have made important health care decisions after accessing the Internet, the quality of Internet-based health information becomes crucial. Another recent study shows that, not less than 60% of Europeans go online when looking for health information [5]. Six out of 10 (60%) Europeans who have found health-related information online thought the information came from a trustworthy source although it remains unclear what they deemed as trustworthy [6].

However, taking into account the quantity of the health-related information available on the Internet in the form of health-related websites or scientific articles, users are often overwhelmed with the quantity of the information available. Recently, efforts have been taken to automatically label online health pages according to the information quality provided on them [7,8]. These research studies remain connected to a certain health domain and to quality criteria defined by study authors. Studies indicate that the quality of the health information found on the Internet is extremely variable [9,10]. Readers have exceeding difficulty in discerning trustworthy from nontrustworthy website content. One approach to this dilemma is to annotate websites that willingly comply to content quality with easily visible badges or icons. This is the approach taken by the Health on the Net (HON) Foundation in HONcode certification [11]. The HONcode is a code of conduct consisting of 8 procedural principles (ie, authority, complementarity, privacy, attribution, justification, contact details, financial disclosure, and advertising policy) that a health website must follow to gain certification [12]. The goal of this process is to create a pool of quality health information available to the general public [13,14]. The HONcode helps the Web user to judge if she/he can trust the information found on the Internet [15,16]. However, because obtaining HONcode certification requires a website manager to voluntarily submit a request for HON review, the scope of existing HONcode certification remains limited.

Search engines represent the source most frequently used. In one survey, 77% of online health advice seekers began their last session at a generalized search engine such as Google, Bing, or Yahoo [17]. A recent European study shows that between 82% and 87% of those who searched for health-related information online used search engines to do so [6]. These search engines typically list results according to popularity rather than quality or trustworthiness. Thus, the first few options they display may not be the best sources of health information. People become confused and anxious after accessing inappropriate health information [18]. Ideally, search engine developers would modify the search engine to promote the most reliable and validated sources of health information. Within the European project KHRESMOI (2010-2014, project No. 2575284), researchers have recently developed tools to automatically assess how well a given website complies with the HONcode principles. Complementing the authors' and our colleagues' work in developing the algorithm [19,20], this study presents an evaluation comparing automated detection of HONcode principle compliance with expert assessments for 30 health websites.

Methods

Overview

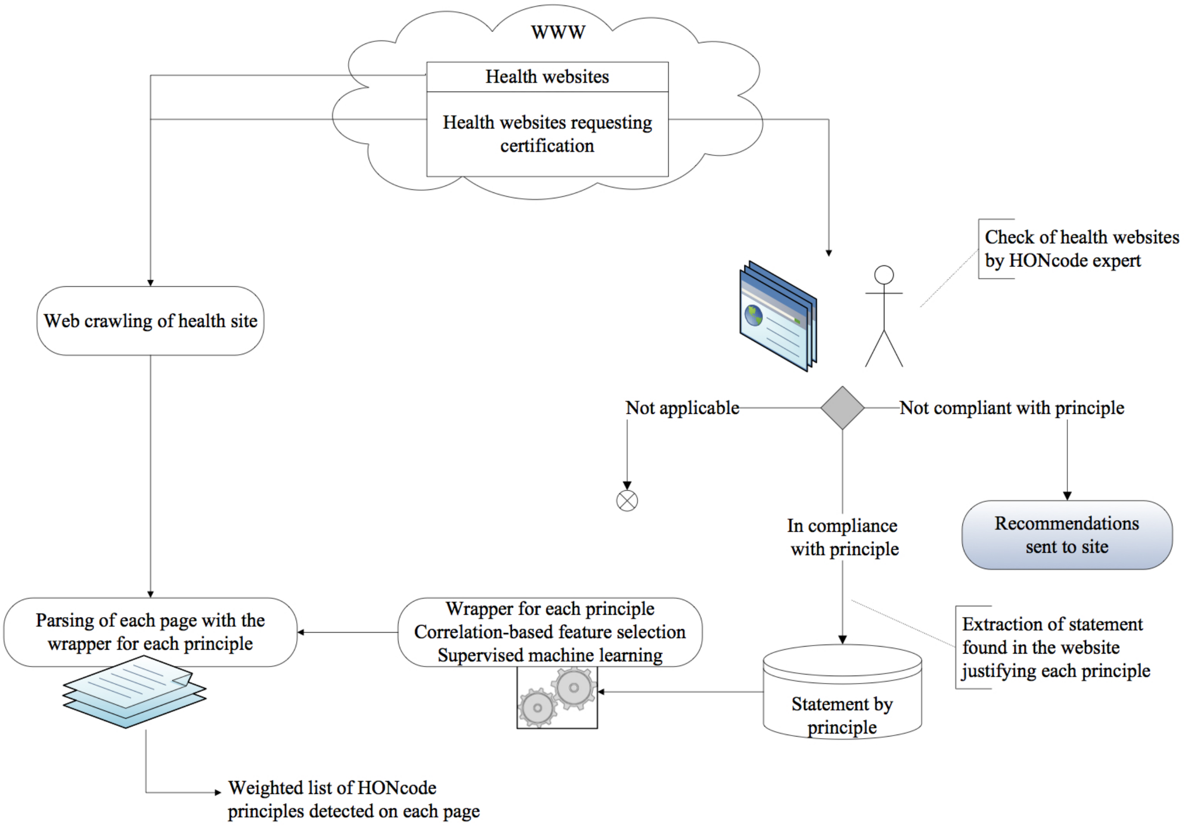
In this study, the authors compared the results of the automated system detection for HONcode principles for a selection of 30 health websites to the ones obtained during the standard manual HONcode process conducted by senior HONcode experts (eg, an expert with more than 10 years' experience in HONcode certification). The senior HONcode expert has a medical background; he/she is responsible for training of new HONcode reviewers and deals with complex certification cases.

HONcode Certification Process

Once a site has requested HONcode certification, the expert navigates the pages of the site to identify if the site respects each of the HONcode principles [12].

When principle justification is found on a page (ie, the site conforms to the given principle), the extract and the Web address are added to the HONcode file and stored in a database. When a principle is not respected either totally or partially, recommendations are sent to the site editor. The manual HONcode certification is described in Figure 1.

Figure 1. HONcode manual and automated detection processes.



HONcode Interreviewer Agreement

With the goal of measuring the level of agreement between expert reviewers and estimating the likelihood of an expert giving a false assessment, we compared the assessments done by 3 senior reviewers for a total of 36 websites. Each criterion was rated by all 3 reviewers.

Description of Automated System for HONcode Detection

Automated detection of HONcode principle compliance consisted of the following steps as illustrated in Figure 1:

1. For a given health-related website, a WebCrawler retrieved a maximal set of its accessible Web pages. This proceeded from the website home page and followed the internal links.
2. The system extracted “meaningful content” from the retrieved Web pages within a given website. This content consisted of textual information within the pages.
3. The content extracted from each Web page was then checked by the automated system for compliance with each HONcode principle. The automated system embodied the machine learning framework described in Williams and Calvo [21]. Separate classifiers were built for each of the HON criteria. The classifiers reviewed the Web page material independently because when a document indicated compliance with 1 HONcode principle, it did not exclude the possibility that the document complied with other HON principles (“any-of” classification) [22]. The process of automated HONcode detection was designed in this way to mimic the current manual certification process. However, the automated system systemically checks all the sites’ webpages retrieved unlike the manual system that stops once criterion compliance is detected. The extracts justifying principle compliance collected during HONcode certification formed the training set for the aforementioned classifiers. HONcode certification is multilingual; 34% of certified websites are in English, 28% in French, 10% in Spanish, and 7% in German. However, this study was limited to the English language only. The number of training documents varied from 872 for the criteria “justifiability” classifier to 2861 for the “contact details” classifier. The general classifier system enabled the user to select from different machine learning algorithms, such as naive Bayes, support vector machine (SVM), and others through various parameter settings [23]. The system also enabled choice of different feature types, such as bag-of-words, bag-of-stems, co-occurrence, etc. Additionally, the system implemented a user-configurable variety of feature selection algorithms (term weighting schemata). In this study, the authors specified use of the naive Bayes algorithm for each of the 8 HONcode principles. The algorithm as implemented checked the page content according to 9 different criteria because 1 of the 8 individual HONcode principles (“attribution”) was divided into 2 parts, “references” and “date,” for this study based on previously validated reasons [19].

To specify conversion of document word counts into vector values, the authors used 2 weighting schemes, namely tfc and tfx, in which *t*, *f*, *c*, and *x* represent document frequency, inverse document frequency, cosine normalization, and none, respectively [24]. The document frequency (t) represents the number of occurrences of the given term within the document being classified. The inverse document frequency (f) is calculated as f=log(N/D), where N is total number of documents within the collection and D represents the number of documents in the collection that contain the given term. Thus, a higher importance is given to a term found in a smaller number of documents within the collection, supposing that the more the documents the term is found in, the less important it is. The final variables indicate whether cosine normalization occurs (c) versus none (x). This parameter gives more importance to the term occurrences within shorter documents. Thus, the tfc conversion additionally normalizes the score by the document length (c).

Automated System Detection Results Compared to the Manual Evaluation Results

The authors selected a convenience sample of 30 health care websites for the comparative evaluation (automated detection vs manual rating by a senior HONcode expert). However, only 27 of 30 websites could be processed by the automated system, so study results used the sample of 27 sites. The convenience sample was selected to broadly cover HONcode potential and actual sites as follows:

1. New potentially certifiable websites (n=9): the HONcode experts estimated that these websites did conform to HONcode, but they had not yet been certified.
2. Likely noncertifiable websites (n=9): the HONcode experts estimated that these websites would not conform to HONcode principles when fully analyzed.
3. Newly certified websites (n=4): these websites had been recently certified for the first time.
4. Previously certified HONcode sites (n=5): these websites were chosen because they were awaiting annual reassessment.

For the purpose of the evaluation, the senior HONcode expert manually reviewed each of the 27 websites described. Simultaneously, the automated system for HONcode detection reviewed the 27 websites for each evaluation criterion [19]. The results obtained by the automated system were then compared to the baseline obtained by the expert. Figure 2 shows the evaluation methods.

Figure 3 gives a sample page conforming to the “complementarity” criterion. On this page, the information the expert was looking for in the process of manual evaluation is marked in yellow. Additionally, the terms that the automated system identified as important for this criterion are boxed in different colors depending on their level of importance (eg, red=most important, green=least important).

Figure 2. Comparison of the automated HONcode detection evaluation to manual evaluation.

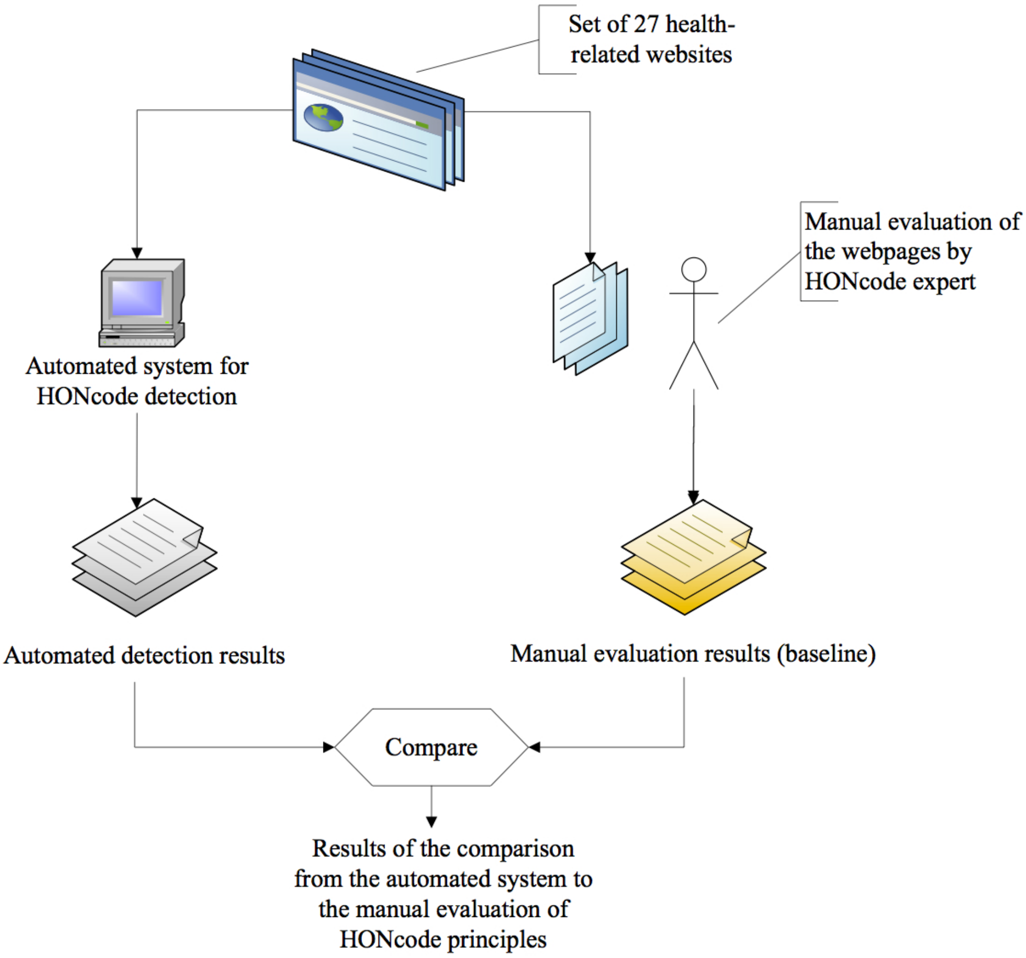
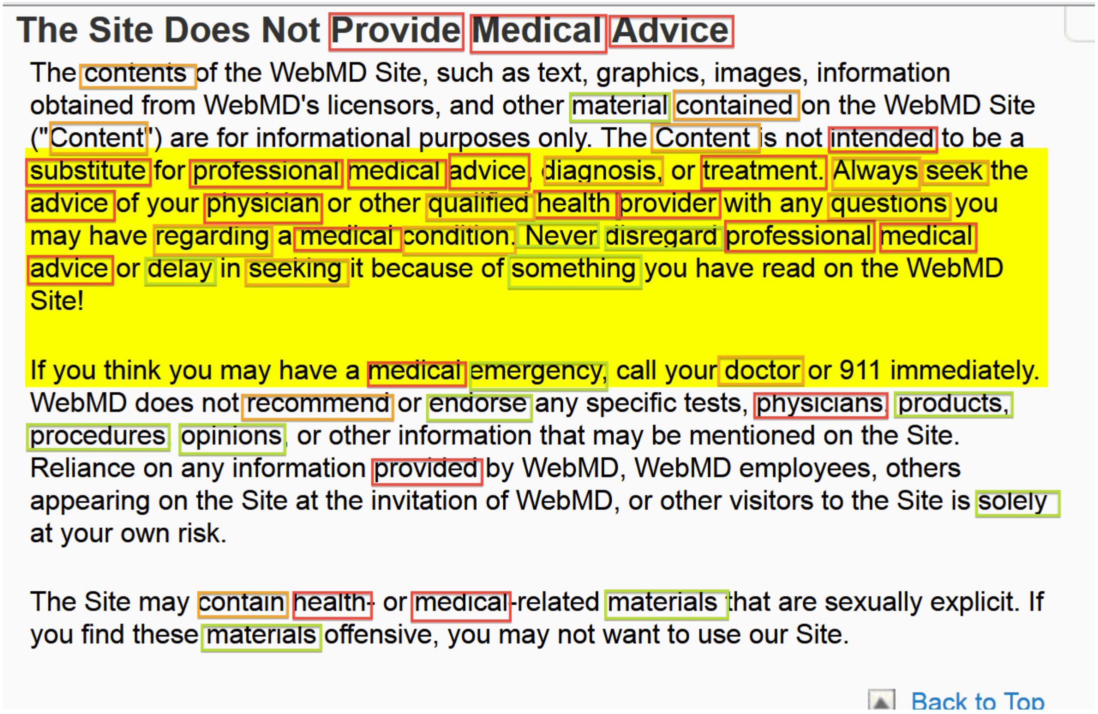


Figure 3. Assessment of “complementarity” criterion with terms detected by the expert (highlighted in yellow) and the automated system (colored boxes with red=most important and green=least important).



The Site Does Not Provide Medical Advice

The contents of the WebMD Site, such as text, graphics, images, information obtained from WebMD's licensors, and other material contained on the WebMD Site ("Content") are for informational purposes only. The Content is not intended to be a substitute for professional medical advice, diagnosis, or treatment. Always seek the advice of your physician or other qualified health provider with any questions you may have regarding a medical condition. Never disregard professional medical advice or delay in seeking it because of something you have read on the WebMD Site!

If you think you may have a medical emergency, call your doctor or 911 immediately. WebMD does not recommend or endorse any specific tests, physicians, products, procedures, opinions, or other information that may be mentioned on the Site. Reliance on any information provided by WebMD, WebMD employees, others appearing on the Site at the invitation of WebMD, or other visitors to the Site is solely at your own risk.

The Site may contain health- or medical-related materials that are sexually explicit. If you find these materials offensive, you may not want to use our Site.

Results

HONcode Principles

Table 1 gives the results of the comparison between manual evaluation and the automated system’s conformity assessment for each of the HONcode principles. When neither manual nor automated analysis found justification of conformity to a given criterion, it was considered a true negative. If a website passed both manual and automated reviews for meeting the specific criterion, it was considered a true positive. The websites where the automated detection system determined the criterion was satisfied but the expert manual evaluation disagreed were considered false positives. The websites where the automated system failed to detect a criterion even though manual review detected it were considered false negatives.

Table 1. Manual versus automated (using tfc and tfx weightings) evaluation (N=27).

Criteria	Manual	Automated									
		tfc					tfx				
		True[a]		False[b]		Other[c]	True[a]		False[b]		Other[c]
		–	+	–	+		–	+	–	+	
Authority	21	4	1	10	2	10	4	4	14	2	3
Complementarity	26	1	0	21	0	5	1	2	23	0	1
Privacy	24	1	14	9	2	1	3	14	10	0	0
Reference (attribution)	16	5	0	6	6	10	8	4	7	3	5
Justifiability	6	14	1	1	7	4	8	3	3	13	0
Contact details	26	1	6	16	0	4	1	15	8	0	3
Financial disclosure	17	8	1	9	2	7	9	0	16	1	1
Advertising policy	16	9	0	13	2	3	10	1	13	1	2
Date (attribution)	21	6	0	16	0	5	6	0	21	0	0

[a] True negative: both manual and automated did not find criterion was satisfied; true positive: both manual and automated did find criterion was satisfied.

[b] False negative: automated system did not find criterion was satisfied but manual review did; false positive: automated system did find criterion was satisfied but manual review did not.

[c] Criterion detected on a Web page different to the one designated in the manual review.

For 23 websites, the automated system failed to detect the criterion complementarity with tfx, even though manual review detected it. In this setup, the false negatives can be interpreted as silence, whereas the false positives represent the noise.

Table 2 gives the results of the evaluation using precision and recall. In order to present the results in this form, the authors made the assumption that the results that were found on different pages between automated and manual evaluations were seen as positive detections.

Table 2. Precision and recall of automated HONcode detection.

Criteria	tfc		tfx	
	Precision	Recall	Precision	Recall
Authority	0.85 (11/13)	0.52 (11/21)	0.78 (7/9)	0.33 (7/21)
Complementarity	1.00 (5/5)	0.19 (5/26)	1.00 (3/3)	0.12 (3/26)
Privacy	0.88 (15/17)	0.63 (15/24)	1.00 (14/14)	0.58 (14/24)
Reference (attribution)	0.63 (10/16)	0.63 (10/16)	0.75 (9/12)	0.56 (9/16)
Justifiability	0.42 (5/12)	0.83 (5/6)	0.19 (3/16)	0.50 (3/6)
Contact details	1.00 (10/10)	0.39 (10/26)	1.00 (18/18)	0.69 (18/26)
Financial disclosure	0.80 (8/10)	0.47 (8/17)	0.50 (1/2)	0.06 (1/17)
Advertising policy	0.60 (3/5)	0.19 (3/16)	0.75 (3/4)	0.19 (3/16)
Date (attribution)	1.00 (5/5)	0.24 (5/21)	0.00 (0/0)	0.00 (0/21)

As described in Boyer and Dolamic [19], this study took the entire specific Web page as the unit of evaluation. Even though the results presented in Boyer and Dolamic indicated a high precision for automated detection of the "date" criterion, this study's comparison had a high number of false negatives using the automated system. For this reason, the authors carried out an additional evaluation using each sentence as the evaluation unit. Table 3 gives the results of this evaluation for criteria "privacy" and "date." Table 4 gives the results of the evaluation using precision and recall.

Table 3. Privacy and date criteria using sentences versus the whole document approach (N=27).

Criteria	Manual, n	Automated (tfc), n									
		Document					Sentence				
		True[a]		False[b]		Other[c]	True[a]		False[b]		Other[c]
		–	+	–	+		–	+	–	+	
Privacy	24	1	14	9	2	1	0	21	2	3	1
Date (attribution)	21	6	0	15	0	6	0	11	1	6	9

[a] True negative: both manual and automatic did not find criterion was satisfied; true positive: both manual and automated did find criterion was satisfied.

[b] False negative: automated system did not find criterion was satisfied but manual review did; false positive: automated system did find criterion was satisfied but manual review did not.

[c] Criterion detected on a Web page different to the one designated in the manual review.

Table 4. Precision and recall of document and sentence automated HONcode detection.

Criteria	Document		Sentence	
	Precision	Recall	Precision	Recall
Privacy	0.88 (15/17)	0.63 (15/24)	0.88 (22/25)	0.92 (22/24)
Date (attribution)	1.00 (6/6)	0.24 (6/21)	0.77 (20/26)	0.95 (20/21)

Results on the HONcode Interreviewer Agreement Level

A total of 36 websites were evaluated for each HONcode criterion by 3 HONcode senior reviewers. The results of the evaluated interrater agreement using both percent agreement and Fleiss' kappa [25] for each of the HONcode principles are presented in Table 5.

Table 5. Interrater agreement, percent versus Fleiss' kappa (κ).

Criteria	Percent agreement (%)	Fleiss' κ	Interpretation
Authority	92.59	.745	Substantial agreement
Complementarity	79.63	–.113	Poor agreement
Privacy	85.19	.614	Substantial agreement
Reference (attribution)	88.89	.756	Substantial agreement
Justifiability	74.07	.463	Moderate agreement
Contact details	95.37	.471	Moderate agreement
Financial disclosure	87.04	.716	Substantial agreement
Advertising policy	85.19	.691	Substantial agreement
Date (attribution)	79.63	.492	Moderate agreement

Discussion

Principal Findings

The automated system performed the most poorly when detecting the "justifiability" criterion. Manual expert review indicated that only 6 of 27 websites fulfilled this criterion. The automated system detected this criterion for only 1 website when tfc weighting was used (eg, precision 0.42 with 4 detections on a different page), and for 3 websites with tfx (eg, precision only 0.19). Additionally, the automated system returned a large number of false positives: 7 and 13 for tfc and tfx, respectively. The poor performance of the automated system in detecting the compliance to this criterion might be explained by the fact that the data set used as a benchmark for training natural language processing algorithms for the automated detection is rather small for this criterion (eg, only 872 documents were available). In certain cases, the certain criterion might be not applicable for a given website. In that case, the website conforms to HONcode but the criteria justification will be missing from the collection. This represents the main reason of the small documents set.

When the automated system detected the criterion satisfaction on a different website page than that marked by the expert, additional manual expert review verified that the system was often correct in doing so. For example, for one website [26] the manual evaluation detected the criterion complementarity on the page [27], whereas the automated system detected it on a different page. Manual reexamination of the page on which the criterion justification was detected by the automated system confirmed that it also contained justification for satisfaction of this criterion. Even though the concept of the automated system is such that it tries to perform as close to manual evaluation as possible, a main difference exists. In the case of manual evaluation, once the criterion (eg, complementarity) is detected, not all the other pages of the website are checked. Contrarily, with the automated system, all pages are crawled before the evaluation step. Thus, the coverage can be much more important. This can also explain the detection of the criteria on other pages than that designated by the expert.

There were certain criteria, such as "date," in which the automated system performance was unexpectedly poor. For this reason, the study examined an alternative approach using the sentence instead of the document as the classification unit (Table 3). The number of automated system detections for the criterion "date" was increased when the sentence was used as the classification unit. Similar improvements occurred using sentence-level analysis for the privacy criterion. Further studies must determine if such increases obtained using variant methods are statistically significant and should be incorporated permanently into the automatic detection algorithms. Manual analyses detected previously unknown technical problems in automated privacy criterion recognition. For one website, this particular criterion was deemed satisfied on 99% of the site's Web pages, in addition to the page marked as correct by the expert. This did not occur when documents were used as the classification units. Another technical problem occurred when the automated system was unable to detect the date on the pages where this information was displayed using only numbers (eg, 07/07/2012) without any accompanying explanatory text. The main source of this problem was the system tokenization approach, which ignores numbers. However, changing the preprocessing and keeping the numbers in the tokenization process would not be beneficial for this criterion detection. A number can represent not only a date but also other information, which could result in a number of false positives for this and for other criteria.

As seen in Table 1, the automated system performed capably for certain criteria. The level of agreement between the manual and automated approaches elevated to 70% (eg, contact details with tfc). Such a level of agreement, approaching the 72% human agreement [28], speaks in favor of the automated system as an alternative to the manual approach. However, the system performed poorly in detecting HON principle satisfaction for funding, complementarity, date, and authority.

The privacy criterion is easy to detect for the automated system and humans. In our previous study, the automated detection of the privacy criterion showed precision of more than 92% with good recall of more than 91% [19]. However, during manual evaluation for this criterion, the expert is not only looking for the privacy statement but also verifies its implementation (eg, cookies). The automated system has to rely only on the privacy statement.

For the privacy criterion, the automated system scored 15 correct (of 24 websites that respected this criteria) for the tfc weighting scheme. Fourteen of these were true positives. It also detected criterion satisfaction on a different page than that designated by the expert for 1 website. For 2 websites, the automated system mistakenly detected privacy as satisfied. For 9 websites, the automated system failed to detect privacy satisfaction when the manual expert did so. This behavior is expected because our automated system is tuned to create less possible noise (false positives). The results described here reinforce the previous deduction of privacy criterion being the "easy" one to detect by the automated system.

Changing the weighting scheme to tfx for the privacy criterion resulted in a seeming performance enhancement. The correct results were returned for 17 websites, with no incorrect detections. This might represent random variation in study results or might suggest that the tfx method better detects the privacy criterion satisfaction.

Manual Evaluation Interreviewer Agreement Level

In Table 5, the values of Fleiss' kappa are rather small when compared to percent agreement. Although the values of .745 for authority and .756 for reference can be interpreted as substantial agreement, they still remain small when compared to percent agreement for these criteria. For the complementarity criterion, the kappa value of -.113 indicates disagreement in contrast to the percent agreement of 79.63% for this criterion. Two effects have been documented that might cause the misrepresentation of the interrater reliability by kappa [29]. The prevalence problem appears when one observation is coded more often than others, resulting in kappa estimation being very low, which is the case for the complementarity criterion in our study. Taking into account the particularity of the data for this criterion, kappa would not be the correct statistic to use. With a kappa value of .463, the criterion justifiability shows moderate agreement between raters (percent agreement 74%). These results show that even during the manual evaluation by experts, the criterion justifiability remains difficult to agree on. These results show that the probability of the expert giving an incorrect evaluation is quite low especially for "easy" criteria such as contact details. However, this probability is somewhat higher for more complicated criteria, such as the justifiability criterion, which further confirms the complexity of this criterion. So, this brief study identifying the level of agreement between expert reviewers shows that the automatic system behaves somewhat similarly to the manual reviewers.

Limitation

In this evaluation, the authors compare automated HONcode conformity assessment to assessments done by a senior HONcode expert. Doing so introduces a bias. It assumes that the experts never improperly assess the presence or absence of HONcode principle satisfaction in documents. Although a HONcode expert has lower likelihood of making a false assessment than other reviewers or other automated systems, we recognize that expert assessments are not always correct, which is shown by the interrater agreement level.

Conclusions

This study analyzed the effectiveness of an automated HONcode criteria compliance detection system. A total of 27 websites chosen with different completion statuses with respect to HONcode certification were included in the evaluation. Study results indicate a relatively high level of agreement between automatic and manual assessments for some of the HONcode criteria. Nevertheless, for other criteria, the manual approach was clearly superior. Study results suggest that "tuning" the automated detection system through future studies for each specific HONcode criterion may improve the system's ability to detect individual criterion satisfaction. Study results also indicate that correcting a small number of technical issues in the automated system, such as the problem of not detecting the date criterion on pages displaying this information, may also improve future system performance. Incorporating third-party libraries or systems that have already proven their ability to detect and extract this kind of information [30,31] might be a solution for this issue. This approach is part of future development for this system.

The KHRESMOI project has attempted to develop a health search engine dedicated to the general public's needs. "KHRESMOI for Everyone" (K4E) [32] is a multilingual, multimodal search and access system for biomedical information and documents. Because K4E is a specialized search engine for health information, it has specialized tools to help users to discern good quality health information from the poor quality information. K4E offers automatic detection of the 8 HONcode principles with additional trustability levels given as a percentage integrated into the search results. It also identifies the HONcode principles that are currently not being respected by the website as estimated by automatic detection so that the reader is aware of the extent to which the website can or cannot be trusted and which HONcode principle is concerned. This interface is described in detail in Pletneva et al [33]. K4E can be used in the future after further research and development based on study results conducted within the European project Kconnect [34] as a specialized quality health search engine or Web service to target trustworthy health information enabling readers to directly access this information without having to wade through multiple pages of dubious material to get there.

Another potential outcome to this study is further development of the automated detection system to assist in conducting the HONcode certification process. The present manual HONcode certification process is time consuming. Even though the level of agreement between the manual and automated systems is somewhat lower than that of 3 experts (eg, 70% vs 95% for contact details), the authors estimate that HONcode automatic detection systems might provide a first screening; thus, helping in the certification process. In summary, the future of identifying quality, trustworthy health information on the Internet will depend on development of advanced search engines with fine-tuned criterion-matching abilities that can guide users to reliable health information websites.

Acknowledgments

The research is conducted in the scope of the European project Kconnect and funded by this project (2015-2018, project No. 644753) in the continuation of the European project KHRESMOI (2010-2014, project No. 257528).

Conflicts of Interest

None declared.

References

1. Fox S. The social life of health information. Washington, DC: Pew Internet & American Life Project; 2011 May 12. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2014-08-14] [WebCite Cache ID 6Rovu1D1m]
2. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the Internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
3. Cooley DL, Mancuso AM, Weiss LB, Coren JS. Health-related Internet use among patients of osteopathic physicians. J Am Osteopath Assoc 2011 Aug;111(8):473-482. [Medline: 21862755]
4. Fox S, Duggan M. Health online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2014-08-14] [WebCite Cache ID 6Row43ynP]
5. European Commission. 2014 Nov 28. Europeans becoming enthusiastic users of online health information URL: https://ec.europa.eu/digital-agenda/en/news/europeans-becoming-enthusiastic-users-online-health-information [accessed 2014-12-04] [WebCite Cache ID 6UZ6MC42s]
6. European Commission. 2014 Nov. European citizens' digital health literacy URL: http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf [accessed 2014-12-04] [WebCite Cache ID 6UZ6Rf0sD]
7. Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis CF. Text categorization models for high-quality article retrieval in internal medicine. J Am Med Inform Assoc 2005;12(2):207-216 [FREE Full text] [doi: 10.1197/jamia.M1641] [Medline: 15561789]
8. Griffiths KM, Tang TT, Hawking D, Christensen H. Automated assessment of the quality of depression websites. J Med Internet Res 2005;7(5):e59 [FREE Full text] [doi: 10.2196/jmir.7.5.e59] [Medline: 16403723]
9. Rudin JL, Littleton D. Searching for information on the World Wide Web--a guide for dental health professionals: Part 1. Compend Contin Educ Dent 1997 Sep;18(9):930-2, 934, 936. [Medline: 9533369]
10. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. Australas Med J 2014;7(1):24-28 [FREE Full text] [doi: 10.4066/AMJ.2014.1900] [Medline: 24567763]
11. Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health Websites. Comput Biol Med 1998 Sep;28(5):603-610. [Medline: 9861515]
12. Health on the Net Foundation. HON Code of Conduct principles and guidelines URL: http://www.hon.ch/HONcode/Webmasters/Guidelines/guidelines.html [accessed 2014-09-01] [WebCite Cache ID 6SGy8KKtZ]
13. Boyer C, Baujard V, Nater T, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for Medical and Health-related Web sites: three years on. J Med Internet Res 1999;1(Suppl 1):e99. [doi: 10.2196/jmir.1]
14. Boyer C, Geissbuhler A. A decade devoted to improving online health information quality. Stud Health Technol Inform 2005;116:891-896. [Medline: 16160371]
15. Laversin S, Baujard V, Gaudinat A, Simonet MA, Boyer C. Improving the transparency of health information found on the internet through the honcode: a comparative study. Stud Health Technol Inform 2011;169:654-658. [Medline: 21893829]
16. Fallis D, Frické M. Indicators of accuracy of consumer health information on the Internet: a study of indicators relating to information for managing fever in children in the home. J Am Med Inform Assoc 2002 Feb;9(1):73-79 [FREE Full text] [Medline: 11751805]
17. Health Fact Sheet. Washington, DC: Pew Internet & American Life Project URL: http://www.pewinternet.org/fact-sheets/health-fact-sheet/ [accessed 2014-08-14] [WebCite Cache ID 6RoyERgF5]
18. White R, Horvitz E. Microsoft Research. Cyberchondria: studies of the escalation of medical concerns in Web search URL: http://research.microsoft.com/pubs/76529/TR-2008-178.pdf [accessed 2014-08-14] [WebCite Cache ID 6RoyoFIYT]
19. Boyer C, Dolamic L. Feasibility of automated detection of HONcode conformity for health-related websites. IJACSA 2014;5(3):69-74. [doi: 10.14569/IJACSA.2014.050309]
20. Gaudinat A, Grabar N, Boyer C. Combination of heterogeneous criteria for the automatic detection of ethical principles on health web sites. AMIA Annu Symp Proc 2007:264-268 [FREE Full text] [Medline: 18693839]
21. Williams K, Calvo RA. A framework for document categorization. In: Proceedings of the Seventh Australasian Document Computing Symposium. 2002 Dec Presented at: Seventh Australasian Document Computing Symposium; December 16, 2002; Sydney, Australia.
22. Manning CD, Raghavan P, Schutze H. Introduction to Information Retrieval. New York: Cambridge University Press; 2008.

23. Baeza-Yeats RA, Riberio-Neto B. Modern Information Retrieval. Reading, MA: Addison-Wesley Longman; 1999.
24. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management 1988 Jan;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]
25. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons; 1981.
26. PTINR.com. URL: http://www.ptinr.com/ [accessed 2015-05-22] [WebCite Cache ID 6Yi2ijMqD]
27. PTINR.com. Terms and conditions URL: http://216.244.100.92/terms-and-conditions [accessed 2015-05-21] [WebCite Cache ID 6YgmqsyUm]
28. Jones R, Mccallum A, Nigam K, Riloff E. Bootstrapping for text learning tasks. In: IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications. 1999 Presented at: IJCAI-99; July 31-August 6, 1999; Stockholm, Sweden p. 52-58.
29. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [Medline: 22833776]
30. Apache OpenNLP. URL: http://opennlp.apache.org/documentation.html [accessed 2015-02-02] [WebCite Cache ID 6W2cIdNSd]
31. The Stanford Natural Language Processing Group. Stanford Named Entity Recognizer (NER) URL: http://nlp.stanford.edu/software/CRF-NER.shtml [accessed 2015-02-02] [WebCite Cache ID 6W2cOtwWr]
32. Khresmoi for Everyone. URL: http://everyone.khresmoi.eu/hon-search/ [accessed 2015-05-21] [WebCite Cache ID 6Ygu38h3z]
33. Pletneva N, Kriewel S, Kritz M. D8.5.2: Prototype of second search system based on feedback. 2013 Dec 08. URL: http://www.khresmoi.eu/assets/Deliverables/WP8/KhresmoiD852.pdf [accessed 2014-08-14] [WebCite Cache ID 6Rp0HW9sm]
34. Kconnect. URL: http://www.kconnect.eu/ [accessed 2015-05-21] [WebCite Cache ID 6YguE8wZI]

Abbreviations

HON: Health on the Net
K4E: KHRESMOI for Everyone
SVM: support vector machine

Edited by G Eysenbach; submitted 02.09.14; peer-reviewed by N Grabar, R Hardikar; comments to author 06.11.14; revised version received 16.12.14; accepted 19.04.15; published 02.06.15.

Please cite as:
Boyer C, Dolamic L
Automated Detection of HONcode Website Conformity Compared to Manual Detection: An Evaluation
J Med Internet Res 2015;17(6):e135
URL: http://www.jmir.org/2015/6/e135/
doi:10.2196/jmir.3831
PMID:26036669

Original Paper

The Development of Online Doctor Reviews in China: An Analysis of the Largest Online Doctor Review Website in China

Haijing Hao[1], PhD

University of Massachusetts Boston, Department of Management Science and Information Systems, Boston, MA, United States

Corresponding Author:
Haijing Hao, PhD
University of Massachusetts Boston
Department of Management Science and Information Systems
100 Morrissey Boulevard
Boston, MA, 02125
United States
Phone: 1 6172877706
Fax: 1 6172877887
Email: haohaijing@gmail.com

Abstract

Background: Since the time of Web 2.0, more and more consumers have used online doctor reviews to rate their doctors or to look for a doctor. This phenomenon has received health care researchers' attention worldwide, and many studies have been conducted on online doctor reviews in the United States and Europe. But no study has yet been done in China. Also, in China, without a mature primary care physician recommendation system, more and more Chinese consumers seek online doctor reviews to look for a good doctor for their health care concerns.

Objective: This study sought to examine the online doctor review practice in China, including addressing the following questions: (1) How many doctors and specialty areas are available for online review? (2) How many online reviews are there on those doctors? (3) What specialty area doctors are more likely to be reviewed or receive more reviews? (4) Are those reviews positive or negative?

Methods: This study explores an empirical dataset from Good Doctor website, haodf.com—the earliest and largest online doctor review and online health care community website in China—from 2006 to 2014, to examine the stated research questions by using descriptive statistics, binary logistic regression, and multivariate linear regression.

Results: The dataset from the Good Doctor website contained 314,624 doctors across China and among them, 112,873 doctors received 731,543 quantitative reviews and 772,979 qualitative reviews as of April 11, 2014. On average, 37% of the doctors had been reviewed on the Good Doctor website. Gynecology-obstetrics-pediatrics doctors were most likely to be reviewed, with an odds ratio (OR) of 1.497 (95% CI 1.461-1.535), and internal medicine doctors were less likely to be reviewed, with an OR of 0.94 (95% CI 0.921-0.960), relative to the combined small specialty areas. Both traditional Chinese medicine doctors and surgeons were more likely to be reviewed than the combined small specialty areas, with an OR of 1.483 (95% CI 1.442-1.525) and an OR of 1.366 (95% CI 1.337-1.395), respectively. Quantitatively, traditional Chinese medicine doctors ($P<.001$) and gynecology-obstetrics-pediatrics doctors ($P<.001$) received more reviews than the combined small specialty areas. But internal medicine doctors received fewer reviews than the combined small specialty areas ($P<.001$). Also, the majority of quantitative reviews were positive—about 88% were positive for the doctors' treatment effect measure and 91% were positive for the bedside manner measure. This was the case for the four major specialty areas, which had the most number of doctors—internal medicine, gynecology-obstetrics-pediatrics, surgery, and traditional Chinese medicine.

Conclusions: Like consumers in the United States and Europe, Chinese consumers have started to use online doctor reviews. Similar to previous research on other countries' online doctor reviews, the online reviews in China covered almost every medical specialty, and most of the reviews were positive even though all of the reviewing procedures and the final available information were anonymous. The average number of reviews per rated doctor received in this dataset was 6, which was higher than that for doctors in the United States or Germany, probably because this dataset covered a longer time period than did the US or German dataset. But this number is still very small compared to any doctor's real patient population, and it cannot represent the reality of that population. Also, since all the data used for analysis were from one single website, the data might be biased and might not be a representative national sample of China.

(J Med Internet Res 2015;17(6):e134) doi:10.2196/jmir.4365

KEYWORDS

online doctor reviews; China health system; quantitative review; qualitative review; patient empowerment; physician quality

Introduction

Overview

Online doctor reviews have been happening across the world since the Internet Web 2.0 came into use in the early 2000s, and they have attracted health care researchers' attention about how these reviews have been used in different countries [1]. Based on a study conducted in 2012, about 17% of physicians had been rated on the Internet in the United States [2]. In the United Kingdom, about 61% of family physicians who are posted on the National Health Service website have been rated [3]. In Germany, 37% of German physicians were rated on the jameda website in 2012 [4]. Also, the difference between traditional patient reviews of their doctors and the online doctor reviews has been discussed [4], as well as what type of information the online doctor reviews could provide [5]. Further, some research also examined the online reviews in different medical specialty areas, such as dentistry [6] and orthopedics [7]. Some research also raised concerns about online doctor reviews, which may be subject to manipulation or could damage physicians' reputations [8-12]. At the same time, studies about how health care consumers used the online doctor reviews have been conducted. A cross-sectional survey in Germany showed that about 32% of respondents heard of physician-rating websites, and about 25% had already used a website to search for a physician [13]. A survey conducted via the most popular online social network in the Netherlands found that about one-third of the Dutch population searched for ratings of health care providers [14]. A representative sample of citizens who were at least 15 years old from seven European countries were surveyed. The results showed that among the people who use the Internet for health care-related purposes, on average, more than 40% of people considered the information of these eHealth services to be important when choosing a new doctor [15]. A 2012 survey in the United States showed that 17% of Internet users have consulted physician-rating sites, and 4% of people posted a review online of a doctor [16]. A 2012 study comprised of a nationally representative sample of US citizens found that 59% of the survey respondents said that online doctor ratings are "somewhat important" for them and 19% said they are "very important" for them when they search for a physician [17]. A study also examined what factors may affect consumers' decisions to adopt online doctor reviews [18].

We can see that various studies regarding online doctor reviews, either based on secondary data on how many doctors have been reviewed online or based on the first-hand survey data on how patients look at those online doctor reviews, have been emerging in the United States and Europe, but there has been no study about whether Chinese consumers use the Internet to review their health care providers. Considering the fact that China has the largest population of Internet users in the world [19], and China is already known for having more than 1 million online doctor reviews by international media [20], this study wants to examine further the current status of online doctor reviews in China.

Without a mature primary care system in China, most Chinese consumers now largely have to self-refer to any health care provider they can afford or who they believe is good based on little to no information [21]. Internet technology brings a new option for Chinese consumers and, particularly, Web 2.0 technology brings the interactive form of information sharing online. The first online local service review website in China was established in 2003—General Public Review Web. At the end of 2014, there were more than 60 million public reviews on General Public Review Web and most of those reviews were about local restaurants or movie theaters [22]. Based on the searches conducted on Google and Baidu, the number one search engine in China [23], and personal meetings with medical professionals in Beijing, the author found a few online doctor review websites in China. Chinese Traditional Medicine Review Web (zydp.org) [24], established in December 2013, focuses completely on reviewing traditional Chinese medicine doctors. Schedule Appointment website (guahao.com) [25], established in 2010, claims that its goal is to help patients to schedule doctor appointments online. At the same time, it also provides a feature to review doctors online. Hao Dai Fu, or the Good Doctor website (haodf.com)—hao means "good" and dai fu means "doctor" in Chinese [26]—is the earliest online doctor review website in China, established in 2006. Its purpose is to provide an online doctor reviewing system for Chinese consumers to review their doctors, and it also helps consumers to select a good doctor for their health care concerns based on online reviews. Also, the Good Doctor website already had more than 1 million reviews by the summer of 2014 [20].

Thus, the purpose of this study is to examine the following research questions about the current status of online doctor reviews in China based on the Good Doctor website: (1) How many doctors and how many specialty areas are available to be rated online on the Good Doctor website? (2) Which medical specialties are most likely to be rated? (3) Which medical specialties receive more reviews? (4) How are the quantitative rating scores distributed? and (5) What are the developing trends of online doctor reviews on the Good Doctor website?

Background

According to meetings with Mr Hang Wang, the founder of the Good Doctor website, the original purpose in establishing this website was to help Chinese consumers to find good and appropriate specialists for their health care problems based on online reviews, after he personally experienced difficulty in finding a good specialist doctor in Beijing. In 2006, the Good Doctor website was launched in Beijing, China, and for the first time an online doctor review system became available for Chinese consumers.

Since its establishment, the Good Doctor company staff has been manually collecting information on Chinese doctors and their hospitals by various means—in China, a majority of doctors work at public hospitals where they are employees and have a responsibility for both inpatient service and outpatient service. The staff collected information by visiting hospital campuses in person, making phone calls to hospitals, or searching hospital websites, for those hospitals that had them. The staff then posted the collected information about the doctors on the Good Doctor website for consumers to browse and review for free. They knew that they would not have a national database to rely on and, as a start-up, they had limited human resources. Therefore, their strategy was to start with the largest and most famous hospitals in Beijing and Shanghai, then gradually cover the remaining parts of China, since the most reputable Chinese hospitals are concentrated mostly in Beijing or Shanghai. According to Fudan University's hospital ranking system, which was based on a peer-review system on hospitals' medical practice, quality of care, and research [27], 26 out of the 100 best hospitals are in Beijing, and 19 are in Shanghai. The posted information includes each doctor's name, short biography, specialty area, technical title, and hospitals where they work. Chinese doctors have a technical title system and the title is assigned through an evaluation process. The rank is nationally unified as four levels—from junior to senior levels—from Resident Physician, Attending Physician, Associate Physician, to Chief Physician. On average, every 5 years a doctor can move up one level in this title track. Thus, a title primarily indicates a doctor's work experience and technical skills, which also determines the consultation fee of a doctor. The Good Doctor website also posts the hospital information where doctors work, such as the hospital name, address, and grade of the hospital. China's hospital grades are evaluated by a government agency—the National Health Department at the provincial level—and the evaluation standards are based on the hospital facilities, the number of beds, technical equipment, quality of care, the doctors' skills, etc [28].

Once a doctor's information is posted on the Good Doctor website, patients can anonymously review those doctors online based on their inpatient or outpatient experiences with the doctor. There are two types of reviews on the Good Doctor website: one is a quantitative review with two measures, *treatment effect* and *bedside manner*, which have to be done together, and the other is a narrative open-ended textual review, which can be done separately from the quantitative reviews. Both of the quantitative measures use a 5-level rating scale, from Very Unsatisfied (1), Unsatisfied (2), OK (3), Satisfied (4), to Very Satisfied (5). Over a few years of development, besides the doctor rankings by specialty area based on patients' reviews, the Good Doctor website has also developed other features. These features include the following: a doctor's personal webpage on the Good Doctor website where the doctor can post medical articles or health care advice, a doctor's personal forum where patients can post questions or initiate discussions with doctors that they choose, daily updates of a doctor's outpatient schedule, appointment scheduling online, telephone consultation, and membership in private patient clubs, etc.

Methods

Data

Based on the application programming interface (API) provided by the Good Doctor website, this study collected data on 314,624 doctors and their associated 3091 hospitals from the website as of April 11, 2014. After data cleaning by removing the records with missing values or abnormal values, there were 731,316 quantitative reviews, including both *treatment effect* and *bedside manner*, and 772,979 narrative textual reviews on 117,624 doctors across China from almost every specialty area. A total of 731,264 records had both quantitative and qualitative reviews. This study focuses on the two 5-scale quantitative reviews only.

Based on the dataset from the Good Doctor website, there are nine major specialty areas and one specialty area called "others," which groups all the uncommon, small specialty areas not listed separately on the Good Doctor website. Table 1 shows that traditional Chinese medicine, gynecology-obstetrics-pediatrics, internal medicine, and surgery are the four top specialty areas which have the most number of doctors and had the most number of reviews, excluding "others" because it is not a single specialty area. Also, those four major specialty areas consist of 8.7%, 12.1%, 21%, and 18.3% of the total doctor population on the Good Doctor website, respectively, which is similar to the national composition of the doctors by percentage of those four specialty areas—16%, 15%, 20.7%, and 12.9%, respectively [29]. Therefore, this study selected those four specialty areas as the major focus for analysis.

Table 1. Specialty areas, number of doctors, and number of reviews from the Good Doctor website.

Specialty areas	Total doctors, n	Doctors receiving reviews, n (%)	Total reviews, n	Average reviews per rated doctor
Cancer	2781	1317 (47.36)	7018	5.3
Traditional Chinese medicine[a]	27,299	12,011 (44.00)	85,649	7.1
Gynecology-obstetrics-pediatrics[a]	38,099	16,506 (43.32)	122,073	7.4
Infectious disease	1122	483 (43.05)	2869	5.9
Internal medicine[a]	66,162	22,345 (33.77)	96,892	4.3
Orthopedics	1008	495 (49.11)	3592	7.3
Others	112,483	36,038 (32.04)	226,823	6.3
Psychiatry	2848	1050 (36.87)	5800	5.5
Oral health	5106	2671 (52.31)	15,690	5.9
Surgery[a]	57,716	24,708 (42.81)	164,910	6.7
Total	314,624	117,624 (37.39)	731,316	6.2

[a] Specialty area that is among the four top specialty areas, which have the most number of doctors.

Table 2 shows the number of reviews for each specialty area in each year. This study ignored 2006 and 2014 data because those two years were not complete calendar years in the dataset. We can see that the number of reviews per year has been increasing over time for all specialty areas, except for 2010 and 2013, both of which had a little dip for all but one specialty area, oral health.

Table 2. Number of reviews for each specialty per year.

Year	Cancer	TCM[a]	G-OB-P[b]	ID[c]	IM[d]	OP[e]	Others	Psychiatry	Oral health	Surgery
2006	0	20	115	1	142	0	155	5	13	108
2007	387	3351	6745	143	6335	143	11,055	261	642	7912
2008	691	10,114	16,008	415	12,508	334	26,006	659	1392	17,177
2009	1000	12,930	18,619	454	13,336	476	33,003	809	1849	21,623
2010	790	10,349	13,788	309	10,081	361	25,393	655	1900	18,830
2011	1181	15,186	20,185	520	14,755	671	38,650	954	2571	27,702
2012	1365	17,676	22,491	531	18,568	846	43,997	1212	3207	33,057
2013	1292	13,561	19,033	388	16,536	605	38,705	968	3222	30,134
2014	312	2462	5089	108	4631	156	9859	277	894	8367
Total	7018	85,649	122,073	2869	96,892	3592	226,823	5800	15,690	164,910

[a] Traditional Chinese medicine (TCM)

[b] Gynecology-obstetrics-pedicatrics (G-OB-P)

[c] Infectious disease (ID)

[d] Internal medicine (IM)

[e] Orthopedics (OP)

Statistical Analysis

In order to further examine the research question of which types of doctors are more likely to be rated, a binary logistic regression model was constructed as follows:

$\text{Logit}(\text{Rated}_i) = \text{Specialty Area}_i + \text{Physician Title}_i + \text{Hospital Level}_i + \text{Beijing}_i + \text{Shanghai}_i$ (1)

Rated_i equals 1 if doctor i has been rated, otherwise it is 0. Specialty Area_i is a categorical variable which differentiates the four major specialty areas from the rest of the combined specialty areas, combined specialties. The combined specialties combined the other five specialty areas listed by the Good Doctor website—infectious disease, orthopedics, psychiatry, oral health, and cancer—with the “others” for concision. Physician Title_i is a categorical variable, too, which indicates doctor i’s technical title from one of the four levels that was discussed earlier. There are three levels of hospital grades—Level 3 is the highest with more beds, better equipment, more highly skilled doctors, etc. This model also controls for Beijing and Shanghai because higher-ranking hospitals are more concentrated in these two cities than in other areas in China.

The following model examines which specialty area doctors would receive more ratings by using a multivariate linear regression model:

$\text{Rating\_count}_i = \text{Specialty Area}_i + \text{Physician Title}_i + \text{Hospital Level}_i + \text{Beijing}_i + \text{Shanghai}_i + \text{error}_i$ (2)

The dependent variable, Rating_count_i , is the number of ratings doctor i received. The independent variables are similar to those in model (1) for doctor i's specialty area, technical title, hospital level, and whether the hospital is in Beijing, Shanghai, or another area.

Results

Regarding the first research question of how many doctors and how many specialty areas are available for review, from Table 1 we can see that there are 314,624 doctors from nine major specialty areas and many small specialty areas combined that are available for online review on the Good Doctor website. Among them, 117,624 doctors have been reviewed since 2006, which is 37.39% of the total doctors available for review. Among those nine major specialty areas, internal medicine has the lowest review percentage at 33.77%, and oral health has the highest review percentage at 52.31%. But since the total number of doctors in oral health is small—only 5106—this study mainly focused on the four major specialty areas, which include the most numbers of doctors: traditional Chinese medicine, gynecology-obstetrics-pediatrics, internal medicine, and surgery. Except for internal medicine, which has a review rate of 33.77%, the other three specialty areas all have a review rate of about 43.32% to 44.00%.

Table 3 shows the binary logistic regression results. We can see that doctors from traditional Chinese medicine, gynecology-obstetrics-pediatrics, and surgery were all about 1.5 times more likely to be reviewed compared to doctors from combined specialties, which is the reference group of the model. Doctors from internal medicine were less likely to be reviewed compared to the doctors from combined specialties. Also, chief physicians were about 4.6 times more likely to be reviewed, associate physicians were about 2.5 times more likely to be reviewed, and attending physicians were 1.6 times more likely to be reviewed than resident physicians. Doctors from Level 3 hospitals were 2 times more likely to be reviewed than doctors from Level 1 hospitals, and doctors from Level 2 hospitals were 1.5 times more likely to be reviewed than doctors from Level 1 hospitals. Doctors in Beijing and Shanghai were 1.5 times and 2 times more likely, respectively, to be reviewed than doctors from other areas of China. All of the estimated odds ratios are statistically significant at a 95% Wald confidence level.

Table 3. Results from the binary logistic regression (n=314,624).

Effect (independent variable)[a]	Odds ratio point estimate[b,c,d]	95% Wald CI
Traditional Chinese medicine	1.483	1.442-1.525
Gynecology-obstetrics-pediatrics	1.497	1.461-1.535
Internal medicine	0.940	0.921-0.960
Surgery	1.366	1.337-1.395
Chief physician	4.648	4.525-4.774
Associate physician	2.592	2.526-2.661
Attending physician	1.624	1.576-1.673
Level 3 hospital	2.047	1.995-2.100
Level 2 hospital	1.590	1.548-1.633
Beijing	1.532	1.486-1.579
Shanghai	2.102	2.035-2.172

[a] *Combined specialties* is the reference group for specialty areas, *resident physician* is the reference group for technical title, *Level 1 hospital* is the reference group for hospital grade, and *other areas* is the reference group for Beijing and Shanghai.

[b] Pseudo R^2 = .115.

[c] The dependent variable is *reviewed or not*.

[d] 5% significance level for all values.

Table 4 exhibits the linear regression results for which types of doctor would receive more reviews quantitatively. We can see that traditional Chinese medicine, gynecology-obstetrics-pediatrics, and surgeon had positive associations with the number of reviews a doctor received, but internal medicine was negatively associated with the number of reviews a doctor received. A chief physician, on average, can have 6 more reviews than a resident physician, which was the largest impact in this model. There are also positive impacts of being an associate physician or an attending physician on the number of reviews a doctor would receive, but the quantitative scale is smaller than that of a chief physician when comparing all of these to a resident physician. Also, being a doctor in Beijing or Shanghai is associated with 3 or 5 more reviews, respectively, than being a doctor in other areas of China. Being a doctor in a Level 3 hospital was associated with more reviews compared to being a doctor in a Level 1 hospital. But interestingly, a doctor working in a Level 2 hospital may receive fewer reviews compared to a doctor working in a Level 1 hospital. All of the estimates, except for surgery, are statistically significant at a 5% level.

Table 4. Results for linear regression for doctors in different areas receiving reviews (n=117,624).

Independent variable[a]	Parameter coefficient estimate[b,c,d]	Standard error	t_{11}	P
Intercept	2.17	0.24	9.14	<.001
Traditional Chinese medicine	0.62	0.17	3.58	<.001
Gynecology-obstetrics-pediatrics	0.97	0.15	6.34	<.001
Internal medicine	-2.44	0.14	-17.69	<.001
Surgery	0.26	0.13	1.95	.052
Chief physician	6.43	0.19	33.59	<.001
Associate physician	2.54	0.19	13.29	<.001
Attending physician	0.93	0.22	4.22	<.001
Level 3 hospital	0.53	0.17	3.10	.002
Level 2 hospital	-1.29	0.18	-7.15	<.001
Beijing	3.23	0.18	18.01	<.001
Shanghai	5.37	0.18	29.28	<.001

[a] *Combined specialties* is the reference group for specialty areas, *resident physician* is the reference group for technical title, *Level 1 hospital* is the reference group for hospital grade, and *other cities* is the reference group for Beijing and Shanghai.

[b] Adjusted R^2=.0353.

[c] The dependent variable is the *number of reviews*.

[d] 5% significance level for all values.

Figures 1 and 2 show the distribution of the number quantitative ratings a doctor received by the four major specialty areas, in absolute numbers and relative percentages. The distributions for the four specialty areas were quite similar—about 37% to 45% of doctors received 1 review, about 16% to 19% of doctors received 2 reviews, and about 19% to 21% of doctors received 3 to 5 reviews. In a few extreme cases, some doctors had received more than 500 reviews. Therefore, the number of quantitative reviews a doctor received was quite spread out.

Figures 3 and 4 show the quantitative rating score distribution among the four major specialty areas. We can see, regardless of the specialty area, that most quantitative ratings were positive—82% to 95% of reviews had responses of either Satisfied or Very Satisfied on either the *treatment effect* or *bedside manner* measure.

Figures 5 and 6 display the quantitative rating score distribution for the other small five specialty areas for *treatment effect* and *bedside manner*, respectively. Again, the quantitative reviews highly concentrate at the positive end of the rating scores.

Based on Table 2 and Figure 7, which shows the number of ratings over time, we can see that the number of reviews on the Good Doctor website has been growing for all specialty areas over the years, except for a little dip in years 2010 and 2013. The number of rated doctors had been growing, then stayed relatively stable in 2012 and 2013, with a similar little dip in 2010, as Figure 8 shows. The average number of ratings per doctor over time was relatively stable, within the range of 1.8 to 3.4, as seen in Figure 9. Since this study is based on a secondary dataset, further investigation is needed to determine the reason for those specific dips.

Figure 1. Distribution of the number of ratings across four major specialty areas (absolute number).

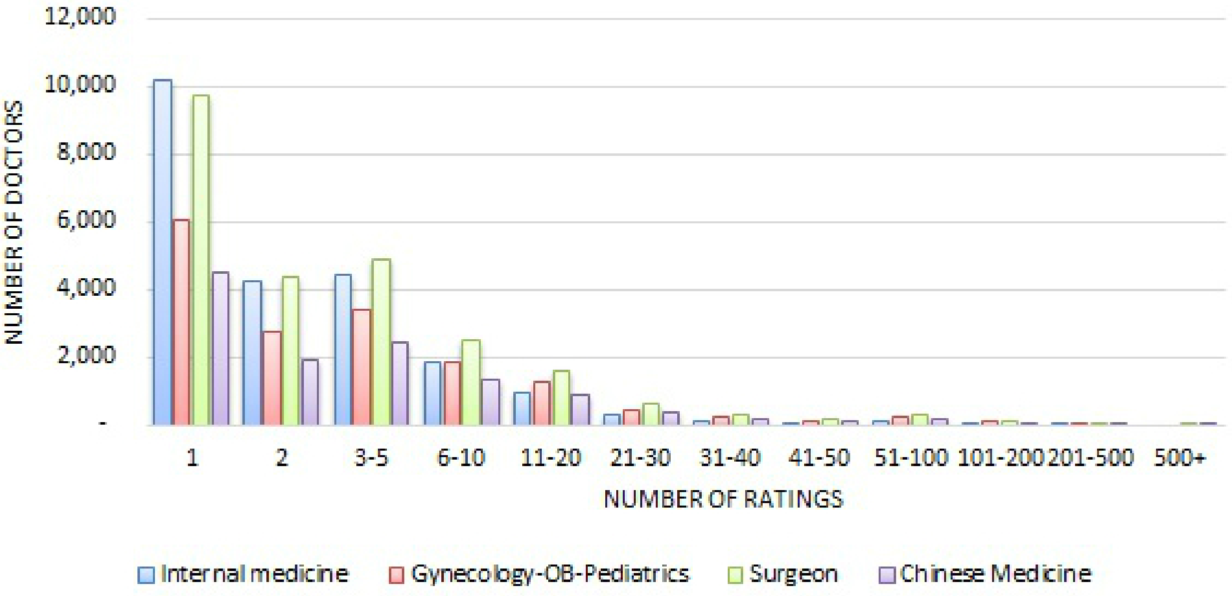


Figure 2. Distribution of the number of ratings across four major specialty areas (relative percentage).

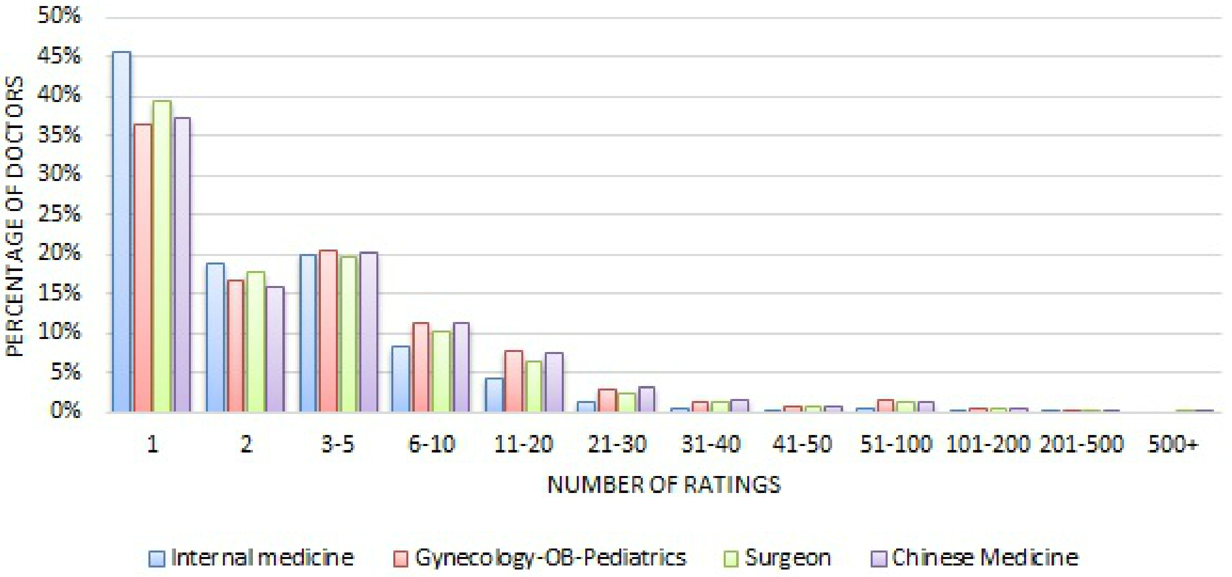


Figure 3. Distribution of treatment effect ratings across four major specialty areas.

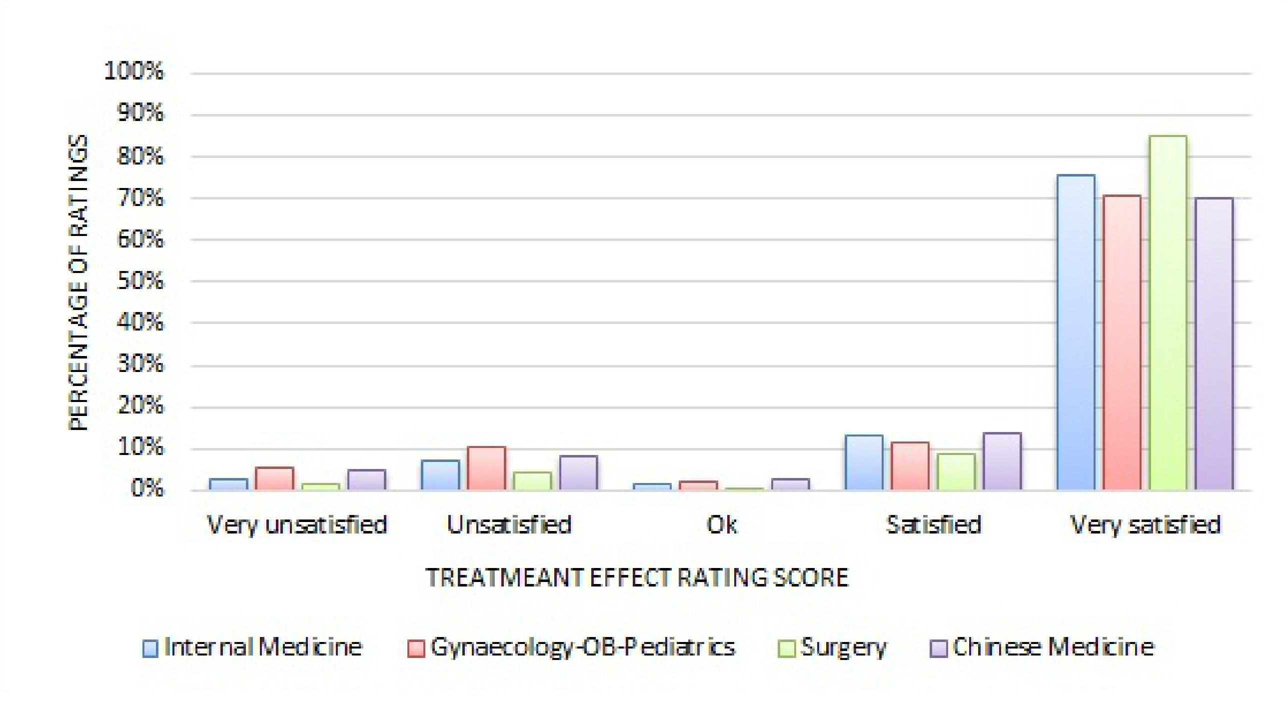


Figure 4. Distribution of bedside manner ratings across four major specialty areas.

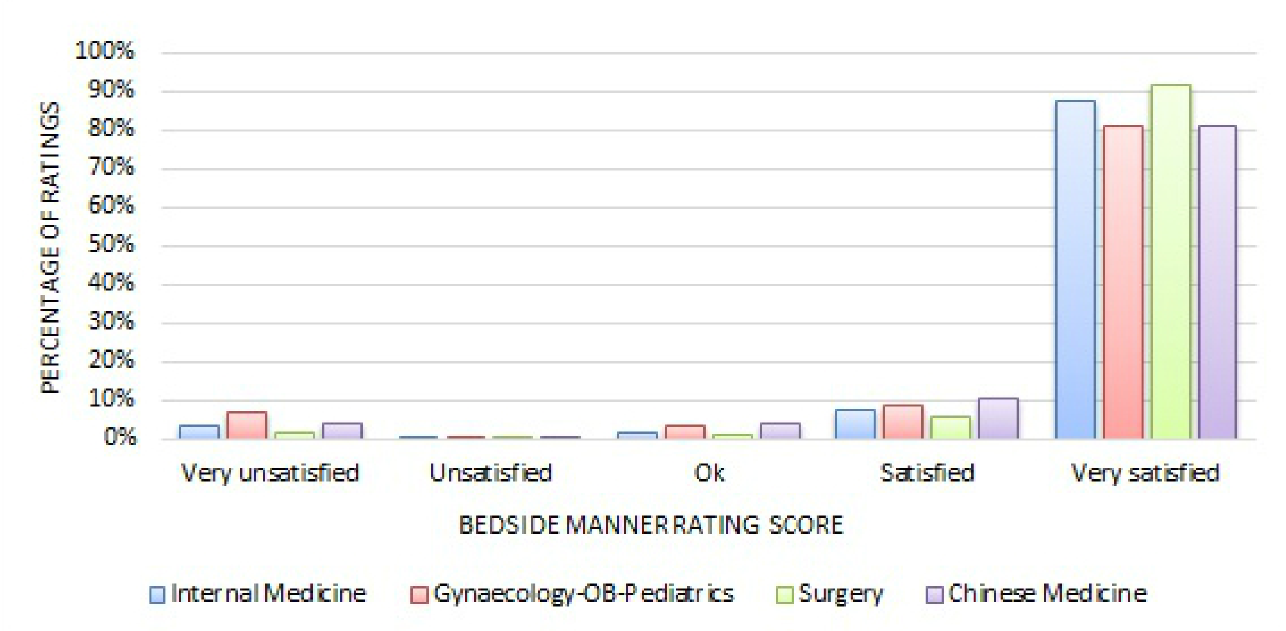


Figure 5. Distribution of treatment effect ratings across small specialty areas.

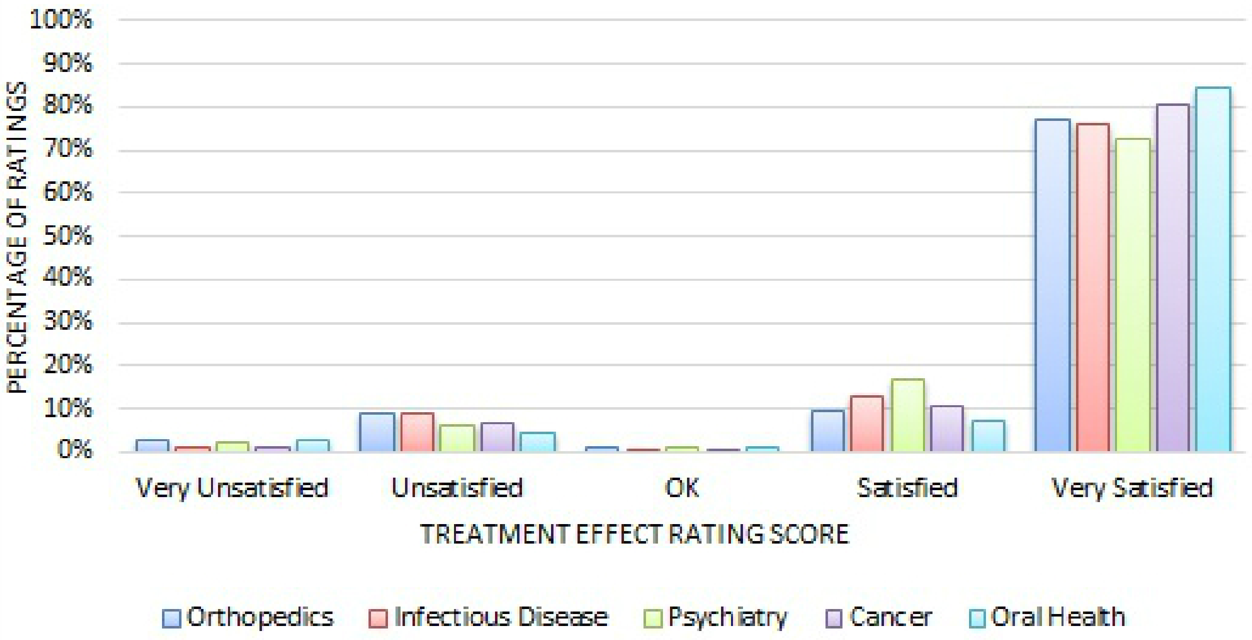


Figure 6. Distribution of bedside manner ratings across small specialty areas.

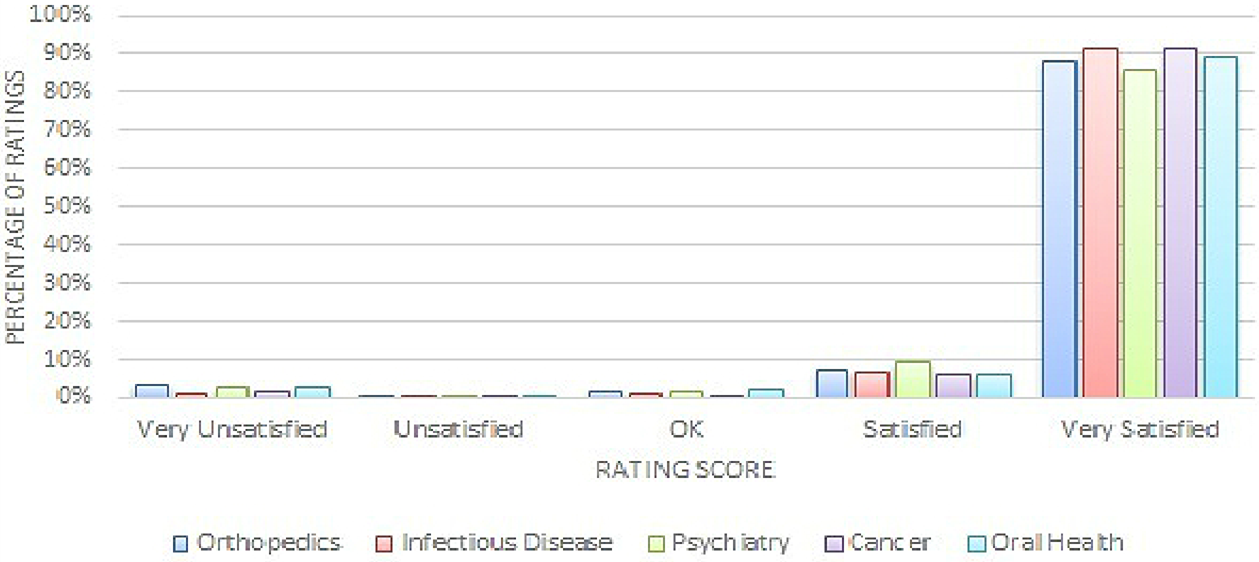


Figure 7. Number of ratings over time among four major specialty areas.

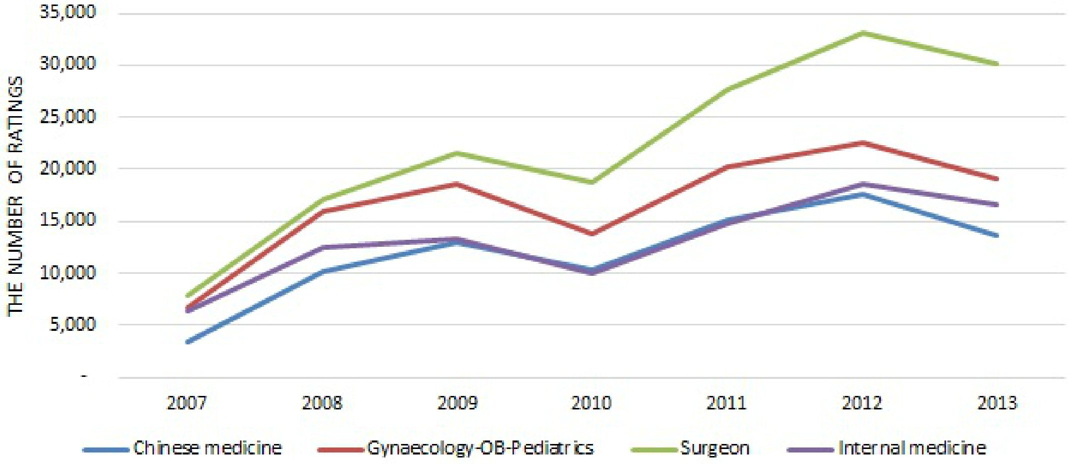


Figure 8. Number of doctors rated over time among four major specialty areas.

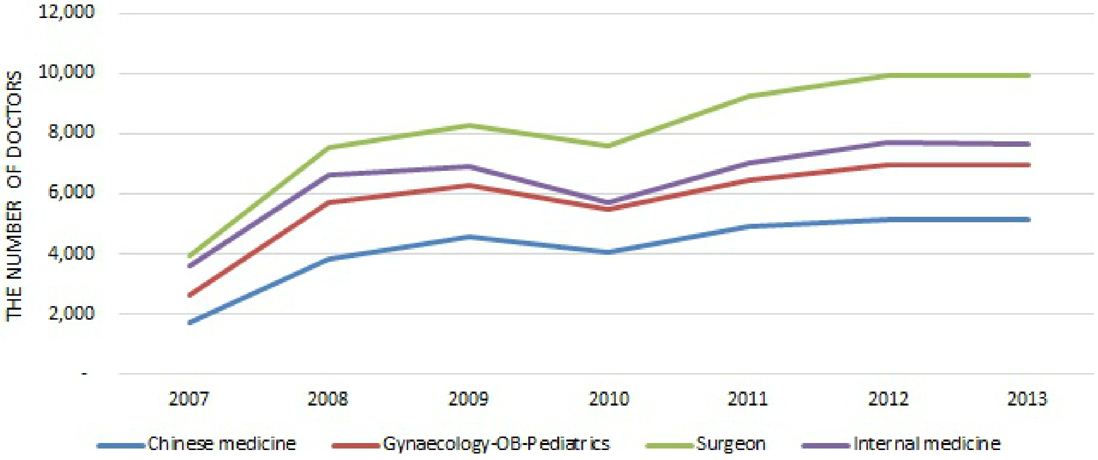
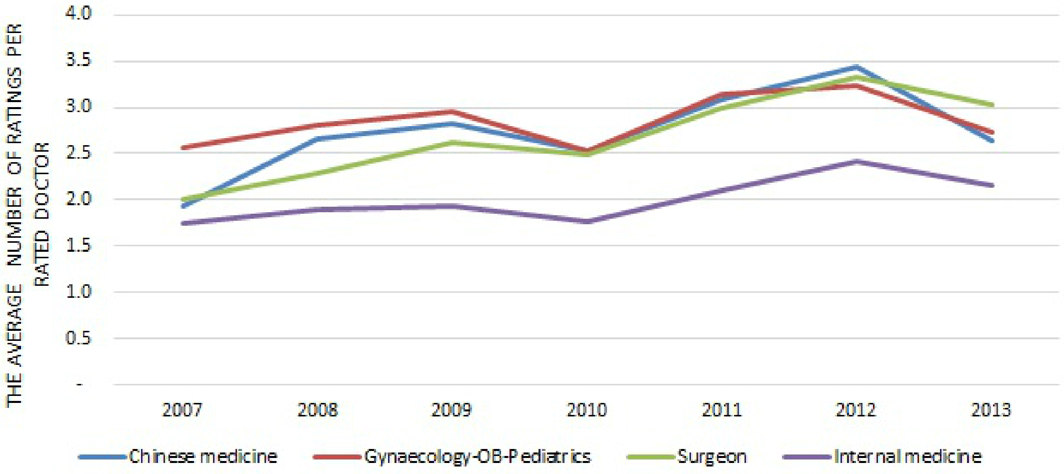


Figure 9. Average number of ratings per rated doctor over time among four major specialty areas.



Discussion

Principal Findings

First, we should realize that this study examines a dataset from a single website in China. Compared to developed countries, such as the United States, the United Kingdom, or Germany, the disadvantage of a dataset from a developing country like China is a lack of an official database which could be accessed to obtain the number of doctors in each specialty area at either the hospital level, the provincial level, or the national level. Thus, it is difficult to stratify this dataset to make a nationally representative sample. But the current dataset is an empirical dataset from the earliest and the largest available online doctor review website in China, and the four major specialty areas' percentage compositions are close to the national level aggregated data. Therefore, this empirical dataset will assist us to understand the current status of online doctor reviews in China.

The average review rate of all doctors over about 8 years on the Good Doctor website was 37.4%, which is close to the national review rate in Germany in 2012 of 37% [4], higher than the review rate between 2005 and 2010 in the United States of 17% [2], and lower than the review rate of family practice physicians in the United Kingdom of 61% between 2009 and 2012 [3]. But it should be noted throughout this study that it was difficult to compare across those countries because the datasets from different countries were collected by different means, and the sample sizes or time periods of those datasets were different.

On the Good Doctor website, 43% of gynecology-obstetrics-pediatrics doctors had been reviewed, which was higher than the review rate of obstetricians in the United States of 33% [2], but lower than that of gynecologists in Germany of 56.9% [4]. We should note that each country's categorization for specialty area is a little different. The Good Doctor website groups gynecology, obstetrics (OB), and pediatrics as one specialty area. Germany studied gynecology and the United States focused on obstetrics. Thus, that might be a reason why the review rate in China is higher because it covers more specialty areas compared to the United States or Germany. If we look further, statistically, at which specialty areas were more likely to be reviewed by using the logistic regression model, we can see that gynecology-obstetrics-pediatrics was the specialty area most likely to be reviewed in China. It was 1.5 times more likely to be reviewed than all of the other small medical specialties combined, which is similar to a study in the United States that showed the OB specialty was more likely to be rated, compared to other specialty areas [2]. This may indicate, as that study suggested, that obstetrics patients—similarly for gynecology or pediatrics patients in this study—are a mostly young and female population who are more likely to log on and use the Internet [2]. A study in Germany also showed that more women than men had used online doctor review websites [30]. Another possible reason that Chinese gynecology-obstetrics-pediatrics doctors received more reviews might be that children are always the focus of a family and the extended family in Chinese culture, thus consumers paid more attention to the quality of care by these doctors. Previous research also showed that the length of the relationship between a doctor and their patients plays a role in online doctor reviews, and patients who have had a longer relationship with their doctors would be more likely to review their doctors [31]. Not surprisingly, most patients would have a relatively stable and longer relationship with their gynecology-obstetrics-pediatrics doctors than with doctors of other medical specialties, hence they would be more likely to review these doctors. The Chinese dataset used in this study also showed that other major specialty areas—surgery and traditional Chinese medicine—received a similar, higher review rate as that of gynecology-obstetrics-pediatrics, which is a little different from the US and the German data. Surely, traditional Chinese medicine is a mainstream medical specialty only in China. Also, traditional Chinese medicine doctors mainly practice in herbal medicine, which usually has a longer treatment time and is not used for acute diseases. Thus, higher review rates might be due to the length of the patient-doctor relationship, too. Surgeons received more reviews probably because they usually have a longer and more interactive relationship with their patients. But the real reasons why doctors from those specialty areas were more likely to be reviewed need further investigation with richer data.

Doctors from the Beijing or Shanghai areas were more likely to be reviewed and were also likely to receive more reviews than doctors from other areas in China. This might be because, first, Beijing and Shanghai have more famous hospitals than other parts of China [27] which attract not only local patients, but patients nationwide. If a patient is nonlocal, the patient probably has a more serious or uncommon disease requiring them to travel to Beijing or Shanghai. If so, these types of patients are more serious about their doctors and probably more likely to review their doctors. Second, local residents of the metropolitan areas of Beijing or Shanghai have the highest Internet accessibility in China—75% and 70%, respectively—compared to other areas where Internet accessibility is lower than 66% [32]. Therefore, local patients have more Internet access and may be more likely to review their doctors. Thus, the doctors from Beijing or Shanghai would be more likely to be reviewed.

Doctors from Level 3 hospitals were more likely to be reviewed, and were likely to receive more reviews than doctors from the Level 1 hospitals. Level 3 hospitals usually have more beds, better equipment, more highly skilled doctors, and deal with more challenging diseases. Again, this might suggest that patients with more serious health care problems and probably a longer interaction time are more likely to review their doctors. Interestingly, doctors from Level 2 hospitals were more likely to be reviewed than doctors from Level 1 hospitals, but received fewer reviews quantitatively compared to Level 1 hospitals. Specific reasons for this phenomenon needs research and data. But one thing that we should realize is that the Good Doctor website intentionally started their data collection from Level 3 hospitals or famous hospitals from large metropolitan areas in order to help consumers find good specialists. This strategy may have resulted in sample selection problems because on the Good Doctor website, 54% of the doctors were from Level 3 hospitals and 38% of the doctors were from Level 2 hospitals. Compared to the national data, where 49% of doctors work in Level 3 hospitals and 51% of doctors work in Level 2 hospitals (there are no Level 1 hospitals in the national aggregated data), the aggregated compositions are different. In other words, the national statistics indicate that the number of doctors in Level 3 and Level 2 hospitals are close in quantity but the Good Doctor website collected more doctors' information from Level 3 hospitals to post online for patients to review.

Although, in total, there were more than 700,000 reviews in either the quantitative or the qualitative review sets, the average number of reviews per rated doctor was about 6.2, compared to that of the United States, which is 3 [2], and Germany, which is 2.37 [4]. We should point out that the dataset from the Good Doctor website covers a longer time period than the datasets from either the United States or Germany. If we look at the distribution of the number of reviews that doctors received for the four major specialty areas only (Figure 2), we can see that about 37% to 45% of doctors who received reviews received only 1 review, which is a little lower than that of Germany where 49.7% of physicians received a single review [4]. That also means that among the doctors who received reviews, a higher percentage of Chinese doctors compared to German doctors received more than 1 review. About 74% to 84% of Chinese doctors received 1 to 5 reviews, which is also a lower rate than that of Germany, where 93.4% of doctors received 1 to 5 reviews. This is also consistent with the fact that German doctors received a lower average number of reviews per doctor than Chinese doctors.

There might be a couple of possible reasons why Chinese consumers would like to review their doctors online. First, without a mature primary care system, Chinese consumers rely more on online doctor reviews to search for a doctor than their Western counterparts. Second, as a developing country, China usually has no formal pen-and-paper-based postvisit surveys to let patients review their health care providers. Some hospitals or clinics may provide a pen-and-paper-based “comment book” in the hospital lobby for patients to leave comments. But this is very informal and most patients ignore the comment book because most hospital lobbies are busy and crowded. Therefore, online reviewing may be the only way, or may be the first time, a Chinese patient can feel free to comment on their doctors with a structured measure. Also, the Good Doctor website was the only online doctor review website in China for a number of years, which may have allowed the Good Doctor website to accumulate more data. Again, further evidence and research are needed to answer the questions of why Chinese patients review their doctors online and how accurate the reviews are.

There has been no study about how Chinese consumers use or look at online doctor reviews or online health care information, and what factors may affect Chinese patients to participate in online doctor reviews. Some research has shown that 59% of American adults used the Internet for health care information and 16% of American adult Internet users have consulted online doctor reviews [33]. A cross-sectional survey conducted in a town in the United Kingdom suggested that the relationship between doctors and their patients may play a role in the patients’ intention to use online doctor review websites [34]. We should expect that more and more studies, either qualitative or quantitative, will investigate what Chinese consumers think about online doctor reviews and how they use them.

Many studies have found that most online doctor reviews are very positive [7,35-38]. Similarly, the majority of the online doctor reviews on the Good Doctor website in China were very positive, too. As Figures 3-6 exhibit, on the Good Doctor website across the four major medical specialties, 88% of the treatment effect evaluations and 91% of the bedside manner evaluations were positive. As well, 75% and 86% of ratings were of the highest level for treatment effect and for bedside manner, respectively. Similar distributions were seen for the five small medical specialties. These were all higher than those of the United States [2] or Germany [4], where 50% and 80% of evaluations, respectively, were in the two best rating categories. Different specialty areas may have variations, but these variations were small on the Good Doctor website. A qualitative study based on a randomly selected sample from online doctor reviewing websites in the United States showed that the majority of online doctor reviews were positive, and in addition to the direct interaction between doctors and patients, staff, access, and convenience all affected patients’ reviews of their doctors [37]. Another study also showed that the UK National Health Service Choice website allowed patients to evaluate their family physicians online and the ratings for all the questions were also quite positive [3]. Thus, Chinese consumers are not different in the positively dominated online reviews of their doctors compared to their Western counterparts, but do seem to give a higher number of, and more positive, evaluations of their doctors. An experimental study in Germany suggested that more reviews may lead to more positive perspectives of a doctor [39]. One research study conducted in a metropolitan area in the United States also showed that a physician’s bedside manner and professional knowledge would significantly lead to a higher rating [7]. Reasons why Chinese patients give more positive reviews could be because of cultural differences or the website review procedures. On the Good Doctor website, although the online evaluation is anonymous to the public, the reviewers are requested to leave a phone number so the webmaster can confirm the truthfulness of the review, if needed. This might lead some conservative people to be cautious, thus they may not leave negative reviews. It’s possible that even the name of the website, Good Doctor, may indicate some signal to the consumers as to the nature of the review they should leave. Again, further studies are needed to investigate why the majority of Chinese online doctor reviewers tend to give very positive online doctor reviews.

Conclusions

In summary, many Chinese consumers have reviewed their doctors online as their Western counterparts have done. By April 11, 2014, 314,624 doctors from almost every medical specialty in China were listed on the Good Doctor website for Chinese consumers to review. There were 731,316 records of quantitative review, including both treatment effects and bedside manner, and 772,979 records of narrative textual review on 117,624 doctors from nine major specialty areas and many small unlisted specialty areas. The first contribution of this study is that it is the first, or one of the first, studies to examine the current status of online doctor reviews in China. Second, empirically, this study shows that like other countries, online doctor reviews in China covered almost all major medical specialties. Gynecology-obstetrics-pediatrics, surgery, and traditional Chinese medicine were more likely to be reviewed than the combined uncommon specialty areas, and gynecology-obstetrics-pediatrics and traditional Chinese medicine received more reviews than the combined specialty areas. But another major specialty area, internal medicine, was less likely to be reviewed than the combined specialty areas. All of the model estimates, except for surgery for the quantitative reviews, were statistically significant at the 5% level. Third, again like other countries, the majority of online doctor reviews were positive on the Good Doctor website. And finally, this study shows that the number of doctors may reach a stable level on the Good Doctor website and the number of reviews has been increasing.

Limitations

This research has limitations. First, all the data used for analysis were from one single website, the Good Doctor, although this website is the largest and the first online doctor reviewing website in China. The website’s design change, database change, and strategy change may affect consumers’ decisions to post a review or not, or to post a positive or negative review. Second, the online doctor reviews were anonymous and there was no way to verify the truthfulness, hence, some of the reviews could have possibly been manipulated with some intention. However, the Good Doctor website does have a policy to check the reliability of reviews by asking the reviewers to leave a phone number on the website, which is not available to the public but only to the webmaster, in order to do random callbacks to verify the truthfulness of the reviews. Third, many doctors received a very limited number of reviews, on average 6 reviews per rated doctor, and those small numbers of reviews may not reflect the reality, or may only partially reflect the reality, of the doctors’ patient populations. Fourth, although China has the largest Internet population in the world, Internet accessibility is still low compared to developed countries around the world—45.8% of the Chinese population has Internet accessibility versus 84% in Germany, 84.2% in the United States, and 89.8% in the United Kingdom [19]. Also, Internet accessibility within China is not equally distributed. On average, about 71% of Internet users are city residents and about 29% are from the countryside, in contrast to 53% of the population being city residents and 47% living in the countryside [32]. Therefore, the digital divide may be preventing many consumers in the countryside from reviewing their doctors online in China.

Acknowledgments

I wish to give great thanks for the longstanding support from Mr Hang Wang, the founder of the Good Doctor website, and his technical team for this research.

Conflicts of Interest

None declared.

References

1. 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]
2. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
3. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]
4. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
5. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
6. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
7. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: analysis of 2185 reviews. Am J Orthop (Belle Mead NJ) 2014 Aug;43(8):359-363. [Medline: 25136868]
8. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009;338:b1033. [Medline: 19293224]
9. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]
10. O'Connor D. American Medical News. 2011 Aug 08. Rated negatively online? What's a physician to do? URL: http://www.amednews.com/article/20110808/profession/308089945/5/ [accessed 2015-04-21] [WebCite Cache ID 6Xxpqls6n]
11. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
12. de Barra M, Eriksson K, Strimling P. How feedback biases give ineffective medical treatments a good reputation. J Med Internet Res 2014;16(8):e193 [FREE Full text] [doi: 10.2196/jmir.3214] [Medline: 25147101]
13. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]

14. Van de Belt TH, Engelen LJ, Berben SA, 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 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
15. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki C, Kummervold PE, Rasmussen J, et al. Online communication between doctors and patients in Europe: status and perspectives. J Med Internet Res 2010;12(2):e20 [FREE Full text] [doi: 10.2196/jmir.1281] [Medline: 20551011]
16. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/~/media//Files/Reports/PIP_HealthOnline.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqOIj8tY]
17. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
18. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]
19. The World Bank. 2013. Internet users (per 100 people) URL: http://data.worldbank.org/indicator/IT.NET.USER.P2 [accessed 2015-02-22] [WebCite Cache ID 6WWUrs3OY]
20. The Economist. 2014 Jul 26. DocAdvisor: Patients around the world are starting to give doctors a piece of their mind. The result should be better care URL: http://www.economist.com/news/international/21608767-patients-around-world-are-starting-give-doctors-piece-their-mind-result [accessed 2015-04-26] [WebCite Cache ID 6Y48Z8Rqm]
21. Eggleston K, Ling L, Qingyue M, Lindelow M, Wagstaff A. Health service delivery in China: a literature review. Health Econ 2008 Feb;17(2):149-165. [doi: 10.1002/hec.1306] [Medline: 17880024]
22. General Public Review Web. Shanghai, China: Dianping URL: http://www.dianping.com/aboutus [accessed 2015-03-27] [WebCite Cache ID 6XLX9TVq7]
23. Barboza D. The New York Times. New York, NY: The New York Times Company; 2010 Jan 13. Baidu's gain from departure could be China's loss URL: http://www.nytimes.com/2010/01/14/technology/companies/14baidu.html?_r=1 [accessed 2015-04-21] [WebCite Cache ID 6Xxt1DKC8]
24. Chinese Traditional Medicine Review Web. URL: http://www.zydp.org/ [accessed 2015-05-18] [WebCite Cache ID 6YckppQG6]
25. Schedule Appointment. URL: http://www.guahao.com [accessed 2015-05-18] [WebCite Cache ID 6Ycl0j8Bz]
26. Hao Dai Fu (The Good Doctor). URL: http://www.haodf.com/ [accessed 2015-05-18] [WebCite Cache ID 6Yclb3j5p]
27. Hospital Management Institute, Fudan University. 2013. Top 100 hospitals ranking in China URL: http://www.fudanmed.com/institute/news2013-2.aspx [accessed 2015-03-29] [WebCite Cache ID 6XOLvZ4Ib]
28. Hospital Grade Level Management Evaluation. Beijing, China: National Health and Family Planning Commission; 1998.
29. China Health and Family Planning Statistical Year Book. Beijing, China: National Health and Family Planning Committee; 2013.
30. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
31. Detz A, López A, Sarkar U. Long-term doctor-patient relationships: patient perspective from online reviews. J Med Internet Res 2013;15(7):e131 [FREE Full text] [doi: 10.2196/jmir.2552] [Medline: 23819959]
32. China Internet Network Information Center. Beijing, China: CNNIC; 2014. Annual Report 2014 URL: http://www.cnnic.cn/gywm/zzkw/cnnicndbg/201505/P020150508352283152629.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqQ4Rv3s]
33. Fox S. The Social Life of Health Information. Washington, DC: Pew Internet & American Life Project; 2011. URL: http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2015-05-27] [WebCite Cache ID 6YqQMDvxN]
34. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2:e001493.
35. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
36. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
37. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
38. Black EW, Thompson LA, Saliba H, Dawson K, Black NM. An analysis of healthcare providers' online ratings. Inform Prim Care 2009;17(4):249-253. [Medline: 20359403]
39. Grabner-Kräuter S, Waiguny MK. Insights into the impact of online physician reviews on patients' decision making: randomized experiment. J Med Internet Res 2015;17(4):e93 [FREE Full text] [doi: 10.2196/jmir.3991] [Medline: 25862516]

Abbreviations

API: application programming interface
G-OB-P: gynecology-obstetrics-pedicatrics
ID: infectious disease
IM: internal medicine
OB: obstetrics
OP: orthopedics
OR: odds ratio
TCM: traditional Chinese medicine

Edited by G Eysenbach; submitted 22.02.15; peer-reviewed by M Emmert, D Hanauer; comments to author 15.03.15; revised version received 25.04.15; accepted 10.05.15; published 01.06.15.

Please cite as:
Hao H
The Development of Online Doctor Reviews in China: An Analysis of the Largest Online Doctor Review Website in China
J Med Internet Res 2015;17(6):e134
URL: http://www.jmir.org/2015/6/e134/
doi:10.2196/jmir.4365
PMID:26032933

Corrigenda and Addenda

Correction: A Text Messaging-Based Smoking Cessation Program for Adult Smokers: Randomized Controlled Trial

Michele Ybarra[1], PhD; A Tülay Bağci Bosi[2], PhD; Josephine Korchmaros[1], PhD; Salih Emri[3], MD

[1]Center for Innovative Public Health Research, San Clemente, CA, United States
[2]Hacettepe University, Department of Public Health, Ankara, Turkey
[3]School of Medicine, Department of Chest Diseases, Hacettepe University, Ankara, Turkey

Corresponding Author:
Michele Ybarra, PhD
Center for Innovative Public Health Research
555 N El Camino Real A347
San Clemente, CA, 92672
United States
Phone: 1 8773026858 ext 801
Email: Michele@InnovativePublicHealth.org

Related Article:

Correction of: http://www.jmir.org/2012/6/e172/

(J Med Internet Res 2015;17(6):e125) doi:10.2196/jmir.4655

The authors of “A Text Messaging-Based Smoking Cessation Program for Adult Smokers: Randomized Controlled Trial” (*J Med Internet Res* 2012;14(6):e172) have overlooked errors in the Results section during the submission and proofreading process. The percentage of intervention participants who are married should be 55.3% (42) instead of 68.4% (52) in Table 2. The P value in Table 4 for Females, ITT analysis, should be P=.05; instead of P=.53. In the sentence, “Finally, compared to the national population of smokers in Turkey [2], the study sample was more educated (eg, 32% of smokers in Turkey have a university education, while 66% of trial participants had a university education)”, 66% should be changed to 56%. In addition, the corresponding author no longer has a fax number, therefore this has been removed from the original published paper. These errors have been corrected in the online version of the paper on the JMIR website on June 5, 2015, together with publishing this correction notice. A correction notice has been sent to PubMed and the corrected full-text has been resubmitted to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 15.05.15; this is a non–peer-reviewed article;accepted 15.05.15; published 05.06.15.

Please cite as:
Ybarra M, Bağci Bosi AT, Korchmaros J, Emri S
Correction: A Text Messaging-Based Smoking Cessation Program for Adult Smokers: Randomized Controlled Trial
J Med Internet Res 2015;17(6):e125
URL: http://www.jmir.org/2015/6/e125/
doi:10.2196/jmir.4655
PMID:26047263

Corrigenda and Addenda

Metadata Correction: Comparative and Cost Effectiveness of Telemedicine Versus Telephone Counseling for Smoking Cessation

Kimber P Richter[1], MPH, PhD (Psych); Theresa I Shireman[1], PhD, RPh; Edward F Ellerbeck[1], MD, MPH; A Paula Cupertino[1], PhD; Delwyn Catley[2], PhD; Lisa Sanderson Cox[1], PhD; Kristopher J Preacher[3], PhD; Ryan Spaulding[4], PhD; Laura M Mussulman[1], MA, MPH; Niaman Nazir[1], MBBS, MPH; Jamie J Hunt[5], PhD; Leah Lambart[1], MPH

[1]University of Kansas Medical Center, Department of Preventive Medicine and Public Health, Kansas City, KS, United States
[2]University of Missouri, Kansas City, Department of Psychology, Kansas City, MO, United States
[3]Vanderbilt University, Department of Psychology & Human Development, Nashville, TN, United States
[4]University of Kansas Medical Center, Center for Community Engagement, Kansas City, KS, United States
[5]University of Missouri, Kansas City, School of Nursing and Health Studies, Kansas City, MO, United States

Corresponding Author:
Kimber P Richter, MPH, PhD (Psych)
University of Kansas Medical Center
Department of Preventive Medicine and Public Health
3901 Rainbow Boulevard
Kansas City, KS, 66160
United States
Phone: 1 9135882718
Fax: 1 9135882780
Email: krichter@kumc.edu

Related Article:

Correction of: http://www.jmir.org/2015/5/e113/

(J Med Internet Res 2015;17(6):e124) doi:10.2196/jmir.4642

The authors of “Comparative and Cost Effectiveness of Telemedicine Versus Telephone Counseling for Smoking Cessation” (J Med Internet Res 2015;17(5):e113) inadvertently omitted Delwyn Catley, PhD (University of Missouri, Kansas City, Department of Psychology, Kansas City, MO, United States) from the list of authors during the submission process. The author Catley should have been added after A Paula Cupertino in the original published manuscript. In addition, the affiliation for the author Sherman should have been the same as the author Richter. Last, the telephone and fax numbers were not the most up to date numbers. These should be Phone: 1 9135882718, Fax: 1 9135882780 instead of Phone: 1 9134490157, Fax:1 9134490157. These errors have been corrected in the online version of the paper on the JMIR website on June 15, 2015, together with publishing this correction notice. This was done after submission to PubMed Central and other full-text repositories. This correction notice has been submitted to PubMed, the original paper resubmitted to PubMed Central, and the metadata has been resubmitted to CrossRef with publishing this correction notice.

Edited by G Eysenbach; submitted 11.05.15; this is a non–peer-reviewed article;accepted 15.05.15; published 15.06.15.

Please cite as:
Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, Preacher KJ, Spaulding R, Mussulman LM, Nazir N, Hunt JJ, Lambart L
Metadata Correction: Comparative and Cost Effectiveness of Telemedicine Versus Telephone Counseling for Smoking Cessation
J Med Internet Res 2015;17(6):e124
URL: http://www.jmir.org/2015/6/e124/
doi:10.2196/jmir.4642
PMID:26075409

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org