Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 16 (2014), Issue 1 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Short Paper

Corrigenda and Addenda

Original Paper

Skin Self-Examination Education for Early Detection of Melanoma: A Randomized Controlled Trial of Internet, Workbook, and In-Person Interventions

June K Robinson[1*], MD; Rikki Gaber[1*], MA; Brittney Hultgren[2*], MS; Steven Eilers[1*], BA; Hanz Blatt[1*], MA; Jerod Stapleton[3*], PhD; Kimberly Mallett[2*], PhD; Rob Turrisi[2*], PhD; Jenna Duffecy[4*], PhD; Mark Begale[4*]; Mary Martini[1*], MD; Karl Bilimoria[5*], MD; Jeffrey Wayne[5*], MD

[1]Northwestern University, Department of Dermatology, Feinberg School of Medicine, Chicago, IL, United States
[2]Pennsylvania State University, Prevention Research Center, Alcohol and Skin Cancer Projects, State College, PA, United States
[3]Rutgers Cancer Institute of New Jersey, Rutgers School of Public Health, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States
[4]Center for Behavioral Intervention Technologies (CBITs), Department of Preventive Medicine, Northwestern University, Chicago, IL, United States
[5]Feinberg School of Medicine, Surgical Oncology Division, Northwestern University, Chicago, IL, United States
[*]all authors contributed equally

Corresponding Author:
June K Robinson, MD
Northwestern University
Department of Dermatology
Feinberg School of Medicine
Arkes Family Pavilion, Suite 1260
676 N St Clair St
Chicago, IL, 60611
United States
Phone: 1 312 926 7449
Fax: 1 312 695 9179
Email: june-robinson@northwestern.edu

Abstract

Background: Early detection of melanoma improves survival. Since many melanoma patients and their spouses seek the care of a physician after discovering their melanoma, an ongoing study will determine the efficacy of teaching at-risk melanoma patients and their skin check partner how to conduct skin self-examinations (SSEs). Internet-based health behavior interventions have proven efficacious in creating behavior change in patients to better prevent, detect, or cope with their health issues. The efficacy of electronic interactive SSE educational intervention provided on a tablet device has not previously been determined.

Objective: The electronic interactive educational intervention was created to develop a scalable, effective intervention to enhance performance and accuracy of SSE among those at-risk to develop melanoma. The intervention in the office was conducted using one of the following three methods: (1) in-person through a facilitator, (2) with a paper workbook, or (3) with a tablet device used in the clinical office. Differences related to method of delivery were elucidated by having the melanoma patient and their skin check partner provide a self-report of their confidence in performing SSE and take a knowledge-based test immediately after receiving the intervention.

Methods: The three interventions used 9 of the 26 behavioral change techniques defined by Abraham and Michie to promote planning of monthly SSE, encourage performing SSE, and reinforce self-efficacy by praising correct responses to knowledge-based decision making and offering helpful suggestions to improve performance. In creating the electronic interactive SSE educational intervention, the educational content was taken directly from both the scripted in-person presentation delivered with Microsoft PowerPoint by a trained facilitator and the paper workbook training arms of the study. Enrollment totaled 500 pairs (melanoma patient and their SSE partner) with randomization of 165 pairs to the in-person, 165 pairs to the workbook, and 70 pairs to electronic interactive SSE educational intervention.

Results: The demographic survey data showed no significant mean differences between groups in age, education, or income. The tablet usability survey given to the first 30 tablet pairs found that, overall, participants found the electronic interactive intervention easy to use and that the video of the doctor-patient-partner dialogue accompanying the dermatologist's examination was particularly helpful in understanding what they were asked to do for the study. The interactive group proved to be just as good as the workbook group in self-confidence of scoring moles, and just as good as both the workbook and the in-person intervention groups in self-confidence of monitoring their moles. While the in-person intervention performed significantly better on a skill-based quiz, the electronic interactive group performed significantly better than the workbook group. The electronic interactive and in-person interventions were more efficient (30 minutes), while the workbook took longer (45 minutes).

Conclusions: This study suggests that an electronic interactive intervention can deliver skills training comparable to other training methods, and the experience can be accommodated during the customary outpatient office visit with the physician. Further testing of the electronic interactive intervention's role in the anxiety of the pair and pair-discovered melanomas upon self-screening will elucidate the impact of these tools on outcomes in at-risk patient populations.

ClinicalTrial: ClinicalTrials.gov NCT01013844; http://clinicaltrials.gov/show/NCT01013844 (Archived by WebCite at http://www.webcitation.org/6LvGGSTKK).

(J Med Internet Res 2014;16(1):e7) doi:10.2196/jmir.2883

KEYWORDS

melanoma; early detection of cancer; educational techniques; health education; behavioral research; online education; electronic interactive education

Introduction

Background

Early detection of melanoma improves survival. In those diagnosed with Stage IA melanoma, the survival rate 10 years after diagnosis was estimated at higher than 95%, which declined to less than 60% when diagnosis is delayed (Stage IIB, C). Previous research has found that melanoma patients and their spouses often discover the melanoma and then seek the care of a physician [1,2]. Any enhanced surveillance for melanoma patients has great potential to detect future melanomas in their earliest stages when treatment prognosis is optimal. Robinson et al found that melanoma patients and their skin check partners can learn the skin self-examination (SSE) skills and that these skills can be improved by routine practice [3,4]. The most appropriate delivery method for the educational and skills training intervention remains unclear; thus, we present an interim analysis of an ongoing clinical trial following subjects at 4-month intervals for 2 years to determine the efficacy of teaching at-risk melanoma patients and their skin check partners how to conduct deliberate skin self-examinations.

The purpose of this study is to develop and evaluate an electronic interactive educational program that provides SSE skills training for at-risk melanoma patients and their partners in an effort to help them enhance their participation and performance of SSE. Internet interventions designed to educate patients about health promotion and personal health care are becoming widely available. A meta-analysis of 85 different Internet intervention programs that promoted health behavior change found that those grounded in the theory of planned behavior were among the most efficacious programs [5]. Interventions modeled after the theory of planned behavior often include content related to modeling skills, prevention planning, goal setting, action planning, and feedback on performance [6]. When this theory is applied to the health care education of patients at risk for developing a second melanoma, the education and skills training in performing SSE may prove extremely beneficial in improving the survival of the patient [7].

This study compares different methods of educating both patients with a history of melanoma and their skin check partners about the need and proper techniques to conduct skin examinations and identify clinically suspicious moles. This study used three educational interventions with identical education components delivered in different formats: (1) in-person PowerPoint presentation delivery by a trained facilitator, (2) a self-guided paper workbook [8], and (3) an electronic interactive intervention delivered on a tablet personal computer (PC). We compared the pairs' confidence in performing SSE and knowledge-based performance immediately after receiving the interventions delivered in-person through a facilitator, with a paper workbook, or a tablet device used in the clinical office in order to elucidate differences related to the method of delivery of the intervention.

Objective

An electronic interactive educational program was created to develop a scalable, effective intervention to enhance performance and accuracy of SSE among those at risk to develop melanoma.

Methods

Design

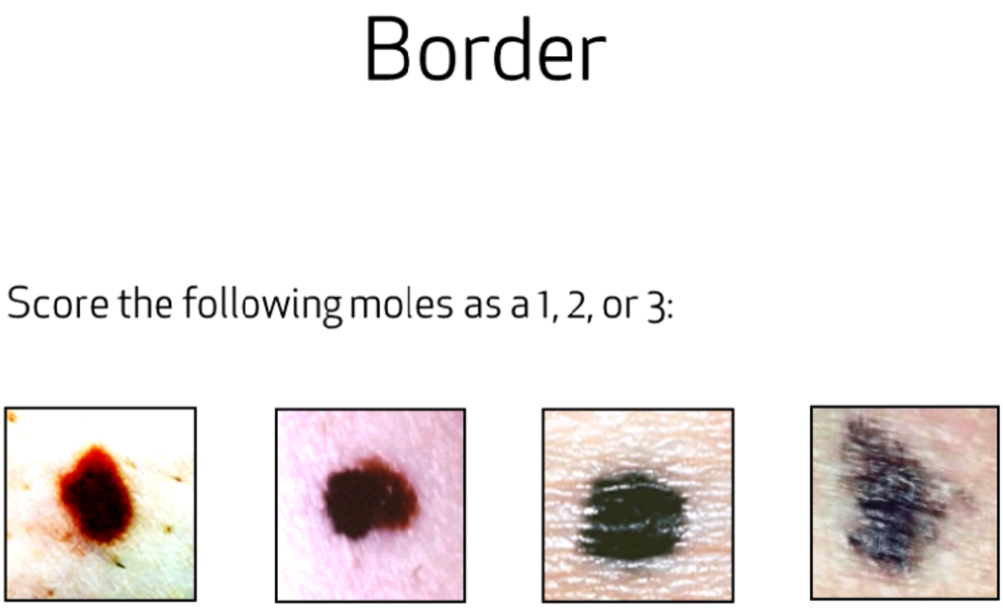
Educational content delivered in each intervention included five elements: modeling of skills, prevention planning, goal setting, action planning, and feedback on performance, which are crucial in creating behavior and attitude change [5]. The three interventions used the following 9 of the 26 behavioral change techniques defined by Abraham and Michie to enhance patient performance and confidence outcomes [9]: (1) information given about melanoma, ability to find it early if the person is trained and actively looking, (2) information about the spread of cancer to other organs, surgery required, mortality statistics, (3) encourage pair to act within 2 weeks of the skills training session, (4) praise selection of correct scores for border, color, and diameter and encourage to keep trying, (5) trained research assistants or physician demonstrated how to correctly perform the behavior, (6) set a goal of checking five moles together each month for the first 4 months, (7) pair uses a body map and scorecard to record their findings, (8) show partners how to check for moles in places patients cannot see themselves, and (9) 4-month follow-up appointments to encourage continued SSE (behavior change). Five moles were selected as the goal to examine each month because in our previous work the majority of melanoma patients had five moles that the physician determined needed to be followed for change [3,8].

The inclusion criteria were having a history of Stage 0 to IIB melanoma and being at least 6 weeks after surgical treatment of melanoma, able to see to read a newspaper, fluent in English, age 21-80, and having significant other person (spouse, partner, close relative) who was willing to participate in the research. Exclusion criteria were a history of Stage III or greater melanoma; ocular, genital, and oral melanoma; being overburdened with comorbid disease; unable to see to read a newspaper; not fluent in English; unable to participate in conversation at a sixth grade language level due to cognitive impairment; and did not have a significant other person (spouse, partner, close relative) who was willing to participate in the research. The Institutional Review Board of Northwestern University approved the study.

Content across all three interventions was controlled for and remained as similar as possible under the restraints of different delivery systems. With respect to delivery, the electronic interactive educational program was delivered on two tablets—one to each member of the pair simultaneously with earphones. The tablet was a Samsung CE 0168 (US $249.00) touch screen device held in the person's hand or set upright on the table for viewing. The tablet PC was used to deliver the intervention in the ambulatory care setting but was not given to the pair to take home. In order to control for dosage effects when comparing with the in-person education, the electronic interactive education was also withheld from the intervention group so it could not be viewed at home. Due of the mode of delivery, the electronic interactive program contained more uniformity in the interactive components than the in-person intervention. While content itself remained similar, the electronic interactive program had more multidimensional learning tools, such as a narrated video presentation and animated graphics to enhance the educational experience. In all three interventions, when learning about scoring the border, color, and diameter of moles, patients and their partners were quizzed on the material and received feedback as to whether they answered correctly (see Figure 1). Finally, pairs also learned about two benign lesions commonly found on skin, seborrheic keratoses and cherry angiomas, and how to differentiate them from moles that may be suspicious (see Figure 2).

One of the modules in the intervention asked pairs to begin their SSE by picking five moles on the patient's body to score and watch for change over the initial 4 months. This action planning was further encouraged by the use of a "skin diary of a body map and a scorecard" to record their observations of their chosen moles, with the goal of having the pair make a commitment to conduct the skin examination each month. In the electronic interactive intervention, use of the skin diary was modeled. Pairs participating in all three interventions were also told that they would review the moles they chose to observe with the doctor at the 4-month visit. In the electronic interactive intervention, the doctor-patient-partner interaction that occurred at the 4-month follow-up visit was demonstrated in a video as a learning experience (see Figure 3). This representation of the demonstration of scoring the border, color, and diameter of the mole with the doctor during the appointment allowed the pair to see what was expected of them at the visit, as well as what they could expect to learn from the doctor during the visit.

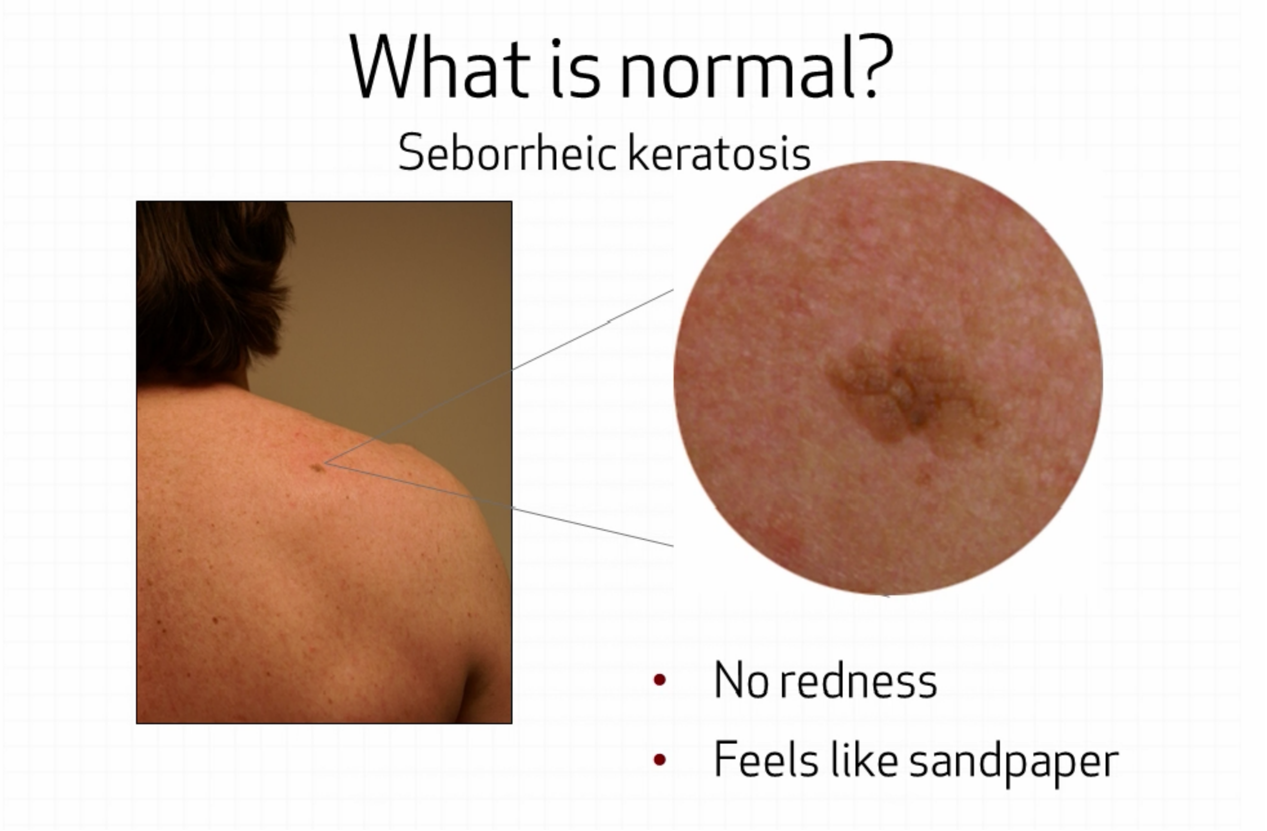
Figure 1. Interactive education on borders of moles.



Border

Score the following moles as a 1, 2, or 3:

Figure 2. Teaching about normal growths.



What is normal?

Seborrheic keratosis

- No redness
- Feels like sandpaper

Figure 3. Video component of education program.

Measures

Participants completed a pre-intervention survey that included demographic questions as well as questions pertaining to their use of Internet and mobile phone technology (see Table 1). After the survey, the pairs were randomized to one of the four arms (three interventions and customary education provided in clinical practice, which was the control condition). Pairs randomized to the control condition did not receive the measures reported in this interim analysis; therefore, there is no presentation of data for the control group. Each individual in the pair completed an evaluation of the intervention consisting of 22 items with a 5-point Likert scale assessing the following domains: clarity of explanation of the ABCDE rule, scoring the three features (border, color, and diameter) and using body maps to locate the mole, confidence in performing SSE, and ease of understanding the content (see Table 2). After participating in the education, the participating pairs were given a skills quiz consisting of life-size photographs of five pigmented lesions printed in color with questions. The pair was asked to discuss and come to agreement on their response (Figure 4). The first 30 pairs in the electronic interactive intervention group also completed a further Tablet Usability survey giving their opinions on the intervention. The Tablet Usability survey was administered to six groups of five pairs, and analysis was performed after each set of five pairs to identify changes that needed to be made to the tablet presentation design, but not content.

Table 1. Demographic information[a].

Characteristics		Workbook (n=165), n (%)	Electronic interactive (n=70), n (%)	In-person (n=165), n (%)
Education				
	No high school	0 (0)	0 (0)	1 (0.6)
	Some high school	2 (1.2)	0 (0)	1 (0.6)
	High school graduate	4 (2.4)	3 (4.3)	9(5.4)
	Some post-high school education	19 (11.5)	5(7.2)	27(15.4)
	College graduate	76 (46.1)	19(27.1)	63(38.2)
	Graduate degree	57(34.5)	26(37.1)	64(38.8)
	Unanswered	2 (1.2)	0(0)	2 (1.2)
Income (USD)				
	<$10,000	1(0.6)	1(1.4)	2 (1.2)
	$10,000-$19,999	1 (0.6)	1(1.4)	1 (0.6)
	$20,000-$34,999	8(4.8)	1(1.4)	8 (4.8)
	$35,000-$50,999	8(4.8)	2(2.8)	12(7.3)
	$51,000-$100,000	45 (27.3)	26(37.1)	43(26.1)
	>$100,000	94(57.0)	22(31.4)	97(58.8)
	Unanswered	3 (1.8)	0 (0)	4 (2.4)
Age (years), mean (SD)		55.19 (14.12)	55.19 (14.12)	54.70 (14.84)

[a] No significant mean differences between groups in age, education, or income.

Table 2. Skills performance[a].

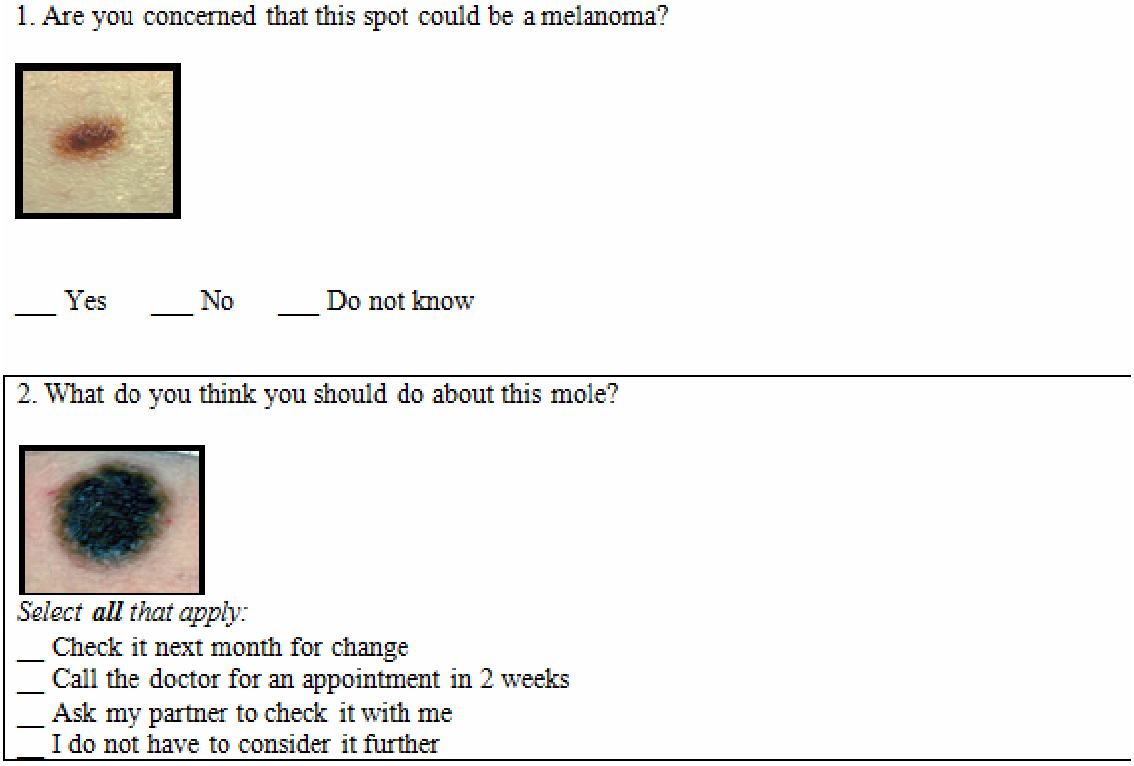
	Workbook (n=165)	Electronic interactive (n=70)	In-person (n=165)
Mean (SD)	3.02 (0.82)[b]	3.25 (0.68)	3.47 (0.69)[c]

[a] 1-way analysis of variance showed a significant difference in means ($P<.001$); $F_{10.31}=14.98$.

[b] Significantly different from In-person group.

[c] Significantly different from Workbook group.

Figure 4. Sample questions from the skills quiz.



1. Are you concerned that this spot could be a melanoma?

___ Yes ___ No ___ Do not know

2. What do you think you should do about this mole?

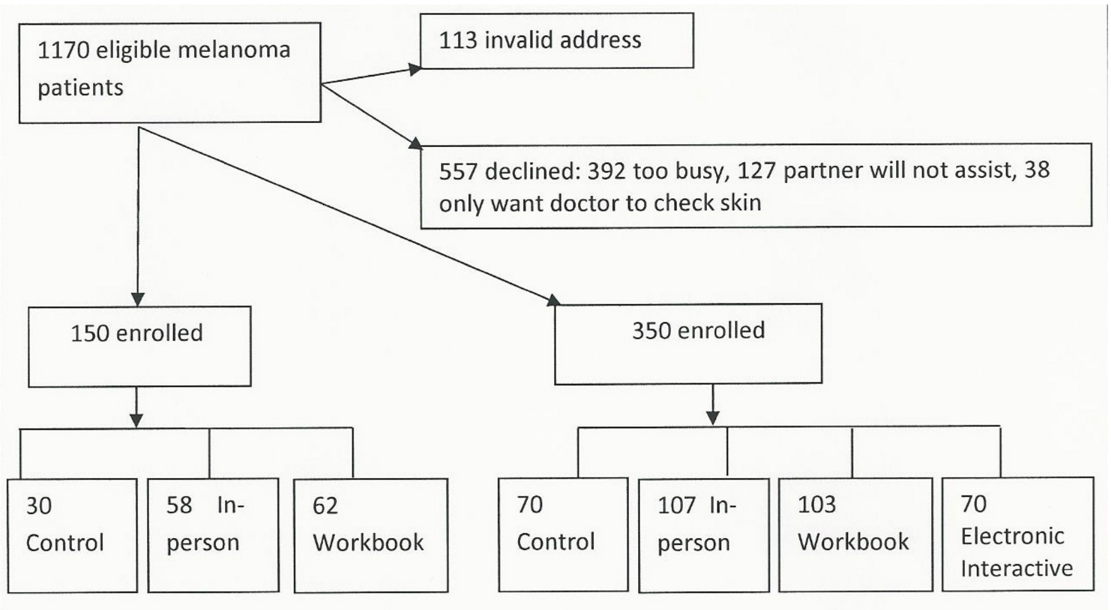
*Select **all** that apply:*

__ Check it next month for change
__ Call the doctor for an appointment in 2 weeks
__ Ask my partner to check it with me
__ I do not have to consider it further

Recruitment

Pairs were recruited through the Northwestern Memorial Faculty Foundation Department of Dermatology at Northwestern University Feinberg School of Medicine in Chicago, IL. Enrollment totaled 500 pairs with randomization of 165 pairs to the in-person, 165 pairs to the workbook, and 70 pairs to the electronic interactive pilot intervention arms, with 100 pairs to the control arm or education as usually performed (see Figure 5). While accruing the initial 150 pairs, we noticed that the pairs reported reading the workbook was a burden; therefore, it was decided to explore the use of an electronic interactive intervention delivered with a tablet during an office visit as another easily disseminated form of the intervention. The sample randomized to the fourth arm (electronic interactive intervention) is smaller than the other two interventions because it was not intended to be subjected to factor analysis, for example, influence of dyadic relationship. Rather, the fourth arm (electronic interactive intervention) will be compared with the others solely on the basis of performance of SSE and accuracy of SSE in comparison with the dermatologist's skin examination over the 2 years.

Figure 5. Flow chart of study recruitment.



Results

The demographic survey data showed no significant mean differences between groups in age, education, or income. In all groups, the majority of individuals had a college degree or higher and had an income of US $51,000 or higher (see Table 1). While over half of study participants randomized to the electronic interactive intervention (39/70, 55%) owned a tablet, 72% (50/70) of study participants had never used their tablet for interactive educational materials. The most common use of the tablet was to read books and email. This research experience with the electronic interactive intervention was a new use of the tablet for most participants.

The Tablet Usability survey given to the first 30 electronic interactive intervention pairs found that, overall, participants found the electronic interactive intervention easy to use and that the video of the doctor-patient-partner dialogue accompanying the dermatologist's examination was particularly helpful in understanding what they would be doing. The initial 15 pairs experienced technical issues in using the tablet, for example, delay in loading images and failure to load chapters without rebooting the device. These technical concerns were resolved before continuing with the rest of the participating pairs.

The duration of the pairs' exposure to the intervention was about 30 minutes for both the in-person and the electronic interactive interventions. Those receiving the workbook intervention spent about 45 minutes reading it. Knowledge of SSE, using the scoring system, and body map diaries were similar across all three interventions. Clarity of the explanation of evolution and scoring in the in-person group were significantly different from the workbook and electronic interactive groups (see Table 3). While the in-person group reported a slightly higher score in the ease of understanding, it was not significantly different from the electronic interactive group, and both the in-person and electronic interactive groups reported a significantly greater ease in understanding than the workbook group. There was a significant difference in the self-efficacy of the pairs receiving the three interventions. The in-person group had significantly higher reports of confidence to monitor and score their moles compared to the workbook group, whereas the electronic interactive group did not significantly differ from either the workbook groups on scoring the moles.

Table 3. Significant differences among interventions[a,b] (1-way analysis of variance showed a significant difference in means; questions were evaluated using a 5-point Likert scale coded between –2 [strongly disagree] and +2 [strongly agree]).

Item	Workbook (n=165), mean (SD)	Electronic interactive (n=70), mean (SD)	In-person (n=165), mean (SD)	*F* (df)
Evolution was explained clearly with useful examples	1.80 (0.40)[c]	1.81 (0.40)[c]	1.96 (0.23)[a,b]	10.23 (2.63)
Scoring was explained clearly with useful examples	1.60 (0.40)[c]	1.68 (0.55)[c]	1.94 (0.26)[a,b]	19.27 (6.91)
I feel that I am now better able to monitor my moles over time	1.53 (0.58)[c]	1.60 (0.49)	1.78 (0.45)[a]	10.09 (4.72)
I feel that I am now better able to score moles accurately	1.41 (0.66)[c]	1.47 (0.58)[c]	1.70 (0.51)[a,b]	10.45 (5.14)
Overall, the information was easy to understand	1.62 (0.57)[c]	1.72 (0.50)	1.79 (0.46)[a]	4.43 (1.26)

[a] Significantly different from Workbook group.

[b] Significantly different from Internet group.

[c] Significantly different from In-person group.

The skills quiz assesses the pairs' use of the ABCDE scoring rules to make correct decisions about five pigmented lesion examples. Two types of questions were used. There were five questions of the first type with each having one correct answer (n=5) and two questions of the second type with one question having two correct responses and one having one correct response (n=3). A higher score indicated better decisions (range 0-5) (see Figure 4). The results of the skills quiz across the three intervention groups showed that the pairs in the in-person group scored significantly higher than the pairs in the workbook group. However, the electronic interactive intervention group scored significantly higher than their counterparts in the workbook group (1-way analysis of variance showed a significant difference in means, *P*=.01) (see Table 2).

Discussion

Principal Findings

In this study, the immediate post-intervention performance skills of participants in the electronic interactive group were no different than those in the in-person education group, but higher than the workbook group. This skills test performance suggested an electronic interactive intervention can be more effective than passively reading information and just as effective as an in-person intervention. The advantages of the electronic interactive intervention were the consistent delivery and the reported understanding of the SSE task, especially evolution. It is possible that the virtual placement of the physician in the videography provided a more personal experience for the patient by modeling the expected interaction between the doctor and the pair. Thus, the electronic interactive intervention may motivate the pair to practice SSE in order to form opinions and questions to discuss with the doctor in a way that cannot be achieved via workbook or the in-person intervention. The electronic interactive intervention and the in-person intervention require less time from the participant (30 minutes) than reading the workbook (45 minutes). In addition to the workbook being a burden to the participant, the electronic interactive intervention will be a more efficient use of personnel and space in the office than an in-person intervention. Overall, this study suggests that an electronic interactive intervention can deliver skills training comparable to other training methods and the experience can be accommodated during the customary outpatient office visit with the physician.

The electronic interactive intervention was effective in teaching skills to patients and their skin check partners. Even pairs who had no prior experience with receiving educational information from a tablet PC or no prior experience with a tablet PC readily began using the device. Thus, it was not essential that the user be familiar with using the technology platform to benefit from education materials on these platforms. This allowed the electronic interactive intervention to be effective across a range of socioeconomic classes or among elderly patients with less experience with technology. However, because the intervention required a secure Internet connection, the feasibility of using this tablet intervention to supplement the usual patient education depended on whether the doctor's office was able to provide Internet access. To help overcome this obstacle, our technology team determined a method to load the program into the tablet's cache, where it was stored and accessible without an Internet connection. An obstacle to this method is that it requires significant coordination with respect to the technology team, the tablet, and the office using the device, as each update to the tablet and other random events can break down the consistent experience observed in the Internet-based format. While this cache-loaded format was not used in the study, the experience is exactly the same as the Internet-based format.

The electronic interactive intervention may prove to be a more consistent method of teaching patients and their skin-check partner how to conduct monthly SSEs than having a conversation with a health care provider (HCP) and could be provided at less cost than delivery by an HCP. Further, the electronic interactive intervention may be made available via the Internet to a wider audience than may be seen by an HCP.

In the future, a technologically more sophisticated and interactive Internet-based intervention may approach that of a face-to-face intervention in promoting self-efficacy and clarity of explanation of content. A caveat is that much of the cost associated with the Internet intervention is incurred at the design and development stage rather than in delivering the intervention to the individual patient. Further, the cost per patient or per life saved decreases significantly when such a program is scaled to a population, as can be accomplished via the Internet.

Limitations

Since the fiscal commitment to developing the electronic interactive intervention was small, the electronic interactive intervention had limited interactivity and personalization. The sample size was limited to one institution and the population was of a slightly higher socioeconomic status (income of US $51,000 and higher) than the national income average (US $42,979), which may limit generalizability of our findings, especially in regards to prior experience with the tablet technology [10]. In the future, this pilot electronic interactive intervention will be redesigned to be more interactive, which is expected to improve the pairs' self-efficacy such that the intervention will be comparable with the in-person education.

As the study has not finished the 2-year period of follow-up with the participating pairs, conclusions cannot be made about the accuracy of SSE as performed by the pairs in comparison with the diagnosis of the pigmented lesion by a dermatologist.

Conclusions

This study suggested that an electronic interactive intervention can deliver skills training comparable to other training methods and the experience can be accommodated during the customary outpatient office visit with the physician. Reducing the burden on the HCP's time to educate patients in person may make it possible to reach large numbers of people at risk for melanoma. As such, this method can be used as the primary intervention, or as an easily accessible Internet-based approach to reinforce education provided in the office. The potential for dissemination of this information to people who are at risk of developing melanoma could greatly improve the rate of early detection of the skin cancer, as well as reduce patient anxiety about reoccurrence as the patient feels more confident and competent at performing SSE.

Acknowledgments

Our study was supported by 5R01 CA-154908 to June K Robinson, MD, from the National Cancer Institute.

The research was supported in part by resources provided by the Northwestern University Skin Disease Research Center (P30AR057216), Chicago, IL, with support from National Institutes of Health (NIH)/National Institute of Arthritis and Musculoskeletal and Skin Disorders (NIAMS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the Northwestern University Skin Disease Research Center or the NIH/NIAMS.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [11].

[PDF File (Adobe PDF File), 987KB - jmir_v16i1e7_app1.pdf]

References

1. Brady MS, Oliveria SA, Christos PJ, Berwick M, Coit DG, Katz J, et al. Patterns of detection in patients with cutaneous melanoma. Cancer 2000 Jul 15;89(2):342-347. [Medline: 10918164]
2. Berwick M, Begg CB, Fine JA, Roush GC, Barnhill RL. Screening for cutaneous melanoma by skin self-examination. J Natl Cancer Inst 1996 Jan 3;88(1):17-23 [FREE Full text] [Medline: 8847720]
3. Robinson JK, Stapleton J, Turrisi R. Relationship and partner moderator variables increase self-efficacy of performing skin self-examination. J Am Acad Dermatol 2008 May;58(5):755-762 [FREE Full text] [doi: 10.1016/j.jaad.2007.12.027] [Medline: 18258332]
4. Robinson JK, Turrisi R. Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. Arch Dermatol 2006 Apr;142(4):447-452. [doi: 10.1001/archderm.142.4.447] [Medline: 16618863]
5. 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]
6. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
7. Titus-Ernstoff L, Perry AE, Spencer SK, Gibson J, Ding J, Cole B, et al. Multiple primary melanoma: two-year results from a population-based study. Arch Dermatol 2006 Apr;142(4):433-438. [doi: 10.1001/archderm.142.4.433] [Medline: 16618861]
8. Robinson JK, Turrisi R, Mallett K, Stapleton J, Pion M. Comparing the efficacy of an in-person intervention with a skin self-examination workbook. Arch Dermatol 2010 Jan;146(1):91-94 [FREE Full text] [doi: 10.1001/archdermatol.2009.350] [Medline: 20083705]
9. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
10. U.S. Social Security Administration. National average wage index. 2012. URL: http://www.ssa.gov/OACT/cola/AWI.html [accessed 2013-08-12] [WebCite Cache ID 6IqRX2Uz2]
11. 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

HCP: health care provider
PC: personal computer
SSE: skin self-examination

Edited by G Eysenbach; submitted 16.08.13; peer-reviewed by R Dellavalle, MD, PhD, MSPH, A Bhatia, MD, PhD, L Ferris, MD, PhD, J Makin; comments to author 10.09.13; revised version received 20.10.13; accepted 06.11.13; published 13.01.14.

Please cite as:
Robinson JK, Gaber R, Hultgren B, Eilers S, Blatt H, Stapleton J, Mallett K, Turrisi R, Duffecy J, Begale M, Martini M, Bilimoria K, Wayne J
Skin Self-Examination Education for Early Detection of Melanoma: A Randomized Controlled Trial of Internet, Workbook, and In-Person Interventions
J Med Internet Res 2014;16(1):e7
URL: http://www.jmir.org/2014/1/e7/
doi:10.2196/jmir.2883
PMID:

Original Paper

Using Ecological Momentary Assessment to Test the Effectiveness of a Web-Based Brief Alcohol Intervention Over Time Among Heavy-Drinking Students: Randomized Controlled Trial

Carmen Voogt[1*], MSc; Emmanuel Kuntsche[1,2*], PhD; Marloes Kleinjan[1*], PhD; Evelien Poelen[1*], PhD; Rutger Engels[1*], PhD

[1]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

[2]Addiction Switzerland, Research Institute, Lausanne, Switzerland

[*]all authors contributed equally

Corresponding Author:
Carmen Voogt, MSc
Behavioural Science Institute
Radboud University Nijmegen
Montessorilaan 3
Nijmegen, 9104 6500 HE
Netherlands
Phone: 31 24 36 12705
Fax: 31 24 36 12776
Email: c.voogt@bsi.ru.nl

Abstract

Background: Web-based brief alcohol interventions are effective in reducing alcohol use among students when measured at limited follow-up time points. To date, no studies have tested Web-based brief alcohol intervention effectiveness over time by using a large number of measurements.

Objective: Testing whether the What Do You Drink (WDYD) Web-based brief alcohol intervention can sustain a reduction in alcohol use among heavy-drinking students aged 18-24 years at 1-, 3-, and 6-month follow-up intervals.

Methods: A purely Web-based, 2-arm, parallel-group randomized controlled trial applying an ecological momentary assessment approach with 30 weekly measurements was conducted in the Netherlands (2010-2011). Participants were recruited offline and online. A total of 907 participants were randomized into the experimental condition (n=456) including the single-session and fully automated WDYD intervention, or into the control condition (n=451) including assessment only. Weekly alcohol consumption and frequency of binge drinking were the self-assessed outcome measures.

Results: Attrition rates of the 907 participants were 110 (12.1%), 130 (14.3%), and 162 (17.9%) at 1-, 3-, and 6-month follow-up intervals, respectively. Latent growth curve analyses according to the intention-to-treat principle revealed that participants in the experimental condition had significantly lower weekly alcohol consumption compared to participants in the control condition that was sustained at 3-month follow-up (intercept=–2.60, *P*<.001; slope=0.16, *P*=.08). Additional linear regression analyses indicated that this intercept difference resulted from significantly higher levels of alcohol units per week for participants in the control condition compared to those in the experimental condition at 1-month (beta=–2.56, SE 0.74, Cohen’s *d*=0.20, *P*=.001), 3-month (beta=–1.76, SE 0.60, Cohen’s *d*=0.13, *P*=.003), and 6-month (beta=–1.21, SE 0.58, Cohen’s *d*=0.09, *P*=.04) follow-up intervals. Latent growth curve analyses further indicated that participants in the experimental condition had a significantly lower frequency of binge drinking compared to participants in the control condition that was sustained at 6-month follow-up (intercept=–0.14, *P*=.01; slope=0.004, *P*=.19). This intercept difference resulted from higher levels in this outcome for participants in the control condition relative to participants in the experimental condition at 1-month (beta=–1.15, SE 0.06, Cohen’s *d*=0.16, *P*=.01), 3-month (beta=–0.12, SE 0.05, Cohen’s *d*=0.09, *P*=.01), and 6-month (beta=–0.09, SE 0.05, Cohen’s *d*=0.03, *P*=.045) follow-up intervals.

Conclusions: The WDYD intervention was shown to be effective in preventing an increase in weekly alcohol consumption and frequency of binge drinking directly after the intervention. This effect was sustained 3 and 6 months after the intervention.

Trial Registration: Netherlands Trial Register NTR2665; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2665 (Archived by WebCite at http://webcitation.org/6LuQVn12M).

(J Med Internet Res 2014;16(1):e5) doi:10.2196/jmir.2817

KEYWORDS

intervention study; drinking; students

Introduction

Given the high prevalence and social and economic costs attributable to heavy drinking among young adults, there is an urgent need for adequate interventions [1-3]. The widespread growth and availability of computer technology and the Internet has provided the opportunity to deliver interventions via the Web [4], which is advantageous for young adults because it allows them to access information at a self-selected time and place while remaining anonymous [5]. Web-based brief alcohol interventions have been found to be effective in reducing the quantity and frequency of alcohol use among heavy-drinking young adults and students [6-9]. However, despite the demonstrated effectiveness of these types of interventions, findings have to be interpreted with caution because of the way alcohol use and intervention effectiveness are assessed. First, alcohol use is typically assessed over relatively long recall periods (eg, 30 days) and participants are often asked to report the average number of alcohol units they consumed in a usual week. This can result in measurement errors because precise recall of alcohol use decreases after 2 or 3 days due to memory deficits [10]. Second, the fluctuating nature of alcohol use among students because of calendar-specific events [1,11] is often overlooked because intervention effectiveness is measured at limited follow-up time points (ie, 1 short and several longer follow-ups of approximately 1-12 months after the intervention). The use of limited follow-up time points not only disregards important drinking events, but also increases the danger of drawing inaccurate conclusions about intervention effectiveness. In a previous study, we recognized the disadvantage of using only 2 follow-up time points (ie, 1 and 6 months) when testing the effectiveness of the What Do You Drink (WDYD) Web-based brief alcohol intervention. Our baseline assessment was completed during a usual drinking period without any remarkable events, whereas our first 1-month follow-up assessment coincided with carnival, a 4-day event associated with excessive drinking. The observed increase in alcohol use was likely the consequence of the selection of this particular week to test intervention effectiveness rather than the WDYD intervention, which aims to detect and reduce heavy drinking in young adults [12]. The single-session and fully automated WDYD intervention was developed in collaboration with the Trimbos Institute (Netherlands Institute of Mental Health and Addiction) by using the intervention mapping protocol [13]. Content is based on motivational interviewing principles [14] and parts of the I-Change model [15] in which knowledge, social norms, and self-efficacy are embedded as the most changeable determinants of behavior change [12]. Despite the sound theoretical background of the WDYD intervention, no significant main effects on alcohol use were found when using 2 follow-up time points with short recall periods of 7 days [5]. Third, individual changes in alcohol use and intervention effectiveness over time remain unnoticed when using limited follow-up time points. Information on these changes can be extremely valuable for determining the time at which the intervention effects have stopped and the time at which booster sessions may be needed to strengthen and/or extend intervention effects.

In evaluating intervention effectiveness, it appears to be important to use short recall periods to reduce measurement errors and to include a large number of measurements to consider the fluctuating nature of alcohol use over time and capture important drinking events. Higher precision in establishing intervention effectiveness can be achieved by means of ecological momentary assessment (EMA) and latent growth curve (LGC) modeling techniques. The ecological aspect of EMA is that data are collected in real-life settings at strategically selected moments in time [16]. The momentary aspect of EMA implies that the assessment of alcohol use focuses on participants' current or recent state. In addition, EMA is characterized by repeated and a large number of measurements over time and often used equivalent to experience sampling methods (ESM), a systematic way for participants to report on their ongoing alcohol use behavior [17,18]. LGC modeling techniques allow for estimation of average growth trajectories (ie, mean intercepts and slopes) of alcohol use over time as well as individual differences in these trajectories (ie, intercept and slope variances) [19,20]. The estimation of variances in growth trajectories increases the reliability of outcome measures. This is not possible with traditional statistical techniques that are often used to test intervention effectiveness, such as repeated-measures ANOVA [21], because they only provide mean growth patterns and treat variances as error [22]. Although the advantages of employing EMA and LGC modeling techniques are evident, most trials on Web-based brief alcohol interventions used long recall periods with limited follow-up time points and traditional techniques [21] to test intervention effectiveness [4]. To our knowledge, this is the first study to test whether a Web-based brief alcohol intervention can sustain a reduction in alcohol use among heavy-drinking students at 1-, 3-, and 6-month follow-up intervals. To test the effectiveness of the WDYD intervention over time, we conducted 30 weekly EMA measurements for 6 months through online surveys and LGC analyses to model individual change in weekly alcohol consumption and frequency of binge drinking at 1-, 3-, and 6-month follow-up intervals by condition. The rationale of reporting over 3 time periods was to gain insight into how long the intervention effects sustained and to limit the chance of reporting outlier trajectories. We hypothesized that participants in the experimental condition would reduce their alcohol use (intercept) compared to participants in the control condition directly after exposure to the WDYD intervention. Based on

Web-based brief alcohol interventions that have produced long-term effects [8,9], it was hypothesized that the reduction would be sustained (slope) at the 6-month follow-up interval.

Methods

Study Design

A 2-arm, parallel-group randomized controlled trial applying an EMA approach with 30 weekly EMA measurements was conducted online in the Netherlands (2010-2011) to test whether the WDYD intervention could sustain a reduction in alcohol use among heavy-drinking students at 1-, 3-, and 6-month follow-up intervals. This trial was purely Web-based because there were no face-to-face components in the intervention and for assessing the outcome measures.

Procedure and Participants

A convenience sampling strategy was used to recruit heavy-drinking students offline by distributing flyers at universities and higher professional education institutions (ie, universities of applied sciences) online by sending emails with information about the study from September to December 2010. The cover story was that students had to evaluate newly developed health education materials addressing alcohol use and that they had to judge these materials to reduce the risk of social desirability bias. Students were blinded to the aim of the study until the end of the EMA study. Interested students were referred to an email address and were sent a detailed description of the study by email. To be included in the study, students had to (1) be between ages 18 and 24 years, (2) report heavy drinking in the past 6 months, (3) be ready to change their alcohol use, (4) have daily access to the Internet (and be literate), and (5) sign an online informed consent form. Heavy drinking was defined as consuming more than 14 (females) or 21 (males) glasses of standard alcohol units per week and/or consuming 5 or more glasses of standard alcohol units per occasion at least 1 day per week [23]. Students reporting a score of 20 or higher on the Alcohol Use Disorders Identification Test (AUDIT) [24], and/or receiving treatment for alcohol-related problems were excluded from the study and advised to seek treatment because the WDYD intervention was developed to reduce heavy drinking rather than problem drinking.

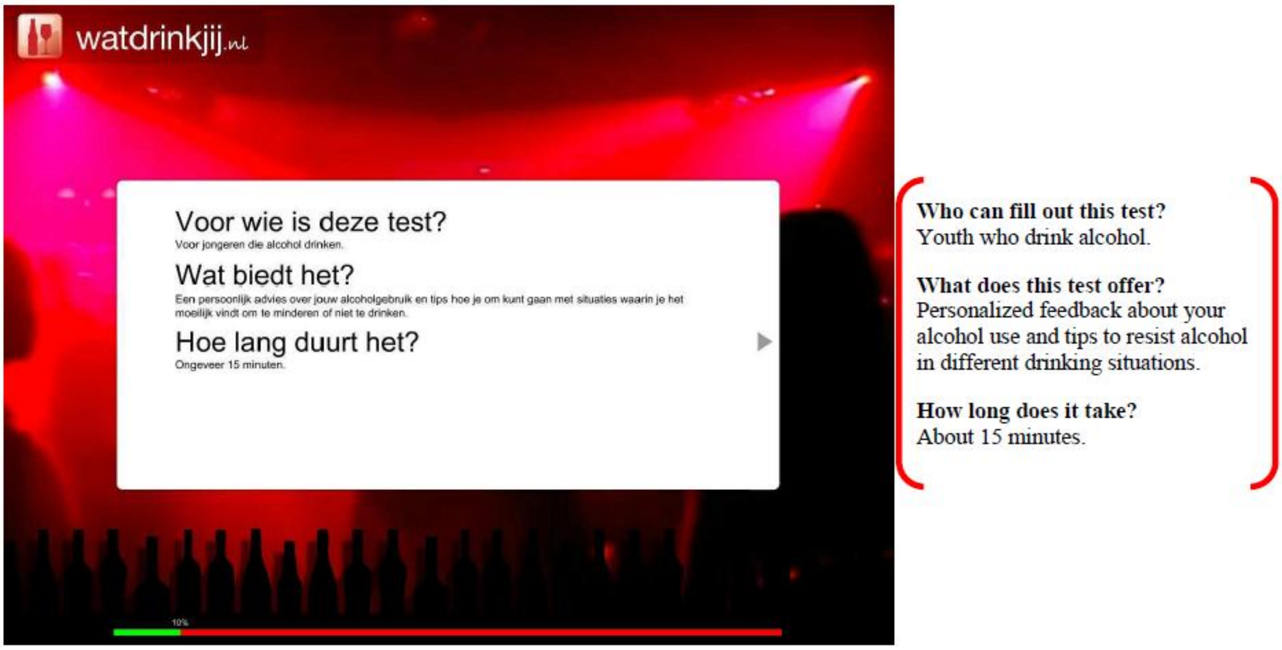
A sample size of 908 participants was necessary to detect an increase in the percentage of participants showing low-risk drinking guidelines after 1 month of 42% in the experimental condition versus 31% in the control condition [25] with a 2-sided 5% significance level and a power of 80%, given an anticipated dropout rate of 30% after randomization. Students who met the inclusion criteria were randomly assigned to the WDYD intervention condition (n=456) or to the control condition (n=451) in blocks of 4 using a computerized random number generator by an independent researcher of the Behavioural Science Institute who could not influence or predict the randomization result. Participants were not blinded to randomization results. Randomization was stratified by sex and educational level before the baseline assessment in January 2011 [23].

In total, 30 weekly EMA measurements were conducted online from January to August 2011, to assess outcome measures with 4 pretests and 26 posttests. After the 4 pretests in January, participants in the experimental condition were exposed to the WDYD intervention, whereas those in the control condition received assessment only. Directly after intervention exposure in the first week of February, participants in both conditions received the first posttest. One week after the intervention, all participants received weekly EMA posttest measurements for 6 months from February to August. EMA measurements were assessed on Monday mornings. All participants received an email with the instructions on the use of the survey, and they were asked to respond to the survey before midnight. Each survey took approximately 10 minutes to complete and contained identical questions about participants' weekly alcohol consumption, frequency of binge drinking, and drinking refusal self-efficacy. In addition, extended surveys were administered at baseline assessment, immediately after the intervention, and at 1 and 6 months after the intervention. These extended surveys included additional questions concerning alcohol-related cognitions, cost-effectiveness, and problem drinking. Completion time of the extended surveys was approximately 20 minutes. Paper-and-pencil surveys with identical content were provided to participants in case they were unable to access the Internet. Participants who failed to complete the survey on Mondays received a short text message on their mobile phones on Tuesdays to remind them. Those who still did not complete the survey on Tuesdays were reminded by a telephone call on Wednesdays. On average, 11% (range 7%-17%) of the surveys were completed on Tuesdays and Wednesdays instead of Mondays. When participants completed at least 28 of 30 surveys, they received €100 as an incentive, as stated in the informed consent. Ethical approval was provided by the Ethical Committee of the Faculty of Social Sciences at Radboud University Nijmegen (ECG30062011). This trial is registered at the Netherlands Trial Register (NTR2665) as mentioned in the trial protocol [23].

Interventions

Participants assigned to the experimental condition used the WDYD intervention. The first part of the WDYD intervention focuses on increasing the users' awareness of the potential problems, consequences, and risks associated with their drinking behavior. It contains a home page and a screening test with personalized feedback delivered in a nonjudgmental, nonconfrontational, and nonaversive way (see Figure 1). The screening test included participants' self-reported name, sex, age, education level, weight, alcohol use, readiness to change alcohol use, average expenses for consumed alcohol beverages, and descriptive social norms. Personalized feedback consisted of advice about drinking according to low-risk drinking guidelines [26], personal drinking profile (quantity-frequency consumed in past year), estimates of calorie intake, increases in weight, money expenses because of drinking, and a comparison of personal use rates with the national norms of same-sex peers to correct misperceptions of descriptive social norms (see Figure 2). The personalized feedback was based on the individuals' personal situation, implying that the WDYD intervention was tailored. The second part of WDYD focused on setting and maintaining drinking goals (see Figure 3) and strengthening users' drinking refusal self-efficacy to succeed and maintain drinking goals by providing tips to resist alcohol in different drinking situations (see Figure 4). Participants were able to track their progress through the WDYD intervention, which took approximately 20 minutes to complete. A full description of the WDYD intervention is given elsewhere [12]. Participants assigned to the control condition received assessment only.

Figure 1. Screenshot and English translation of the What Do You Drink home page.



Voor wie is deze test?
Voor jongeren die alcohol drinken.

Wat biedt het?
Een persoonlijk advies over jouw alcoholgebruik en tips hoe je om kunt gaan met situaties waarin je het moeilijk vindt om te minderen of niet te drinken.

Hoe lang duurt het?
Ongeveer 15 minuten.

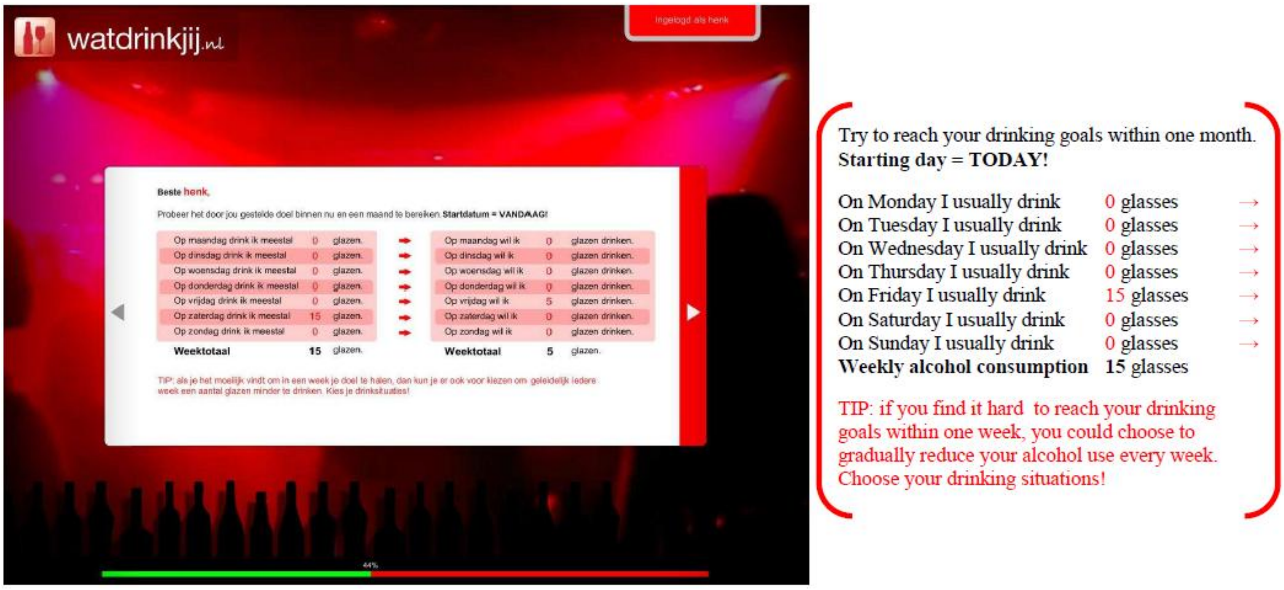
Who can fill out this test?
Youth who drink alcohol.

What does this test offer?
Personalized feedback about your alcohol use and tips to resist alcohol in different drinking situations.

How long does it take?
About 15 minutes.

Figure 2. Screenshot and English translation of personalized feedback on the What Do You Drink website.

Figure 3. Screenshot and English translation of drinking goals on the What Do You Drink website.



Beste henk,

Probeer het door jou gestelde doel binnen nu en een maand te bereiken. **Startdatum = VANDAAG!**

Op maandag drink ik meestal	0	glazen.	➡	Op maandag wil ik	0	glazen drinken.
Op dinsdag drink ik meestal	0	glazen.	➡	Op dinsdag wil ik	0	glazen drinken.
Op woensdag drink ik meestal	0	glazen.	➡	Op woensdag wil ik	0	glazen drinken.
Op donderdag drink ik meestal	0	glazen.	➡	Op donderdag wil ik	0	glazen drinken.
Op vrijdag drink ik meestal	0	glazen.	➡	Op vrijdag wil ik	5	glazen drinken.
Op zaterdag drink ik meestal	15	glazen.	➡	Op zaterdag wil ik	0	glazen drinken.
Op zondag drink ik meestal	0	glazen.	➡	Op zondag wil ik	0	glazen drinken.
Weektotaal	**15**	glazen.		**Weektotaal**	**5**	glazen.

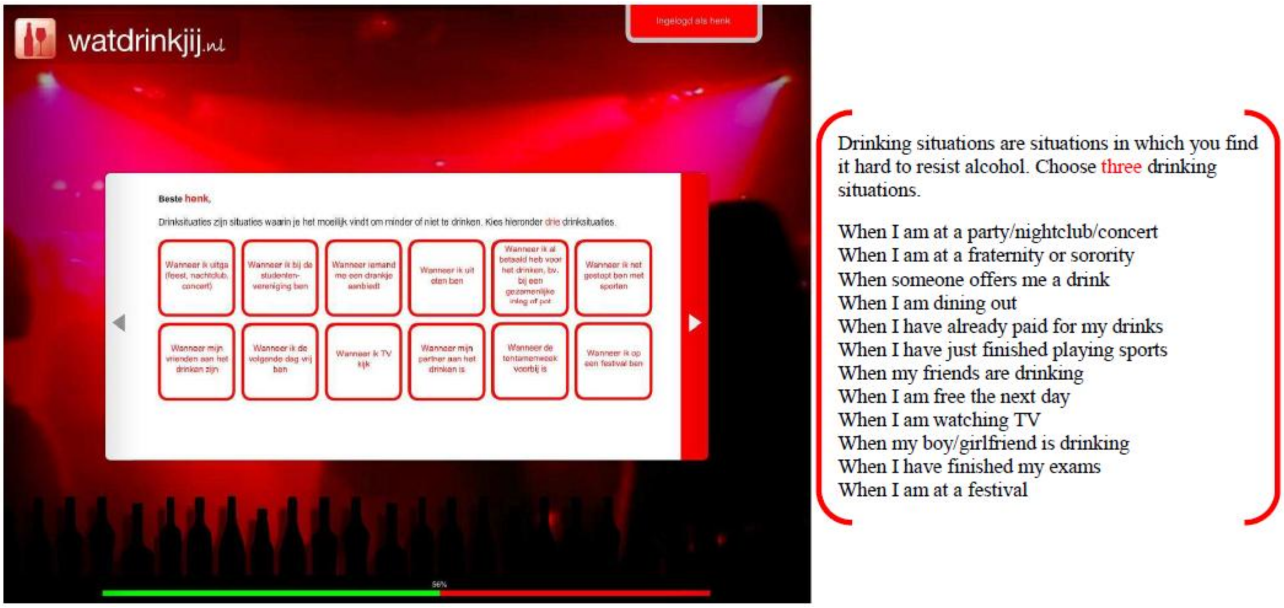
TIP: als je het moeilijk vindt om in een week je doel te halen, dan kun je er ook voor kiezen om geleidelijk iedere week een aantal glazen minder te drinken. Kies je drinksituaties!

Try to reach your drinking goals within one month.
Starting day = TODAY!

On Monday I usually drink	0 glasses	→
On Tuesday I usually drink	0 glasses	→
On Wednesday I usually drink	0 glasses	→
On Thursday I usually drink	0 glasses	→
On Friday I usually drink	15 glasses	→
On Saturday I usually drink	0 glasses	→
On Sunday I usually drink	0 glasses	→
Weekly alcohol consumption	**15** glasses	

TIP: if you find it hard to reach your drinking goals within one week, you could choose to gradually reduce your alcohol use every week. Choose your drinking situations!

Figure 4. Screenshot and English translation of the overview of drinking situations on the What Do You Drink website.



Beste henk,

Drinksituaties zijn situaties waarin je het moeilijk vindt om minder of niet te drinken. Kies hieronder drie drinksituaties.

Wanneer ik uitga (feest, nachtclub, concert)
Wanneer ik bij de studenten-vereniging ben
Wanneer iemand me een drankje aanbiedt
Wanneer ik uit eten ben
Wanneer ik al betaald heb voor het drinken, bv. bij een gezamenlijke inleg of pot
Wanneer ik net gestopt ben met sporten
Wanneer mijn vrienden aan het drinken zijn
Wanneer ik de volgende dag vrij ben
Wanneer ik TV kijk
Wanneer mijn partner aan het drinken is
Wanneer de tentamenweek voorbij is
Wanneer ik op een festival ben

Drinking situations are situations in which you find it hard to resist alcohol. Choose three drinking situations.

When I am at a party/nightclub/concert
When I am at a fraternity or sorority
When someone offers me a drink
When I am dining out
When I have already paid for my drinks
When I have just finished playing sports
When my friends are drinking
When I am free the next day
When I am watching TV
When my boy/girlfriend is drinking
When I have finished my exams
When I am at a festival

Outcome Measures

Weekly alcohol consumption and frequency of binge drinking were the self-assessed primary outcome measures through online surveys. Weekly EMA measurements were used to assess the primary outcome measures over time at 1-, 3-, and 6-month follow-up intervals.

Weekly alcohol consumption, defined as the mean number of glasses of standard alcohol units consumed in the previous 7 days, was assessed using the Dutch version of the Alcohol Weekly Recall [27]. Participants were asked to indicate retrospectively the exact number, size, and type of alcohol beverage they consumed on each day of the previous 7 days. An overview of standard units for various beverages was provided to guarantee standardized responses. In total, 1.47% of the participants scored above 3 standard deviations of the sample mean of weekly alcohol consumption, but they were given a score exactly at 3 standard deviations above the sample mean of weekly alcohol consumption to retain outliers in the analyses (resulting range 0-109) [28] and to handle outliers in accordance with our previous studies [5,29]. Binge drinking frequency, defined as the number of days in the previous week on which participants drank 5 or more glasses of standard alcohol units per occasion [5], was assessed on an 8-point Likert scale ranging from 0=never to 7=every day.

Analyses

Data were analyzed according the intent-to-treat (ITT) principle. Missing data were handled by using multiple imputations using the predictive mean matching method [5,30]. Twenty imputed datasets were evaluated with $P<.05$ as the criterion for statistical significance by averaging the results (ie, pooling). First, descriptive analyses involving *t* tests and chi-square tests were conducted to explore whether the randomization resulted in a balanced distribution of participants' demographic characteristics and alcohol use (ie, weekly alcohol consumption and frequency of binge drinking) across conditions at baseline assessment. Loss to follow-up was also examined with attrition analyses using 1-, 3-, and 6-month follow-up intervals as outcome measures and demographic characteristics, alcohol use, and condition status (intervention vs control) as predictors. Second, LGC analyses were conducted to model individual change in alcohol use over time by condition at 1-, 3-, and 6-month follow-up intervals. LGC analyses were conducted over 3 time periods to limit the chance of reporting outlier trajectories. The models without condition status and baseline levels of alcohol use were tested first. Subsequently, the growth curves were regressed on condition status for weekly alcohol consumption and frequency of binge drinking separately while adjusting for baseline levels of alcohol use. A random-effect parameter for educational institutions was not included in the models since variation in participants between institutions was expected to be limited because all participants needed to meet the inclusion criteria of the study.

Unstandardized intercepts, representing alcohol use directly after the intervention, and unstandardized slopes, representing the change of alcohol use over time, were reported. Global fit indexes were used to assess model fit for each construct: chi-square statistic, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) with a cut-off value of ≥0.90 and ≥0.95 for acceptable fit, and Root Mean Square Error of Approximation (RMSEA) with a cut-off value of ≤0.06 for acceptable fit [31]. In parallel with the LGC analyses, linear regression analyses were conducted for weekly alcohol consumption and frequency of binge drinking at 1-, 3-, and 6-month follow-up intervals and presented as unadjusted and adjusted for baseline levels of alcohol use to provide additional specific tests of a difference between the conditions. For the linear regression analyses, unstandardized coefficients (betas), standard errors (SE), and Cohen’s *d* [32] effect sizes were provided. All analyses were performed using Mplus version 6.0 [20].

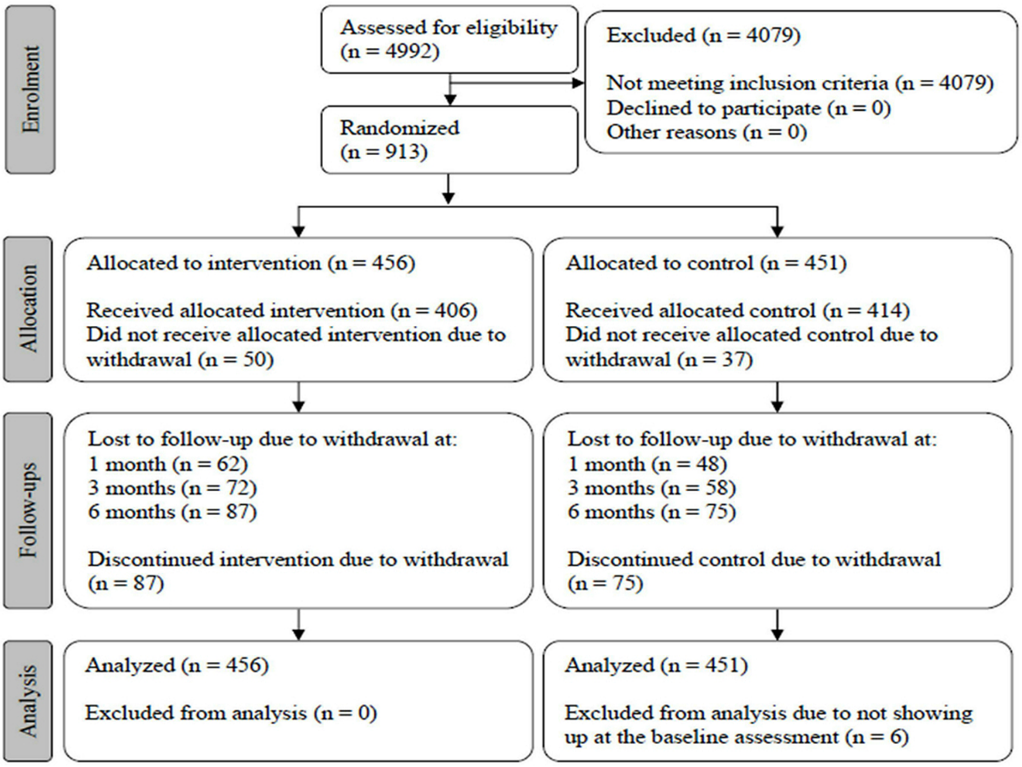
In total, 30 weekly EMA measurements were conducted with 4 pretests and 26 posttests. For the LGC and linear regression analyses, the 4 pretests were aggregated into a baseline score. The first posttest immediately after the intervention was excluded from both analyses because participants reported on outcome measures over the previous week, thereby making it impossible to observe direct intervention effects. Thus, for all analyses, 4 pretests and 25 posttests were used. Only for the additional linear regression analyses, aggregated scores were computed for the 1-month follow-up interval (posttests 1-4), the 3-month follow-up interval (posttests 1-12), and the 6-month follow-up interval (posttests 1-25).

Results

Participant Flow

Figure 5 illustrates the participant flow through the study following the Consolidated Standards of Reporting Trials (CONSORT) guidelines [33] and the data collection with 30 weekly EMA measurements. Originally, 913 students were included in the study. However, 6 students did not fill in the baseline assessment and were excluded from the study. Finally, 907 participants were enrolled in the EMA study, randomized into the experimental condition (n=456, 50.3%) or control condition (n=451, 49.7%), and eligible for the ITT analyses. In total, 82.1% (745/907) completed the baseline assessment and all 25 EMA follow-ups.

Figure 5. Participant flow diagram.



Descriptive Statistics

Of the 907 participants, 547 (60.3%) were male, 667 (73.5%) attended university, and 194 (21.4%) were ready to reduce alcohol use in the near future at baseline assessment. The screening survey was administered between September and December 2010, whereas the baseline assessment was administered in January 2011, which might explain the lower rates of participant's readiness to change alcohol use at baseline assessment. On average, participants were age 20.8 years (SD 1.7). At baseline assessment, participants reported to consume a mean 21.9 (SD 13.5) alcohol units per week and reported to have 1.8 (SD 1.0) occasions in which they drank 5 or more glasses of alcohol units per week (see Table 1). No significant differences emerged between conditions in demographic characteristics and outcome measures at baseline assessment (analyses not shown here).

Table 1. Demographic characteristics and outcome measures at baseline assessment.

Demographic characteristics	Intervention (n=456)	Control (n=451)	Total sample (N=907)
Male, n (%)	275 (60.3)	272 (60.3)	547 (60.3)
Age, mean (SD)	20.9 (1.7)	20.8 (1.7)	20.8 (1.7)
Higher professional education, n (%)	122 (26.8)	118 (26.2)	240 (26.5)
University education, n (%)	334 (73.2)	333 (73.8)	667 (73.5)
Contemplation stage[a], n (%)	93 (20.4)	101 (22.4)	194 (21.4)
Weekly alcohol consumption, mean (SD)	22.2 (12.9)	22.1 (13.8)	21.9 (13.5)
Frequency of binge drinking, mean (SD)	1.8 (1.0)	1.7 (1.1)	1.8 (1.0)

[a] Readiness to change alcohol use was assessed through a question asking the participants which statement applied best to them. Participants selecting "I want to reduce drinking alcohol within the upcoming 6 months" or "I want to reduce drinking alcohol within the upcoming month" were considered to be in the contemplation stage of change, meaning that they were willing to reduce their alcohol use in the near future.

Loss to Follow-Up

Because of lack of interest and time, the attrition rate at the 1-month follow-up interval was 12.1% (110/907; intervention: 62/456, 13.6%; control: 48/451, 10.6%), 14.3% at the 3-month follow-up interval (130/907; intervention: 72/456, 15.8%; control: 58/451, 12.9%), and 17.9% at 6-month follow-up interval (162/907; intervention: 87/456, 19.1%; control: 75/451, 16.6%). Attrition was not related to conditions at 1-, 3-, and 6-month follow-up intervals (χ^2_1=1.9, P=.17; χ^2_1=1.6, P=.21; and χ^2_1=0.9, P=.34). Completers (those who completed the baseline assessment and all 25 EMA follow-ups, n=745) did not differ from noncompleters (n=162) with respect to demographic characteristics (ie, sex: χ^2_1=0.3, P=.56; age: t_{902}=–0.25, P=.80; education: χ^2_1=1.9, P=.17; and readiness to change alcohol use: χ^2_1=0.1, P=.73), and alcohol use (ie, weekly alcohol consumption: t_{903}=0.32, P=.75; frequency of binge drinking: t_{903}=–0.57, P=.57) at baseline assessment. The distribution of the missing values indicated that 87 of 907 participants (9.6%) did not complete the EMA study and that 75 of 907 participants (8.3%) nearly completed the survey (missing 1 or 2 of 30 EMA measurements).

Model Findings

The models for weekly alcohol consumption and frequency of binge drinking without condition status and baseline levels of alcohol use were tested first. The intercept and slope of weekly alcohol consumption were significant (intercept=23.7, P<.001; slope=–0.06, P=.002), indicating that participants consumed 23.7 alcohol units on average and gradually reduced their consumption throughout the 6-month study period (χ^2_{320}=1393.2, P<.001; CFI=0.90; TLI=0.91; RMSEA=0.06). For frequency of binge drinking, a significant intercept and slope was found (intercept=1.9, P<.001; slope=–0.01, P<.001), meaning that the average number of occasions in the previous week that participants had drunk 5 or more glasses of alcohol units was 1.89. Participants' frequency of binge drinking slowly reduced throughout the 6-month study period (χ^2_{320}=904.2, P<.001; CFI=0.92; TLI=0.92; RMSEA=0.05). Next, condition status and baseline levels of alcohol were added to both models. The weekly alcohol consumption model provided an acceptable fit to the data at follow-up assessments, except for the RMSEA at the 1-month follow-up interval. Fit indexes for weekly alcohol consumption were χ^2_9=107.3, P<.001, CFI=0.95, TLI=0.92, and RMSEA=0.11 at the 1-month follow-up interval, χ^2_{93}=644.7, P<.001, CFI=0.91, TLI=0.91, and RMSEA=0.08 at the 3-month follow-up interval, and χ^2_{366}=1451.2, P<.001, CFI=0.91, TLI=0.91, and RMSEA=0.06 at the 6-month follow-up interval.

The LGC analyses revealed that participants in the experimental condition had a significantly lower weekly alcohol consumption compared to participants in the control condition directly after the intervention. The intercept difference in alcohol units between conditions sustained at the 3-month follow-up interval (intercept=–2.60, P<.001; slope=0.16, P=.08), but faded out over time resulting in a significant slope of the LCG at the 6-month follow-up interval (intercept=–2.18, P=.001; slope=0.08, P=.02) (see Table 2 and Figure 6). Linear regression analyses indicated that the intercept difference resulted from significantly higher levels of alcohol units per week for participants in the control condition compared to those in the experimental condition at 1-month (beta=–2.56; SE 0.74; Cohen's d=0.20; P=.001), 3-month (beta=–1.76; SE 0.60;

Cohen’s *d*=0.13; *P*=.003), and 6-month (beta=–1.21; SE 0.58; Cohen’s *d*=0.09; *P*=.04) follow-up intervals (see Table 3).

The frequency of binge drinking model provided an acceptable fit at all 3 follow-up intervals. Fit indexes for frequency of binge drinking were χ^2_9=42.3, *P*<.001, CFI=0.97, TLI=0.95, and RMSEA=0.06 at the 1-month follow-up interval, χ^2_{93}=341.3, *P*<.001, CFI=0.93, TLI=0.93, and RMSEA=0.05 at the 3-month follow-up interval, and χ^2_{366}=956.9, *P*<.001, CFI=0.92, TLI=0.93, and RMSEA=0.04 at the 6-month follow-up interval. According to the LGC analyses, the frequency of binge drinking of participants in the experimental condition was significantly lower compared to participants in the control condition. The intercept difference in frequency of binge drinking was sustained at the 6-month follow-up interval (intercept=–0.14, *P*=.01; slope=0.004, *P*=.19) (see Table 2) and resulted from higher levels in this outcome for participants in the control condition relative to participants in the experimental condition at 1-month, (beta=–1.15; SE 0.06; Cohen’s *d*=0.16; *P*=.01), 3-month (beta=–0.12; SE 0.05; Cohen’s *d*=0.09; *P*=.01), and 6-month (beta=–0.09; SE 0.05; Cohen’s *d*=0.03; *P*=.045) follow-up intervals (see Table 3 and Figure 6).

Table 2. Latent growth curve models presenting alcohol use intercepts and alcohol use slopes of intervention effects on alcohol use at 1-, 3-, and 6-month follow-up intervals (N=907).

Alcohol use intercepts and slopes at follow-up intervals	Weekly alcohol consumption		Binge drinking	
	Unstandardized estimate (SE)	*P*	Unstandardized estimate (SE)	*P*
1 month (posttests 1-4)				
Baseline alcohol use on alcohol use intercept	0.86 (0.03)	<.001	0.71 (0.04)	<.001
Baseline alcohol use on alcohol use slope	0.01 (0.02)	.60	–0.03 (0.02)	.14
Intervention condition on alcohol use intercept	–2.70 (0.89)	.002	–0.21 (0.08)	.01
Intervention condition on alcohol use slope	0.16 (0.44)	.73	0.04 (0.04)	.32
3 months (posttests 1-12)				
Baseline alcohol use on alcohol use intercept	0.87 (0.03)	<.001	0.68 (0.03)	<.001
Baseline alcohol use on alcohol use slope	–0.01 (0.003)	<.001	–0.01 (0.004)	.01
Intervention condition on alcohol use intercept	–2.60 (0.73)	<.001	–0.15 (0.06)	.02
Intervention condition on alcohol use slope	0.16 (0.09)	.08	0.01 (0.01)	.56
6 months (posttests 1-25)				
Baseline alcohol use on alcohol use intercept	0.85 (0.03)	<.001	0.66 (0.03)	<.001
Baseline alcohol use on alcohol use slope	–0.01 (0.001)	<.001	–0.004 (0.002)	.003
Intervention condition on alcohol use intercept	–2.18 (0.65)	.001	–0.14 (0.05)	.01
Intervention condition on alcohol use slope	0.08 (0.04)	.02	0.004 (0.003)	.19

Figure 6. Left: latent growth trajectory for weekly alcohol consumption by condition after 6 months follow-up. Right: latent growth trajectory for frequency of binge drinking by condition after 6 months follow-up.

Table 3. Intervention effects of alcohol use at 1-, 3-, and 6-month follow-ups by condition (intervention vs control): linear regression analyses unadjusted and adjusted for the outcome measures at baseline assessment (N=907).

Intervention effects at follow-up intervals	Group, mean (SD)		Beta (SE)	Cohen's *d*	*P*
	Intervention (n=456)	Control (n=451)			
Weekly alcohol consumption: unadjusted					
Baseline (4 pretests)	22.2 (12.9)	22.1 (13.8)			
1 month (posttests 1-4)	24.0 (15.0)	26.5 (17.4)	–2.44 (1.09)	0.20	.03
3 months (posttests 1-12)	23.1 (13.2)	24.9 (14.7)	–1.66 (0.94)	0.13	.08
6 months (posttests 1-25)	22.9 (13.0)	24.0 (13.7)	–1.11 (0.89)	0.09	.21
Weekly alcohol consumption: adjusted					
Baseline (4 pretests)	22.2 (12.9)	22.1 (13.8)			
1 month (posttests 1-4)	24.0 (15.0)	26.5 (17.4)	–2.56 (0.74)	0.20	.001
3 months (posttests 1-12)	23.1 (13.2)	24.9 (14.7)	–1.76 (0.60)	0.13	.003
6 months (posttests 1-25)	22.9 (13.0)	24.0 (13.7)	–1.21 (0.58)	0.09	.04
Binge drinking: unadjusted					
Baseline (4 pretests)	1.8 (1.0)	1.7 (1.1)			
1 month (posttests 1-4)	1.9 (1.1)	2.0 (1.1)	–0.13 (0.08)	0.16	.08
3 months (posttests 1-12)	1.8 (0.9)	1.9 (1.0)	–0.10 (0.06)	0.09	.12
6 months (posttests 1-25)	1.8 (0.9)	1.8 (0.9)	–0.07 (0.25)	0.03	.25
Binge drinking: adjusted					
Baseline (4 pretests)	1.8 (1.0)	1.7 (1.1)			
1 month (posttests 1-4)	1.9 (1.1)	2.0 (1.1)	–0.15 (0.06)	0.16	.01
3 months (posttests 1-12)	1.8 (0.9)	1.9 (1.0)	–0.12 (0.05)	0.09	.01
6 months (posttests 1-25)	1.8 (0.9)	1.8 (0.9)	–0.09 (0.05)	0.03	.045

Discussion

Principal Results

This study is the first to test whether a Web-based brief alcohol intervention can sustain a reduction in alcohol use among heavy-drinking students at 1-, 3-, and 6-month follow-up intervals by means of an EMA approach with 25 posttests. The WDYD intervention did not reduce weekly alcohol consumption and frequency of binge drinking of participants in the experimental condition compared to participants in the control condition. Instead, the WDYD intervention was shown to be effective in preventing an increase in weekly alcohol consumption and frequency of binge drinking directly after the intervention that was sustained at 3 and 6 months postintervention. Ideally, participants in the experimental condition should reduce their alcohol use and participants in the control condition should stabilize after intervention exposure. However, these results revealed that participants in the experimental condition stabilized, whereas participants in the control condition deteriorated by increasing their alcohol use. Calendar-specific events might explain the increase in alcohol use that occurred among participants in the control condition from the beginning of February. The alcohol use patterns of participants in the control condition were similar to the patterns of binge drinking among freshmen that increased from winter break mid-December to New Year's Eve, subsequently decreased up to the end of January and then increased again to the end of Spring Break in mid-March [34]. In the first week of February, participants in the experimental condition could benefit from the tips of the WDYD intervention to resist alcohol in different drinking situations. Exposure to the WDYD intervention might have led to an increase in drinking refusal self-efficacy, thereby making participants in the experimental condition less susceptible compared to participants in the control condition for calendar-specific events associated with elevated risk of excessive drinking [11]. In addition, binge drinking primarily occurs when students are with friends inside their homes, and outside their homes in bars, at parties, on dates, or during socializing activities [35]. It is reasonable to assume that students perceived the tips to resist alcohol as more relevant when they actually found themselves in drinking situations in which binge drinking occurs. This might explain the short-term preventive effect of the WDYD intervention for weekly alcohol consumption at 3 months postintervention and the long-term preventive effect of binge drinking frequency at 6 months postintervention.

In our previous study, we did not find significant main effects when we tested the effectiveness of the WDYD intervention at 1 and 6 months postintervention [5]. However, by using EMA and LGC modeling techniques, overall significant intervention effects were generated for weekly alcohol consumption and frequency of binge drinking that were sustained 3 and 6 months postintervention. This finding stresses the importance of using a large number of measurements in combination with appropriate statistical techniques to obtain higher precision in intervention effectiveness and minimize the danger of inaccurate conclusions about intervention effectiveness when using limited follow-up time points. Moreover, the use of EMA enables one to examine whether intervention effectiveness on the treatment outcome varied over time and helps determine the time at which the intervention effects stopped and the time at which booster sessions are needed to strengthen and/or extend intervention effects.

Limitations and Strengths

Limitations of this study include the use of a large number of measurements by means of EMA, which might have affected the observed changes in the outcome measures by the act of assessing [36-38], yet participants in both conditions received weekly posttest measurements. If there was assessment reactivity, it could lead to underestimates of the true intervention effect [39]. Also, EMA could impose participant burden and reduce compliance because of the length of the survey entry, the frequency of responses, and the length of the study period [16]. Nonetheless, noncompliance and attrition were low in the current study. It seemed to be important to provide a briefing about the study procedure before the study onset, use short and well-conducted surveys, and offer a monetary incentive after study completion. In addition, the use of EMA might even alleviate sample size requirements because it provides more refined outcome measures that are more sensitive to change, thereby making studies less difficult and less expensive to conduct [16]. Additionally, the effect sizes of the WDYD intervention were small but comparable to those reported in other Web-based brief alcohol interventions [40,41]. Despite the small absolute differences in alcohol use between the conditions, the advantage lies in the inclusion of all the effects of the WDYD intervention over time across a far larger group of heavy drinkers with less serious alcohol-related problems resulting in a greater societal gain than reducing problem drinking among a smaller number of dependent drinkers, known as the prevention paradox, that is used to justify a population strategy of prevention [42]. Further, the representativeness of the study sample might have been affected because of the convenience sampling strategy, although the majority of trials on Web-based brief alcohol interventions have used this type of sampling strategy (eg, [43]) in which participants are selected based on availability. Moreover, contamination between conditions might have occurred if participants in the experimental condition shared the link of the WDYD intervention with participants in the control condition. Nonetheless, WDYD is not yet available online; thereby, it reduces the likelihood of contamination between conditions. Additionally, the EMA measurements relied on self-reported measures with 7-day recall, which still remains subject to measurement errors because data were not collected in the event and precise recall of alcohol use decreases after 2 or 3 days due to memory deficits [10]. True in-the-event measures would be very difficult over long time periods. However, this is the first study using 30 weekly EMA measurements to assess outcome measures, thereby generating outcome measures that are much closer to the actual drinking behavior of individuals than any other trial on Web-based brief alcohol interventions in the current alcohol prevention literature. Another limitation is that the outcome measures of weekly alcohol consumption and frequency of binge drinking have not been validated for online use. However, these outcome measures have been validated in paper-based surveys [27,44]. In addition, research has shown that online survey data can be equal or superior to that of equivalent paper-based survey data [45]. Furthermore, participants were not blinded to the assigned interventions, which is a common limitation in Web-based trials [46]. Participants who are aware that they have been assigned to the experimental condition might have favorable expectations or increased apprehension and participants assigned to the control condition might feel deprived or relieved, which can affect their responses on the outcome measures. Finally, one should be careful in generalizing our findings to students who are not ready to change their alcohol use, individuals younger than 18 years, and those who do not attend or have not attended a college or university.

Strengths of this study included the online weekly EMA methodology for assessing alcohol use and intervention effectiveness over time while maintaining a high retention rate. First, the use of a large number of measurements over time by means of EMA enabled us to assess changes in alcohol use and intervention effectiveness over time while taking into account the fluctuating nature of alcohol use among students. Second, the coverage strategy of EMA minimized recall bias because of a relative short reference period (ie, 1 week), thereby generating more ecologically valid outcome measures of self-reported drinking behaviors [16]. Third, the use of online surveys had the advantage over paper-and-pencil surveys because it reduced the likelihood of entry errors while improving cost-effectiveness [47]. Fourth, the use of EMA in combination with LGC modeling techniques allowed diminishing statistical errors by generating overall intervention effects resulting in more reliable outcome measures and a higher precision in intervention effectiveness. In addition, the WDYD intervention is based upon the intervention mapping protocol, which is a sound framework for theoretical- and evidence-based development, implementation, and evaluation of effective behavior change interventions [13]. Moreover, the WDYD intervention incorporated components (eg, personalized normative feedback) that are successful in reducing heavy drinking among student populations [48].

Future Directions

The findings of the current study suggest that the WDYD intervention can prevent an increase in weekly alcohol consumption and frequency of binge drinking among heavy-drinking students that is sustained at 3 to 6 months postintervention. The collaboration with the Trimbos Institute can ensure an adequate large-scale implementation of the WDYD intervention by incorporating it in their materials and programs [12]. In addition, the findings indicate the relevance of including a large number of measurements by means of EMA for assessing the outcome measures and evaluating the intervention effectiveness to obtain higher precision in future alcohol prevention trials. If a large number of measurements with extremely short reference periods (ie, 2 hours) are used to assess outcome measures, smartphones might be more beneficial than online surveys because they can capture data regardless of time and location of the participant [49]. Moreover, future research should identify whether alcohol-related cognitions (eg, self-efficacy) account for the observed outcomes to help explain why Web-based brief alcohol interventions are effective in reducing or, in our case, preventing an increase in alcohol use among heavy-drinking students, especially considering that most Web-based brief alcohol interventions are designed to affect alcohol-related cognitions that determine heavy drinking in young adults [50,51].

Conclusions

The WDYD Web-based brief alcohol intervention was shown to be effective in preventing an increase in weekly alcohol consumption and frequency of binge drinking directly after the intervention among heavy-drinking students that was sustained at 3 and 6 months postintervention. Moreover, the findings emphasize the strengths of using EMA and statistical techniques, such as LGC, in testing the intervention effectiveness that would otherwise remain undetected.

Acknowledgments

The major funding agency ZonMw, The Netherlands Organization for Health Research and Development, supported this study (grant number 50-50110-96-682).

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [46].

[PDF File (Adobe PDF File), 997KB - jmir_v16i1e5_app1.pdf]

References

1. Maggs JL, Williams LR, Lee CM. Ups and downs of alcohol use among first-year college students: Number of drinks, heavy drinking, and stumble and pass out drinking days. Addict Behav 2011 Mar;36(3):197-202 [FREE Full text] [doi: 10.1016/j.addbeh.2010.10.005] [Medline: 21106298]
2. Ragsdale K, Porter JR, Mathews R, White A, Gore-Felton C, Mcgarvey EL. “Liquor before beer, you're in the clear”: binge drinking and other risk behaviours among fraternity/sorority members and their non-Greek peers. J Subst Use 2011;17:323-339. [doi: 10.3109/14659891.2011.583312]
3. Wicki M, Kuntsche E, Gmel G. Drinking at European universities? A review of students' alcohol use. Addict Behav 2010 Nov;35(11):913-924. [doi: 10.1016/j.addbeh.2010.06.015] [Medline: 20624671]
4. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
5. Voogt CV, Poelen EA, Kleinjan M, Lemmers LA, Engels RC. The effectiveness of the 'what do you drink' Web-based brief alcohol intervention in reducing heavy drinking among students: a two-arm parallel group randomized controlled trial. Alcohol Alcohol 2013;48(3):312-321. [doi: 10.1093/alcalc/ags133] [Medline: 23303466]
6. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of Web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]
7. Hustad JT, Barnett NP, Borsari B, Jackson KM. Web-based alcohol prevention for incoming college students: a randomized controlled trial. Addict Behav 2010 Mar;35(3):183-189 [FREE Full text] [doi: 10.1016/j.addbeh.2009.10.012] [Medline: 19900763]
8. Kypri K, Hallett J, Howat P, McManus A, Maycock B, Bowe S, et al. Randomized controlled trial of proactive Web-based alcohol screening and brief intervention for university students. Arch Intern Med 2009 Sep 14;169(16):1508-1514. [doi: 10.1001/archinternmed.2009.249] [Medline: 19752409]
9. Kypri K, McCambridge J, Vater T, Bowe SJ, Saunders JB, Cunningham JA, et al. Web-based alcohol intervention for Māori university students: double-blind, multi-site randomized controlled trial. Addiction 2013 Feb;108(2):331-338 [FREE Full text] [doi: 10.1111/j.1360-0443.2012.04067.x] [Medline: 22925046]
10. Gmel G, Daeppen JB. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs. J Stud Alcohol Drugs 2007 Mar;68(2):303-310. [Medline: 17286350]
11. Neighbors C, Atkins DC, Lewis MA, Lee CM, Kaysen D, Mittmann A, et al. Event-specific drinking among college students. Psychol Addict Behav 2011 Dec;25(4):702-707 [FREE Full text] [doi: 10.1037/a0024051] [Medline: 21639597]

12. Voogt CV, Poelen EAP, Kleinjan M, Lemmers LACJ, Engels RCME. The development of a Web-based brief alcohol intervention in reducing heavy drinking among college students: An Intervention Mapping approach. Health Promot Int 2013 Mar 22:1-11. [doi: 10.1093/heapro/dat016]
13. Bartholomew K, Parcel G, Kok GS, Gottlieb N. Intervention Mapping: Designing Theory- and Evidence-Based Health Promotion Programs. Mountain View, CA: Mayfield; Oct 15, 2001.
14. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change (second edition). New York: Guilford Press; Apr 12, 2002.
15. De Vries H, Dijkstra M, Kuhlman P. Self-efficacy: The third factor besides attitude and subjective norm as a predictor of behavioral intentions. Health Educ Res 1988;3:273-282. [doi: 10.1093/her/3.3.273]
16. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
17. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
18. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462. [doi: 10.1037/a0017653] [Medline: 19947780]
19. Muthén B, Brown CH, Masyn K, Jo B, Khoo ST, Yang CC, et al. General growth mixture modeling for randomized preventive interventions. Biostatistics 2002 Dec;3(4):459-475 [FREE Full text] [doi: 10.1093/biostatistics/3.4.459] [Medline: 12933592]
20. Muthén LK, Muthén B. Mplus User's Guide (Sixth Edition). Los Angeles, CA: Muthén & Muthén; Apr 2010.
21. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an Internet-based intervention for alcohol abusers. Addiction 2009 Dec;104(12):2023-2032 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
22. Hardy SA, Thiels C. Using latent growth curve modeling in clinical treatment research: An example comparing guided self-change and cognitive behavioral therapy treatments for bulimia nervosa. Int J Clin Hlth Psyc 2009;9(1):51-71.
23. Voogt CV, Poelen EA, Kleinjan M, Lemmers LA, Engels RC. Targeting young drinkers online: the effectiveness of a Web-based brief alcohol intervention in reducing heavy drinking among college students: study protocol of a two-arm parallel group randomized controlled trial. BMC Public Health 2011;11:231 [FREE Full text] [doi: 10.1186/1471-2458-11-231] [Medline: 21492412]
24. Babor T, Higgins-Biddle JC, Saunders J, Monteiro MG. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care (second edition). Geneva: World Health Organization. Department of Mental Health and Substance Dependence; 2001:1-40.
25. Boon B, Risselada A, Huiberts A, Riper H, Smit F. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. J Med Internet Res 2011;13(2):e43 [FREE Full text] [doi: 10.2196/jmir.1695] [Medline: 21719412]
26. Gezondheidsraad. Richtlijnen voor goed voeding 2006 [Guidelines for Healthy Nutrition 2006]. Den Haag, the Netherlands: Gezondheidsraad; 2006.
27. Lemmens P, Tan ES, Knibbe RA. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. J Stud Alcohol 1992 Sep;53(5):476-486. [Medline: 1405641]
28. Tabachnick BG, Fidell LS. Using Multivariate Statistics (5th edition). Boston, MA: Allyn and Bacon; Mar 03, 2007.
29. Voogt CV, Kuntsche E, Kleinjan M, Poelen EAP, Lemmers LACJ, Engels RCME. Using ecological momentary assessment in testing the effectiveness of an alcohol intervention: A two-arm parallel group randomized controlled trial. PLOS ONE 2013 Nov;8(11):e78436. [doi: 10.1371/journal.pone.0078436]
30. Landerman LR, Land KC, Pieper CF. An empirical evaluation of the predictive mean matching method for imputing missing values. Socio Meth Res 1997;26(1):3-33. [doi: 10.1177/0049124197026001001]
31. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
32. Cohen J. Statistical Power Analysis for the Behavioural Sciences (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates; Jul 01, 1988.
33. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869 [FREE Full text] [Medline: 20332511]
34. Beets MW, Flay BR, Vuchinich S, Li KK, Acock A, Snyder FJ, Tobacco Etiology Research Network. Longitudinal patterns of binge drinking among first year college students with a history of tobacco use. Drug Alcohol Depend 2009 Jul 1;103(1-2):1-8. [doi: 10.1016/j.drugalcdep.2008.12.017] [Medline: 19423242]
35. Demers A, Kairouz S, Adlaf EM, Gliksman L, Newton-Taylor B, Marchand A. Multilevel analysis of situational drinking among Canadian undergraduates. Soc Sci Med 2002 Aug;55(3):415-424. [Medline: 12144149]
36. Clifford PR, Maisto SA, Davis CM. Alcohol treatment research assessment exposure subject reactivity effects: part I. Alcohol use and related consequences. J Stud Alcohol Drugs 2007 Jul;68(4):519-528. [Medline: 17568955]

37. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML. Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking. Addiction 2007 Jan;102(1):62-70. [doi: 10.1111/j.1360-0443.2006.01632.x] [Medline: 17207124]
38. Maisto SA, Clifford PR, Davis CM. Alcohol treatment research assessment exposure subject reactivity effects: part II. Treatment engagement and involvement. J Stud Alcohol Drugs 2007 Jul;68(4):529-533. [Medline: 17568956]
39. McCambridge J. [Commentary] Research assessments: instruments of bias and brief interventions of the future? Addiction 2009 Aug;104(8):1311-1312. [doi: 10.1111/j.1360-0443.2009.02684.x] [Medline: 19624324]
40. Carey KB, Scott-Sheldon LA, Elliott JC, Bolles JR, Carey MP. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
41. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: A meta-analysis. Addiction 2011;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x]
42. Poikolainen K, Paljärvi T, Mäkelä P. Alcohol and the preventive paradox: serious harms and drinking patterns. Addiction 2007 Apr;102(4):571-578. [doi: 10.1111/j.1360-0443.2006.01740.x] [Medline: 17286638]
43. Spijkerman R, Roek MA, Vermulst A, Lemmers L, Huiberts A, Engels RC. Effectiveness of a Web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. J Med Internet Res 2010;12(5):e65 [FREE Full text] [doi: 10.2196/jmir.1465] [Medline: 21169172]
44. Hibell B, Andersson B, Bjarnasson T, Ahlström S, Balakivera O, Kokkevi A, et al. The ESPAD Report 2003: Alcohol and other drug use among students in 35 European countries. Stockholm, Sweden: The Swedish Council for Information on Alcohol and other Drugs; 2004. URL: http://www.espad.org/Uploads/ESPAD_reports/2003/The_2003_ESPAD_report.pdf [accessed 2013-12-19] [WebCite Cache ID 6Lz0vneu4]
45. Chang TZ, Vowles N. Strategies for improving data reliability for online surveys: A case study. IJECS 2013;4(1):121-130. [doi: 10.7903/ijecs.1121]
46. 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]
47. Miller ET, Neal DJ, Roberts LJ, Baer JS, Cressler SO, Metrik J, et al. Test-retest reliability of alcohol measures: is there a difference between Internet-based assessment and traditional methods? Psychol Addict Behav 2002 Mar;16(1):56-63. [Medline: 11934087]
48. Carey KB, Scott-Sheldon LA, Carey MP, DeMartini KS. Individual-level interventions to reduce college student drinking: a meta-analytic review. Addict Behav 2007 Nov;32(11):2469-2494 [FREE Full text] [doi: 10.1016/j.addbeh.2007.05.004] [Medline: 17590277]
49. Burger C, Riemer V, Grafeneder J, Woisetschlager B, Vidovic D, Hergovich A. Reaching the mobile respondent determinants of high-level mobile phone use among a high-coverage group. Soc Sci Comput Rev 2009 Dec 02;28(3):336-349. [doi: 10.1177/0894439309353099]
50. Cullum J, Armeli S, Tennen H. Drinking norm-behavior association over time using retrospective and daily measures. J Stud Alcohol Drugs 2010 Sep;71(5):769-777 [FREE Full text] [Medline: 20731984]
51. Danielsson AK, Wennberg P, Tengström A, Romelsjö A. Adolescent alcohol use trajectories: Predictors and subsequent problems. Addict Behav 2010 Sep;35(9):848-852. [doi: 10.1016/j.addbeh.2010.05.001] [Medline: 20626071]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CFI: Comparative Fit Index
CONSORT: Consolidated Standards of Reporting Trials
EMA: ecological momentary assessment
ESM: experience sampling methods
ITT: intent-to-treat
LGC: latent growth curve
RMSEA: Root Mean Square Error of Approximation
TLI: Tucker-Lewis Index
WDYD: What Do You Drink

Edited by G Eysenbach; submitted 05.07.13; peer-reviewed by L Le Garjean, D Ho, J Merrill; comments to author 18.09.13; revised version received 18.10.13; accepted 04.11.13; published 08.01.14.

Please cite as:
Voogt C, Kuntsche E, Kleinjan M, Poelen E, Engels R
Using Ecological Momentary Assessment to Test the Effectiveness of a Web-Based Brief Alcohol Intervention Over Time Among Heavy-Drinking Students: Randomized Controlled Trial
J Med Internet Res 2014;16(1):e5
URL: http://www.jmir.org/2014/1/e5/
doi:10.2196/jmir.2817
PMID:24401555

Original Paper

Effectiveness of Web-Based Tailored Advice on Parents' Child Safety Behaviors: Randomized Controlled Trial

Mirjam Elisabeth Johanna van Beelen[1*], PhD; Tinneke Monique Jozef Beirens[1,2*], PhD; Paul den Hertog[3*], PhD; Eduard Ferdinand van Beeck[1*], PhD; Hein Raat[1*], MD, PhD

[1]Erasmus MC – University Medical Centre Rotterdam, Department of Public Health, Rotterdam, Netherlands

[2]Dutch Association for Youth Health Care Physicians, Utrecht, Netherlands

[3]Consumer Safety Institute, Amsterdam, Netherlands

[*]all authors contributed equally

Corresponding Author:
Hein Raat, MD, PhD
Erasmus MC – University Medical Centre Rotterdam
Department of Public Health
Postbox 2040
Rotterdam, 3000 CA
Netherlands
Phone: 31 10 703 8580
Fax: 31 10 703 8474
Email: h.raat@erasmusmc.nl

Abstract

Background: Injuries at home are a major cause of death, disability, and loss of quality of life among young children. Despite current safety education, required safety behavior of parents is often lacking. To prevent various childhood disorders, the application of Web-based tools has increased the effectiveness of health promotion efforts. Therefore, an intervention with Web-based, tailored, safety advice combined with personal counseling (E-Health4Uth home safety) was developed and applied.

Objective: To evaluate the effect of E-Health4Uth home safety on parents' safety behaviors with regard to the prevention of falls, poisoning, drowning, and burns.

Methods: A randomized controlled trial was conducted (2009-2011) among parents visiting well-baby clinics in the Netherlands. Parents were randomly assigned to the intervention group (E-Health4Uth home safety intervention) or to the control condition consisting of usual care. Parents in the intervention condition completed a Web-based safety behavior assessment questionnaire; the resulting tailored safety advice was discussed with their child health care professional at a well-baby visit (age approximately 11 months). Parents in the control condition received counseling using generic safety information leaflets at this well-baby visit. Parents' child safety behaviors were derived from self-report questionnaires at baseline (age 7 months) and at follow-up (age 17 months). Each specific safety behavior was classified as safe/unsafe and a total risk score was calculated. Logistic and linear regression analyses were used to reveal differences in safety behavior between the intervention and the control condition at follow-up.

Results: A total of 1292 parents (response rate 44.79%) were analyzed. At follow-up, parents in the intervention condition (n=643) showed significantly less unsafe behavior compared to parents in the control condition (n=649): top of staircase (23.91% vs 32.19%; OR 0.65, 95% CI 0.50-0.85); bottom of staircase (63.53% vs 71.94%; OR 0.69, 95% CI 0.53-0.88); top and bottom of staircase (68.94% vs 78.28%; OR 0.62, 95% CI 0.48-0.81); storage of cleaning products (30.33% vs 39.91%; OR 0.67, 95% CI 0.53-0.85); bathing of the child (23.46% vs 32.25%; OR 0.65, 95% CI 0.51-0.84); drinking hot fluids (34.84% vs 41.73%; OR 0.76, 95% CI 0.61-0.96); using rear hotplates (79.34% vs 85.27%; OR 0.67, 95% CI 0.50-0.90); and the total risk score in which a higher score indicates more unsafe behavior (mean 13.63, SD 6.12 vs mean 15.34, SD 6.07; beta –1.59, 95% CI –2.26 to –0.93). There were no significant differences for other specific behaviors between the two study conditions.

Conclusions: Compared to generic written materials, the E-Health4Uth home safety intervention seems more effective in promoting parents' safety behavior for safe staircases, storage of cleaning products, bathing, drinking hot fluids, and cooking. This study supports the application of Web-based, tailored, safety advice for the prevention of unintentional injuries in the youth health care setting.

Trial Registration: Nederlands Trial Register: NTR1836; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1836 (Archived by WebCite at http://www.webcitation.org/6MPIGQxpx).

(J Med Internet Res 2014;16(1):e17) doi:10.2196/jmir.2521

KEYWORDS

child; eHealth; injury; parent; prevention; primary care; RCT; safety

Introduction

Background

Unintentional injuries are a major cause of death, and a major source of morbidity and loss of quality of life among children aged 0-4 years [1-4]. In children aged 5 years or younger, more than 90% of unintentional injuries occur in and around the home [1]. Although the type and cause of injury varies by age, the most common injuries of children aged 0 to 4 years are falls, poisoning, drowning, and burns [1,5]. Each year in the Netherlands, 18 children aged between 0 and 4 years die because of injuries in and/or around the home [6]. Moreover, an additional 46,000 children aged between 0 and 4 years are medically treated because of home injuries [6]. To reduce the number of injuries, the Dutch Consumer Safety Institute introduced the use of safety information leaflets at preventive youth health care centers to provide safety education to parents of children aged between 0 and 4 years [7]. These leaflets are successfully employed in preventive youth health care and appear to have a modest effect on parental behavior [8,9]. Many countries have installed preventive youth health care, which refers to various activities to improve and protect the health, growth, and development of young people, and to prevent illness and disability in early life. These activities include a system of maternal and child health care, which serves children from birth to age 18 years [10,11]. The preventive youth health care also plays a significant role in injury prevention [12].

In the Netherlands, all parents are invited to regularly attend (free of charge) scheduled well-child visits at their well-baby clinic. During these visits, the growth and development of the child is monitored and relevant health information and vaccinations are provided. In the Netherlands, approximately 93% of parents attend 1 or more well-baby visits when their child is aged 4 years and younger; the attendance rates range from approximately 50% to 93% between the specific age-related scheduled visits [13]. Parents receive health information on various topics, including nutrition, growth, and child home safety [14]. Currently, this safety information is provided to parents by using generic information leaflets that they receive at their regular visits to the well-baby clinic. Nevertheless, the required safety behavior of parents is often lacking, causing unnecessary risk of injury to young children [15-17].

To prevent other childhood disorders, the application of Web-based tailored tools (eHealth) has increased the effectiveness of health promotion effects [18-20]. The field of eHealth, health services and information delivered or enhanced through the Internet and related technologies [21], is broad and emerging at the intersection of medical informatics, public health, and business. It involves the use of information and communications (especially the Internet) to improve or enable health and health care [22]. It could also be used to provide information to parents on several health topics, including home safety. Because tailored information combined with counseling, which can be provided by using eHealth, is personalized, parents could find the information more useful than general information materials [23]. Furthermore, parents may be more inclined to change their behavior when the information they receive is perceived as personally relevant [24,25].

A home safety intervention with Web-based, tailored, safety information was developed and applied (E-health4Uth). It uses Web-based, tailored, safety information in combination with personal counseling at well-baby clinics for safety behaviors required with a child at home. A pilot study showed that most parents found this new safety information to be useful and applicable, and that child health care professionals were enthusiastic about the eHealth intervention [26]. However, no information is available about the effects of the new Internet-based, tailored, safety information on parents' child safety behaviors compared to the older method of safety education. Tailored information is thought to promote behavior change by providing personally relevant feedback. Tailoring is defined as "any combination of information or change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and have been derived from an individual assessment" [27]. Although (online) computer-tailored interventions seem to have a positive effect on adult behaviors compared to generic information or to no information [17,28], there is no evidence for the efficacy of a tailored intervention on parents' specific child safety behaviors for the prevention of unintentional injuries.

Objective of the Study

This study evaluates the effect of Web-based, tailored, safety information combined with personal counseling on parents' child safety behaviors for the prevention of falls, poisoning, drowning, and burns. The hypothesis is that parents in the E-Health4Uth home safety intervention condition will show less unsafe behavior and will have a lower total risk score 6 months postintervention compared to parents in the control condition with usual care. In addition, the use and application of the E-Health4Uth home safety module and the well-baby visit, including the use of the tailored safety advice, will be evaluated.

Methods

Study Design

The E-Health4Uth home safety study (BeSAFE study) is a randomized controlled trial (NTR1836) with a baseline measure point before the intervention and a follow-up measure point 6 months after the intervention; the study is described in detail elsewhere [29]. The Medical Ethics Committee of the Erasmus Medical Center gave a declaration of no objection for this study (MEC-2008-370).

Participants and Procedures

Overview

A flow diagram of the enrollment and follow-up of study participants is schematically described in Figure 1.

Managers of an opportunity sample of 26 youth health care organizations in the Netherlands were informed about the study and invited to participate. A total of 5 youth health care organizations in the mixed urban-rural provinces of Zuid-Holland, Noord-Brabant, and Zeeland volunteered to participate, with a total of 30 well-baby clinics.

All parents with a child aged between 5 and 8 months (1 parent per family) who were eligible for a routine well-baby visit at their well-baby clinic from June 2009 until December 2010 received written information about the study and were invited to provide informed consent to participate (n=3147). Parents who provided informed consent were invited to complete the baseline questionnaire.

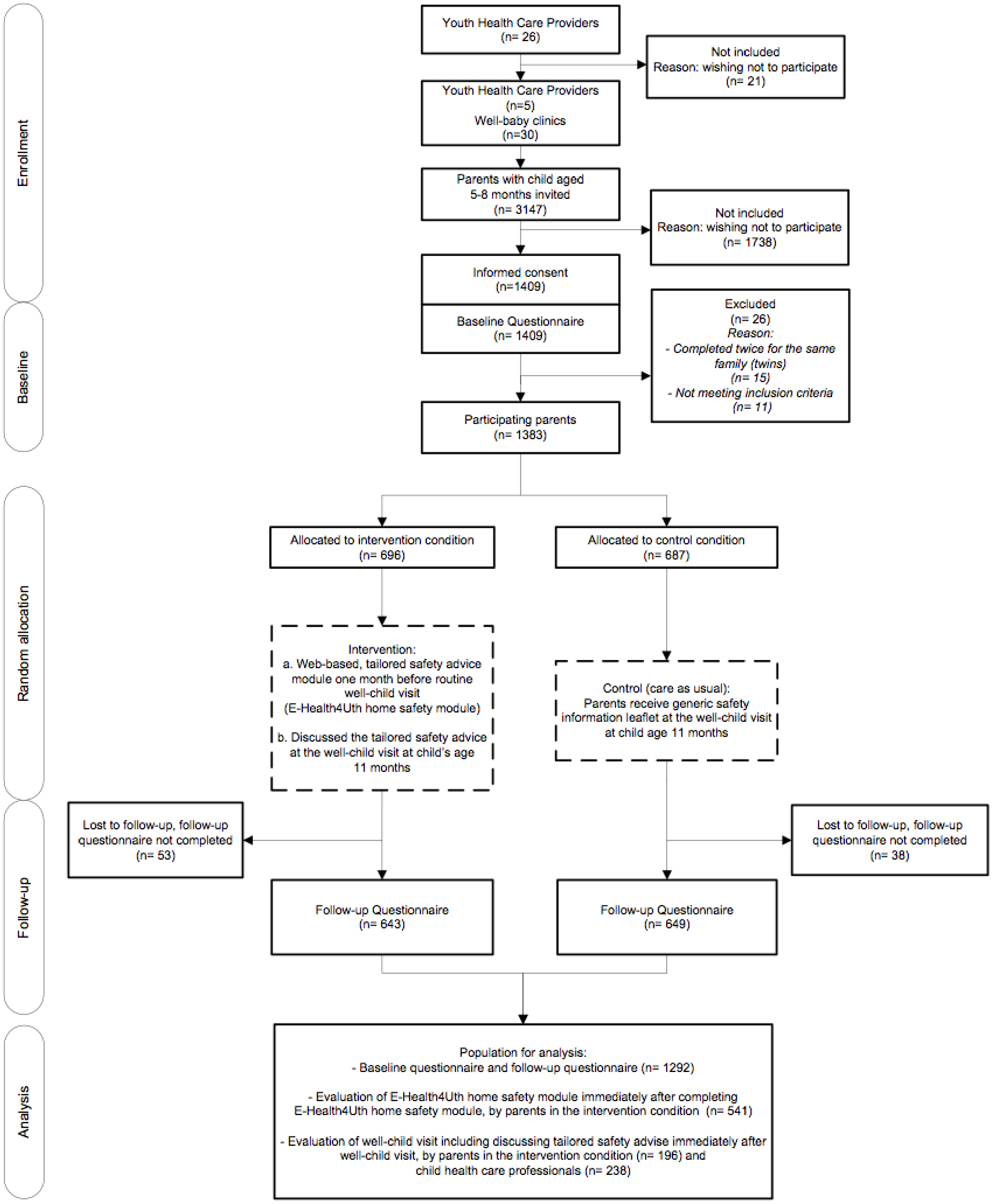
Subsequently, parents were randomly assigned to one of two conditions: (1) Web-based, tailored, safety advice module combined with discussion of the tailored safety advice at the well-baby visit (E-Health4Uth home safety intervention condition), or (2) care as usual (ie, received a generic written safety information leaflet at the well-baby visit; control condition). Randomization was done using a computerized random allocation generator.

Parents in the intervention condition were invited to complete the E-Health4Uth home safety module when their child was approximately 10 months old (ie, 1 month before their routine well-baby visit at the well-baby clinic). The intervention is described in detail in the following section. Parents in the control condition also visited their well-baby clinic when their child was approximately 11 months of age (see control condition described subsequently). All parents received a follow-up questionnaire when their child was approximately 17 months (6 months postintervention). The baseline and follow-up data were collected from June 2009 until July 2011.

Parents received a maximum of 2 regular mail reminders for completing the questionnaires. Parents who did not respond to the invitations to complete the follow-up questionnaire received a telephone call to motivate them to complete the intervention or the questionnaire. Parents in the intervention condition received a maximum of 2 reminders to complete the E-Health4Uth home safety module. If they did not respond, they received a telephone call to motivate them to complete the E-Health4Uth home safety module.

Parents in both conditions who completed the baseline questionnaire received a gift voucher of €15. Parents who also completed the follow-up questionnaire received a second gift voucher of €10.

Figure 1. Flow diagram of the enrollment and follow-up of study participants.



E-Health4Uth Home Safety Intervention Condition

The E-Health4Uth home safety intervention aims at 4 major topics with regard to safety in/around the home of children aged between 12 and 24 months: prevention of falls, poisoning, drowning, and burns [7,8,17,30-34]. The components of the 4 safety topics of the intervention are shown in Multimedia Appendix 1.

Parents allocated to the E-Health4Uth home safety intervention condition received a personal log-in name and password by email when their child was approximately 10 months. Parents were asked to complete the E-Health4Uth home safety module before their next routine well-baby visit at approximately 11 months of age. Parents could complete the E-Health4Uth home safety module wherever they wished (eg, at home, at work) as long as Internet was available. As a first step, parents completed a safety assessment questionnaire. The answers to this assessment questionnaire were used to generate tailored safety advice, which parents could read immediately online. The tailored safety advice was personalized with the child's name and consisted of messages tailored to the parent's current situation and safety behavior (Multimedia Appendices 2-4). This included sections with general information on the importance and relevance of the injury area. A total of 114 messages were developed for this tailored safety advice, which could be combined in various ways based on the parent's answers to the assessment questionnaire.

When parents had completed reading their personal safety advice, they were invited to formulate an implementation-intention plan. In this implementation-intention plan, parents planned specific actions (ie, what, when, and where to improve their safety behavior and implement these in their home situation at a specified time) [35,36].

The tailored safety advice and implementation-intention plan of each parent was sent by email to both the parent and the child health care professional to prepare for the routine well-baby visit at age 11 months. At the well-baby visit, the child health care professional discussed the tailored safety advice and the implementation-intention plan with the parent using motivational interviewing techniques [35-37]. Preceding the start of the study, health care professionals received training from the researchers. In this training, the study was explained and examples of the tailored safety advice were used to give instructions on how the intervention should be delivered to ensure integrity of delivery of the tailored safety advice.

Parents in the intervention condition received the E-Health4Uth home safety intervention, but could also receive the generic safety information leaflet as a part of usual care.

Approximately 4 weeks after the well-baby visit, parents received a reminder about their tailored safety advice and the implementation-intention plan by email to strengthen the message.

The content and development of the E-Health4Uth home safety module was not changed during the study. The intervention software (TailorBuilder) was developed by OverNite Software Europe (OSE, Sittard, the Netherlands).

Control Condition

Parents in the control condition received care as usual; that is, parents received a generic safety information leaflet (for children aged 12-24 months) published by the Dutch Consumer Safety Institute [7,9,10,34] during their routine well-baby visit at approximately 11 months of age. During this well-baby visit, the child health care professional discussed the safety in and around the home with the parents using the generic safety information leaflet and motivated parents to adopt safety measures in their home.

The safety information leaflet contained relevant information on the prevention of toddler injuries in and/or around the home, such as information on safety and advice about the prevention of falls (ie, window protection, stair gates, practice walking down the stairs), poisoning (ie, safe storage of cleaning products and medicines), drowning (ie, ponds), and burns (ie, hot fluids, hot pans) [7].

Outcomes/Measures

Overview

Data on demographic factors and parents' child safety behaviors were collected at enrollment at approximately age 7 months (baseline) and at 6 months postintervention at approximately age 17 months (follow-up) by self-report questionnaires.

Parents' Child Safety Behaviors

In the two study conditions, specific parents' child safety behaviors for the prevention of falls, poisoning, drowning, and burns were assessed. Some behaviors were assessed only when they were applicable to the situation of the parent. Each specific safety behavior was first classified as being safe or unsafe. Both the E-Health4Uth home safety intervention and the generic safety information leaflet covered the same topics with regard to the prevention of falls, burns, drowning, and poisoning.

Prevention of falls was assessed by the safety of staircases (only if a staircase was present) and window safety (only if there was a window a child could reach present). Safety of staircases was assessed with 4 items: presence of a stair gate at the top of the stairs (yes/no); closing a stair gate at the top of the stairs (always/often/sometimes/rarely/never); presence of a stair gate at the bottom of the stairs (yes/no); and closing a stair gate at the bottom of the stairs (always/often/sometimes/rarely/never). Safe behavior for the top of the staircase was defined as having a stair gate at the top of the staircase and always closing the stair gate. Safe behavior for the bottom of the staircase was defined as having a stair gate at the bottom of the staircase and always closing the stair gate. Additionally, safe behavior regarding the top/bottom of the staircase was defined as having a stair gate at the top/bottom of the staircase and always closing the stair gates. Safety of windows was assessed with 1 item: having window guards on windows a child can reach (yes/no). Safe behavior was defined as having a window guard on windows a child can reach.

Prevention of poisonings was assessed by the storage of cleaning products and medicines. Storage of cleaning products was assessed with 12 items: presence of cleaning products at different storage locations (yes/no). Safe storage of cleaning products was defined as storing them on a shelf or in a drawer or cabinet without a lock higher than 1.50 m or storing them in a drawer or cabinet with a lock. Storage of medicines was assessed with 13 items: presence of medicines at different storage locations (yes/no). Safe storage of medicines was defined as storing them on a shelf or in a drawer or cabinet without a lock higher than 1.50 m or storing them in a drawer or cabinet with a lock.

Prevention of drowning was assessed with regard to bathing (only if the child was bathed), safety around ponds (only if a pond was present), safety around private swimming pools (only if a swimming pool was present), and swimming (only if the child swam). Safety of bathing was assessed with 1 item: how often the child is left unsupervised in the bathtub, even for a short period (very often/often/sometimes/rarely/never). Safe bathing was defined as never leaving the child unsupervised in the bathtub. Safety of a pond was assessed with 1 item: presence of a fence around the pond (yes, fence higher than 1.20 m/yes, fence lower than 1.20 m/no). A safe pond was defined as having a fence higher than 1.20 m around the pond. Safety of a private swimming pool was assessed with 1 item: presence of a fence around the swimming pool (yes, fence higher than 1.20 m/yes, fence lower than 1.20 m/no). A safe private swimming pool was defined as having a fence higher than 1.20 m around the swimming pool. Swimming was assessed with 2 items: whether the child wears a flotation device (always/often/sometimes/rarely/never) and how often the child is left alone in the swimming pool (very often/often/sometimes/rarely/never). Safe swimming was defined as the child always wears a flotation device in the swimming pool and is never left alone in the swimming pool (either a private swimming pool or a small inflatable swimming pool).

Prevention of burns was assessed from hot water safety in the bath and shower (depending on what is present in the home), drinking hot fluids with child on parent's lap, and safety during cooking. Presence of thermostatic-controlled taps was assessed with 2 items: does the hot water tap in the bath have a thermostatic-controlled tap (yes/no) and does the hot water tap in the shower have a thermostatic-controlled tap (yes/no). Safe hot water taps in bath/shower was defined as having a thermostatic-controlled tap present. Drinking hot fluids was assessed with 1 item: how often the parent drinks hot fluids with the child on their lap (always/often/sometimes/rarely/never). Safe drinking of hot fluids was defined as never drinking hot fluid with their child on their lap. Safe cooking was assessed with 4 items: presence of a stove guard (yes/no); child present in the kitchen during cooking (always/often/sometimes/rarely/never); use of rear burners during cooking (always/often/sometimes/rarely/never); and turning pan handles away during cooking (always/often/sometimes/rarely/never). Safe cooking was defined as a stove guard present, the child never in the kitchen during cooking, always using rear burners during cooking, and always turning panhandles away during cooking.

Subsequently, a total risk score was calculated for each parent by summing each specific parent's safety behavior for the 4 topics assessed (according to allocated scores shown in Multimedia Appendix 5). A higher score indicated more unsafe behavior. When a situation was not applicable, a score of zero was assigned. A maximum score of 53 points could be obtained. The individual scores of the total risk score were based on previous literature [17] and expert consultation (Consumer Safety Institute, the Netherlands).

Demographic Factors

Parents' gender, age, educational level, employment status, and ethnicity were assessed in the baseline questionnaire. Educational level was categorized as high, intermediate, or low. High level was defined as higher professional education or academic higher education; intermediate level as senior secondary vocational education, senior general secondary education, or university preparatory education; and low educational level as preparatory secondary vocational education or lower [38]. Employment status was defined as unemployed if they did not have a part-time or a full-time job. Parents' ethnicity was determined based on their own parents' country of birth (grandparents of the infant). A parent was of Dutch ethnicity if both grandparents were born in the Netherlands. If one of the grandparents was born in another Western country, a parent was of other Western ethnicity. If both grandparents were born in another Western or non-Western country, ethnicity was determined by the grandmother's country of birth [38].

Family situation, number of children, child's gender, age, and ability to crawl or walk were reported. Family situation was defined as single parent, or living with child or children and other parent or caregiver. Number of children in the family was dichotomized as first child in the family or second child or more children in the family. Crawling of the child was assessed and defined as an infant's ability to crawl on hands and knees and/or crawl on their tummy and/or shuffle on their bottom (yes/no).

Parent Evaluation of the E-Health4Uth Home Safety Module

Parents in the E-Health4Uth home safety intervention were asked to evaluate the E-Health4Uth home safety intervention module immediately after completing the module. Unless stated otherwise, all evaluation items were assessed on a 5-point Likert scale ranging from totally disagree (=1) to totally agree (=5).

An objective measure of parents' exposure to the intervention was obtained from the log-in data from the intervention registration, which stored information on parents' use of the intervention, including receipt of the tailored safety advice and completion of an implementation-intention plan.

Parents' evaluation of the E-Health4Uth home safety intervention was assessed immediately after receiving the tailored safety advice and formulating an implementation-intention plan, by using a Web-based evaluation form. Parents reported the following items on the evaluation forms: the reading of the Web-based tailored safety advice (read completely, read partly, or did not read their advice); whether they found the advice to be reliable, understandable, relevant, useful, and motivating to take action; their intention to change their behavior with regard to falls, poisoning, drowning, and burns (yes/no); whether it was easy or difficult to complete an implementation-intention plan (from very difficult to very easy); time needed to answer the questions and read the safety advice (in minutes); whether they perceived this time to be long or short (from very long to very short); the ease of use of the E-Health4Uth home safety intervention (from very difficult to very easy); and whether they perceived it as being a pleasant source of information (from very unpleasant to very pleasant). Furthermore, they rated the E-Health4Uth home safety intervention on a scale from 1 (most negative) to 10 (most positive).

Parent and Health Care Professional Evaluation of the Well-Baby Visit

Parents and child health care professionals in the E-Health4Uth home safety intervention were asked to evaluate the well-baby visit, including the discussion of the tailored safety advice, immediately after the well-baby visit at approximately age 11 months.

Parents reported their satisfaction with the information they discussed at the well-baby visit, whether discussing the tailored safety advice was a valuable supplement to receiving tailored safety advice, the overall satisfaction with the well-baby visit, and they rated the well-baby visit on a scale from 1 (most negative) to 10 (most positive).

Child health care professionals reported the time they needed for the well-baby visit (in minutes); the time needed to discuss the safety at home (in minutes); whether they gave a safety information leaflet to the parent (yes/no); whether the tailored safety advice was present in the child's dossier (yes/no); whether the tailored safety advice was brought to the well-baby visit by the parent (yes/no); and whether the tailored safety advice was discussed with the parent during the well-baby visit (yes/no). Furthermore, the evaluation assessed whether the tailored safety advice was useful to discuss safety at home during the well-baby visit, the satisfaction with the information given, and the overall satisfaction with the well-baby visit. Child health care professionals rated the well-baby visit on a scale from 1 (most negative) to 10 (most positive).

Statistical Analyses

Intention-To-Treat Analysis

An intention-to-treat analysis was applied [39]. Parents who were randomly assigned to either the intervention condition or the control condition were analyzed as such, regardless of whether they received the intervention or not. Cases with complete data on outcomes at follow-up were analyzed on the effectiveness of the intervention compared to the control condition.

Descriptive statistics were used to describe the characteristics of parents, children, and housing in the two study conditions. Differences between the intervention and control condition, as measured at baseline, were tested with an independent-samples t test or the Mann-Whitney U test (continuous variables) and chi-square test (categorical variables).

Effect Evaluation

The effectiveness of the E-Health4Uth home safety intervention was studied by means of logistic regression analyses (for all specific safety behaviors) and linear regression analyses (for total risk score). Regression analyses were performed with unsafe behavior of total risk score as dependent variable and condition (E-Health4Uth home safety intervention condition vs control condition) as independent variables. All regression analyses were adjusted for demographic factors that showed a significant difference between the two study conditions at baseline (P<.05).

Subsequently, it was determined whether the number of children, parents' educational level, and parents' ethnicity moderated the effects of the E-Health4Uth home safety intervention on unsafe behavior. This was done by adding an interaction term (group × demographic factor) to the regression analysis. If these interaction terms were significant at P<.05, stratified analyses were conducted. Results with a P value <.05 were considered to be statistically significant. All analyses were performed using SPSS 20.0 (IBM Corp, Armonk, NY, USA).

Results

Participants

A total of 1409 parents of the 3147 initially invited provided informed consent and completed the baseline questionnaire with a response rate of 44.79% (Figure 1). A total of 26 parents were excluded because they completed the questionnaire twice for the same family (1 questionnaire was removed from the database at random), or they did not meet the inclusion criteria of child's age ≤12 months. After completing the baseline questionnaire, 696 parents were allocated to the E-Health4Uth home safety condition and 687 parents to the control condition. A total of 1292 parents completed the follow-up questionnaire (dropout rate 6.60%). Dropout was higher among mothers with a low educational level, unemployed mothers, and parents of non-Western ethnicity (P<.05). No other differences were observed between parents who completed the follow-up questionnaire and parents who were lost to follow-up. A study population of 1292 parents and their child were used in the analyses. Table 1 shows the family, child, and housing characteristics of the participants in the two study conditions. Most participants were mothers (93.58%), mean age 32.06 (SD 4.63) years, 15.19% had a low educational level, 83.44% were employed, and 88.46% were of Dutch ethnicity. Father's mean age was 34.51 (SD 5.17) years; 22.40% had a low educational level, 95.67% were employed, and 87.94% were of Dutch ethnicity. In the present study, 2.26% of families included a single parent and 48.14% had 1 child. Of all children, 51.32% were boys, mean age 7.21 (SD 1.07) months, and 33.98% could crawl and 0.47% could walk. A main staircase was present in 87.52% of the homes, 36.41% had a window a child could reach, 11.07% had a pond present, and 3.18% had a private swimming pool at the home.

Table 1. Descriptive characteristics of the study sample[a] at baseline (N=1292).

Characteristics	Total N=1292	Intervention n=643	Control n=649	*P* value[b]
Family characteristics				
Mother is respondent, n (%)	1209 (93.58)	597 (92.85)	612 (94.30)	.16
Parent age [c]				
Mother’s age (years), mean (SD)	32.06 (4.63)	32.08 (4.60)	32.04 (4.67)	.92
Mother’s age (years), range	19.00-48.00	20.00-48.00	19.00-47.00	
Father’s age (years), mean (SD)	34.51 (5.17)	34.50 (5.14)	34.52 (5.21)	.96
Father’s age (years), range	21.00-56.00	22.00-56.00	21.00-55.0	
Mother’s educational level, n (%) [d]				.26
High	524 (40.62)	254 (39.56)	270 (41.67)	
Intermediate	570 (44.19)	280 (43.61)	290 (44.75)	
Low	196 (15.19)	108 (16.82)	88 (13.58)	
Father’s educational level, n (%) [e]				.17
High	470 (36.69)	217 (34.17)	253 (39.16)	
Intermediate	524 (40.91)	268 (42.20)	256 (39.63)	
Low	287 (22.40)	150 (23.62)	137 (21.21)	
Parent employment, n (%) [f]				
Mother is employed	1068 (83.44)	512 (80.25)	556 (86.60)	.002
Father is employed	1214 (95.67)	599 (94.93)	615 (96.39)	.20
Mother’s ethnicity, n (%) [g]				.67
Dutch	1142 (88.46)	570 (88.65)	572 (87.27)	
Other Western	60 (4.65)	32 (4.98)	28 (4.32)	
Non-Western	89 (6.89)	41 (6.38)	48 (7.41)	
Father’s ethnicity, n (%) [h]				.43
Dutch	1130 (87.94)	566 (88.58)	564 (87.31)	
Other Western	65 (5.06)	34 (5.32)	31 (4.80)	
Non-Western	90 (7.00)	39 (6.10)	51 (7.89)	
Single parent, n (%)[d]	29 (2.26)	13 (2.03)	16 (2.49)	.58
First child in family, n (%)	622 (48.14)	317 (49.30)	302 (47.00)	.41
Child characteristics [i]				
Gender (boys), n (%)	663 (51.32)	317 (49.30)	346 (53.31)	.15
Age (months), mean (SD)	7.21 (1.07)	7.26 (1.08)	7.17 (1.07)	.17
Age (months), range	4.73-11.56	4.73-11.56	4.76-11.47	
Child can crawl, n (%)	438 (33.98)	231 (35.93)	207 (32.04)	.14
Child can walk, n (%)	6 (0.47)	3 (0.47)	3 (0.47)	.99
Housing characteristics [j]				
Main staircase present, n (%)	1129 (87.52)	564 (87.71)	565 (87.33)	.83
Windows a child can reach, n (%)	470 (36.41)	241 (37.48)	229 (35.34)	.42
Pond present, n (%)	127 (11.07)	65 (11.40)	62 (10.75)	.72
Private swimming pool present, n (%)	41 (3.18)	23 (3.58)	18 (2.79)	.42

[a] Participants with complete data available at baseline and follow-up.

[b] Differences between intervention condition and control condition, as measured at baseline, tested with independent-samples *t* test (continuous variables) and chi-square test (categorical variables).

[c] Missing data: n=20

[d] Missing data: n=2

[e] Missing data: n=11

[f] Missing data (mother): n=12; missing data (father): n=23

[g] Missing data: n=1

[h] Missing data: n=7

[i] Missing data (crawl): n=3; missing data (walk): n=2

[j] Missing data (staircase): n=2; missing data (window): n=1; missing data (pond): n=5; missing data (pool): n=4

E-Health4Uth Home Safety Intervention Effects

Because the proportion of employed mothers (86.60%) was significantly higher in the control condition compared to those in the intervention condition (80.25%, *P*=.002), regression analyses were adjusted for mother's employment status.

Concerning the prevention of falls, parents in the intervention condition showed significantly less unsafe behavior at follow-up for the top of the staircase (23.91% vs 32.19%; OR 0.65, 95% CI 0.50-0.85), the bottom of the staircase (63.53% vs 71.94%; OR 0.69, 95% CI 0.53-0.88), and the top and bottom of the staircase (68.94% vs 78.28%; OR 0.62, 95% CI 0.48-0.81) compared to parents in the control condition (Tables 2 and 3).

For the prevention of poisoning, parents in the intervention condition showed significantly less unsafe behavior with regard to the storage of cleaning products (30.33% vs 39.91%; OR 0.67, 95% CI 0.53-0.85) compared to parents in the control condition. For the prevention of drowning, parents in the intervention condition showed significantly less unsafe behavior with regard to bathing of the child (23.46% vs 32.25%; OR 0.65, 95% CI 0.51-0.84) compared to parents in the control condition.

For the prevention of burns, parents in the intervention condition showed significantly less unsafe behavior with regard to drinking hot fluids (34.84% vs 41.73%; OR 0.76, 95% CI 0.61-0.96) and using rear hotplates on the stove (79.34% vs 85.27%; OR 0.67, 95% CI 0.50-0.90) compared to parents in the control condition. There were no significant differences with regard to other specific behaviors between the two study conditions.

From baseline to follow-up, the prevalence of unsafe behavior in bathing of the child increased in both the intervention (5.74%-23.46%) and the control condition (6.29%-32.25%). Furthermore, from baseline to follow-up, the prevalence of unsafe behavior for children present in the kitchen increased in both the intervention (64.27%-91.24%) and the control condition (64.81%-93.18%). All other unsafe behaviors showed a decrease between baseline and follow-up.

At follow-up, parents in the intervention condition had a significantly lower total risk score (mean 13.63, SD 6.12; range 1.00-33.00) compared to parents in the control condition (mean 15.34, SD 6.07; range 0.00-37.00; beta coefficient=–1.59, 95% CI –2.26 to –0.93).

Explorative interaction analyses showed no significant interactions between number of children, parents' educational level, and parents' ethnicity with the intervention and control condition on unsafe behavior; therefore, stratified analyses were not conducted.

Table 2. Descriptive statistics of parents' child safety behavior as measured at baseline and at follow-up for intervention and control condition (N=1292).

Behavior	Baseline, %			Follow-up, %		
	Intervention condition n=643	Control condition n=649	*P* value[a]	Intervention condition n=643	Control condition n=649	*P* value[a]
Falls [b]						
Unsafe top of staircase[c]	407 (72.29)	404 (72.14)	.96	137 (23.91)	187 (32.19)	.002
Unsafe bottom of staircase[c]	497 (88.12)	504 (89.52)	.46	364 (63.53)	418 (71.94)	.002
Unsafe top and bottom of staircase[c]	510 (90.43)	516 (91.65)	.47	395 (68.94)	454 (78.28)	< .001
Unsafe windows a child can reach[d]	139 (57.68)	125 (54.59)	.50	97 (46.86)	101 (51.53)	.35
Poisoning						
Unsafe storage of cleaning products	387 (60.19)	401 (61.98)	.30	195 (30.33)	259 (39.91)	.001
Unsafe storage of medicines	247 (38.41)	255 (39.41)	.90	193 (30.02)	221 (34.16)	.14
Drowning [b]						
Unsafe bathing of the child[e]	36 (5.74)	40 (6.29)	.68	141 (23.46)	198 (32.25)	.001
Unsafe pond[f]	54 (88.52)	49 (81.67)	.29	41 (77.36)	41 (77.36)	.99
Unsafe swimming pool[g]	19 (82.61)	15 (83.33)	.95	6 (50.00)	11 (78.57)	.13
Unsafe swimming[h]	253 (64.38)	256 (65.14)	.82	288 (51.61)	299 (52.73)	.71
Burns						
Unsafe hot water taps in bath/shower	188 (30.18)	176 (27.76)	.35	167 (26.09)	162 (25.16)	.70
Unsafe drinking hot fluids	351 (54.67)	328 (50.85)	.17	224 (34.84)	270 (41.73)	.01
Unsafe cooking (not using a stove guard)	628 (97.82)	619 (96.12)	.08	591 (92.49)	610 (94.57)	.13
Unsafe cooking (child in kitchen)	412 (64.27)	418 (64.81)	.84	583 (91.24)	601 (93.18)	.19
Unsafe cooking (not using rear burners on stove)	589 (91.89)	576 (89.86)	.21	507 (79.34)	550 (85.27)	.005
Unsafe cooking (not turning pan handles away)	334 (52.11)	321 (50.00)	.45	174 (27.27)	183 (28.33)	.67

[a] Mann-Whitney *U* test for continuous outcome, chi-square test for binominal outcomes.

[b] Only when applicable, such as a staircase is present (n=1154), a window a child can reach is present (n=404), the child is bathed (n=1220), a pond is present (n=106), a private swimming pool is present (n=26), or the child swims (n=1134).

[c] Baseline intervention: n=564, baseline control: n=565; follow-up intervention: n=573, follow-up control: n=581.

[d] Baseline intervention: n=241, baseline control: n=229; follow-up intervention: n=207, follow-up control: n=197.

[e] Baseline intervention: n=629, baseline control: n=637; follow-up intervention: n=601, follow-up control: n=619.

[f] Baseline intervention: n=65, baseline control: n=62; follow-up intervention: n=53, follow-up control: n=53.

[g] Baseline intervention: n=23, baseline control: n=18; follow-up intervention: n=12, follow-up control: n=14.

[h] Baseline intervention: n=402, baseline control: n=395; follow-up intervention: n=564, follow-up control: n=570.

Table 3. Outcomes of logistic regression analyses and linear regression analysis of the effect of E-Health4Uth home safety intervention on unsafe behavior at follow-up, with control condition as reference (N=1292).

Behavior	OR (95% CI)[a]	*P* value
Falls [b]		
Top of staircase	0.65 (0.50, 0.85)	.001
Bottom of staircase	0.69 (0.53, 0.88)	.003
Top and bottom of staircase	0.62 (0.48, 0.81)	< .001
Windows a child can reach	0.86 (0.58, 1.27)	.44
Poisoning		
Storage of cleaning products	0.67 (0.53, 0.85)	.001
Storage of medicines	0.88 (0.69, 1.12)	.30
Drowning [b]		
Bathing of the child	0.65 (0.51, 0.84)	.001
Pond	1.12 (0.44, 2.83)	.82
Private swimming pool	0.27 (0.05, 1.51)	.14
Swimming	0.98 (0.77, 1.24)	.87
Burns		
Hot water taps in bath/shower	1.05 (0.82, 1.35)	.73
Drinking hot fluids	0.76 (0.61, 0.96)	.02
Cooking (using a stove guard)	0.76 (0.48, 1.20)	.23
Cooking (child in kitchen)	0.81 (0.54, 1.23)	.33
Cooking (using rear burners)	0.67 (0.50, 0.90)	.008
Cooking (turning pan handles away)	0.94 (0.74, 1.20)	.63

[a] Logistic regression analyses with unsafe behavior as dependent variable and group (intervention condition vs control condition) as independent variable, adjusted for mother’s employment status.

[b] Only when applicable, such as a staircase is present (n=1154), a window a child can reach is present (n=404), the child is bathed (n=1220), a pond is present (n=106), a private swimming pool is present (n=26), or the child swims (n=1134).

[c] Linear regression analyses with unsafe behavior as dependent variable and group (intervention condition vs control condition) as independent variable, adjusted for mother’s employment status.

Parent Evaluation of the E-Health4Uth Home Safety Module

Of all parents in the intervention condition (n=643), 587 completed the E-Health4Uth home safety module (91.29%). The Web-based evaluation form of the E-Health4Uth home safety module was completed by 541 of 643 parents (84.14%) immediately after completing the E-Health4Uth home safety module (Table 4).

The Web-based evaluation forms showed that 72.07% (369/541) of parents had read the tailored safety advice completely, 24.41% (125/541) had read it partly, and 3.52% (18/541) had not read their advice. Parents evaluated the received tailored safety advice as being reliable (mean 4.19, SD 0.75), understandable (mean 4.36, SD 0.60), relevant (mean 3.53, SD 0.92), useful (mean 3.90, SD 0.77), and motivating to take action with regard to safety at home (mean 3.60, SD 0.90).

An implementation-intention plan was completed by 68.80% (322/541) of parents; a second implementation-intention plan was completed by 31.20% (146/541) of parents. Parents positively evaluated the ease of completing an implementation-intention plan for their own situation (mean 4.08, SD 0.79). Parents spent a mean time of 14.44 (SD 7.08) minutes to answer the questions and read the safety advice; they evaluated this as being a short time (mean 3.20, SD 0.56). Parents positively evaluated the use of the E-Health4Uth home safety intervention (mean 4.05, SD 0.62) and found the intervention to be a pleasant source of information (mean 3.67, SD 0.78). Parents rated the E-Health4Uth home safety intervention with a mean score of 7.28 (SD 1.14).

Table 4. Evaluation of the E-Health4Uth home safety module by parents in the intervention condition immediately after completing the module (n=541).

Subject	n (%)/mean (SD)
Reading of the Web-based, tailored safety advice, n (%)	
Have read their advice completely	369 (72.07)
Have read their advice partly	125 (24.41)
Have not read their advice	18 (3.52)
Tailored safety advice, mean (SD)	
The safety advice was reliable[a]	4.19 (0.75)
The safety advice was understandable[a]	4.36 (0.60)
The safety advice was relevant[a]	3.53 (0.92)
The safety advice was useful[a]	3.90 (0.77)
The safety advice motivated to take action[a]	3.60 (0.90)
Implementation-intention plan	
Completed an implementation-intention plan, n (%)	322 (68.80)
Completed a second implementation-intention plan, n (%)	146 (31.20)
Was it easy to complete an implementation-intention plan?, mean (SD)[a]	4.08 (0.79)
E-Health4Uth home safety intervention, mean (SD)	
Minutes spent to answer the questions and read the safety advice[a]	14.44 (7.08)
Did you think that the time spent to answer the questions and read the safety advice was (very) long or (very) short?[a]	3.20 (0.56)
Was the intervention easy to use?[a]	4.05 (0.62)
Was the intervention a pleasant source of information?[a]	3.67 (0.78)
Rating for the Web-based, tailored safety advice intervention[b]	7.28 (1.14)

[a] Scores on a 5-point Likert scale ranging from 1 (most negative) to 5 (most positive).

[b] Scores from 1 (most negative) to 10 (most positive).

Parent and Health Care Professional Evaluation of the Well-Baby Visit

During the well-baby visit, the tailored safety advice was discussed with 48.9% of the parents, was not discussed with 18.9%, and for 32.2% of the parents it was unclear whether the advice was discussed because no evaluation form was available and child health care professionals could not recall whether or not they had discussed this advice with the parent.

Parents (n=196) and child health care professionals (n=238) completed written evaluation forms immediately after the well-baby visit at which the tailored safety advice was discussed with the parent (Table 5).

Parents had a positive evaluation of the information discussed during the well-baby visit (mean 4.38, SD 0.62), rated discussing the tailored safety advice as a valuable supplement (mean 3.82, SD 0.87), and were satisfied overall with the well-baby visit (mean 4.38, SD 0.62). Parents rated the well-baby visit, including discussing the tailored safety advice, with a mean score of 8.20 (SD 0.87).

Child health care professionals reported that the mean total time spent in the well-baby visit was 20.40 (SD 4.51) minutes, with a mean of 5.70 (SD 2.27) minutes used for discussing safety at home. In addition to receiving tailored safety advice, the generic safety information leaflet was given to 72.03% (170/238) of the parents. The tailored safety advice was present in 87.82% (209/238) of the child dossiers and it was brought to the well-baby visit by 21.61% (51/238) of parents. Child health care professionals positively evaluated the tailored safety advice with regard to its usefulness to discuss safety at home during the well-baby visit (mean 3.77, SD 0.77), were satisfied with the information given to parents (mean 3.98, SD 0.64), and had an overall satisfaction with the well-baby visit (mean 4.01, SD 0.62). They rated the well-baby visit, including discussing the tailored safety advice, with a mean score of 7.30 (SD 0.79).

Table 5. Evaluation of the well-baby visit including discussing the tailored safety advice by parents in the intervention condition (n=196) and health care professionals (n=238) immediately after the well-baby visit.

Subject	n (%)/mean (SD)
Parents, mean (SD)	
Satisfaction with information discussed[a]	4.38 (0.62)
Discussing the tailored safety advice was a valuable supplement to the tailored safety advice?[a]	3.82 (0.87)
Overall satisfaction with the well-baby visit[a]	4.38 (0.62)
Rating for the well-baby visit[b]	8.20 (0.87)
Child health care professionals	
Total time for well-baby visit (min), mean (SD)	20.40 (4.51)
Time for safety during well-baby visit (min), mean (SD)	5.70 (2.27)
Safety information leaflet given to the parent, n (%)	170 (72.03)
Tailored safety advice present in dossier, n (%)	209 (87.82)
Tailored safety advice brought by parent, n (%)	51 (21.61)
Was the tailored safety advice useful to discuss safety at home during the well-baby visit?, mean (SD)[a]	3.77 (0.77)
Satisfaction with information given, mean (SD)[a]	3.98 (0.64)
Overall satisfaction with the well-baby visit, mean (SD)[a]	4.01 (0.62)
Rating for the well-baby visit, mean (SD)[b]	7.30 (0.79)

[a] Scores on a 5-point Likert scale ranging from 1 (most negative) to 5 (most positive).

[b] Scores from 1 (most negative) to 10 (most positive).

Discussion

Principal Findings

This study evaluated the effect of Web-based, tailored, safety advice combined with personal counseling on parents' child safety behaviors. Compared to counseling with generic written materials, the E-Health4Uth home safety intervention appeared to be effective in promoting several relevant parents' child safety behaviors. As hypothesized, parents in the intervention condition showed significantly less unsafe behavior with regard to safe staircases, storage of cleaning products, bathing, drinking hot fluids, and cooking compared to parents who received counseling with generic written safety information. There were no significant differences for other specific behaviors between the two study conditions. At follow-up, parents in the intervention condition also showed a significantly lower total risk score compared to parents in the control condition.

Parents in the intervention group were positive about the E-Health4Uth home safety module and its use in well-baby visits was positively evaluated by both parents and child health care professionals.

This study confirms the results of previous studies that showed that applying techniques of computer-tailored safety education in a primary care setting was effective in adopting safety behaviors of parents when compared to receiving generic safety advice [17,40]. The present study focused on total risk scores and also investigated the effects of the E-Health4Uth home safety intervention on specific safety behaviors. This approach allowed for clarification of which specific safety behaviors the intervention is or is not effective for. These insights may guide the development and evaluation of additional approaches to improve parental safety behavior. To our knowledge, this is the first experimental study on the effectiveness of computer-tailored education to change parents' safety behaviors.

These results support the use of Web-based tailored methods to help increase the effectiveness of parental safety advice. For the prevention of poisoning, the E-Health4Uth home safety intervention was effective on the storage of cleaning products. Although we anticipated that the intervention would have a similar effect on the storage of medicines, no difference was found in the unsafe storage of medicines between the intervention and control condition. At baseline, the prevalence of unsafe storage of medicines was, in fact, lower than the prevalence of unsafe storage of cleaning products. It is possible that the content of the intervention did not sufficiently increase parents' motivation to store their medicines in a more safe way.

The E-Health4Uth home safety intervention was not effective on behaviors for window safety, storage of medicines, ponds, swimming pools and swimming, hot water taps of bath/showers, and some items of burns prevention. It is possible that the intervention did not sufficiently address these specific determinants of safe behavior. An explanation why the E-Health4Uth home safety intervention showed no effect on the behavior for window safety, ponds, and swimming pools could be the low numbers of households that had a pond or swimming pool. This affects statistical power so the results for these behaviors should be interpreted with care.

Despite a lower prevalence in the E-Health4Uth intervention condition, some behaviors appeared to be more unsafe at follow-up compared to baseline (eg, bathing of the child and cooking with the child present in the kitchen). This result was found in both the intervention and the control condition, and might be because of the change in both age and development of the child between baseline and follow-up (Multimedia Appendix 6). For example, with increasing age parents might assume they can leave the child alone in the bathtub, or it may be more difficult for the parent to keep the child out of the kitchen. However, such behavior is not recommended and current safety advice (either generic or individually tailored) still seems suboptimal. Moreover, despite the decrease of unsafe behavior in the two study conditions between baseline and follow-up, the prevalence of many unsafe behaviors remained high. The prevalence of unsafe behavior for the top/bottom of staircases, ponds, swimming pools, and some items of burns prevention was still over 70%. This indicates that the content of the tailored safety advice for these behaviors needs to be improved.

When parents reported at follow-up (child age approximately 17 months) that their child swam, they reported leaving their child unsupervised in the swimming pool sometimes (0.81%), rarely (4.19%), and never (95.00%; data not shown). Despite the fact that most parents never left their child unsupervised in the swimming pool, 5% still left their child unsupervised, although these were very young children. These children are at risk of drowning and should never be left unsupervised because they do not have swimming skills yet.

This study used a tailored home safety intervention delivered in a primary care setting. Next to primary care settings, computer-tailored home safety information can be applied to other health care settings, such as emergency departments [41]. This study also shows that computer-tailored home safety information is effective in improving parents' child safety behavior.

Parents and youth health care professionals are positive about the tailored safety advice and the use of the tailored safety advice in well-baby visits. On the other hand, the intervention effect may have been diluted by suboptimal uptake of the novel method by parents and/or health care workers.

The E-Health4Uth home safety intervention consists of a home safety assessment questionnaire, Web-based tailored safety advice, an implementation-intention plan, and discussing the tailored safety advice with the child health care professional. However, the evaluations of the child health care professionals showed that the tailored safety advice was discussed with only 48.9% of the parents.

Unfortunately, we only received evaluation forms from approximately one-third of the health care professionals, so it is unknown whether the health care professionals discussed the tailored safety advice with the other parents. According to the evaluation forms we did receive from the health care professionals, there was a considerable chance of not having discussed the tailored safety advice. Examining this issue as a possible confounder in the logistic regression models showed similar results of the effect of the intervention on parents' child safety behaviors compared to our findings without adjustment for the discussion of the tailored advice (data not shown). The main reasons for not discussing the safety advice with the parents were that (1) the well-baby visit was made by another child health care professional from another well-baby clinic who was not familiar with the study, (2) parents indicated that discussing the tailored safety advice was not necessary, and (3) the tailored safety advice was not present in the child's dossier. However, although the tailored safety advice was not discussed with approximately 50% of the parents and uptake among parents needs improvement, a positive effect on parents' child safety behavior was shown.

In daily practice, all parents (in both the intervention or control condition) received care as usual: the generic safety information leaflet. Parents in the intervention condition received the E-Health4Uth home safety intervention, but could also receive the generic safety information leaflet as a part of usual care.

Strengths and Limitations

Our focus on the effect of a tailored intervention on both specific parents' child safety behaviors and on an overall safety risk score is a major strength of this study. Other strengths include the randomized controlled design, the large number of participants (N=1292) and the small number lost to follow-up: only 6.6% of the participants failed to complete the follow-up questionnaire. However, dropout was higher among mothers with a low educational level, unemployed mothers, and parents of non-Western ethnicity, which could affect the generalizability of the results. In addition, the participation rate was 45%.

We may have recruited parents who were more receptive to this way of providing safety education; in this case, this could have led to an overestimation of the intervention effect. On the other hand, the study population was a reasonable reflection of the general population in the Netherlands [42].

Because we had low numbers of missing data and participants lost to follow-up (6.60%), missing data was not imputed. Given these low numbers, it is not likely that missing data lead to loss of power of the study [43].

The intervention was developed for use on computers with connection to the Internet. The intervention was not tested for functionality on mobile phones and tablets. Perhaps in the near future when implementing the intervention, it could be made accessible on all mobile devices that have access to the Internet.

Receiving gift vouchers may cause recall bias because parents could expect to receive further incentives in the future. This may have positively biased total effectiveness for both the experimental and control condition although the magnitude of any effects would be small. Despite this possible recall bias, the E-Health4Uth intervention is effective in specific parents' child safety behaviors compared to receiving the control condition.

A high percentage of youth health care organizations declined to participate in the study. Of the 26 youth health care organizations that were initially invited to participate in the study, 5 volunteered to take part in the study. The main reason why health care providers did not wish to participate was that they were already involved in 1 or more other studies.

Finally, the high prevalence of unsafe parental behaviors might even be an underestimation of the real child safety situation. Because the present study relied on self-report of safety behavior by parents, misclassification might have occurred if parents gave socially desirable answers in order to look good (ie, overstating their safe behavior) [44-46]. Furthermore, self-report can be subject to recall bias or inaccurate responses. We tried to minimize the occurrence of socially desirable answers by ensuring confidentiality of the questionnaires. Earlier validation studies showed that there is an acceptable agreement between parents reported safety behavior and observations in homes of the parent [46]. Future studies with smaller samples assessing specific behaviors could include the use of home observations.

Implications and Future Research

Findings from this study support the use of a tailored education approach involving the provision of tailored safety information. The tailored safety information was found to be more effective than generic safety information in promoting preventive behavior. Providing tailored safety information before a visit to the well-baby clinic might be more efficient because parents and child health care professionals can better prepare for this visit in which safety at home is discussed [47-50]. Moreover, the parents receive information that is more specific because it is tailored to the personal situation of the parent [23]. However, because the prevalence of unsafe behavior remains relatively high, additional approaches to improve parental safety behavior need to be developed.

To improve parents’ child safety behaviors, various cognitions (eg, perceived self-efficacy, perceived response efficacy, perceived vulnerability, and perceived severity) could be addressed [51-53]. Changing these cognitions about injury prevention behavior could possibly lead to more safe behavior. More insight is needed into why the Web-based, tailored, safety advice intervention is effective for some parents and not for others. Perhaps different determinants are correlated with different safety behaviors.

It is possible that parents have different motivations for change for different injury mechanisms. This is supported by the finding that the intervention is not equally effective for all parents’ child safety behaviors. Therefore, this issue needs to be further explored.

Future studies should also investigate the effect of discussing the tailored safety advice during the well-baby clinic visit in a larger sample, as well as other approaches to increase the effectiveness of the E-Health4Uth intervention. Also, more insight is needed on the effect of the intervention among various subgroups (eg, based on ethnicity or educational level).

Conclusions

Compared to counseling with generic written materials, the E-Health4Uth home safety advice combined with counseling is effective in promoting parents’ child safety behavior for safe staircases, storage of cleaning products, bathing, drinking hot fluids, and cooking. There were no significant differences for other specific behaviors between the two study conditions.

Parents were positive about the E-Health4Uth home safety module and its use in well-baby visits was positively evaluated by both parents and child health care professionals. The results of this study support the application of Web-based, tailored, safety advice for the prevention of unintentional injuries in the youth health care setting.

Acknowledgments

The authors thank the parents and the staff of the preventive youth health care centers for their participation in this study. This study is funded by a grant from ZonMw, the Netherlands Organization for Health Research and Development (project no: 50-50205-98-25028000). The study’s sponsors had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, approval, or submission of the manuscript.

Authors' Contributions

HR had the original idea for the study and its design, and was responsible for acquiring the study grant. MB is responsible for the data collection, data analysis, and reporting of the study results. MS helped coordinate the study and participated in data collection. EB, PH, MS, and TB provided expert input during the study. HR and TB supervised the study. All authors regularly participated in discussing the design and protocols used in the study. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Safety advice of the intervention based on safety behavior in and around the home.

[PDF File (Adobe PDF File), 56KB - jmir_v16i1e17_app1.pdf]

Multimedia Appendix 2

Sample page 1 of tailored safety advice.

[JPG File, 406KB - jmir_v16i1e17_app2.jpg]

Multimedia Appendix 3

Sample page 2 of tailored safety advice.

[JPG File, 511KB - jmir_v16i1e17_app3.jpg]

Multimedia Appendix 4

Sample page 3 of tailored safety advice.

[JPG File, 441KB - jmir_v16i1e17_app4.jpg]

Multimedia Appendix 5

Risk scores assigned to injury safety behaviors.

[PDF File (Adobe PDF File), 202KB - jmir_v16i1e17_app5.pdf]

Multimedia Appendix 6

Child characteristics at baseline and follow-up divided in intervention condition and control condition.

[PDF File (Adobe PDF File), 302KB - jmir_v16i1e17_app6.pdf]

Multimedia Appendix 7

CONSORT-EHEALTH V1.6.2 checklist [54].

[PDF File (Adobe PDF File), 1002KB - jmir_v16i1e17_app7.pdf]

References

1. Peden M, Oyegbite K, Ozanne-Smith J, Hyder AA, Branche C, Fazlur Rahman AKM. World Report on Child Injury Prevention. Geneva: WHO; 2008.
2. Polinder S, Haagsma JA, Toet H, Brugmans MJ, van Beeck EF, EUROCOSTAPOLLO reference groups. Burden of injury in childhood and adolescence in 8 European countries. BMC Public Health 2010;10:45 [FREE Full text] [doi: 10.1186/1471-2458-10-45] [Medline: 20113463]
3. Polinder S, Meerding WJ, Toet H, Mulder S, Essink-Bot ML, van Beeck EF. Prevalence and prognostic factors of disability after childhood injury. Pediatrics 2005 Dec;116(6):e810-e817 [FREE Full text] [doi: 10.1542/peds.2005-1035] [Medline: 16322139]
4. Maskell J, Newcombe P, Martin G, Kimble R. Psychosocial functioning differences in pediatric burn survivors compared with healthy norms. J Burn Care Res 2013;34(4):465-476. [doi: 10.1097/BCR.0b013e31827217a9] [Medline: 23702857]
5. Theurer WM, Bhavsar AK. Prevention of unintentional childhood injury. Am Fam Physician 2013 Apr 1;87(7):502-509. [Medline: 23547592]
6. VeiligheidNL. Letsel informatie systeem. Letsel Informatie Systeem (Lis) [Dutch Injury Surveillance System Amsterdam] 2011 Dec 31:1-1.
7. VeiligheidNL. Veilig Groot Worden: Veiligheidskaarten Voor Kinderen Van 0-6 Maanden, 6-12 Maanden, 1-2 Jaar, 2-4 Jaar [Growing up Safe; Safety Information Leaflets for Children Aged 0-6 Months, 6-12 Months, 1-2 Years, 2-1 Years]. Amsterdam: VeiligheidNL; 2002.
8. van der Burg-Beijk P. Veiligheidskaarten, Een Waardevol Hulpmiddel [Safety information leaflets, a valuable resource]. T Jeugdgezondheidszorg 1990:67-68.
9. VeilgheidNL. Factsheet 4: Onderzoek Naar De Effectiviteit Van De Veiligheidsinformatiekaarten [Factsheet 4: Study of the Effectiveness of the Safety Information Leaflets]. Amsterdam: VeilgheidNL; 2007.
10. Basistakenpakket Jeugdgezondheidszorg 0-19 Jaar Youth Health Care Tasks 0-19 Years. Basistakenpakket Jeugdgezondheidszorg. Ministerie van Volksgezondheid, Welzijn en Sport 2002:1-1.
11. Verbrugge HP. Youth health care in The Netherlands: a bird's eye view. Pediatrics 1990 Dec;86(6 Pt 2):1044-1047. [Medline: 2243736]
12. Höllwarth ME. Prevention of unintentional injuries: a global role for pediatricians. Pediatrics 2013 Jul;132(1):4-7. [doi: 10.1542/peds.2013-0571] [Medline: 23776116]
13. Centraal Bureau voor de Statistiek. 2010 Mar 16. Statline: Gebruik medische voorzieningen; t/m 2009 [Use of medical facilities; t/m 2009] URL: http://statline.cbs.nl/StatWeb/publication/

?VW=T&DM=SLNL&PA=7042mc&D1=390-425&D2=a&HD=111003-1301&HDR=T&STB=G1 [WebCite Cache ID 6D0Kw0MN1]

14. US Preventive Services Task Force. Guide to Clinical Preventive Services 2006: Recommendations of the US Preventive Services Task Force. Baltimore: Williams & Wilkins; 2006.
15. Beirens TM, Brug J, van Beeck EF, Dekker R, Juttmann RE, Raat H. Presence and use of stair gates in homes with toddlers (11-18 months old). Accid Anal Prev 2007 Sep;39(5):964-968. [doi: 10.1016/j.aap.2007.01.001] [Medline: 17324371]
16. Beirens TM, van Beeck EF, Dekker R, Brug J, Raat H. Unsafe storage of poisons in homes with toddlers. Accid Anal Prev 2006 Jul;38(4):772-776. [doi: 10.1016/j.aap.2006.02.007] [Medline: 16545327]
17. Nansel TR, Weaver N, Donlin M, Jacobsen H, Kreuter MW, Simons-Morton B. Baby, Be Safe: the effect of tailored communications for pediatric injury prevention provided in a primary care setting. Patient Educ Couns 2002 Mar;46(3):175-190 [FREE Full text] [Medline: 11932115]
18. Mangunkusumo R, Brug J, Duisterhout J, de Koning H, Raat H. Feasibility, acceptability, and quality of Internet-administered adolescent health promotion in a preventive-care setting. Health Educ Res 2007 Feb;22(1):1-13 [FREE Full text] [doi: 10.1093/her/cyl010] [Medline: 16714366]
19. Mangunkusumo RT, Brug J, de Koning HJ, van der Lei J, Raat H. School-based internet-tailored fruit and vegetable education combined with brief counselling increases children's awareness of intake levels. Public Health Nutr 2007 Mar;10(3):273-279. [doi: 10.1017/S1368980007246671] [Medline: 17288625]
20. Crane LA, Asdigian NL, Barón AE, Aalborg J, Marcus AC, Mokrohisky ST, et al. Mailed intervention to promote sun protection of children: a randomized controlled trial. Am J Prev Med 2012 Oct;43(4):399-410. [doi: 10.1016/j.amepre.2012.06.022] [Medline: 22992358]
21. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
22. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
23. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
24. Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med 1999;21(4):290-298. [Medline: 10721435]
25. Short CE, James EL, Plotnikoff RC, Girgis A. Efficacy of tailored-print interventions to promote physical activity: a systematic review of randomised trials. Int J Behav Nutr Phys Act 2011;8:113 [FREE Full text] [doi: 10.1186/1479-5868-8-113] [Medline: 21999329]
26. van Beelen ME, Vogel I, Beirens TM, Kloek GC, den Hertog P, van der Veen MD, et al. Web-Based eHealth to Support Counseling in Routine Well-Child Care: Pilot Study of E-health4Uth Home Safety. JMIR Res Protoc 2013;2(1):e9. [doi: 10.2196/resprot.1862] [Medline: 23611794]
27. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
28. Lemmens VE, Oenema A, Klepp KI, Henriksen HB, Brug J. A systematic review of the evidence regarding efficacy of obesity prevention interventions among adults. Obes Rev 2008 Sep;9(5):446-455. [doi: 10.1111/j.1467-789X.2008.00468.x] [Medline: 18298429]
29. van Beelen ME, Beirens TM, Struijk MK, den Hertog P, Oenema A, van Beeck EF, et al. 'BeSAFE', effect-evaluation of internet-based, tailored safety information combined with personal counselling on parents' child safety behaviours: study design of a randomized controlled trial. BMC Public Health 2010;10:466 [FREE Full text] [doi: 10.1186/1471-2458-10-466] [Medline: 20696070]
30. Injury prevention program. URL: http://www.aap.org/FAMILY/tippmain.htm [accessed 2014-01-05] [WebCite Cache ID 6MPJIsoxy]
31. Kloek G, Raat H. Het Begin Van Gezond Leven; Interactief Advies-Op-Maat Voor Ouders Van Jonge Kinderen in Combinatie Met Persoonlijke Advisering Op Het Consultatiebureau [Interactive Tailored Advice for Parents of Young Children Combined with Personal Advice at the Child Health Clinic]. Rotterdam: Afdeling Maatschappelijke gezondheidszorg. Erasmus MC - Universitair Medisch Centrum Rotterdam; 2007.
32. Kendrick D, Barlow J, Hampshire A, Polnay L, Stewart-Brown S. Parenting interventions for the prevention of unintentional injuries in childhood. Cochrane Database Syst Rev 2007(4):CD006020. [doi: 10.1002/14651858.CD006020.pub2] [Medline: 17943875]
33. Gielen AC, Wilson ME, McDonald EM, Serwint JR, Andrews JS, Hwang WT, et al. Randomized trial of enhanced anticipatory guidance for injury prevention. Arch Pediatr Adolesc Med 2001 Jan;155(1):42-49. [Medline: 11177061]
34. VeiligheidNL. Protocol Veilig Groot Worden, Methodiek Veiligheidsinformatiekaarten [Protocol Growing up Safe, Methodology Safety Information Leaflets]. Amsterdam: VeiligheidNL; 2007.
35. Sheeran P. Intention-Behavior Relations: A Conceptual and Empirical Review. European Review of Social Psychology 2002;12:1-1. [doi: 10.1002/0470013478.ch1]

36. Sheeran P, Silverman M. Evaluation of three interventions to promote workplace health and safety: evidence for the utility of implementation intentions. Soc Sci Med 2003 May;56(10):2153-2163. [Medline: 12697204]
37. Sindelar HA, Abrantes AM, Hart C, Lewander W, Spirito A. Motivational interviewing in pediatric practice. Curr Probl Pediatr Adolesc Health Care 2004 Oct;34(9):322-339. [doi: 10.1016/j.cppeds.2004.06.003] [Medline: 15383822]
38. Centraal Bureau voor de Statistiek. Definities Demographische Gegevens [Definitions Demographics] URL: http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm [accessed 2012-12-18] [WebCite Cache ID 6D0M8KCS2]
39. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999 Sep 11;319(7211):670-674 [FREE Full text] [Medline: 10480822]
40. McDonald EM, Solomon B, Shields W, Serwint JR, Jacobsen H, Weaver NL, et al. Evaluation of kiosk-based tailoring to promote household safety behaviors in an urban pediatric primary care practice. Patient Educ Couns 2005 Aug;58(2):168-181. [doi: 10.1016/j.pec.2004.08.015] [Medline: 16009293]
41. Shields WC, McDonald EM, McKenzie L, Wang MC, Walker AR, Gielen AC. Using the pediatric emergency department to deliver tailored safety messages: results of a randomized controlled trial. Pediatr Emerg Care 2013 May;29(5):628-634. [doi: 10.1097/PEC.0b013e31828e9cd2] [Medline: 23603653]
42. Centraal Bureau voor de Statistiek [Statistics Netherlands]. URL: http://www.cbs.nl/nl-NL/menu/home/default.htm [accessed 2014-01-05] [WebCite Cache ID 6MPJWujdA]
43. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393 [FREE Full text] [Medline: 19564179]
44. Chen LH, Gielen AC, McDonald EM. Validity of self reported home safety practices. Inj Prev 2003 Mar;9(1):73-75 [FREE Full text] [Medline: 12642564]
45. Morrongiello BA, Kiriakou S. Mothers' home-safety practices for preventing six types of childhood injuries: what do they do, and why? J Pediatr Psychol 2004 Jun;29(4):285-297 [FREE Full text] [Medline: 15148351]
46. Watson M, Kendrick D, Coupland C. Validation of a home safety questionnaire used in a randomised controlled trial. Inj Prev 2003 Jun;9(2):180-183 [FREE Full text] [Medline: 12810749]
47. Paperny DM, Hedberg VA. Computer-assisted health counselor visits: a low-cost model for comprehensive adolescent preventive services. Arch Pediatr Adolesc Med 1999 Jan;153(1):63-67. [Medline: 9895001]
48. Fotheringham MJ, Owies D, Leslie E, Owen N. Interactive health communication in preventive medicine: internet-based strategies in teaching and research. Am J Prev Med 2000 Aug;19(2):113-120. [Medline: 10913902]
49. Patrick K, Sallis JF, Prochaska JJ, Lydston DD, Calfas KJ, Zabinski MF, et al. A multicomponent program for nutrition and physical activity change in primary care: PACE+ for adolescents. Arch Pediatr Adolesc Med 2001 Aug;155(8):940-946. [Medline: 11483123]
50. Sciamanna CN, Novak SP, Houston TK, Gramling R, Marcus BH. Visit satisfaction and tailored health behavior communications in primary care. Am J Prev Med 2004 Jun;26(5):426-430. [doi: 10.1016/j.amepre.2004.02.003] [Medline: 15165659]
51. Beirens TM, Brug J, van Beeck EF, Dekker R, Juttmann RE, Raat H. Presence and use of stair gates in homes with toddlers (11-18 months old). Accid Anal Prev 2007 Sep;39(5):964-968. [doi: 10.1016/j.aap.2007.01.001] [Medline: 17324371]
52. van Beelen ME, Beirens TM, den Hertog P, van Beeck EF, Raat H. First-time parents are not well enough prepared for the safety of their infant. PLoS One 2013;8(3):e58062 [FREE Full text] [doi: 10.1371/journal.pone.0058062] [Medline: 23483969]
53. van Beelen ME, van Beeck EF, den Hertog P, Beirens TM, Raat H. Correlates of unsupervised bathing of infants: a cross-sectional study. Int J Environ Res Public Health 2013 Mar;10(3):856-866 [FREE Full text] [doi: 10.3390/ijerph10030856] [Medline: 23459215]
54. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Edited by G Eysenbach; submitted 10.01.13; peer-reviewed by D Schwebel, G Floros; comments to author 12.05.13; revised version received 08.07.13; accepted 12.08.13; published 24.01.14.

Please cite as:
van Beelen MEJ, Beirens TMJ, den Hertog P, van Beeck EF, Raat H
Effectiveness of Web-Based Tailored Advice on Parents' Child Safety Behaviors: Randomized Controlled Trial
J Med Internet Res 2014;16(1):e17
URL: http://www.jmir.org/2014/1/e17/
doi:10.2196/jmir.2521
PMID:24463421

Original Paper

Effects of a Web-Based Tailored Multiple-Lifestyle Intervention for Adults: A Two-Year Randomized Controlled Trial Comparing Sequential and Simultaneous Delivery Modes

Daniela N Schulz[1], MSc; Stef PJ Kremers[2], PhD; Corneel Vandelanotte[3], PhD; Mathieu JG van Adrichem[1], MSc; Francine Schneider[1], MSc; Math JJM Candel[4], PhD; Hein de Vries[1], PhD

[1]CAPHRI School for Public Health and Primary Care, Department of Health Promotion, Maastricht University, Maastricht, Netherlands
[2]Nutrition and Toxicology Research Institute Maastricht (NUTRIM), Department of Health Promotion, Maastricht University, Maastricht, Netherlands
[3]Institute for Health and Social Science Research, Centre for Physical Activity Studies, Central Queensland University, Rockhampton, Australia
[4]CAPHRI School for Public Health and Primary Care, Department of Methodology and Statistics, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Daniela N Schulz, MSc
CAPHRI School for Public Health and Primary Care
Department of Health Promotion
Maastricht University
P Debyeplein 1
Maastricht, 6229 HA
Netherlands
Phone: 31 43 388 2832
Fax: 31 43 367 1032
Email: dn.schulz@maastrichtuniversity.nl

Abstract

Background: Web-based computer-tailored interventions for multiple health behaviors can have a significant public health impact. Yet, few randomized controlled trials have tested this assumption.

Objective: The objective of this paper was to test the effects of a sequential and simultaneous Web-based tailored intervention on multiple lifestyle behaviors.

Methods: A randomized controlled trial was conducted with 3 tailoring conditions (ie, sequential, simultaneous, and control conditions) in the Netherlands in 2009-2012. Follow-up measurements took place after 12 and 24 months. The intervention content was based on the I-Change model. In a health risk appraisal, all respondents (N=5055) received feedback on their lifestyle behaviors that indicated whether they complied with the Dutch guidelines for physical activity, vegetable consumption, fruit consumption, alcohol intake, and smoking. Participants in the sequential (n=1736) and simultaneous (n=1638) conditions received tailored motivational feedback to change unhealthy behaviors one at a time (sequential) or all at the same time (simultaneous). Mixed model analyses were performed as primary analyses; regression analyses were done as sensitivity analyses. An overall risk score was used as outcome measure, then effects on the 5 individual lifestyle behaviors were assessed and a process evaluation was performed regarding exposure to and appreciation of the intervention.

Results: Both tailoring strategies were associated with small self-reported behavioral changes. The sequential condition had the most significant effects compared to the control condition after 12 months (T1, effect size=0.28). After 24 months (T2), the simultaneous condition was most effective (effect size=0.18). All 5 individual lifestyle behaviors changed over time, but few effects differed significantly between the conditions. At both follow-ups, the sequential condition had significant changes in smoking abstinence compared to the simultaneous condition (T1 effect size=0.31; T2 effect size=0.41). The sequential condition was more effective in decreasing alcohol consumption than the control condition at 24 months (effect size=0.27). Change was predicted by the amount of exposure to the intervention (total visiting time: beta=–.06; *P*=.01; total number of visits: beta=–.11; *P*<.001). Both interventions were appreciated well by respondents without significant differences between conditions.

Conclusions: Although evidence was found for the effectiveness of both programs, no simple conclusive finding could be drawn about which intervention mode was more effective. The best kind of intervention may depend on the behavior that is targeted or on personal preferences and motivation. Further research is needed to identify moderators of intervention effectiveness. The results need to be interpreted in view of the high and selective dropout rates, multiple comparisons, and modest effect sizes. However, a large number of people were reached at low cost and behavioral change was achieved after 2 years.

Trial Registration: Nederlands Trial Register: NTR 2168; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2168 (Archived by WebCite at http://www.webcitation.org/6MbUqttYB).

(J Med Internet Res 2014;16(1):e26) doi:10.2196/jmir.3094

KEYWORDS

multiple behavior change; Web-based intervention; computer tailoring; effectiveness; physical activity; fruit consumption; vegetable consumption; alcohol intake; smoking

Introduction

Since the development of the Internet, several types of Web-based interventions have been offered to the population to modify unhealthy lifestyle behaviors. An unhealthy lifestyle can be described as one that is not compliant with the guidelines for different prominent health risk behaviors, such as being insufficiently active, eating insufficient fruit and vegetables, drinking too much alcohol, and using tobacco [1,2]. Unhealthy lifestyle habits are among the main causes of mortality and morbidity [3]. Noncommunicable chronic diseases, such as heart diseases, cancer, diabetes, and chronic respiratory diseases [1], are associated with a limited number of common modifiable health behaviors [4].

Given the high prevalence of unhealthy lifestyle habits [5-8], it is reasonable to offer interventions that can be disseminated among large numbers of people at low cost [9]. Because individuals with multiple unhealthy lifestyle behaviors are at the greatest risk of developing chronic diseases leading to increased health care costs [10], Web-based interventions with a focus on different lifestyle behaviors and integrated within one intervention seem to be an appropriate choice. To increase changes in health behavior, tailored interventions using a computerized expert system to select the best-fitting messages to generate personal relevant feedback messages have been developed [11]. Although inconsistent findings are reported [12,13], many Web-based computer-tailored interventions have proven to be an effective tool for improving health-related behaviors (eg, [14-16]). Different interventions based on the I-Change model have shown positive results [17-19]. As shown by Webb et al [20], the use of theory results in larger effect sizes. Moreover, Web-based computer-tailored interventions that contain relevant and attractive information adapted to the respondents' individual characteristics and needs have proven to be cost-effective (eg, [21,22]) and have been evaluated more positively than general information [8].

It has been suggested that interventions that focus on multiple behaviors have a greater impact on public health than single-behavior interventions [8,23]. However, such interventions are more extensive and, thus, require more engagement, time, and effort from the respondent. There is limited and inconsistent evidence about how best to accomplish multiple behavior change when using Web-based computer-tailored interventions. One strategy is to intervene in a single behavior at a time (sequential approach); another approach is to intervene in all health risk behaviors at the same time (simultaneous approach) [8,24-30]. In earlier studies, no consistent findings were reported regarding the most effective strategy. For example, King et al [25] and Hyman et al [24] reported that their simultaneous intervention mode was superior to their sequential intervention mode. The first aimed at changes in diet and physical activity; the latter aimed at smoking cessation and improvements in diet and physical activity. A study by Vandelanotte et al [30] aimed at lowering fat intake and increasing physical activity found no differences between the sequential and the simultaneous condition.

Focusing on different behaviors sequentially at different points in time is associated with less cognitive effort during the individual visits; however, respondents may experience lower levels of autonomy because of the limited choices during the individual visits [31,32], and repeated participation is necessary to receive information about multiple behaviors. Intervening in all behaviors simultaneously has the advantage that respondents receive all relevant information during the first visit; however, such a program may lead to ego depletion by overwhelming respondents with too much information [33,34] which may lead to a more negative feeling regarding the intervention and immediate dropout. In previous research, dose-response relationships have been found between exposure to an intervention (number and duration of exposures) and behavior change outcomes [35,36]. This may imply that repeated participation and thus repeated exposure to the Web-based program can be beneficial for realizing substantial behavioral change [37]. Appreciation of the intervention may lead to increased use, which, in turn, may improve intervention effectiveness [18,38,39]. Thus, both delivery modes have potential advantages and disadvantages that may influence their effectiveness, use, and the appreciation of the different types of interventions. Current evidence regarding the effectiveness of sequential and simultaneous delivery modes is inconsistent and none of these studies to date have tested the 2 delivery modes within an intervention targeting 5 different health behaviors.

In summary, randomized controlled trials assessing the effects, especially the long-term effectiveness (after more than 1 year) [40], of computer-tailored multiple-lifestyle interventions using different strategies (sequential vs simultaneous) among adults are scarce. Therefore, the main aim of the current study was to assess whether a multisession, Web-based, tailored lifestyle intervention was effective in enhancing multiple lifestyle behaviors (ie, physical activity, vegetable consumption, fruit consumption, alcohol intake, and smoking) in the long term. First, potential differences in effects of lifestyle change in general were assessed between the sequential and simultaneous delivery mode and a control condition at 12-month and 24-month follow-up. Second, we evaluated whether there were differences between the 3 groups with regard to effects on adherence to individual guidelines for each of the 5 behaviors being assessed. As a tertiary objective, a process evaluation was executed by studying the influence of intervention exposure on effectiveness and by summarizing the appreciation of the intervention.

Methods

Overview

A detailed description of the study protocol has been published elsewhere [27]; therefore, a summary of study methodology and procedures is provided subsequently.

Participants, Procedure, and Study Design

We conducted a randomized controlled trial (Dutch Trial Register NTR2168), involving 2 experimental conditions and a control condition. The Web-based intervention, focusing on unhealthy lifestyle behaviors in the general population, was conducted in the Netherlands from November 2009 to July 2012. Follow-up measurements took place 12 (T1) and 24 months (T2) after the first intervention visit. In the first instance, adult participants were recruited via 4 Dutch Regional Health Authorities that had conducted the quadrennial Adult Health Monitor 2009 of inhabitants of the provinces of North-Brabant and Zeeland (N=96,388). This monitoring tool is used to assess general health (eg, physical and mental health) and cover health-related topics (eg, social and physical environment) among representative samples of the Dutch population [41]. Of the 41,155 (42.70%) respondents who completed the Monitor, 24,215 (58.84%) filled out the written version and 16,940 (41.16%) filled out the online version (see also [42]). Our intervention was partly integrated into the online version of the Monitor. At the end of this Web-based questionnaire, participants were invited to take part in our intervention study. They had to give informed consent and provide their email address for participation and the handling of their data. Three weeks later, the study sample received an email containing an invitation and a link to the intervention website. After 1 month, a reminder email was sent to the individuals in the sample who had not responded to the first invitation. The website was also open to the general public, so it was also possible to register directly on the website without having to complete the Monitor. Randomization to 1 of the 3 study groups took place by means of a computer software randomization system. No block or cluster randomization was applied; the randomization was done at the individual level. The following inclusion criteria were established for this study: being between ages 18 and 65 years, having a computer with Internet access, having basic Internet literacy, and having a valid email address. The study was approved by the Medical Ethics Committee of Maastricht University and the University Hospital Maastricht (MEC 09-3-016/NL27235.068.09).

Intervention Program

Overview

The intervention program, called myHealthyBehavior (Dutch: *mijnGezondGedrag*), was a Web-based computer-tailored program targeting adults. The main aim of the intervention was to motivate participants to enhance 5 health behaviors. The theoretical framework for the development of the intervention was the I-Change model [43].

The Health Risk Appraisal

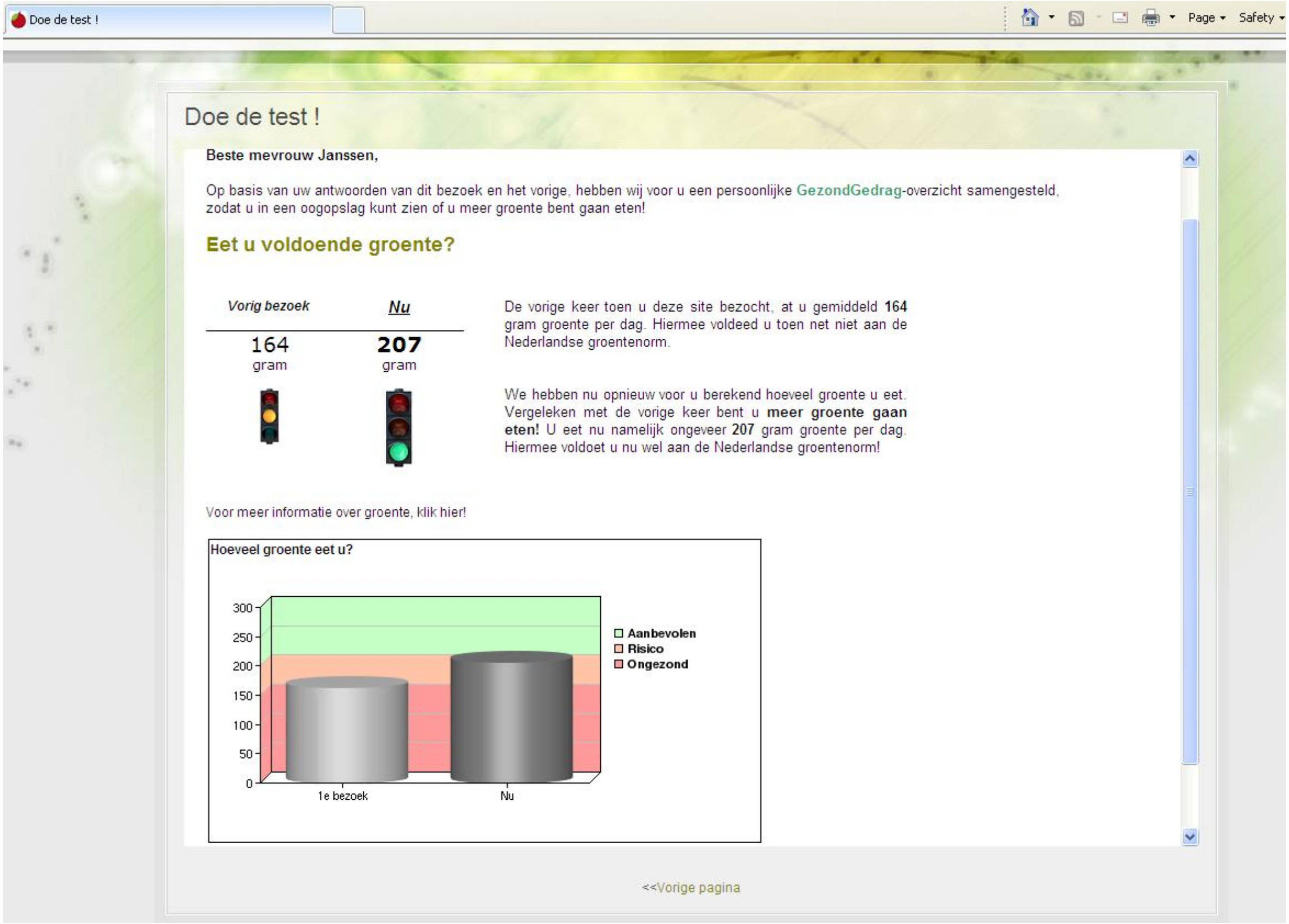
In the first part of the intervention, all groups received a health risk appraisal (HRA) in which the 5 behaviors were placed in the context of the following Dutch public health guidelines for each of the different behaviors: being moderately physically active for 30 minutes on at least 5 days a week [44], eating 200 g of vegetables per day [45], eating 2 pieces of fruit per day [45], not drinking more than 1 (women) or 2 (men) glasses of alcohol a day [45], and not smoking [46]. Figures with pictures of traffic lights were presented for every behavior to indicate whether respondents met (green), almost met (orange), or did not meet (red) the guidelines.

Sequential and Simultaneous Delivery Modes

The second part of the intervention was only available for the 2 experimental conditions: (1) the sequential condition and (2) the simultaneous condition. In this part, self-assessed questionnaires were used to measure the psychosocial concepts of the I-Change model [43]. The feedback messages were based on these questionnaires and were revealed on the respondents' computer screens immediately after the completion of the surveys. Respondents who did not comply with at least 1 guideline were invited to change the behavior which they had not complied with. They were invited to complete additional questions regarding motivational constructs related to a chosen behavior (sequential condition) or all relevant behaviors (simultaneous condition) within 1 or more modules of the intervention, respectively. Motivational feedback was given that was related to the relevant behaviors. Feedback on each person's perception of the pros and cons of the health behavior (attitudinal feedback) was given as the first step of the program, followed by feedback on perceived social influences as the second step. For example, information was given on how the social environment could help the respondent to live healthily. In the third step, the concept of preparatory planning was addressed. Feedback was provided on how to prepare for behavior change, for example, by planning to be physically active at fixed times, adding the plans to an agenda, trying new/different kinds of sports, having enough vegetables in stock at home, eating fruit or vegetables as a snack, and (for smokers) planning a quit date. In the final step, the focus was placed on how to cope with difficult situations to also increase self-efficacy. Several tailoring strategies were used in the feedback messages; for example, respondents were addressed by their name, normative and ipsative feedback were given (ie, during revisits current scores were compared to scores of previous visits; see Figure 1), graphs and bar charts were included, and a personal tone and empathy was applied.

It was a multisession program, in which respondents were encouraged to revisit the website on an unlimited basis. After 1 year, respondents in the sequential condition had the opportunity to choose a second module and to receive feedback on another lifestyle behavior. Respondents in the simultaneous condition received feedback on all behaviors for which they did not meet the guideline simultaneously at baseline and after 12 months. When visiting the website in the meantime, the sequential group received feedback on the chosen behavior at baseline or at 12-month follow-up, respectively; the simultaneous group received feedback on all unhealthy behaviors they reported at that moment. After 24 months, an invitation to complete the last follow-up questionnaire was sent to all respondents, followed by 2 reminders to increase the response rate.

Figure 1. Screenshot of the intervention website showing the vegetable part of the health risk appraisal. The traffic lights show that the respondent did not comply with the vegetable guideline at her first visit (orange traffic light), but that she complied with the guideline at her second visit (green traffic light). The graph shows the scores (amount of vegetables) at all visits.



Doe de test !

Beste mevrouw Janssen,

Op basis van uw antwoorden van dit bezoek en het vorige, hebben wij voor u een persoonlijke GezondGedrag-overzicht samengesteld, zodat u in een oogopslag kunt zien of u meer groente bent gaan eten!

Eet u voldoende groente?

Vorig bezoek	***Nu***
164 gram	**207** gram

De vorige keer toen u deze site bezocht, at u gemiddeld **164** gram groente per dag. Hiermee voldeed u toen net niet aan de Nederlandse groentenorm.

We hebben nu opnieuw voor u berekend hoeveel groente u eet. Vergeleken met de vorige keer bent u **meer groente gaan eten!** U eet nu namelijk ongeveer **207** gram groente per dag. Hiermee voldoet u nu wel aan de Nederlandse groentenorm!

Voor meer informatie over groente, klik hier!

Hoeveel groente eet u?

<<Vorige pagina

Questionnaires

Demographic Information

We assessed 8 demographic variables: age, gender (male=1; female=2), educational level (low=1, no education or primary education; medium=2, secondary education; high=3, tertiary education), net household income (euros per month), employment situation, marital status, number of persons in household, and native country (The Netherlands=1; other country=2).

Health Status

We assessed different kinds of diseases, such as cardiovascular diseases, diabetes, cancer, and high blood pressure. The 12-item short form (SF-12) Health Survey [47,48] was used to assess quality of life. The Kessler Psychological Distress Scale (K10) [49] was included to assess symptoms of depression and anxiety. The items regarding height and weight were used to estimate the body mass index (BMI).

Lifestyle Behavior

Physical activity was measured by the short questionnaire to assess health-enhancing physical activity (SQUASH) [41,50], and guideline adherence was assessed using comparable procedures developed by Ainsworth et al [51]. Fruit consumption was measured using a 4-item food frequency questionnaire (FFQ) assessing weekly fruit and fruit juice intake [41]. Vegetable consumption was measured using a 4-item FFQ assessing the weekly consumption of boiled or baked vegetables, as well as salads or raw vegetables [41]. Alcohol intake was measured by the 5-item Dutch Quantity-Frequency-Variability (QFV) questionnaire [52].

Current smoking behavior was assessed by asking participants if they smoked, what they smoked (cigarettes, cigars, or pipe tobacco), and how much they smoked per day (cigarettes) or per week (cigars or pipe tobacco) [41]. The answers were converted into an overall score for tobacco consumption (expressed in number of cigarettes): 1 cigar=4 cigarettes and 1 package of pipe tobacco (50 g)=50 cigarettes [53].

Social Cognitive Variables

Based on earlier studies [8,15], the following social cognitive variables were assessed regarding 1 or more of the 5 different lifestyle behaviors (dependent on the study condition). These variables were used to compose the personalized advice.

Intention was assessed by 1 item per behavior using an extended version of the stage of change concept [54,55]. For example, for the question “Do you intend to be physically active for 30 minutes on at least 5 days a week?” answers included no, I don’t intend to do so (=1); I never thought about it (=2); I thought about it, but I don’t know yet (=3); yes, but not within the next 5 years (=4); yes, within 1 to 5 years (=5); yes, within 6 to 12 months (=6); yes, within 3 to 6 months (=7); yes, within 1 to 3 months (=8); yes, within a month (=9); yes, and I’m already doing so (=10).

Attitude was assessed by 3 pros per behavior, such as “regular physical activity is good for my health” with responses ranging from totally disagree (=1) to totally agree (=5); and by 3 cons, such as “regular physical activity costs a lot of time” from totally disagree (=1) to totally agree (=5).

Social influence was assessed by social norms (1 item), social modeling (1 item) and social support (1 item), such as “According to the people in my direct environment...” with answers ranging from I certainly should smoke (=1) to I certainly should not smoke (=5); “How many people in your direct environment smoke?” with answers ranging from nobody (=1) to everybody (=5); and “People from my direct environment support me not to smoke” with answers ranging from yes, they support me a lot (=4) to no, they do not support me at all (=1).

Preparatory plans were assessed by 3 items per behavior, such as “I intend to allow time for physical activity” with answers ranging from no, definitely not (=1) to yes, definitely (=5).

Self-efficacy was assessed by 6 items per behavior regarding difficult social, emotional, and routine situations, such as “I am able to meet the alcohol guideline...when I am at a party;...when I feel stressed or nervous;...during a meal,” or “I am able to eat sufficient vegetables when I have other delicious food at home” with answers ranging from no, definitely not (=1) to yes, definitely (=5).

Coping plans were assessed by 6 items per behavior, such as “I have made a plan to drink no more than 2 glasses of alcohol when I feel stressed or nervous” with answers ranging from totally disagree (=1) to totally agree (=5).

Process Evaluation

To report user statistics, we recorded the number of log-ins to the tailored intervention per respondent and the time respondents spent on the tailored intervention during each visit. At baseline, after 12 months, and after 24 months, respondents evaluated the HRA on a scale from very bad (=1) to excellent (=10). At 24-month follow-up, all respondents were invited to complete 4 more questions measuring appreciation of the program by assessing the user-friendliness of the website and satisfaction with the layout, the HRA in general, and the use of traffic lights in particular. Additionally, a subsample completed a separate questionnaire, including items evaluating the website (6 items), the HRA (13 items), and the personalized advice (15 items).

Power Analyses

Linear mixed model analysis (lifestyle factor is the outcome variable) was the main analysis and power analysis suggested that a total sample of 1182 respondents was needed (correcting for possible attrition) based on P=.05, a power of 80%, a 2-sided test, and an effect size (ES) of 0.20. For the logistic analyses, a sample of 882 respondents was needed.

Statistical Analyses

The data were analyzed using SPSS software version 20 (IBM Corp, Armonk, NY, USA). To examine whether the randomization had been successful, the 3 groups were compared in terms of demographics, health status, and lifestyle behavior; ANOVAs were executed for continuous variables and chi-square tests for discrete variables. In the case of significant differences, variables were included as covariates (ie, potential confounders) in subsequent analyses.

Descriptive statistics were used to describe the characteristics of the study sample and the dropout rate within the groups. Dropout analyses, including a comparison between respondents lost to follow-up and T0-T1 completers (ie, respondents who completed the baseline and the 12-month follow-up measurement) and T0-T2 completers (ie, respondents who completed the baseline and the 24-month follow-up measurement), respectively, and fully complete cases were done using ANOVAs and chi-square tests.

We calculated a risk factor score by summing up all risky/unhealthy behaviors defined by the guideline status; the value for the risk factor score could range from zero (adhering to all guidelines) to 5 (adhering to no guideline). The 3 study groups were compared in terms of their lifestyle behavior at the follow-up measurements. First, ES were calculated (Cohen’s *d*). Those ES below 0.30 were considered small, whereas those between 0.30 and 0.80 were considered medium, and those larger than 0.80 were regarded as large [56]. Second, repeated measures analyses using the top-down procedure were conducted to study changes during the study period (time) and differences in changes between the study groups (time × condition). Linear mixed model analyses were used for the analyses with the risk factor score as outcome measure. Logistic mixed model analyses were used with the guidelines status for the 5 lifestyle behaviors as outcome measures. These kinds of analyses allow for inclusion of all cases (despite missing values of the outcome variable), and are valid in case the missing values satisfy the missingness at random assumption [57].

For the sensitivity analyses, differences in effect between the groups were explored by means of linear and logistic regression analyses by using the top-down procedure. The dependent variables were the risk factor score and complying with the physical activity guideline (yes=1; no=0), complying with the vegetable guideline (yes=1; no=0), complying with the fruit guideline (yes=1; no=0), complying with the alcohol guideline (yes=1; no=0), and complying with the smoking guideline (yes=1; no=0). All analyses were done for both the 12-month and 24-month follow-up measurement. To increase power and external validity, these regression analyses were first performed for T0-T1 completers and T0-T2 completers, respectively. Next, these analyses were also performed based on fully complete cases (ie, respondents who completed both follow-up measurements). The results of the sensitivity analyses are outlined in Multimedia Appendices 1-2.

Among the experimental conditions, linear regression analyses were performed to study the predictive value of the total visiting time of the intervention and the total number of visits during the study period on the risk factor score after 24 months. Descriptive statistics were used to describe the evaluation/appreciation of the intervention at different time points.

Tests were performed at alpha=.05 for the intervention factor and alpha=.10 for covariates [58].

Results

Participation and Sample Characteristics

Figure 2 shows the flow of the respondents from enrollment in the study to allocation to the 3 different conditions, and revisits after 12 and 24 months. In total, 5055 respondents were included in the analyses of this study, of which 4833 (95.61%) were participants of the Adult Health Monitor. A description of the study sample is shown in Table 1. We found some baseline differences between the 3 study groups. The age of respondents in the control condition was significantly higher when compared to the age in the sequential condition (*P*=.03). More respondents in the experimental conditions suffered from heart attacks (*P*=.01), but fewer people reported high blood pressure (*P*=.002). Compliance rates regarding vegetable intake were higher in the simultaneous group than in the control condition (*P*=.07) although this did not reach statistical significance; and respondents in the sequential condition reported smoking more cigarettes than respondents in the control condition (*P*=.04).

Figure 2. Flowchart of study participants. Completed: respondents who adhered to all study protocols; temporary loss to follow-up: respondents who did not complete the follow-up at 12 months, but did at 24 months; started: respondents who logged into intervention, but did not complete assessment.

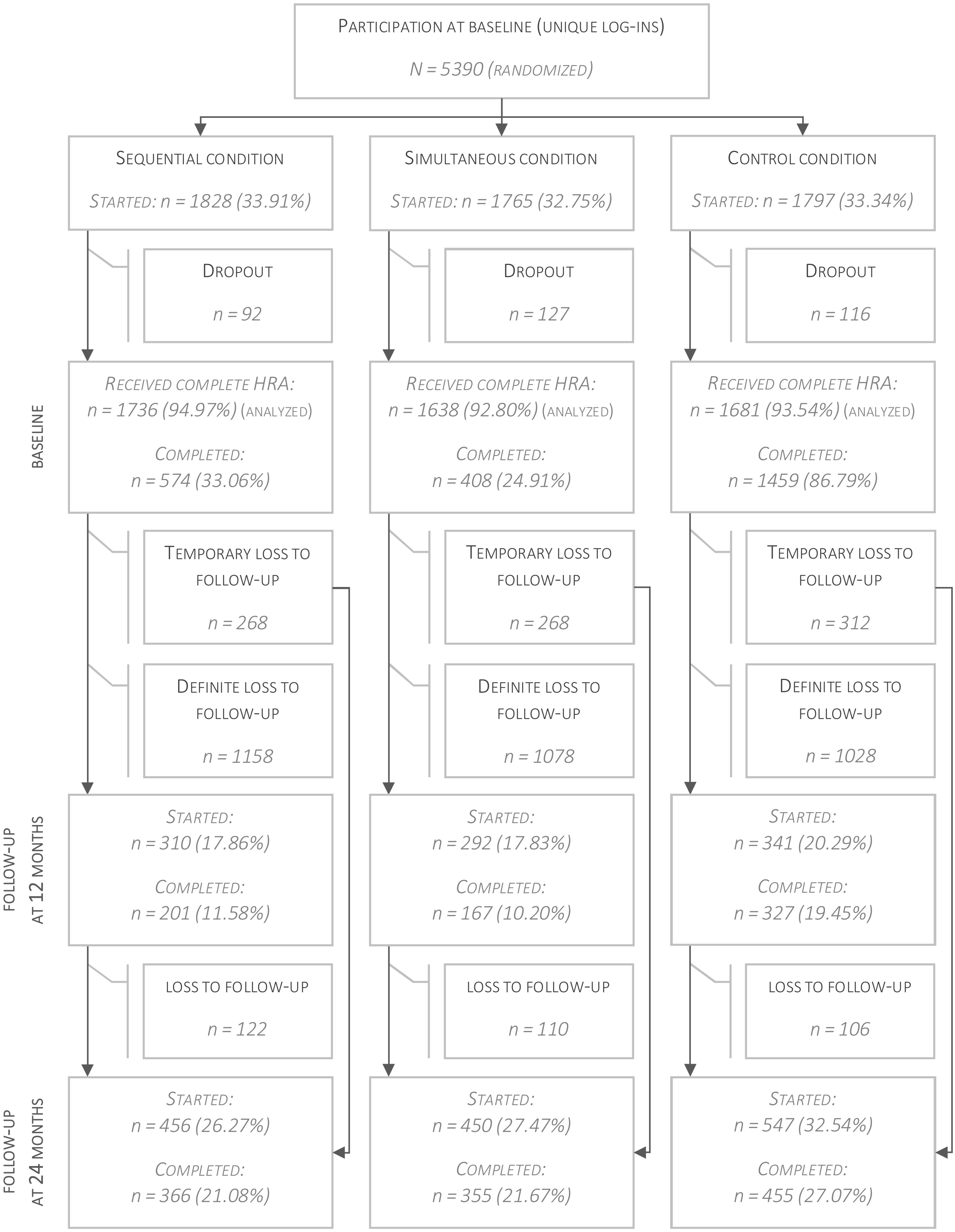


Table 1. Demographics, health status, and lifestyle behavior of the study sample at baseline.

Variable	Total N=5055	Sequential n=1736	Simultaneous n=1638	Control n=1681	*P*
Demographics					
Age (range 19-65 years), mean (SD)	44.15 (12.67)	43.70 (12.62)	43.94 (12.57)	44.82 (12.81)	.03[a]
Gender, n (%)					
Male	2661 (52.64)	893 (51.44)	878 (53.60)	890 (52.94)	
Female	2394 (47.36)	843 (48.56)	760 (46.40)	791 (47.06)	
Education (n=4961), n (%)					
Low	515 (10.38)	167 (9.88)	174 (10.84)	174 (10.45)	
Medium	2334 (47.05)	809 (47.84)	731 (45.55)	794 (47.69)	
High	2112 (42.57)	715 (42.28)	700 (43.61)	697 (41.86)	
Income (€/month; n=4970), n (%)					
<1750	1165 (23.44)	399 (23.51)	373 (23.23)	393 (23.58)	
1751-3050	1688 (33.96)	573 (33.77)	543 (33.81)	572 (34.31)	
>3051	1397 (28.11)	465 (27.40)	466 (29.02)	466 (27.95)	
Not reported	720 (14.49)	260 (15.32)	224 (13.95)	236 (14.16)	
Employment situation (n=4970), n (%)					
Job (paid employment)	3788 (76.22)	1290 (75.97)	1240 (77.26)	1258 (75.46)	
No job	1182 (23.78)	408 (24.03)	365 (22.74)	409 (24.54)	
Marital status (n=4957), n (%)					
Relationship	3775 (76.15)	1292 (76.27)	1215 (75.94)	1268 (76.25)	
Single	1182 (23.85)	402 (23.73)	385 (24.06)	395 (23.75)	
Persons in household (n=4980), mean (SD)	2.9 (1.42)	2.93 (1.47)	2.89 (1.37)	2.88 (1.42)	
Native country (n=4973), n (%)					
The Netherlands	4727 (95.05)	1613 (94.99)	1531 (95.27)	1583 (94.90)	
Other	246 (4.95)	85 (5.01)	76 (4.73)	85 (5.10)	
Health status					
BMI (range 13.82-58.11; n=5012), mean (SD)	25.20 (4.03)	25.26 (4.01)	25.15 (3.99)	25.17 (4.09)	
Quality of life (range 15-48; n=4925), mean (SD)	40.08 (5.24)	40.02 (5.28)	40.19 (5.09)	40.03 (5.34)	
Psychological distress (range 12-50; n=4944), mean (SD)	44.77 (5.74)	44.69 (5.83)	44.86 (5.62)	44.76 (5.76)	
Disease (n=4950), n (%)					
Diabetes	144 (2.91)	48 (2.84)	44 (2.74)	52 (3.14)	
Brain hemorrhage, TIA	19 (0.38)	4 (0.24)	8 (0.50)	7 (0.42)	
Heart attack	39 (0.79)	22 (1.30)	11 (0.68)	6 (0.36)	.003[a]; .001[b]; .08[c]
Other serious heart disease	61 (1.23)	21 (1.24)	21 (1.31)	19 (1.15)	
Cancer	67 (1.35)	24 (1.42)	19 (1.18)	24 (1.45)	
High blood pressure	618 (12.4)	200 (11.79)	174 (10.83)	244 (14.69)	.01[a]; .001[b]
Asthma, COPD	333 (6.73)	109 (6.45)	110 (6.84)	114 (6.88)	

Variable	Total N=5055	Sequential n=1736	Simultaneous n=1638	Control n=1681	*P*
One or more diseases	1048 (21.17)	348 (20.58)	324 (20.16)	376 (22.76)	
Lifestyle behavior					
Number of risk factors (n=4965), n (%)					
0	547 (11.02)	196 (11.48)	178 (11.05)	173 (10.51)	
1	1413 (28.46)	476 (27.87)	458 (28.43)	479 (29.10)	
2	1779 (35.83)	589 (34.48)	576 (35.75)	614 (37.30)	
3	935 (18.83)	334 (19.56)	298 (18.50)	303 (18.41)	
4	262 (5.28)	102 (5.97)	90 (5.59)	70 (4.25)	
5	29 (0.58)	11 (0.64)	11 (0.68)	7 (0.43)	
Number of risk factors, mean (SD)	1.81 (1.07)	1.83 (1.10)	1.81 (1.08)	1.78 (1.03)	
Physical activity (n=5053)					
Minutes per day, mean (SD)	159.32 (160.84)	158.78 (160.12)	155.57 (159.05)	163.53 (163.30)	
Noncompliance, n (%)	651 (12.88)	235 (13.54)	220 (13.44)	196 (11.67)	
Vegetable consumption (n=5018)					
Number of grams, mean (SD)	172.49 (85.07)	171.19 (83.61)	173.97 (82.94)	172.40 (88.58)	
Noncompliance, n (%)	3419 (68.13)	1166 (67.63)	1085 (66.56)	1168 (70.19)	.03[b]
Fruit consumption (n=5019)					
Pieces of fruit, mean (SD)	1.94 (1.29)	1.94 (1.27)	1.93 (1.33)	1.96 (1.28)	
Noncompliance, n (%)	2710 (53.99)	946 (54.90)	892 (54.86)	872 (52.22)	
Alcohol intake (n=5034)					
Number of drinks, mean (SD)	1.26 (1.57)	1.22 (1.46)	1.31 (1.64)	1.25 (1.61)	
Noncompliance, n (%)	1405 (27.91)	488 (28.19)	453 (27.77)	464 (27.75)	
Smoking (n=5055)					
Number of cigarettes, mean (SD)	2.25 (6.47)	2.48 (7.34)	2.33 (6.05)	1.94 (5.86)	.09[a]
Noncompliance, n (%)	897 (17.74)	321 (18.49)	302 (18.44)	274 (16.30)	

[a] Sequential vs control.

[b] Simultaneous vs control.

[c] Sequential vs simultaneous.

Dropout Analyses

Respondents followed up and respondents lost to follow-up differed on a number of variables. Dropout was associated with some demographic factors (eg, a younger age), a better health status as indicated by fewer diseases and lower BMI, but an unhealthier lifestyle. More detailed information can be found in Multimedia Appendix 3.

Intervention Effects

We assessed the effects of the intervention on the overall lifestyle risk factor. The higher the score on the risk factor, the more a respondent did not comply with the Dutch guidelines concerning the lifestyle behaviors. The results of the linear mixed model analyses show that the risk factor score changed favorably and significantly over time (Figure 3) and that there is a statistically significant difference between the experimental conditions and the control condition (Table 2). After 12 months, the sequential condition was significantly more effective in reducing the risk factor score compared to the control condition. A similar but not statistically significant effect (*P*=.08) was found for the simultaneous condition compared to the control condition. After 24 months, only the simultaneous condition showed a statistically significant effect compared to the control condition, revealing a significantly lowered risk score for participants in the simultaneous condition. On both follow-up measurements, there were no statistically significant differences regarding the risk factor score between the sequential and the simultaneous condition.

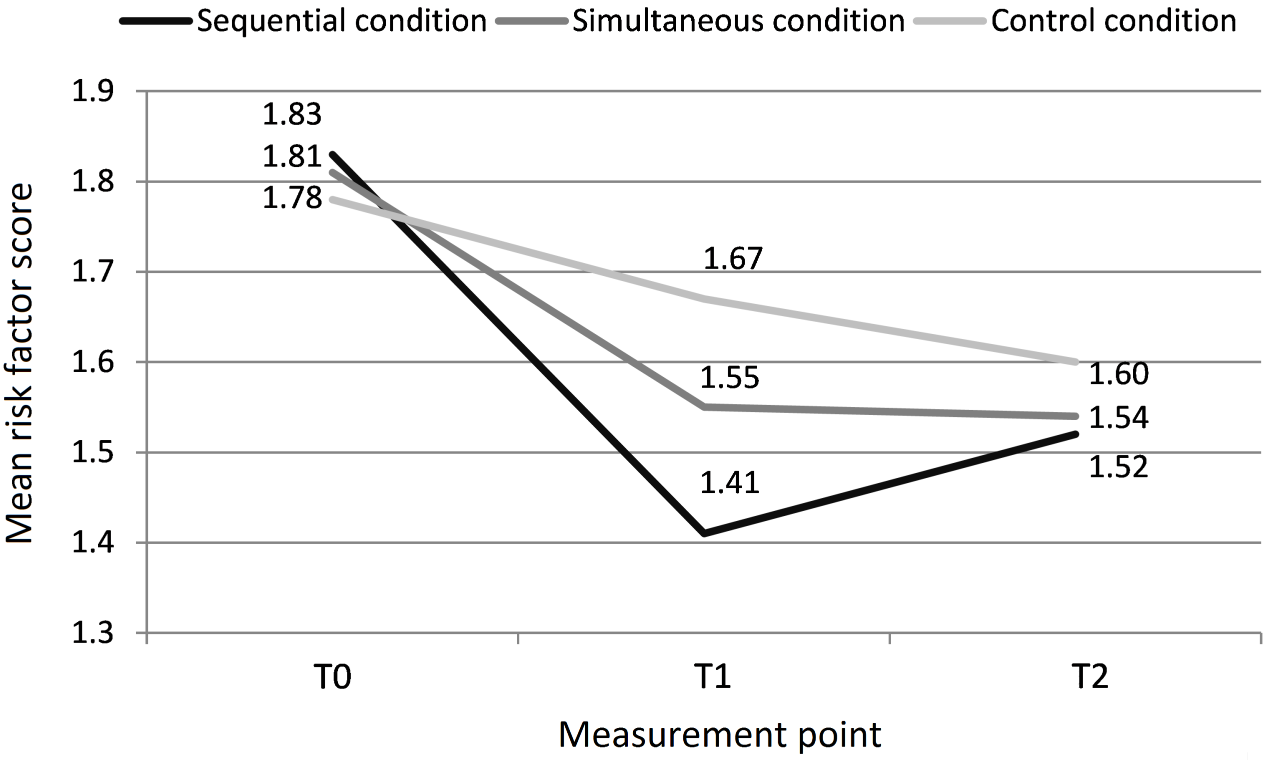
The sensitivity analyses showed similar results (see Multimedia Appendices 1 and 2), except for 2 differences: (1) among the T0-T1 completers, both simultaneous and sequential interventions were effective in reducing the risk factor score after 12 months compared to the control condition, and (2) among fully complete cases, both simultaneous and sequential interventions revealed significant effects in reducing the risk factor score in comparison to the control condition after 24 months.

Table 2. Results of linear mixed model analyses (top-down procedure)[a] with the risk factor score after 12 and 24 months as outcome measure.

Condition and time	*P*	Effect size
Type III tests		
Condition × time	.04	—
Time	<.001	—
After 12 months (T1)		
Sequential vs control	.008	0.28
Simultaneous vs control	.08	0.19
Sequential vs simultaneous	.39	0.10
After 24 months (T2)		
Sequential vs control	.13	0.14
Simultaneous vs control	.048	0.18
Sequential vs simultaneous	.68	0.04

[a] All variables regarding demographics, health status, and lifestyle behavior were included in the most extensive model.

Figure 3. Mean number of risk factors among the different conditions at baseline (T0) and at 12-month (T1) and 24-month (T2) follow-ups.



Differences in Lifestyle Behaviors

We conducted in-depth analysis to assess the effects of the interventions for each lifestyle behavior separately. Table 3 summarizes the changes in the 5 different health risk behaviors according to the guideline status. The logistic mixed model analyses showed that all 5 lifestyle behaviors changed over time, but only a few effects differed significantly between the conditions. At both follow-up measurements, the sequential condition was found to result in significant changes in smoking abstinence in comparison to the simultaneous condition. After 24 months, the sequential condition had greater effect than the control condition, although this did not meet statistical significance (*P*=.06). After 24 months, the sequential condition was more effective in decreasing alcohol consumption than the control condition. The differences between both conditions, although not statistically significant (*P*<.10), indicates that the simultaneous condition might have been more effective in increasing the amount of physical activity compared to the sequential condition after 12 months, and in increasing fruit intake (after 12 and 24 months) compared to the control condition.

However, when performing the sensitivity analyses, no consistent pattern could be found. Differences in statistical significance were found, especially in comparisons between groups with regard to fruit and vegetable intake, alcohol consumption, and smoking. More information can be found in Multimedia Appendices 1 and 2.

Table 3. Results of the logistic mixed model analyses (top-down procedure)[a] with the specific lifestyle behavior (guideline status) after 12 and 24 months as outcome measure.

Condition and time	Outcome measure									
	Physical activity		Vegetable		Fruit		Alcohol		Smoking	
	P	ES	*P*	ES	*P*	ES	*P*	ES	*P*	ES
Fixed effects										
Condition × time	.30	—	.65	—	.21	—	.27	—	.045	—
Time	.04	—	.046	—	.13	—	<.001	—	<.001	—
After 12 months (T1)										
Sequential vs control	.23	0.28	.26	0.20	.61	0.08	.66	0.07	.42	0.13
Simultaneous vs control	.52	0.17	.41	0.15	.06	0.29	.24	0.19	.23	0.18
Sequential vs simultaneous	.07	0.45	.78	0.05	.18	0.21	.49	0.12	.048	0.31
After 24 months (T2)										
Sequential vs control	.43	0.18	.62	0.07	.30	0.14	.048	0.27	.06	0.27
Simultaneous vs control	.48	0.16	.66	0.07	.07	0.24	.20	0.17	.26	0.15
Sequential vs simultaneous	.94	0.02	.37	0.14	.46	0.10	.49	0.10	.004	0.41

[a] All variables regarding demographics, health status, and lifestyle behavior were included in the most extensive model.

Process Evaluation

Exposure to the Intervention

When comparing the total visiting time and the total number of visits in the intervention during the 24 months, statistically significant differences between the 3 groups were found. On average, respondents in the sequential condition visited the intervention for 31 (SD 40) minutes, respondents of the simultaneous condition stayed on the website for 28 (SD 36) minutes, and respondents in the control condition visited the website for 16 (SD 21) minutes ($F_{2,1124}$=23.78; sequential vs control: P<.001; simultaneous vs control: P<.001; sequential vs simultaneous: P=.31). The mean number of visits in the sequential condition was 2.04 (SD 1.35), in the simultaneous condition this was 2.01 (SD 1.45), and in the control condition this was 1.85 (SD 0.93; $F_{2,1124}$=2.84; sequential vs control: P=.75; simultaneous vs control: P=.16; sequential vs simultaneous: P=.91). The regression analyses conducted among respondents in the sequential and the simultaneous conditions only showed that the risk factor after 24 months was predicted by the total visiting time (beta=–.06; P=.01) and the total number of visits during the study period (beta=–.11; P<.001). Longer visits and a greater number of visits predicted more favorable risk factor changes.

Appreciation of the Intervention

The HRA was evaluated at all 3 measurement points. At baseline, the HRA score was evaluated as mean 7.2 (SD 1.3; n=2441), after 12 months the HRA was evaluated as mean 7.3 (SD 1.1; n=368), and after 24 months it was evaluated as mean 7.6 (SD 1.0; n=1176). No differences existed between the experimental conditions; however, at baseline, the HRA was more positively evaluated by the sequential condition (mean 7.4, SD 1.1; n=574) and the simultaneous condition (mean 7.3, SD 1.4; n=408) than the control condition respondents (mean 7.1, SD 1.3; n=1459; $F_{2,2438}$=16.48; P<.001). Of those respondents who completed the last follow-up measurement, 84.43% (998/1182) stated that the HRA gave a good overview of their lifestyle, 77.58% (917/1182) liked the use of traffic lights in the HRA, 72.33% (852/1178) liked the layout, and 76.74% (904/1178) experienced website use as user friendly.

Additional evaluations of the intervention, filled out by 305 respondents, revealed no statistically significant differences between the conditions with regard to the website and the HRA. The evaluation of the tailored advice among respondents of the sequential and simultaneous condition revealed that the respondents evaluated the personalized advice as relevant (75.4%, 86/114), credible (76.5%, 88/115), informative (70.4%, 81/155), well arranged (84.3%, 97/115), clear (85.1%, 97/114), interesting (71.3%, 82/115), and with an attractive layout (70.0%, 77/115). The personalized advice was evaluated as mean 6.9 (SD 1.3), with no statistically significant differences between conditions. Almost 80% (79.8%, 91/114) reported having read all parts of the advice, 40.9% (47/115) reported that they wanted to live a healthier life because of the received advice, and 16.5% (19/115) found the advice too long. One statistically significant difference was found: respondents in the sequential condition indicated more often that they missed information in the advice than respondents in the simultaneous condition ($t_{111.49}$=2.01; P=.047). Some suggestions were given for improving the intervention; for example, attention should be paid to personal circumstances by asking more specific questions about reasons for not eating sufficient vegetables or for being insufficiently physically active, and the advice could be made more personal by giving more concrete examples.

Discussion

Effectiveness of the Intervention

The primary aim of this study was to evaluate the effectiveness of 2 computer-tailored Web-based interventions compared to a control condition with regard to lifestyle improvement. Overall, the computer-tailored intervention for multiple health behaviors resulted in favorable lifestyle changes. Compared to the control condition, the sequential delivery mode was found to be most effective after 12 months, whereas the simultaneous delivery mode was most effective after 24 months. The sensitivity analyses yielded comparable results and suggested slightly stronger effects for both delivery modes, which may be because those who were motivated to fill out more postmeasurements were also those who were more motivated to change. The effect sizes were small which is common for computer-tailored interventions, but they can still result in a large public health impact when widely implemented [14,59]. For instance, for smoking cessation a small ES is considered clinically significant [60], which may also be relevant for other behaviors addressed in our study. Moreover, our control condition received a minimal intervention (ie, the tailored HRA). Personalized information regarding one's lifestyle behavior may be sufficient to facilitate change and improve lifestyle behaviors [61]. Further studies are needed that compare a sequential and simultaneous condition for a group that receives general information or no information at all.

Regarding the overall lifestyle behavior changes, no differences were found between the sequential and the simultaneous condition. This is in-line with the findings Vandelanotte et al [30] reported in their study aimed at lowering fat intake and increasing physical activity. We assume that only those respondents with the highest motivational level to change remained in our study at follow-up. Therefore, it might be that no differential effects were found between the sequential and simultaneous condition.

When analyzing the separate behaviors, the largest changes were found for smoking cessation, followed by lower alcohol intake and increased fruit consumption. However, these findings were only partly replicated in the sensitivity analyses. The results of the sensitivity analyses supposed that the effect on the overall risk factor change can primarily be ascribed to changes in fruit consumption, vegetable consumption, alcohol intake, and probably in tobacco use. Hence, no firm conclusion can be drawn with regard to the differential effects on separate lifestyle behaviors. Further research is needed to investigate whether the optimum tailoring strategy (ie, a sequential, simultaneous, or even single-behavior approach) may depend on the behavior(s) being targeted.

Dropout Rates

An important limitation of our study is the high dropout rates. Mixed model analyses were performed to increase internal and external validity; however, it is not unlikely that informative dropout occurred (ie, the dropout process depends on the unobserved measurements) [57,62]. This would be a violation of the missing at randomness assumption underlying the mixed model analyses. The results of the sensitivity analyses, using regression techniques, might not be generalizable to the general population, but only to those people who remain participating in such a study over a longer period of time. As shown in our attrition analyses study [63], younger people and those with an unhealthier lifestyle were more likely to drop out.

We made use of a low-intensity implementation in the current study (recruitment via Regional Health Authorities, sending emails without face-to-face or telephone contact for the entire study). This implementation strategy may have led to relatively high attrition [64,65]. Yet, our dropout levels are comparable with those of other website-delivered studies with similar protocols [66-68] and some participants indicated (via email) that they did not revisit the intervention because they had complied with all 5 guidelines and found that they did not need more information. Furthermore, factors of the intervention itself may have caused dropout, such as technical issues or problems of navigating through the intervention website. As hypothesized in the Model of Internet Interventions, website characteristics, such as appearance, behavioral prescriptions, burdens, content, delivery, message, participation, and assessment, may influence website use and effectiveness [69].

Exposure Rates

There is growing evidence that the level of engagement in eHealth plays an important role in explaining the use and effectiveness of Web-based interventions. It is challenging to develop eHealth interventions focused on multiple health behaviors that engage participants sufficiently to revisit the website on more than 1 occasion. As our results suggest, higher usage is related to higher effects, which has also been observed in other studies [70,71]. With regard to the time respondents spent on the website, it is striking that many really short visits (<20 seconds) were recorded. This is in-line with the findings of Brouwer et al [72] who reported that more than half of their visitors left the website within 30 seconds. Internet interventions have the disadvantage that clicking off the website is easy. Respondents may have opened the link without knowing what they were opening (in our case, not reading the email invitations to visit our intervention) resulting in loss of interest after opening the website. On the other hand, some quite long visits were recorded. The visiting time as recorded by the system may be not fully reliable because it might be that respondents opened the website, but did not use it while it remained open. Moreover, visiting time also depends on the bit rate of the individual respondent's Internet.

In our intervention, exposure to the program was similar for both the sequential and the simultaneous group, although respondents in the simultaneous condition had the possibility of receiving more parts of the program than respondents in the sequential condition did. It may be that respondents in the simultaneous condition scanned the advice and did not read it carefully, resulting in comparable visiting times for the 2 intervention modes. In our sequential intervention, the modules were delivered sequentially over time so respondents could only choose 1 lifestyle module in the first year [27]. Exposure to multiple behaviors directly at the beginning of an intervention could have advantages for those people who drop out early [25]. Those participating in this way will already have received more information during their first visit than those taking part in a single-behavior intervention in which the people who drop out after their first visit can only receive information about a single behavior. Thus, it may be better to give respondents the choice of immediately initiating more than 1 module. A study by Brouwer et al [72] showed that more than half of their respondents selected 2 of 3 modules and approximately one-third initiated all 3 modules. A preference-based tailoring strategy [73,74]—a combination of our sequential and simultaneous condition—could be used in which respondents can select the modules they want to complete without any limitations on the number of behavior modules. The finding that revisits are uncommon in interventions [37,75]—our respondents visited the intervention on average only twice—is a further argument against delivering the modules, or even sections of particular modules, sequentially over time.

Strengths and Limitations

To our knowledge, this study is unique in comparing the effectiveness of sequential and simultaneous interventions in addressing the 5 lifestyle behaviors of physical activity, fruit consumption, vegetable consumption, alcohol intake, and smoking. The intervention was effective with a small ES and was appreciated well. Our intervention met a number of criteria related to higher effectiveness of health behavior change interventions offered via the Internet, such as extensive use of theory and behavior change techniques [20]. However, several limitations should be kept in mind: dropout rates were high resulting in a small sample size; the findings were based on self-reports which may have resulted in recall bias (eg, the high proportion of respondents who reported being sufficiently physically active at baseline [87%] may represent an overestimation of their actual level of physical activity; however, another reason for this high proportion might be that the intensity of the different kinds of physical activities was not measured in the short version of the SQUASH that we used [41]); we cannot guarantee that a representative sample of the Dutch population was reached by our recruitment strategies; and a selective group filled out the follow-up questionnaires. Thus, the results may not be generalizable and may be biased. An additional limitation of our study is that respondents in the sequential condition had the possibility of choosing a maximum of 2 behavioral modules, whereas respondents in the simultaneous condition could receive personal feedback on more than 2 behaviors. Thus, the possible number of behavior modules that could be completed differed between the sequential and the simultaneous group, which implies that the sequential condition might have been more effective if respondents had the opportunity to choose more than 2 modules. Finally, the control condition received a minimal intervention, which might have led to improved lifestyle behaviors in this condition too.

Implications for Future Interventions and Research

The high prevalence of people engaging in multiple health risk behaviors calls for the development of multiple behavior change interventions. Future studies should examine the number of behaviors that can be addressed in a multiple behavior change intervention without overloading the respondents [76]. Research is needed to identify strategies to stimulate exposure and participation, in particular for Internet interventions with multiple sessions, such as those involving email/phone contact with visitors and updates of the intervention website [39,42,77]. In addition to the necessity of research regarding the optimal delivery mode of these kinds of interventions, future research should focus on attracting, engaging, and retaining participants. In future interventions, other features could be added, such as short message service (SMS) text messaging, which can increase intervention effectiveness [20], or real-time sensors, such as accelerometer apps so that real-time motivational feedback can be provided [78]. Online community features, as a kind of self-help option, could be integrated in Web-based lifestyle programs to reduce attrition [79]. Moreover, interventions should be accessible on all different kinds of channels, such as desktop computers, laptops, smartphones, and tablets. Research is needed to assess the additional effects of these elements.

Conclusions

Both sequential and simultaneous strategies were effective in improving lifestyle in a Web-based computer-tailored intervention. Because no crucial differences have been found with regard to dropout rates and appreciation of the interventions, providers can use the strategy that suits their particular circumstances best. However, the best kind of intervention may be dependent on the behavior that is targeted or other personal factors (eg, motivational level to change).

Acknowledgments

This study was funded by ZonMw, the Netherlands Organization for Health Research and Development (grant number: 120610012). Intervention development and data analysis took place at Maastricht University. Data collection was done in collaboration with the Regional Health Authorities of the Dutch provinces of North-Brabant (GGD Brabant-Zuidoost; GGD Hart voor Brabant; and GGD West-Brabant) and Zeeland (GGD Zeeland). The authors wish to thank the project partners at the Regional Health Authorities for their assistance in implementing the intervention and recruiting the participants.

Conflicts of Interest

Hein de Vries is scientific director of Vision2Health, a company that licenses evidence-based, innovative, computer-tailored health communication tools. No other authors report any conflicts of interest.

Multimedia Appendix 1

T0-T1 completers and T0-T2 completers.

[PDF File (Adobe PDF File), 250KB - jmir_v16i1e26_app1.pdf]

Multimedia Appendix 2

Fully complete cases.

[PDF File (Adobe PDF File), 207KB - jmir_v16i1e26_app2.pdf]

Multimedia Appendix 3

Dropout analyses.

[PDF File (Adobe PDF File), 12KB - jmir_v16i1e26_app3.pdf]

Multimedia Appendix 4

CONSORT-EHEALTH checklist V1.6.2 [80].

[PDF File (Adobe PDF File), 997KB - jmir_v16i1e26_app4.pdf]

References

1. Australian Institute of Health and Welfare. Chronic Diseases and Associated Risk Factors in Australia, 2001. Canberra: AIHW; 2002.
2. Smeets T. Towards a Healthy Lifestyle: the Development and Evaluation of a Computer-Tailored Lifestyle Intervention. PhD thesis. Maastricht: Maastricht University; 2006.
3. World Health Organization. The World Health Report 2002: Reducing risks, promoting healthy life. Geneva: World Health Organization; 2002. URL: http://www.who.int/whr/2002/en/whr02_en.pdf [accessed 2014-01-17] [WebCite Cache ID 6MfzxCKXz]
4. Ezzati M, Hoorn SV, Rodgers A, Lopez AD, Mathers CD, Murray CJ, Comparative Risk Assessment Collaborating Group. Estimates of global and regional potential health gains from reducing multiple major risk factors. Lancet 2003 Jul 26;362(9380):271-280. [Medline: 12892956]
5. Fine LJ, Philogene GS, Gramling R, Coups EJ, Sinha S. Prevalence of multiple chronic disease risk factors. 2001 National Health Interview Survey. Am J Prev Med 2004 Aug;27(2 Suppl):18-24. [doi: 10.1016/j.amepre.2004.04.017] [Medline: 15275670]
6. Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. Prev Med 2007 Feb;44(2):124-128. [doi: 10.1016/j.ypmed.2006.10.006] [Medline: 17157369]
7. Schuit AJ, van Loon AJ, Tijhuis M, Ocké M. Clustering of lifestyle risk factors in a general adult population. Prev Med 2002 Sep;35(3):219-224. [Medline: 12202063]
8. Vries HD, Kremers S, Smeets T, Reubsaet A. Clustering of diet, physical activity and smoking and a general willingness to change. Psychology & Health 2008 Apr;23(3):265-278. [doi: 10.1080/14768320701349107]
9. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
10. Edington DW. Emerging research: a view from one research center. Am J Health Promot 2001;15(5):341-349. [Medline: 11502015]
11. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015]
12. Neville LM, O'Hara B, Milat A. Computer-tailored physical activity behavior change interventions targeting adults: a systematic review. Int J Behav Nutr Phys Act 2009;6:30 [FREE Full text] [doi: 10.1186/1479-5868-6-30] [Medline: 19490649]
13. Ryan P, Lauver DR. The efficacy of tailored interventions. J Nurs Scholarsh 2002;34(4):331-337. [Medline: 12501736]
14. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
15. Smeets T, Kremers SP, Brug J, de Vries H. Effects of tailored feedback on multiple health behaviors. Ann Behav Med 2007 Apr;33(2):117-123. [doi: 10.1080/08836610701307801] [Medline: 17447863]
16. de Vries H, Kremers SP, Smeets T, Brug J, Eijmael K. The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors. Am J Health Promot 2008;22(6):417-425. [doi: 10.4278/ajhp.22.6.417] [Medline: 18677882]

17. Smit ES, de Vries H, Hoving C. Effectiveness of a Web-based multiple tailored smoking cessation program: a randomized controlled trial among Dutch adult smokers. J Med Internet Res 2012;14(3):e82 [FREE Full text] [doi: 10.2196/jmir.1812] [Medline: 22687887]
18. Elfeddali I, Bolman C, Candel MJ, Wiers RW, de Vries H. Preventing smoking relapse via Web-based computer-tailored feedback: a randomized controlled trial. J Med Internet Res 2012;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]
19. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. Efficacy of two tailored interventions promoting physical activity in older adults. Am J Prev Med 2009 Nov;37(5):405-417. [doi: 10.1016/j.amepre.2009.07.009] [Medline: 19840695]
20. 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]
21. Smit F, Lokkerbol J, Riper H, Majo MC, Boon B, Blankers M. Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. J Med Internet Res 2011;13(3):e56 [FREE Full text] [doi: 10.2196/jmir.1694] [Medline: 21840836]
22. Smit ES, Evers SM, de Vries H, Hoving C. Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. J Med Internet Res 2013;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
23. Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. Prev Med 2008 Mar;46(3):181-188 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.001] [Medline: 18319098]
24. Hyman DJ, Pavlik VN, Taylor WC, Goodrick GK, Moye L. Simultaneous vs sequential counseling for multiple behavior change. Arch Intern Med 2007 Jun 11;167(11):1152-1158. [doi: 10.1001/archinte.167.11.1152] [Medline: 17563023]
25. King AC, Castro CM, Buman MP, Hekler EB, Urizar GG, Ahn DK. Behavioral impacts of sequentially versus simultaneously delivered dietary plus physical activity interventions: the CALM trial. Ann Behav Med 2013 Oct;46(2):157-168. [doi: 10.1007/s12160-013-9501-y] [Medline: 23609341]
26. Oenema A, Brug J, Dijkstra A, de Weerdt I, de Vries H. Efficacy and use of an internet-delivered computer-tailored lifestyle intervention, targeting saturated fat intake, physical activity and smoking cessation: a randomized controlled trial. Ann Behav Med 2008 Apr;35(2):125-135. [doi: 10.1007/s12160-008-9023-1] [Medline: 18363076]
27. Schulz DN, Kremers SP, van Osch LA, Schneider F, van Adrichem MJ, de Vries H. Testing a Dutch web-based tailored lifestyle programme among adults: a study protocol. BMC Public Health 2011;11:108 [FREE Full text] [doi: 10.1186/1471-2458-11-108] [Medline: 21324181]
28. Vandelanotte C, De Bourdeaudhuij I, Sallis JF, Spittaels H, Brug J. Efficacy of sequential or simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2005 Apr;29(2):138-146. [doi: 10.1207/s15324796abm2902_8] [Medline: 15823787]
29. Vandelanotte C, De Bourdeaudhuij I, Brug J. Two-year follow-up of sequential and simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. Ann Behav Med 2007 Apr;33(2):213-219. [doi: 10.1080/08836610701310086] [Medline: 17447874]
30. Vandelanotte C, Reeves MM, Brug J, De Bourdeaudhuij I. A randomized trial of sequential and simultaneous multiple behavior change interventions for physical activity and fat intake. Prev Med 2008 Mar;46(3):232-237. [doi: 10.1016/j.ypmed.2007.07.008] [Medline: 17707079]
31. Markland D, Vansteenkiste M. Self determination theory: motivational interviewing in exercise. In: Hagger MS, Chatzisarantis NLD, editors. Intrinsic Motivation and Self-Determination in Exercise and Sport. Champaign, IL: Human Kinetics; 2007:87-99.
32. Zuckerman M, Porac J, Lathin D, Deci EL. On the importance of self-determination for intrinsically-motivated behavior. Personality and Social Psychology Bulletin 1978 Jul 01;4(3):443-446. [doi: 10.1177/014616727800400317]
33. Baumeister RF, Bratslavsky E, Muraven M, Tice DM. Ego depletion: is the active self a limited resource? J Pers Soc Psychol 1998 May;74(5):1252-1265. [Medline: 9599441]
34. Hagger MS, Wood C, Stiff C, Chatzisarantis NL. Ego depletion and the strength model of self-control: a meta-analysis. Psychol Bull 2010 Jul;136(4):495-525. [doi: 10.1037/a0019486] [Medline: 20565167]
35. Rodondi N, Humair JP, Ghali WA, Ruffieux C, Stoianov R, Seematter-Bagnoud L, et al. Counselling overweight and obese patients in primary care: a prospective cohort study. Eur J Cardiovasc Prev Rehabil 2006 Apr;13(2):222-228. [doi: 10.1097/01.hjr.0000209819.13196.a4] [Medline: 16575276]
36. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]
37. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
38. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]

39. Crutzen R, Cyr D, de Vries NK. Bringing loyalty to e-Health: theory validation using three internet-delivered interventions. J Med Internet Res 2011;13(3):e73 [FREE Full text] [doi: 10.2196/jmir.1837] [Medline: 21946128]
40. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
41. Nederland GGD, ActiZ, RIVM. Lokale en Nationale Monitor Gezondheid gezondheid [Local and national Health Monitor Indicators of the Public Health Monitor]. 2005. URL: https://www.monitorgezondheid.nl/volksindicatoren.aspx [accessed 2013-11-08] [WebCite Cache ID 6KyXUfR1j]
42. Schneider F, Schulz DN, Pouwels LH, de Vries H, van Osch L. The use of a proactive dissemination strategy to optimize reach of an internet-delivered computer tailored lifestyle intervention. BMC Public Health 2013;13:721 [FREE Full text] [doi: 10.1186/1471-2458-13-721] [Medline: 23914991]
43. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626 [FREE Full text] [Medline: 14572020]
44. Hildebrandt VH, Chorus AMJ, Stubbe JH. Trendrapport Bewegen en Gezondheid 2008-2009 [Trend Report on Physical Activity and Health]. Leiden: TNO; 2010.
45. Gezondheidsraad. Richtlijnen goede voeding 2006 [Guidelines for good nutrition 2006]. Den Haag: Gezondheidsraad; 2006.
46. Kwaliteitsinstituut voor de Gezondheidszorg CBO. Richtlijn behandeling van tabaksverslaving [Guidelines for treating tobacco addiction]. Utrecht/Den Haag: BVO/Partnership Stop met Roken; 2004.
47. Ware JE, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston: The Health Institute, New England Medical Center; 1995.
48. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]
49. Kessler R, Mroczek D. Final Versions of our Non-Specific Psychological Distress Scale [memo dated 10/3/94]. Ann Arbor, MI: Survey Research Center of the Institute for Social Research; 1994.
50. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]
51. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [Medline: 10993420]
52. Lemmens P, Tan ES, Knibbe RA. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. J Stud Alcohol 1992 Sep;53(5):476-486. [Medline: 1405641]
53. Mudde AN, Willemsen MC, Kremers S, De Vries H. Meetinstrumten: voor onderzoek naar roken en stoppen met roken. Den Haag: Stivoro voor een rookvrije toekomst; 2006.
54. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
55. Prochaska JO, Redding CA, Evers KE. The transtheoretical modelstages of change. In: Glanz K, editor. Health Behaviour and Health Education: Theory, Research, and Practice. 3rd ed. San Francisco: Jossey-Bass; 1997:99-120.
56. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd ed). Hillsdale, NJ: L Erlbaum Associates; 1988.
57. Enders CK. Applied missing data analysis. In: Applied Missing Data Analysis (Methodology In The Social Sciences). New York: The Guilford Press; 2010.
58. Rosnow RL, Rosenthal R. Statistical procedures and the justification of knowledge in psychological science. American Psychologist 1989;44:1276-1284.
59. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
60. West R. The clinical significance of "small" effects of smoking cessation treatments. Addiction 2007 Apr;102(4):506-509. [doi: 10.1111/j.1360-0443.2007.01750.x] [Medline: 17362283]
61. Parekh S, King D, Owen N, Jamrozik K. Spousal concordance and reliability of the 'Prudence Score' as a summary of diet and lifestyle. Aust N Z J Public Health 2009 Aug;33(4):320-324. [doi: 10.1111/j.1753-6405.2009.00402.x] [Medline: 19689591]
62. Little RJA, Rubin DB. Statistical Analysis with Missing Data (2nd ed). Hoboken, NJ: Wiley; 2002.
63. Schulz DN, Schneider F, de Vries H, van Osch LA, van Nierop PW, Kremers SP. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
64. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

65. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
66. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
67. Kosma M, Cardinal BJ, McCubbin JA. A pilot study of a web-based physical activity motivational program for adults with physical disabilities. Disabil Rehabil 2005 Dec 15;27(23):1435-1442. [doi: 10.1080/09638280500242713] [Medline: 16418058]
68. Woolf SH, Krist AH, Johnson RE, Wilson DB, Rothemich SF, Norman GJ, et al. A practice-sponsored Web site to help patients pursue healthy behaviors: an ACORN study. Ann Fam Med 2006;4(2):148-152 [FREE Full text] [doi: 10.1370/afm.522] [Medline: 16569718]
69. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
70. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
71. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful web-based smokeless tobacco cessation program. Addiction 2008 Oct;103(10):1706-1712 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02295.x] [Medline: 18715238]
72. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]
73. Kwak L, Kremers SP, Werkman A, Visscher TL, van Baak MA, Brug J. The NHF-NRG In Balance-project: the application of Intervention Mapping in the development, implementation and evaluation of weight gain prevention at the worksite. Obes Rev 2007 Jul;8(4):347-361. [doi: 10.1111/j.1467-789X.2006.00304.x] [Medline: 17578384]
74. Kwak L, Kremers SP, Candel MJ, Visscher TL, Brug J, van Baak MA. Changes in skinfold thickness and waist circumference after 12 and 24 months resulting from the NHF-NRG In Balance-project. Int J Behav Nutr Phys Act 2010;7:26 [FREE Full text] [doi: 10.1186/1479-5868-7-26] [Medline: 20370934]
75. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
76. Taylor WC, Hepworth JT, Lees E, Cassells A, Gousse Y, Sweeney MM, et al. Readiness to change physical activity and dietary practices and willingness to consult healthcare providers. Health Res Policy Syst 2004 Jun 10;2(1):2 [FREE Full text] [doi: 10.1186/1478-4505-2-2] [Medline: 15191611]
77. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
78. Vankipuram M, McMahon S, Fleury J. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. AMIA Annu Symp Proc 2012;2012:931-939 [FREE Full text] [Medline: 23304368]
79. 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]
80. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BMI: body mass index
ES: effect size
FFQ: food frequency questionnaire
HRA: health risk appraisal
K10: Kessler psychological distress scale (10 items)
QFV: quantity-frequency-variability
SQUASH: short questionnaire to assess health-enhancing physical activity

Edited by G Eysenbach; submitted 11.11.13; peer-reviewed by DJ Hyman, S Peters, W Evans; comments to author 03.12.13; revised version received 11.12.13; accepted 23.12.13; published 27.01.14.

Please cite as:
Schulz DN, Kremers SPJ, Vandelanotte C, van Adrichem MJG, Schneider F, Candel MJJM, de Vries H
Effects of a Web-Based Tailored Multiple-Lifestyle Intervention for Adults: A Two-Year Randomized Controlled Trial Comparing Sequential and Simultaneous Delivery Modes
J Med Internet Res 2014;16(1):e26
URL: http://www.jmir.org/2014/1/e26/
doi:10.2196/jmir.3094
PMID:24472854

Original Paper

Perceptions of Successful Cues to Action and Opportunities to Augment Behavioral Triggers in Diabetes Self-Management: Qualitative Analysis of a Mobile Intervention for Low-Income Latinos With Diabetes

Elizabeth R Burner[1], MD, MPH; Michael D Menchine[1], MD, MPH; Katrina Kubicek[2,3], PhD; Marisela Robles[2], MS; Sanjay Arora[1], MD

[1]Keck School of Medicine of the University of Southern California, Department of Emergency Medicine, Los Angeles, CA, United States
[2]SC Clinical & Translational Science Institute, Community Engagement, University of Southern California, Los Angeles, CA, United States
[3]Children’s Hospital Los Angeles, The Saban Research Institute, Los Angeles, CA, United States

Corresponding Author:
Elizabeth R Burner, MD, MPH
Keck School of Medicine of the University of Southern California
Department of Emergency Medicine
1200 N State St
Room 1011
Los Angeles, CA, 90033
United States
Phone: 1 3232266667
Fax: 1 3232265464
Email: eburner@usc.edu

Abstract

Background: The increasing prevalence of diabetes and the associated cost of managing this complicated disease have a significant impact on public health outcomes and health expenditures, especially among resource-poor Latino patients. Mobile health (mHealth) may be the solution to reaching this group and improving their health.

Objective: In this qualitative study, we examined nuances of motivation, intention, and triggers to action effected by TExT-MED (Trial to Examine Text Messaging for Emergency Department patient with Diabetes), an mHealth intervention tailored to low-income, urban Latinos with diabetes. TExT-MED is a fully-automated, text message-based program designed to increase knowledge, self-efficacy, and subsequent disease management and glycemic control.

Methods: We conducted 5 focus group interviews with 24 people who participated in TExT-MED. We employed a modified grounded theory analytic approach—an iterative process of coding and immersion in the data used to recognize the patterns and links between concepts voiced by the participants. We coded data to identify themes of participant experiences, motivations, and responses to the program. We organized themes into a theory of TExT-MED’s action.

Results: Participants enjoyed their experience with TExT-MED and believed it improved their diabetes management. Through analysis of the transcripts, we identified that the strengths of the program were messages that cued specific behaviors such as medication reminders and challenge messages. Our analysis also revealed that increasing personalization of message delivery and content could augment these cues.

Conclusions: This in-depth qualitative analysis of TExT-MED shows that low-income Latino patients will accept text messages as a behavioral intervention. This mHealth intervention acts as a behavioral trigger rather than an education platform. Personalization is an opportunity to enhance these cues to action and further research should be conducted on the ideal forms of personalization.

(J Med Internet Res 2014;16(1):e25) doi:10.2196/jmir.2881

KEYWORDS

diabetes; Latinos; cellular phone

Introduction

The increasing prevalence of diabetes and the associated cost of managing this complicated disease have a significant impact on public health outcomes and health expenditures [1]. This is especially true among Latino patients, who have a higher rate of diabetes and diabetic complications than non-Hispanic White patients [2]. Additionally, the need for improved diabetes management is especially pronounced among resource-poor, low-income populations, as they fare worse than other patients [3]. Innovative and cost-effective solutions must be explored to improve health behaviors and health outcomes for these high-risk, low-income populations.

Mobile health (mHealth) may be the solution to reaching this group and improving their health. mHealth is the use of mobile devices to provide public health interventions and medical care. mHealth can consist of text messages, smartphone applications (apps), or Web-based interfaces and continues to evolve on other platforms. While mHealth has been shown to be effective in improving diabetes outcomes [4-6], it has not been widely tested in low-income, resource-poor populations, and is largely untested in Spanish-speaking populations. TExT-MED (Trial to Examine Text Messaging for Emergency Department patient with Diabetes) is the first program designed specifically for low-income Latinos. It involves is a six-month, fully-automated, text message-based program designed to increase knowledge, self-efficacy, and subsequent disease management and glycemic control [7]. The twice-daily text messages consisted of (1) educational/motivational messages, (2) medication reminders, (3) trivia questions, and (4) healthy living challenges. A randomized controlled trial of TExT-MED improved medication adherence compared to controls and was highly accepted by patients. Additionally, Spanish-speaking patients significantly improved their glycosylated hemoglobin (a measure of long-term blood glucose control) [7].

To better understand the findings of TExT-MED and to be able to develop and improve future interventions, we need to examine more than empirical data on outcomes. We must examine nuances of motivation, intention, and triggers to action. These are difficult to measure in novel interventions such as mHealth and among disadvantaged patients with poor health care access. Scales developed for patients from the mainstream culture may not correspond to the same concepts when delivered to low-income, low health-literacy, and non-English speaking populations [8]. Qualitative analytic techniques allow for a deeper understanding of the patient experience with mHealth, allowing us to maximize the benefits of this and future mHealth interventions. Qualitative analysis allows the participants to identify the components of the program they found most successful at achieving behavior change and improving their health rather than limit the responses to our preconceived ideas and hypotheses that are necessary for quantitative analysis. We conducted a series of focus groups and analyzed this data using a modified grounded theory approach to uncover those components of TExT-MED that participants perceived as most beneficial.

Methods

In this qualitative study, we conducted a series of focus group interviews with people with diabetes who participated in TExT-MED. All 47 patients who had completed the earlier TExT-MED intervention and 6-month assessment were invited to participate in the focus groups through a series of phone calls and text messages. All 24 patients who agreed to participate were compensated for their time and travel. Focus groups were stratified by language and by gender to improve the comfort level of participants [9]. At the start of each focus group, an anchor survey was administered (Multimedia Appendix 1) and dinner provided to ease the start of the focus group [9]. Each session lasted 90 minutes to 2 hours, and a semi-structured question guide was used to prompt conversation (Multimedia Appendix 2). Experienced, bilingual team members facilitated the focus groups that were audio recorded. Focus groups were professionally transcribed verbatim and translated in the case of the Spanish language groups. These transcripts were uploaded to Dedoose, a Web-based qualitative analysis program [10]. This study was approved by the Institutional Review Board.

We used a modified grounded theory analysis approach [11], an iterative process of coding and immersion in the data to recognize the patterns and links between the concepts voiced by the participants. Two project team members independently completed line-by-line open coding of two of the transcripts in order to maintain the detail of the data. An audit trail including memoing [11] was maintained to document all analytic decisions to increase trustworthiness (the qualitative analogue of validity). Through an iterative process of coding transcripts and discussing the coding schema, we developed a set of codes from the recurring ideas presented in the data, using the Health Belief Model as a sensitizing theoretical lens (Figure 1) [12]. After the development of the preliminary set of codes, we conducted multiple rounds of co-coding until consensus on code definitions was achieved. A test of coders' agreement on application of these codes resulted in a Cohen's pooled Kappa coefficient of greater than .7, showing good inter-coder reliability [13]. We then coded all transcripts through focused coding with the final codebook, while also checking to see if new codes needed to be created to capture new emerging themes. In order to better understand the relationships between codes, we then used a technique known as axial coding, comparing the situations where codes overlapped and the cases where these connections did not occur. This resulted in a grounded description of TExT-MED's effect on participants. Saturation of codes was achieved by the fourth focus group. Our final codebook contained 27 codes and 24 sub-codes. We produced 327 pages of transcripts, which were read by three members of the research team (EB, MR, and KK.) In total, 759 individual excerpts were coded with between 1 and 11 codes each.

Figure 1. Health Belief Model.

Results

Participant Summary

A total of 24 participants were successfully recruited and participated in a total of 5 focus groups (2 in English and 3 in Spanish; 2 for men and 3 for women). Study participants were between the ages of 26 and 76 years. They were 67% (16/24) female and 75% (18/24) Latino. A majority (58%, 14/24) spoke primarily Spanish. Almost all study participants (92%, 22/24) were uninsured. They had known of their diabetes diagnosis for between 1 to 45 years (Table 1). These characteristics did not differ significantly between the focus group participants and those patients that did participate in the focus groups (see Table 2).

Overall, the participants enjoyed the experience with TExT-MED and believed it had helped them improve their diabetes management and their “way of living”. Participants wanted the program and text-messages to continue and no one said TExT-MED had gone on too long. Many wanted the program to be expanded to more people and extended in length in future versions. Although not required, participants could add friends or family members to the program to receive messages. Many did, and the participants reported their friends and family found the program effective and enjoyable. Exemplary quotes included:

> *I am living well with this program. It has helped me enormously, enormously…I come because it’s worked*

> *for me. It has worked for me. A lot.* [translated from Spanish]

> *It has been a very good program for all the people with diabetes, and with this program we are better controlling our lives, our way of living.* [translated from Spanish]

> *I feel that it has helped me, it has helped me. And also the people that I signed up—I signed up five or six—it also helped them and they say, "When there is another program, invite us. It has worked."* [translated from Spanish]

In general, participants were excited about this mHealth intervention as a novel way to learn about and manage their diabetes. In sum, the participants reported that TExT-MED helped them to take control of their diabetes and to make the behavior choices they believed would benefit their diabetes. Through the analysis, two specific strengths of the program and two opportunities for development emerged.

Table 1. Focus group participant characteristics (n=24).

Characteristic		Years or n (%)
Age		26-76 years (IQR[a] 49-60 years)
Gender		
	Female	16 (67%)
Ethnicity		
	Latino	18 (75%)
	Non-Latino	6 (25%)
Primary language		
	Spanish	14 (58%)
	English	10 (42%)
Insurance status		
	Uninsured	22 (92%)
	Insured	2 (8%)
Number of years participants have known about diagnosis		1-45 years (IQR 6-15 years)

[a] IQR: interquartile range

Table 2. Focus group participant and non-participant characteristics.

Characteristic	Participants (n=24)	Non-participants (n=68)	*P* value
Age, mean (SD)	52.88 (10.25)	50.74 (10.42)	.387
Spanish language preference, n (%)	14 (58%)	50 (73%)	.171
Latino ethnicity, n (%)	18 (75%)	58 (85%)	.787
Female gender, n (%)	16 (62%)	46 (67%)	.667
Number of years participants have known about diagnosis, mean (SD)	11.6 (11.36)	9.8 (7.33)	.466

Specific Strengths of Program: Medication Reminders

In every group, medication reminder messages were the most prominent and salient benefit mentioned. TExT-MED included various kinds of automated messages: educational/motivational messages, medication reminders, and healthy living challenges. Medication reminders consisted of specific prompts to take medication, as well as cues to have medications refilled or to take medications with them if they left the house. When discussing which messages most impacted their behavior, a participant stated:

> *Before you go to bed, it [the TExT-MED message] told me, like, "Don't forget your medications and don't eat nothing heavy before you go to bed". And that was good.*

And in this conversation between respondents:

> *That is what I like about the program, about the messages because they're reminding me always to get our medicines.* [Respondent 1]

> *So that we won't run out of medicines.* [Respondent 2]

> *Exactly. It helps us a lot because in reality, they're good because they're reminding us—* [Respondent 1]

> *They remind us not to forget to take the medicine and talk to the Doctor. Be on time with your medicines, do not wait to the last minute until we run out. We see all of this in those…in the messages.* [Respondent 3; translated from Spanish]

Participants found the specific behavior cued by the medication reminder messages easy to follow and to remember. They found both the direct cues to take medications, as well as the other tips to improve medication adherence, helpful and positive. While the participants generally were positive about the program, this specific component was the most frequently and emphatically mentioned as having the most influence on their diabetes management.

Specific Strengths of Program: Healthy Living Challenges

Another major category of messages participants found impactful was the healthy living challenge. Participants received a message with a specific healthy behavior to perform (diet choice, physical activity, and mental health/relaxation) twice a week. There was no repeat reminder message to the challenge and participants did not have to respond that they performed the challenge. Participants mentioned these as particularly motivating. In regards to challenge messages, participants stated:

> *The challenges are good. The challenges you send us. One imagines that I see the message, and when I read the challenge, those are my challenges for the day. The challenges. It's not every day but when they send challenges, they helped me a lot. I don't answer them but I read them and I say, "I have to do this". I motivate myself, like if I'm going to go walk. Or if I'm going to go eat a salad.* [translated from Spanish]
>
> *It's like the little thing you put, you have in your head all the time, just thinking about it. It makes you think actually about what you're doing to yourself actually. You know, if you're not doing the proper thing. You know it's like a head, a head reminder all the time. You wanna live longer, you know. It's not just a little message, it's what they, it's what stays in your head about 'em. And that's the way I see it. More like a challenge.*

Participants noted that these concrete behavior challenges were able to effect diet and exercise behavior and did not note this about educational messages. The participants perceived a need for a boost to live the healthier lives they already wanted. The healthy living challenges provided this boost. The emphasis on a specific behavior enabled participants to focus on one healthy change to make for that day. It seems that when this one goal was able to take root for that day, participants believed that it helped them in future diabetes management as well.

Specific Development Opportunities: Increase Personalization of Message Delivery

In the design of TExT-MED, we opted to have all participants receive the same messages and to have automated message delivery. The messages did not include the patient name and the delivery schedule was not specified to anyone's schedule. We did this to maximize the scalability of the program. However, participants specifically requested personalization of message delivery, including using their names and timing. Participants said they would have felt more cared for if the messages were more personalized. They also noted that medication reminders would come too early or too late to be most effective. They elaborated:

> *Can we ask for something more personal? Like "Hi, [states her own name],"…Computers can do so much now, nowadays; if you just said, "oh [own name]…just texting about your diabetes"…You see that it's a text from you guys and say, "Hi, [own name], how are you doing today?"*
>
> *I think that as technology advances, it will improve more-- in my case when I would receive messages about my pills, well I already have…I think that as technology improves, we can also see on the Internet, putting the hours at which we have to take our pills so that it can tell us exactly at what hours we need to do that. Because right now it is just a general announcement…But I think as time goes by and it becomes more personalized, it will help much, much more.* [translated from Spanish]

In summary, when asked about problems with the program, a lack of personalization was the most recurrent and prominent theme. Participants made the aforementioned suggestions about how their user experience could be improved. Additionally, they believed that this would make the program more effective. Specific development opportunities are to personalize messages with the patient name and to adjust the message delivery schedule according to patient preference and needs.

Specific Development Opportunities: Increase Personalization of Message Content

Participants also wanted a program designed for their specific experiences with diabetes. They needed information that was relevant to the complications that they faced. Participants also noted that people at different stages of disease have different needs in establishing or changing disease management regimens:

> *Well, this is very personal, but the thing is we have to differentiate the levels of diabetes…For example in my case, when they took my, they cut my foot off. So then I had to look for, on my own, what types of exercises I have to do to help me to manage my diabetes and also learn how to live without my foot. So little things like that, like I say, everything is in general terms, because it's a pilot program. But as it becomes more personalized, well maybe, we know that certain things affect this man and we can send him specific information for him.* [translated from Spanish]

The participants detailed several other areas that might vary between individuals including knowledge of food, exercise preferences, family, social support, and mental health needs. Tailoring an mHealth program to the specific needs of individuals would make it more impactful to their daily lives and health behaviors.

Discussion

Principal Findings

mHealth holds great promise for reaching resource-poor populations, especially low-income, urban Latinos. Text-messaging is a low-cost and feasible way to reach this high-risk group. More than three-quarters of Latinos in the United States own a cellular phone and, of these cell phone owners, 72% already use them to send or receive text messages [14]. Additionally, the potential for scalability of mHealth offers an opportunity to reach these patients cost-effectively. mHealth interventions have previously been shown effective as a tool in chronic diabetes management [4-6,15,16]. Currently studied mHealth programs for diabetes range from the highly intensive two-way and personalized messages [5,15-18] to one-way broadcast messages with no personalization at all [7,19,20]. While these reports are encouraging, neither the mechanism of mHealth action nor the optimal design of mHealth interventions is fully understood. A recent Cochrane review was conducted of text-messaging to improve self-care in patients with chronic disease [21]. This in-depth qualitative analysis of a low-cost, scalable mHealth program shows that patients will accept and enjoy uni-directional and broadcast messages. The nuances of motivation and behavior change that we uncovered using qualitative techniques gives us a greater understanding of how and why this mHealth intervention functioned. We identified specific cues to action as the most impactful messages and ways to make these cues more meaningful to patients. This allows us to create new programs that are more persuasive and effective.

Participants identified the most impactful and motivational messages to be the medication reminders and challenge messages—the two types of messages that cued concrete behavioral responses. According to the Health Belief Model applied to diabetes [12], individual patient factors affect the perceived threat of diabetes, the perceived benefits to acting to manage diabetes, and the barriers to taking these actions. Interventions can increase the perceived threat through cues to action, increase the perceived benefits of healthy behaviors, or decrease the perceived barriers to diabetes self-management. As behavior triggers were noted to be the most persuasive messages, this mHealth intervention appears to be most effective as a novel cue to action rather than educating patients to increase perceived benefits or decrease perceived barriers. Prior mHealth studies have shown success in affecting concrete behavior changes, suggesting mHealth takes on a different role than the traditional public health interventions that focus on educating patients to improve healthy behaviors [7,22,23]. Patients do not only need to be told why they should perform health behaviors, they need to be reminded, urged, and persuaded to undertake them. This is consistent with the model of persuasive design proposed by Fogg [24], that the likelihood of a behavior is a function of a person’s motivation and ability to perform that behavior, as well as the trigger to behavior provided by an intervention. The Fogg Behavior Model, while not specific to health interventions, helps to explain our findings. In this mHealth intervention, the behavior triggers were the most effective component.

Our analysis also uncovered ways to make these cues to action more meaningful; we found that patients desire some personalization of messages. This personalization may improve acceptability of an mHealth program for patients with diabetes when deployed in a larger population. Some of the personalization options identified by participants are feasible in an automated system, especially as the platforms for delivery of mHealth messages continue to advance. Using patient names and timing messages to the personal schedule of each patient would require a negligible increase in time in the initial set-up of messages for each patient, and could continue to run on an automated system. Tailoring messages to specific complications or years with diabetes may not be as easy to implement, but may make messages more persuasive and useful. However, this possible increase in effectiveness must be weighed against the cost of developing individualized programs for smaller and smaller subsets of patients. Further research should be conducted into how modules increasing a sense of personalization could be developed on a cost-effective basis and what types of personalization will result in more persuasive cues.

Limitations

This qualitative analysis has several limitations. Participants were selected for their experience in the TExT-MED program and the groups were completed when all available participants had participated in a focus group rather than the common qualitative technique of terminating data collection when saturation of codes and concepts has been reached [11]. However, after the fourth group, no new codes were developed so we do not believe that we lost data through our sampling process. The structure of data collection (focus groups rather than individual interviews) may have limited the participants’ ability to express individual opinions. However, focus groups were selected over individual interviews due to the disadvantaged background of most of the participants, as Magill [25] asserts that focus groups allow the disenfranchised to voice disagreement with authority, which was critical to discovering areas of the program that required improvement. The perceptions and experiences of the respondents in this study came from a group of low-income, Latino adults who were largely uninsured. The nature of qualitative work limits the generalizability of findings to patients in different settings. Future researchers should consider focusing on younger populations as well as suburban and rural communities. Last, the use of a sensitizing concept in our modified grounded theory technique may have blinded us to other important themes. However, the use of the Health Belief Model as a theoretical lens was critical to organizing our understanding of the relationship between patient factors and the behavioral cues of TExT-MED. Data was carefully coded by multiple coders and memoing was employed to be certain that the themes developed came from the data and not from the sensitizing concepts.

Conclusions

In spite of these limitations, this qualitative analysis has uncovered important information for further development of mHealth interventions for patients with chronic diseases, including diabetes. This analysis shows that a fully automated mHealth intervention tailored for low-income Latinos with diabetes is acceptable and persuasive for the target audience. This low-cost, scalable mHealth intervention acts as a behavioral trigger rather than a patient education platform. Increased personalization is an opportunity to enhance the strength of these cues to action. mHealth developers should focus on creating messages that are concrete behavioral triggers and personalizing the experience of users rather than in developing extensive educational messages. Future researchers should investigate whether increased personalization, especially in behavioral cues, results in improved health outcomes.

Acknowledgments

This study was supported by a mobilizing for health grant from the McKesson foundation (Drs Arora and Menchine) and a pilot grant from the Southern California Clinical and Translational Science Institute (SC CTSI) (NIH/NCRR/NCATS Grant # UL1TR000130). Dr Burner’s time on this study was supported by a NRSA F32 grant from the National Institute of Diabetes and Digestive and Kidney Diseases, Grant # 1F32DK094547 and the SC CTSI (NIH/NCRR/NCATS) Grant # KL2TR000131.

We are grateful to the patients who participated so willingly in the study. We thank Dr Benjamin Henwood, MSW, PhD from the Department of Social Work, University of Southern California for his helpful reading and editing of this manuscript. We acknowledge Adrianna Berumen for assistance with data collection and organization of focus groups.

Conflicts of Interest

The intellectual property rights to the TExT-MED program have been purchased from the University of Southern California by Agile Health, LLC. After study completion and data analysis, Drs Arora and Menchine have been hired by Agile Health, LLC to serve as consultants on the development of future text-based mobile health products. No other authors reported disclosures. The McKesson Foundation, The National Institutes of Health, and Agile Health LLC had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Multimedia Appendix 1

Anchor survey used at beginning of focus groups.

[PDF File (Adobe PDF File), 22KB - jmir_v16i1e25_app1.pdf]

Multimedia Appendix 2

Interview guide used for focus group sessions.

[PDF File (Adobe PDF File), 27KB - jmir_v16i1e25_app2.pdf]

References

1. American Diabetes Association. Economic costs of diabetes in the US In 2007. Diabetes Care 2008 Mar;31(3):596-615. [doi: 10.2337/dc08-9017] [Medline: 18308683]
2. Caballero A. Type 2 Diabetes in the Latino population in the US: a growing challenge. Avances en Diabetología 2008;24(4):283-290.
3. Banks J, Marmot M, Oldfield Z, Smith JP. Disease and disadvantage in the United States and in England. JAMA 2006 May 3;295(17):2037-2045. [doi: 10.1001/jama.295.17.2037] [Medline: 16670412]
4. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]
5. Kim HS. A randomized controlled trial of a nurse short-message service by cellular phone for people with diabetes. Int J Nurs Stud 2007 Jul;44(5):687-692. [doi: 10.1016/j.ijnurstu.2006.01.011] [Medline: 16618486]
6. Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res 2007;9(5):e36 [FREE Full text] [doi: 10.2196/jmir.9.5.e36] [Medline: 18166525]
7. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to Examine Text Message-Based mHealth in Emergency Department Patients With Diabetes (TExT-MED): A Randomized Controlled Trial. Ann Emerg Med 2013 Nov 5:1. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
8. Kagawa-Singer M, Emmons KM. Behavioral theory in a diverse society: moving our field forward. Health Educ Behav 2009 Oct;36(5 Suppl):172S-176S. [Medline: 19830890]
9. Krueger R. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, CA: Sage Publications (CA); 2009.
10. Dedoose. A web application for managing, analyzing, and presenting qualitative and mixed method data. Los Angeles, CA: SocioCultural Resources Counsultants LLC; 2012.

11. Charmaz K. Constructing grounded theory: a practical guide through qualitative analysis. London: SAGE; 2006.
12. Becker MH, Janz NK. The Health Belief model applied to understanding diabetes regimen compliance. The Diabetes Educator 1985 Mar 01;11(1):41-47. [doi: 10.1177/014572178501100108]
13. De Vries H, Elliott MN, Kanouse DE, Teleki SS. Using pooled kappa to summarize interrater agreement across many items. Field Methods 2008 Mar 03;20(3):272-282. [doi: 10.1177/1525822X08317166]
14. Livingston G. Pew Hispanic Center. 2011 Feb 09. Latinos and digital technology, 2010 URL: http://www.pewhispanic.org/2011/02/09/latinos-and-digital-technology-2010/ [accessed 2013-08-14] [WebCite Cache ID 6IsM9CkQ3]
15. 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]
16. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
17. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz DL. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract 2008 Jun;14(3):465-469. [doi: 10.1111/j.1365-2753.2007.00881.x] [Medline: 18373577]
18. Rami B, Popow C, Horn W, Waldhoer T, Schober E. Telemedical support to improve glycemic control in adolescents with type 1 diabetes mellitus. Eur J Pediatr 2006 Oct;165(10):701-705. [doi: 10.1007/s00431-006-0156-6] [Medline: 16670859]
19. Hanauer DA, Wentzell K, Laffel N, Laffel LM. Computerized Automated Reminder Diabetes System (CARDS): e-mail and SMS cell phone text messaging reminders to support diabetes management. Diabetes Technol Ther 2009 Feb;11(2):99-106. [doi: 10.1089/dia.2008.0022] [Medline: 19848576]
20. Bell AM, Fonda SJ, Walker MS, Schmidt V, Vigersky RA. Mobile phone-based video messages for diabetes self-care support. J Diabetes Sci Technol 2012 Mar;6(2):310-319 [FREE Full text] [Medline: 22538140]
21. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12:CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
22. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
23. McGillicuddy JW, Gregoski MJ, Brunner-Jackson BM, Weiland AK, Patel SK, Rock RA, et al. Proof of concept findings with uncontrolled hypertensives and kidney transplant recipients. In: Proceedings of the conference on Wireless Health. 2012 Presented at: Wireless Health; 2012; San Diego, California p. 1-9. [doi: 10.1145/2448096.2448108]
24. Fogg, B. BJ Fogg's Behavior Model. A behavior model for persuasive design URL: http://bjfogg.com/fbm_files/page4_1.pdf [accessed 2014-01-09] [WebCite Cache ID 6MVPyb0Q7]
25. Magill RS. Focus groups, program evaluation, and the poor. J. Soc. & Soc. Welfare 1993;20:103-114.

Abbreviations

TExT-MED: Trial to Examine Text Messaging for Emergency Department patient with Diabetes

Edited by G Eysenbach; submitted 14.08.13; peer-reviewed by P Foley, F Khan, K Pal; comments to author 10.09.13; revised version received 20.11.13; accepted 21.12.13; published 29.01.14.

Please cite as:
Burner ER, Menchine MD, Kubicek K, Robles M, Arora S
Perceptions of Successful Cues to Action and Opportunities to Augment Behavioral Triggers in Diabetes Self-Management: Qualitative Analysis of a Mobile Intervention for Low-Income Latinos With Diabetes
J Med Internet Res 2014;16(1):e25
URL: http://www.jmir.org/2014/1/e25/
doi:10.2196/jmir.2881
PMID:24476784

Original Paper

Mind the Gap: Social Media Engagement by Public Health Researchers

Brett Keller[1], MPA, MSPH; Alain Labrique[2], MHS, MS, PhD; Kriti M Jain[1], MSPH; Andrew Pekosz[3], PhD; Orin Levine[4], PhD

[1]Johns Hopkins School of Public Health, Baltimore, MD, United States

[2]Johns Hopkins School of Public Health, Department of International Health and Department of Epidemiology, Baltimore, MD, United States

[3]Johns Hopkins School of Public Health, Department of Molecular Microbiology and Immunology and Department of Environmental Health Sciences, Baltimore, MD, United States

[4]Bill & Melinda Gates Foundation, Seattle, WA, United States

Corresponding Author:
Alain Labrique, MHS, MS, PhD
Johns Hopkins School of Public Health
Department of International Health and Department of Epidemiology
615 North Wolfe St, E5543
Baltimore, MD, 21205
United States
Phone: 1 443 287 4744
Fax: 1 410 510 1055
Email: alabriqu@jhsph.edu

Abstract

Background: The traditional vertical system of sharing information from sources of scientific authority passed down to the public through local health authorities and clinicians risks being made obsolete by emerging technologies that facilitate rapid horizontal information sharing. The rise of Public Health 2.0 requires professional acknowledgment that a new and substantive forum of public discourse about public health exists on social media, such as forums, blogs, Facebook, and Twitter.

Objective: Some public health professionals have used social media in innovative ways: to surveil populations, gauge public opinion, disseminate health information, and promote mutually beneficial interactions between public health professionals and the lay public. Although innovation is on the rise, most in the public health establishment remain skeptical of this rapidly evolving landscape or are unclear about how it could be used. We sought to evaluate the extent to which public health professionals are engaged in these spaces.

Methods: We conducted a survey of professorial- and scientist-track faculty at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA. We asked all available faculty via email to complete a 30-question survey about respondent characteristics, beliefs about social media, and usage of specific technologies, including blogs, Facebook, Twitter, and YouTube.

Results: A total of 181 (19.8%) of 912 professor- and scientist-track faculty provided usable responses. The majority of respondents rarely used major social media platforms. Of these 181 respondents, 97 (53.6%) had used YouTube, 84 (46.4%) had used Facebook, 55 (30.4%) had read blogs, and 12 (6.6%) had used Twitter in the prior month. More recent degree completion was the best predictor of higher usage of social media. In all, 122 (67.4%) agreed that social media is important for disseminating information, whereas only 55 (30.4%) agreed that social media is useful for their research. In all, 43 (23.8%) said social media was helpful for professional career advancement, whereas 72 (39.8%) said it was not. Only 43 (23.8%) faculty said they would employ a full- or part-time social media consultant, and 30 (16.6%) currently employed one.

Conclusions: Despite near-universal appreciation of the potential for social media to serve as a component of public health strategy, a small minority are actually engaged in this space professionally, whereas most are either disinterested or actively opposed to professional engagement. Social media is seen by most as more useful for spreading information than obtaining it. As public discourse on a number of critical health topics continues to be influenced and sometimes shaped by discussions online from Twitter to Facebook, it would seem that greater discourse is needed about when and how public health professionals should engage in these media, and also how personal, institutional, and professional barriers to greater use of social media may be overcome.

(J Med Internet Res 2014;16(1):e8) doi:10.2196/jmir.2982

KEYWORDS

Internet; social media; public health; blogging

Introduction

Background

Over the past 2 decades, the Internet has become an important source of public information. Recently, with the growth and global penetration of social networks, a wide range of online platforms have become important forums for public dialog about health and health care. Across the globe, we posit that a sea change is occurring, characterized not by seekers of static reference information from Internet sources, but by those looking to engage in interactive, bidirectional communication with global communities of like individuals sharing common health aspirations or challenges. As proposed by Evgeny Morozov [1] in his recent book *To Save Everything, Click Here*, we find ourselves rapidly ascending to dizzying heights of "technologic solutionism," characterized by the devolution of trust in traditional sources of public health information and a growing reliance on the "wisdom of crowds and the marketplace of ideas." Social media has taken on a new and important role in public discourse and debate, ranging from the mundane to issues of public health significance. Social media platforms, Web- or mobile-based, facilitate interaction by letting users create, share, and view user-generated content. Users can be transformed from passive consumers into content producers [2].

Even before the rise in popularity of social media, patients were increasingly accustomed to seeking medical information online. In a survey of patients from a primary care internal medicine private practice, 53.5% of respondents said they used the Internet to find medical information [3]. A further 60% of those who used the Internet to find medical information believed the quality of information that they found online was the same as or better than what their doctors provided [3]. A 2013 online survey confirmed that health professionals and patients use social media for different purposes, with patients seeking knowledge, garnering social support, and exchanging advice, whereas health professionals communicate with colleagues and market their services [4].

In the past decade, however, many reviews of the quality of medical information available on the Internet based on reviews by clinical experts have rated information as poor and often potentially dangerous [5]. A 2007 systematic review of online information about inflammatory bowel disease identified many sources, but these were characterized by highly variable quality [6]. Eysenbach [7], in his 1998 *BMJ* review, proposed that although the Internet's inherent "anarchic nature" was essential for uncensored debate, this same attribute can allow poor quality and even dangerous misinformation to proliferate.

When the interactive element of social media is added to the ease of finding information (and potential misinformation) online, the speed and reach of public health debate is often unprecedented, especially in areas that are fraught with controversy. The tools required to create sophisticated online content are widely accessible and information that seems to be authoritative or scientifically valid can be generated at low cost. The result may be that distinguishing science from opinion is much more challenging today for the lay public than it may have been a decade ago. Civil society organizations ranging from patient and special interest groups to so-called "citizen advocacy" groups are increasingly driving public discussion around the introduction of new vaccines [8]. Social media is being explored as a means to deliver some interventions, although the full potential of these strategies remain to be demonstrated [9,10]. The capacity of social networking and social media technologies to effect sea changes in society were made evident by the events of the 2011 "Arab Spring," demonstrating how ideas and information could spread virally across population [11].

Still, a strong health professional view persists that the user-generated content of social media sites are little more than backchannels, which serve mainly to spread "misinformation and rumor" [3]. This divide is further illustrated by recent findings that patients are more likely than physicians to use social media sites to access or discuss health information [12]. A review of social media usage by local health departments found that communication has been mostly one-way, from departments to the public, but that dialog and engagement are increasing [13].

We argue that social media can be seen as a new landscape for dialog and public health insight, where researchers can gather health information, disseminate research findings, and provide guidance. Opportunities for research abound because social media users are surprisingly open in discussing their own health [14]. Patient-centered online communities, such as PatientsLikeMe.com, have been used to gain insight into rare disease conditions, patient-reported drug-related side effects, and even to validate new instruments [15-17]. Researchers have used sophisticated natural language processing and big data tools to analyze conversations on Twitter, and have been able to predict annual influenza and other epidemic diseases with remarkable accuracy [14]. Scanfield and colleagues [18] used Twitter in 2010 as a means of exploring antibiotic misuse and sharing. As early as 2008, Collier and colleagues [19] had developed Web-based text mining systems to identify and map "public health rumors" into a system they coined BioCaster. These examples illustrate how social media sites, ranging from Twitter to Facebook and even the media-sharing site YouTube, offer novel platforms for health information exchange. Despite initial skepticism about the reach or impact of these platforms, it has become clear that these networks continue to grow, and younger populations are likely to increase their reliance on these sources for public health and medical guidance [20].

Objective

Although social media sites seem to be increasingly important tools for personal health information exchange, relatively few empirical studies have examined characteristics of those who use social media or the potential health effects of accessing user-generated content [21]. There is even less published information about how and why public health researchers and practitioners use social media professionally. As the public discourse continues to mature within these virtual spaces, it will be critical to identify opportunities to engage health professionals in the dialog. Although it may be overly ambitious to seek to balance this conversation, further study may identify ways to improve the dissemination of valid information and influence positive behavior change [18].

We sought to illustrate this potential by evaluating the extent to which public health researchers are engaged in these spaces. To that end, we conducted a survey of professorial- and scientist-track faculty at the Johns Hopkins Bloomberg School of Public Health, the oldest and largest school of public health in the United States, to begin filling in this gap in public health knowledge. We investigated Johns Hopkins faculty beliefs about social media and their use of various social media tools. A follow-up survey is planned in 2014 to track changes in opinions and activities over time.

Methods

In April 2011, we conducted an online survey of social media beliefs and practices among faculty at the Johns Hopkins School of Public Health, Baltimore, MD, USA, a leading public health institution in the United States with a large faculty with diverse research interests. To our knowledge, this is the first survey of its kind.

The Johns Hopkins Bloomberg School of Public Health regularly updates a public listing of its 1564 full- and part-time faculty and researchers. There are 2 primary tracks for faculty—professor or scientist—in addition to positions of varying permanence. We restricted our survey to the 912 faculty in professorial- and scientist-track positions.

We requested all members of an updated faculty email list to complete an anonymous online survey with questions about respondent characteristics, beliefs about social media, and usage of specific technologies, including blogs, Facebook, Twitter, and YouTube. The 30-question survey was administered using a commercial survey website, SurveyMonkey. Three email invitations were sent during a 6-week period in March-April 2011.

We constructed a Social Media Usage Index (SMUI) score with a possible range of 0-16, weighting each social media service equally. For each of the 4 social media services mentioned in the survey (blogs, Facebook, Twitter, and YouTube), respondents were assigned zero points for never having heard of a service, 1 point for having heard of but never using the service, 2 points for using the service but not in the past month, 3 points for using the service once or twice in the past month, and 4 points for using the service 3 or more times in the past month.

Responses were analyzed with Stata statistical software version IC 11.2 (StataCorp LP, College Station, TX, USA). The Johns Hopkins School of Public Health Institutional Review Board reviewed the study and declared it exempt.

Results

We received 181 usable responses by professor- and scientist-track faculty out of a total of 912 potential respondents in those positions, for a response rate of 19.8%. Unusable responses included incomplete responses and responses by individuals on the faculty email list but not part of the professor- or scientist-track faculty. Respondent characteristics are described in Table 1.

Most respondents rarely used major social media platforms. Respondents were more likely to have used YouTube (94/181, 51.9%) and Facebook (81/181, 44.8%) than to have read blogs (53/181, 29.3%) or used Twitter (12/181, 6.6%) in the prior month. Awareness of these services was nearly universal: only 1 respondent had not heard of Twitter and YouTube, and all respondents had heard of Facebook and blogs. Respondents were much more likely to use all 4 services for personal reasons, but the proportion of use that was predominantly personal for YouTube (163/181, 90.0%) and Facebook (164/181, 90.6%) was greater than for Twitter (146/181, 80.7%) and blogs (149/181, 82.3%). When restricted to respondents who used the services regularly (≥3 times/week) the proportion of mostly personal use was similar for YouTube (160/181, 88.4%), Facebook (165/181, 91.2%), and blogs (122/181, 67.4%); only 6/181 (3.3%) respondents used Twitter this frequently.

Faculty responses seemed to distribute along a continuum of enthusiasm. One respondent, asked about whether faculty should be engaged in social media discussions, replied, “I very rarely look at YouTube, only if someone sends me something. I write 30,000 emails per year. There is no time for additional media of any kind. I am skeptical of researchers who use social media to increase interest in their work.” In contrast, others expressed supportive opinions of engagement, with a caveat of limiting such interaction, “I think it is helpful to be engaged in some social media in the current age; however, maintaining a formal presence takes time most of us don’t have. It is beneficial that my center organizes it for us, and we work with students to do blogging. Unfortunately, credibility as an academic may be affected when blogging too much.”

Finally, a smaller number of faculty recognized that social media is useful to public health professionals “to get information out when trying to build awareness or change policy. The short, informal nature of social media is critical to reaching certain audiences who don’t have the time to weed through long reports or read journal articles.” Another respondent echoed the sentiment, “I think it is very important for raising awareness of a public health issue, for advocacy purposes, and for dissemination of study results.”

Table 1. Social Media Usage Index (SMUI) score by respondent characteristics.

Characteristics	Respondents n (%)	SMUI mean (95% CI)[a]
Total	181 (100)	
Gender		
Male	83 (45.9)	8.33 (2.58-14.08)
Female	97 (53.6)	8.42 (2.94-13.90)
Age (years)		
≤40	49 (27.1)	9.12 (4.07-14.18)
41-50	47 (26.0)	8.45 (2.91-13.99)
51-60	43 (23.8)	8.70 (2.83-14.57)
≥61	42 (23.2)	7.07 (1.83-12.31)
Years as John Hopkins faculty		
0-10	92 (50.8)	8.89 (3.10-14.68)
11-20	37 (20.4)	8.19 (2.99-13.39)
≥21	52 (28.7)	7.57 (2.41-12.75)
Position		
Assistant scientist	29 (16.0)	8.66 (3.68-13.63)
Associate scientist	9 (5.0)	8.78 (2.75-14.80)
Senior scientist	4 (2.2)	7.50 (2.32-12.68)
Assistant professor	35 (19.3)	9.03 (3.43-14.62)
Associate professor	34 (18.8)	7.71 (2.28-13.13)
Professor	70 (38.7)	8.24 (2.38-14.09)

[a] From bivariate analysis.

Social media was seen, in this sample, as more useful for spreading research results than conducting research: 122 (67.4%) agreed that social media is important for disseminating information, whereas only 55 (30.4%) agreed that social media is useful for their research. Respondents were skeptical when asked whether social media engagement was helpful for career advancement; 43 (23.8%) said social media usage was useful, whereas 72 (39.8%) said it was not. A minority were very involved with social media: 43 (23.8%) faculty said they would employ a full- or part-time social media consultant, and 30 (16.6%) currently employed one. Open responses questions indicated considerable enthusiasm for social media from respondents engaged in the field. Figure 1 is a visual diagram (Wordle) that arranges the top 150 most common words in survey respondents' open response answers, scaling the size of words to match their usage frequency.

We fitted bivariate and multivariate linear regression models with the SMUI score as the dependent variable to identify factors associated with higher social media usage scores (Table 2). Mean SMUI did not differ by gender or age category, but was higher for associate professors. For each additional year since degree completion, the mean SMUI score decreased slightly.

Table 2. Correlates of Social Media Usage Index (SMUI) scores in multivariate analysis.

Correlates in model	Difference in mean SMUI score, (95% CI)	*P* value[a]
Gender		
Male	[Reference]	
Female	0.52 (–0.34, 1.38)	.23
Age (years)		
≤40	[Reference]	
41-50	–0.59 (–1,83, 0.64)	.35
51-60	0.19 (–1.34, 1.73)	.81
≥61	–0.84 (–2.80, 1.11)	.40
Position		
Assistant scientist	[Reference]	
Associate scientist	0.67 (–2.29, 3.62)	.66
Senior scientist	0.47 (–0.89, 1.84)	.50
Assistant professor	0.25 (–1.22, 1.73)	.74
Associate professor	2.40 (0.68, 4.12)	.006
Professor	1.73 (–0.43, 3.90)	.12
Years since degree completion		
0 years	[Reference]	
Per year increase	–0.12 (–0.18, –0.05)	<.001

[a] *P* values calculated with *t* tests.

Figure 1. A “Wordle” diagram of survey responses on current and desired uses of social media in public health research and practice.

Discussion

Our results suggest that despite a substantial appreciation for the potential of social media to serve as a component of public health strategy, only a small minority of public health faculty are actually engaged in this space professionally. The majority of respondents were either disinterested or actively opposed to professional engagement in the social media space. Although the generalizability of these results is limited somewhat by only soliciting responses at a single institution, the Johns Hopkins Bloomberg School of Public Health is the largest of its type and represents a substantial portion of public health faculty in the United States.

As public discourse on a number of critical health topics continues to be influenced and sometimes shaped by discussions online from Twitter to Facebook, public health faculty should seriously consider when and how public health professionals should engage in these media, but also how personal, institutional, and professional barriers to greater use of social media may be overcome. Strategies for overcoming these barriers may start with greater awareness of specific applications of social media for public health practice and research. Textbox 1 lists social media resources from the Centers for Disease Control and Prevention (CDC), the National Information Center on Health Services Research and Health Care Technology, and the Johns Hopkins School of Public Health. Recognition of engagement in social media also needs to be incorporated into the way faculty are recognized for their professional practice. This, however, will require the need for robust metrics to quantify the reach and impact specific faculty or institutions are having in these spaces. Services like Klout, Tweetlevel, Bloglevel, and Export.ly provide analytics with some interpretive filters which attempt to quantify the influence specific individuals or organizations have on others.

Textbox 1. Social media resources.

- CDC Social Media page [22]
- CDC's Health Communicator's Social Media Toolkit [23]
- Social Media Resources listed by the National Information Center on Health Services Research and Health Care Technology [24]
- Social Media Channels at the Johns Hopkins School of Public Health [25]

In the age of social media, information is no longer constrained to vertical channels of authority; ideas are shared freely between citizens who can inform or misinform the public. As Chou et al [20] pointed out in their 2009 characterization of users of social media for health information, age, socioeconomic status, and ethnicity may be important confounders of access to or use of these channels. It will be important to further explore the degree to which social media information sources play a role in individual decision making, such as whether to vaccinate one's children or to choose formula over breast-milk. Public health professionals—from physicians to government officials to academic researchers—should strategically adopt new technologies and styles of communication or risk being excluded from this conversation entirely.

To encourage the use of new approaches to information sharing through social media and engagement in public discourse within this space, we call for increased professional discussion of the benefits and risks of more active engagement in social media by public health professionals, both as a means of gathering new information and to influence ongoing discussions of public health importance.

Acknowledgments

AL, AP, KJ, and OL conceived the study; AL, KJ, and BK developed and disseminated the survey; KJ and BK analyzed the data and wrote the first draft of the manuscript. All authors edited and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey on faculty social media use.

[PDF File (Adobe PDF File), 40KB - jmir_v16i1e8_app1.pdf]

References

1. Morozov E. To Save Everything, Click Here: The Folly of Technological Solutionism. Philadelphia, PA: Public Affairs; 2013.
2. Keim ME, Noji E. Emergent use of social media: a new age of opportunity for disaster resilience. Am J Disaster Med 2011 Feb;6(1):47-54. [Medline: 21466029]
3. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]
4. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
5. Craigie M, Loader B, Burrows R, Muncer S. Reliability of health information on the Internet: an examination of experts' ratings. J Med Internet Res 2002;4(1):e2 [FREE Full text] [doi: 10.2196/jmir.4.1.e2] [Medline: 11956034]

6. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
7. Eysenbach G, Diepgen TL. Towards quality management of medical information on the Internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1500 [FREE Full text] [Medline: 9831581]
8. Laurent-Ledru V, Thomson A, Monsonego J. Civil society: a critical new advocate for vaccination in Europe. Vaccine 2011 Jan 17;29(4):624-628. [doi: 10.1016/j.vaccine.2010.11.004] [Medline: 21081178]
9. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
10. Moniz MH, Hasley S, Meyn LA, Beigi RH. Improving influenza vaccination rates in pregnancy through text messaging: a randomized controlled trial. Obstet Gynecol 2013 Apr;121(4):734-740. [doi: 10.1097/AOG.0b013e31828642b1] [Medline: 23635672]
11. Stepanova E. PONARS Eurasia. 2011 May. The role of information communication technologies in the "Arab Spring" URL: http://ponarseurasia.com/sites/default/files/policy-memos-pdf/pepm_159.pdf [accessed 2013-12-29] [WebCite Cache ID 6MESMmiWG]
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. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
14. Paul M, Dredze M. You are what you tweet: Analyzing Twitter for public health. In: 5th International Conference on Weblogs and Social Media. 2011 Presented at: ICWSM; July 17-21, 2011; Barcelona, Spain.
15. Wicks P, Vaughan TE, Massagli MP. The multiple sclerosis rating scale, revised (MSRS-R): development, refinement, and psychometric validation using an online community. Health Qual Life Outcomes 2012;10:70 [FREE Full text] [doi: 10.1186/1477-7525-10-70] [Medline: 22709981]
16. Nakamura C, Bromberg M, Bhargava S, Wicks P, Zeng-Treitler Q. Mining online social network data for biomedical research: a comparison of clinicians' and patients' perceptions about amyotrophic lateral sclerosis treatments. J Med Internet Res 2012;14(3):e90 [FREE Full text] [doi: 10.2196/jmir.2127] [Medline: 22721865]
17. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
18. 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]
19. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, et al. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 2008 Dec 15;24(24):2940-2941 [FREE Full text] [doi: 10.1093/bioinformatics/btn534] [Medline: 18922806]
20. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
21. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010;15 Suppl 3:216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
22. Social Media at CDC. URL: http://www.cdc.gov/socialmedia/ [WebCite Cache ID 6MLHeOeBc]
23. The Health Communicator's Social Media Toolkit. URL: http://www.cdc.gov/socialmedia/tools/guidelines/pdf/socialmediatoolkit_bm.pdf [WebCite Cache ID 6MLHzCwma]
24. National Information Center on Health Services Research and Health Care Technology (NICHSR). URL: http://www.nlm.nih.gov/hsrinfo/hsrdlist.html [WebCite Cache ID 6MLIMdv70]
25. Johns Hopkins School of Public Health. Social media channels URL: http://www.jhsph.edu/news/social-media-channels/ [WebCite Cache ID 6MLIjs4H8]

Abbreviations

CDC: Centers for Disease Control and Prevention
SMUI: Social Media Usage Index

Edited by G Eysenbach; submitted 26.09.13; peer-reviewed by Z Hajar, M Hamm; comments to author 22.11.13; revised version received 06.12.13; accepted 10.12.13; published 14.01.14.

Please cite as:
Keller B, Labrique A, Jain KM, Pekosz A, Levine O
Mind the Gap: Social Media Engagement by Public Health Researchers
J Med Internet Res 2014;16(1):e8
URL: http://www.jmir.org/2014/1/e8/
doi:10.2196/jmir.2982
PMID:24425670

Review

Caught in the Web: A Review of Web-Based Suicide Prevention

Mee Huong Lai[1], MD; Thambu Maniam[2], MBBS, MPsychMed, FAMM; Lai Fong Chan[2], MD, DrPsych; Arun V Ravindran[2,3], MD, PhD, FRCPC, FRCPsych

[1]Faculty of Medicine & Health Sciences, Department of Psychological Medicine, Universiti Malaysia Sarawak, Kuching, Malaysia
[2]Department of Psychiatry, Universiti Kebangsaan Malaysia Medical Centre, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia
[3]Department of Psychiatry, University of Toronto and Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada

Corresponding Author:
Lai Fong Chan, MD, DrPsych
Department of Psychiatry
Universiti Kebangsaan Malaysia Medical Centre
Universiti Kebangsaan Malaysia
Jalan Yaacob Latif
Bandar Tun Razak, Cheras
Kuala Lumpur, 56000
Malaysia
Phone: 60 3 91455555 ext 6143
Fax: 60 3 91457841
Email: laifchan@gmail.com

Abstract

Background: Suicide is a serious and increasing problem worldwide. The emergence of the digital world has had a tremendous impact on people’s lives, both negative and positive, including an impact on suicidal behaviors.

Objective: Our aim was to perform a review of the published literature on Web-based suicide prevention strategies, focusing on their efficacy, benefits, and challenges.

Methods: The EBSCOhost (Medline, PsycINFO, CINAHL), OvidSP, the Cochrane Library, and ScienceDirect databases were searched for literature regarding Web-based suicide prevention strategies from 1997 to 2013 according to the modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. The selected articles were subjected to quality rating and data extraction.

Results: Good quality literature was surprisingly sparse, with only 15 fulfilling criteria for inclusion in the review, and most were rated as being medium to low quality. Internet-based cognitive behavior therapy (iCBT) reduced suicidal ideation in the general population in two randomized controlled trial (effect sizes, d=0.04-0.45) and in a clinical audit of depressed primary care patients. Descriptive studies reported improved accessibility and reduced barriers to treatment with Internet among students. Besides automated iCBT, preventive strategies were mainly interactive (email communication, online individual or supervised group support) or information-based (website postings). The benefits and potential challenges of accessibility, anonymity, and text-based communication as key components for Web-based suicide prevention strategies were emphasized.

Conclusions: There is preliminary evidence that suggests the probable benefit of Web-based strategies in suicide prevention. Future larger systematic research is needed to confirm the effectiveness and risk benefit ratio of such strategies.

(J Med Internet Res 2014;16(1):e30) doi:10.2196/jmir.2973

KEYWORDS

suicide prevention; Web-based; Internet

Introduction

Suicide is a common and serious problem worldwide. The World Health Organization reported in 2009 that nearly one million people commit suicide yearly, with 3000 deaths a day or 1 suicide every 40 seconds. Suicide is also the second largest cause of mortality in the 10-24 year age group [1].

The emergence and rapid growth of the digital world and its technology tools has had a tremendous influence on people’s lives. World Internet usage has increased 566.4% from 2000 to

2012 [2]. In particular, there has been an obvious and growing interest both in the negative contribution of the Internet to suicidal behaviors, as well as on its potential usefulness as a tool for prevention [3-9].

There are several lines of evidence suggesting that the Internet may negatively influence suicidal behaviors in certain vulnerable groups. First, there have been clusters of suicides after sensational reporting of suicide cases in the media, referred to as the “Werther effect” or copycat suicides [10-14]. The effect of the Internet in reporting and glorifying suicide may be regarded as comparable to that of other conventional media [5]. Sporadic cases of copycat suicide via the Internet have been reported [4,15,16]. Second, pro-suicide websites and chat rooms often facilitate suicide pacts [17,18] and provide detailed models of lethal methods [15,19]. They also frequently discourage constructive help-seeking behaviors and often exert peer pressure on vulnerable Internet users who are ambivalent about committing suicide [16,20].

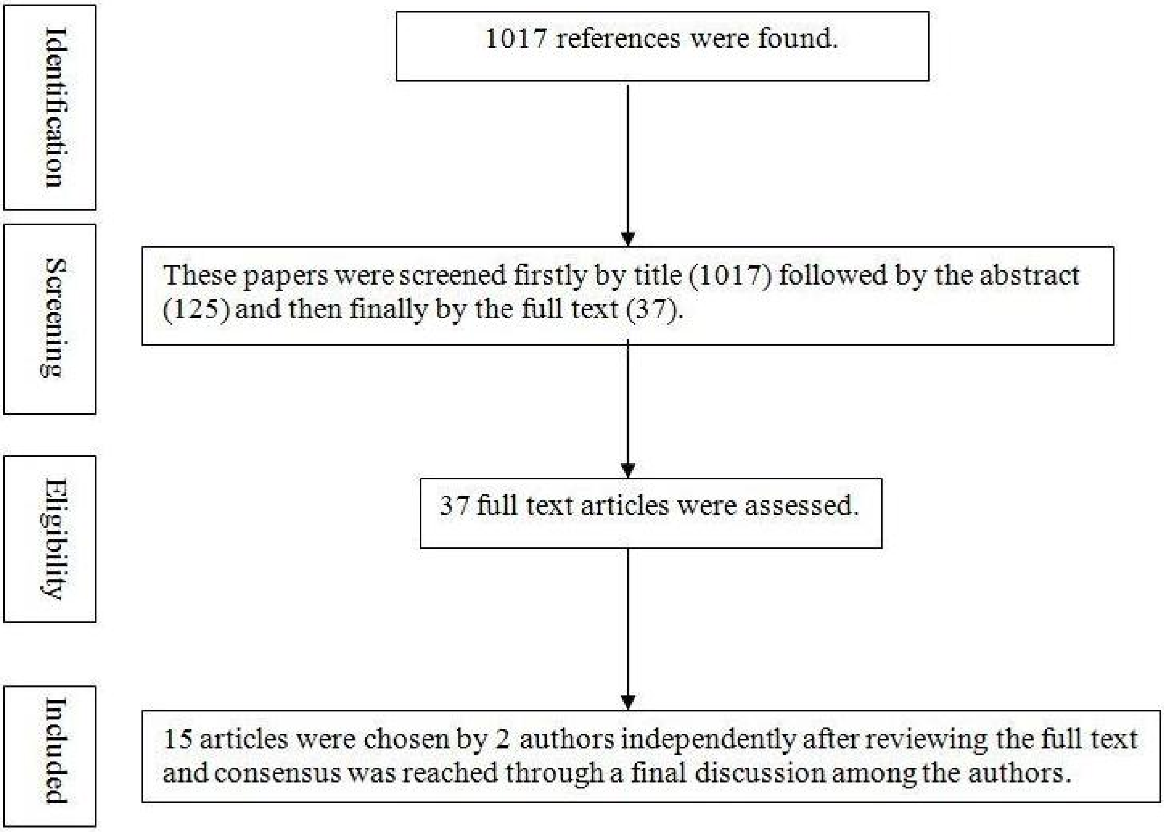
The Internet can be viewed as a double-edged tool [5]. While it is accepted that the Internet may be used to trigger and encourage suicidal behavior, its potential as a tool for suicide prevention has been equally recognized [8,9,21,22]. There is good evidence that populations vulnerable to suicide often access Web-based resources. For example, it has been shown that half of the service users of some online suicide prevention programs were clearly suicidal [23]. Similarly, a Web-based study among people with common mental disorders showed that 53.4 % (268/502 participants) reported some degree of suicidal ideation [24] and that suicide threats were more frequent among users of an online support group than among telephone hotline users [25]. Gould et al (2002) studied help-seeking resources for emotional distress in a community sample of adolescents and found that 18.2% (94/519 participants) used the Internet as their help-seeking tool and approximately one in ten of those logged in the Internet was seeking help for their suicidal thoughts [26]. According to Mohd Daud (2005), about a third of high school students from Malaysian schools with high rates of disciplinary problems used the Internet as a help-seeking tool for their emotional problems, out of which 6.4% (7/110 students) cited suicidal thoughts as being the reason for seeking help [27]. Interestingly, it has been noted that a majority of the hits generated by the search engine Google on the word “suicide” in 2005, 2009, and 2012 had a clear suicide-preventive message [28].

Durkee et al (2011) reviewed different pathways by which suicidal risks and prevention efforts are facilitated through the Internet, particularly in young people [8]. Daine et al (2013) performed a systematic review and concluded that Internet use may exert both positive and negative effects on young people at risk of self-harm or suicide [29]. To the best of our knowledge, there is currently no published systematic review focusing specifically on existing Web-based suicide prevention strategies. Therefore, this paper aims to review the available Web-based suicide prevention strategies and evaluate the evidence for their efficacy, benefits, and challenges.

Methods

The modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [30] was used as the methodological approach for this review. The search engines were EBSCOhost (Medline, PsycInfo, CINAHL), OvidSP, the Cochrane Library, and ScienceDirect from 1997-2013. The keywords were *suicide prevention*, *suicide support group*, *suicide help and Internet* or *suicide help and Web-based* or *suicide help online*, either in the title, abstracts, or keywords of the publications. Only English language papers were included. One criterion for inclusion in this review was that the publication should have a clear discussion of efficacy or benefits or challenge of a specific suicide prevention strategy. Letters to the editor and sample descriptions of study protocols alone were excluded, as well as articles that focused solely on the process of training and education on suicide prevention. The quality of primary research papers (not applicable in the case of review papers) was assessed according to the “Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields” by the Alberta Heritage Foundation for Medical Research [31]. The number of articles that were selected following the PRISMA statement with modification is shown in Figure 1. The initial screening process was done by 1 author (MHL). The selection of 15 articles was based on the independent assessment of the full-text articles by 2 authors (MHL and LFC), and consensus was reached through a final discussion among 3 authors (MHL, LFC, and MT).

Figure 1. Flow chart of identification, screening, eligibility assessment, and final inclusion of number of articles.



Results

The published literature on good-quality Web-based suicide prevention programs was sparse, with 15 articles (from 2007-2013) fulfilling our criteria as assessed independently by 2 authors. Two articles were narrative type review papers [32,33]. For primary research papers, there were two randomized controlled trials (RCTs) [34,35], but one of the RCTs produced two papers [34,36] that evaluated the efficacy and cost-effectiveness of unguided online cognitive behavioral therapy (CBT)-based self-help to reduce suicidal thoughts. Other primary research papers were one pre- and posttreatment case series [37], two cohort studies [38,39], two cross-sectional studies [40,41], one qualitative study [42], and four descriptive reports [23,43-45]. Based on the quality assessment and ratings of the 13 primary research papers, five papers were rated as high quality, five rated as medium quality, and three rated as low quality (Multimedia Appendix 1).

For the evaluation of the efficacy of the reviewed Web-based suicide prevention strategies, two review papers did not find clear evidence to confirm the efficacy of those strategies. Nevertheless, those strategies seemed promising and had the potential for greater opportunities for suicide prevention [32,33]. Three studies on the effectiveness of interventions based on Internet CBT (iCBT) showed some benefit in reducing suicidal ideation [34,35,37]. Spijker's (2012) RCT of 6 weeks in a general population sample aimed to test the efficacy and cost-effectiveness of an iCBT-based self-help intervention that consisted of 6 modules. The control group received access to a website with information on suicidality and common treatment options with links to mental health care institutions. The intervention arm showed a significant improvement in suicidal thoughts via a cost-effective way [36], albeit with a small effect size (*d*=0.28) [34]. Christensen et al's (2013) RCT showed that a 6-week course of iCBT among telephone helpline callers significantly reduced suicidal ideation. However, iCBT with or without telephone call back was not more effective compared to telephone call back alone (participants were contacted for 10 minutes systematically once a week for 6 weeks) or treatment-as-usual control condition (participants could call the crisis line at any time). Effect sizes of iCBT intervention were relatively small ranging from 0.04-0.45 over a 12-month period [35]. Watts et al's (2012) effectiveness study within a non-controlled clinical audit showed a significant reduction in suicidal ideation regardless of sex and age in a sample of primary health care patients who were given a six-lesson course of iCBT for depression [37].

The other papers mainly described and evaluated the existing real-life Web-based suicide prevention strategies without directly studying their efficacy with objective measurable outcomes. Among them, seven publications anecdotally reported some degree of positive benefit for users. These included cohort and descriptive studies among university students with suicidal behavior who seemed to have improved access to face-to-face mental health services after using the Internet as a means of communication on institution-supported Web programs

[38,39,45]. The other two papers reported the formation of support groups; one was qualitative research on understanding the help-seeking experience through an online message board with the theme of “suicide” [42] and the other was a descriptive paper on the model of an online support group [44]. Two other papers reported positive feedback from users who were in crisis [23,40].

One study evaluated the quality of information regarding youth suicide in community accessed websites for youth and found that less than half of such websites contained evidence-based statements [41]. Lester (2008) described the benefits and feasibility of counseling of suicidal individuals via email [43].

The majority of the programs utilized were developed and implemented by four non-governmental organizations (NGOs) [23,42-44], and three others were university-based suicide prevention programs [38,39,45]. Most of the Web-based programs provided information on contributing factors, guidance on seeking help, and information on available resources. In addition, interventional components that included elements of support, counseling, or cognitive therapy were integrated into most of the Web-based programs (Multimedia Appendix 1). The 15 articles have been grouped under the following categories in Multimedia Appendix 1: reviews, RCTs, and descriptive/naturalistic studies.

The literature suggests that three factors—ease of accessibility, degree of anonymity, and the nature of text-based communication—are key components for Web-based suicide prevention strategies. Each of these components provides certain benefits but also poses some challenges (Table 1).

Table 1. Benefits and challenges of the main characteristics of Web-based suicide prevention programs.

Characteristics	Benefits	Challenges
Accessibility	Convenient access to information [23]	Unauthorized access may affect the data accuracy; may lead to breaches of confidentiality [43]
	Providing international services across geographical [42,43], cultural, and physical boundaries [43,44]	Limits access for those without resources [43]
		May create inappropriate counseling interventions due to cultural differences [33]
		Logistical difficulties of intercountry referral process [33]
		Lack of formal environment may reduce efficacy of therapy [43]
Anonymity	Alternative to conventional services that may reduce barriers to help-seeking, especially in vulnerable populations seeking anonymity [23,33,39,43]	Risk of non-genuine message postings (estimated <5% of all communications with SAHAR) and accelerating burnout among helpers [23]
	Decreasing inhibitions with regard to self-disclosure and prompting of ventilation [23]	Risk of “flaming effect” with insults and hostile comments affecting at-risk individuals [33]
	Higher assurance of confidentiality in the context of sharing sensitive personal issues, including suicidal thoughts in a supportive online community [42]	Difficulties in sustaining therapeutic alliance with multiple “therapists” online [43]
		Ethical issues of identifying and monitoring quality of therapists [33,43]
Text-based	Writing has therapeutic effect [33]	Safety concern, ie, delayed response to an acute crisis [33,38]
	Flexibility of time in terms of immediate ventilation during crisis [33] and control in timing of sending message [42]	Risk of Internet addiction in clients who write excessively [43]
		Benefits limited to literate clients [43]
		Technical problems that interrupt communication [23]
	Saved text of counselor’s response can be studied by client [33] and also used to monitor and mentor the therapists and data for research [33,43]	Lack of physical cues [33] and non-immediate feedback [43] may impair communication
	Potentially an ideal communication tool for those with hearing or speech impairment and volunteers with physical impairment [43]	Therapists need to undergo training to be familiar with text-based communication [38]
	Resource-saving in terms of ability of experienced therapists in conversing with a few clients simultaneously [23]	The nature of communication that is saved may increase therapist anxiety [43]

Discussion

Principal Findings

This review confirms the paucity of strong evidence-based literature on the effectiveness of Web-based suicide prevention strategies. After the quality assessment of the primary research papers, a majority (8/13) of them were rated as medium to low quality. The interpretation of published data was difficult due to methodological constraints such as the heterogeneous nature of the strategies, target population, and the lack of systematic outcome evaluations. Greater attention was paid to papers of higher quality.

To date, two published RCTs examined the effectiveness of iCBT-based interventions in reducing suicidal thoughts in a general population [34,35]. Both demonstrated positive benefits of iCBT with some suggestion of changes in hopelessness [34] and depressive symptoms [35] as possible mediating mechanisms, as well as some evidence of cost-effectiveness [36]. However, effect sizes were small and Christensen's (2013) study failed to show superiority of iCBT with or without telephone call back compared to telephone call back alone or treatment-as-usual controls [35]. iCBT may be effective in reducing suicidal ideation in a clinically depressed population, although the level of evidence is less robust in view of the lack of controlled studies [37]. Whether such positive benefits on suicidal ideation can be sustained over time warrants longer-term studies. Well-controlled studies examining outcome after intervention for severe suicidal behavior, such as attempted or completed suicide, do not exist, which highlights the need for such studies in the future.

Nevertheless, there is some preliminary evidence from descriptive reports that the Internet may serve as a good step for individuals with suicidal thoughts to gain better access to face-to-face mental health care services [38,39,45]. The anonymous nature of Internet communication is likely to help reduce barriers in accessing mental health services due to stigma, particularly among adolescents, celebrities [23], and medical personnel [39].

The online suicide prevention approaches reviewed were automated (such as in iCBT), informative, or interactive in nature and mostly comprised a combination of the above. The quality of information, as well as the evidence base for their benefit, appear to vary between programs. For example, evaluated the quality of information on such resource sites and found that less than half of the information on the Canadian suicide prevention websites was truly evidence-based [41]. The target groups for most of the suicide prevention programs were also not often specified. A notable exception was the program outlined by Feigelman et al (2008). It described the demographic and loss-related characteristics of parents in an Internet suicide survivor support group [40].

Interactive approaches of Web-based suicide prevention often target an individual or alternatively lead to the formation of a support group. The means to reach an individual can be via email [23,38,39,43,44], chat software [23,44], or advertisements on websites and in newspapers [23]. Creation of an online support group is often achieved through online forums [23], chat rooms [44], social networking [45], or bulletin boards [42]. User feedback in the published studies indicate that, in safe and successful programs, the support group is created through supervised chat rooms or through a forum involving trained volunteers [42]. In the absence of such supervision, chat rooms or forums are thought to be potentially harmful [33].

The appealing characteristics of Web-based suicide prevention strategies (ie, ease of accessibility, greater degree of anonymity, and component of text-based communication) give rise to the benefits of promoting help-seeking behavior and potentially accessing hard-to-reach at-risk groups [23,33,39,43]. However, those characteristics also pose some potential challenges such as the risk of breach of confidentiality [43], inappropriate counseling derived from cultural differences [33], hostile comments that affect at-risk individuals [33], fake messages that accelerate burnout among volunteers in online groups [23], and safety concerns in an acute crisis [33,38]. Therefore, Miller (2009) suggested that the development of consumer guidelines regarding the safer use of the Internet is paramount [33]. However, systematic research confirming these challenges is still lacking (Table 1).

There is also general agreement [21,32] that the evaluation of Web-based suicide prevention strategies has many challenges. The anonymity of the participants makes valid sampling and measurement outcome more difficult. Therefore, more robust design in fully utilizing the benefits and managing the challenges of the characteristics of the Internet as a communication tool is needed for future study. Further large-scale systematic research is needed to confirm the evidence base for the effectiveness as well as the risk-benefit ratio of these Web-based strategies. According to Boyce (2010), future research also should aim "to understand the path that people with suicidal thoughts travel online, and to work out when and how to intervene" (p. 1890). He further recommended "a mixture of both quantitative and qualitative methods, examining both overall online trends and individual views" (p. 1890) [22]. Efficacy and cost effectiveness of Web-based intervention to reduce suicidality, including comparison between Web-based, telephone hotlines, and face-to-face interventions are other areas for future research.

Limitations

This review has several limitations. The number of publications available and included was small and mostly non-RCT in nature. Furthermore, the pooled information was difficult to evaluate due to the heterogeneous nature of the population, differences in design, and measurement methods. A more detailed search strategy that included studies identifying interventions for depression with suicidal outcomes that were not explicitly mentioned in the title or abstract would have been more comprehensive. In addition, non-English publications were not included in the review. Other publication biases included the following search omissions: grey literature, and contact with experts in the field to find additional unpublished studies. As mentioned, most of the studies were uncontrolled and unblinded, therefore subject to the risk of observer bias in the reported outcomes. In the case of RCTs, blinding of participants was not possible due to the nature of the active intervention delivered, that is, iCBT. Hence, there might be a tendency towards more positive outcomes based on the self-report of participants receiving iCBT due to favorable expectations towards receiving active intervention versus treatment-as-usual control conditions.

Conclusions

In conclusion, there appears to be preliminary evidence that suggests the probable benefit of Internet-based suicide prevention strategies as novel and cost-effective interventions. In particular, iCBT may serve as a means to reduce suicidal ideation. Web-based approaches seem to be advantageous in terms of potentially reaching out to populations at-risk for suicidal behavior for whom conventional methods have limited accessibility due to stigma, physical or psychological limitations, or geographical location. Future research is needed to elucidate the impact of and strategies to overcome the potential challenges such as issues of confidentiality, safety concerns in acute crisis, and feasibility of resources, among others. Findings from such studies would further clarify the risk-benefit ratio in terms of implementation of Web-based suicide prevention strategies as potential adjuncts or alternatives to mainstream interventions in the future.

Acknowledgments

We would like to thank Ms Nur Ahlina Abdullah for her administrative assistance in the preparation of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Web-based suicide prevention strategies—summary of the included articles.

[PDF File (Adobe PDF File), 46KB - jmir_v16i1e30_app1.pdf]

References

1. WHO Media Centre. Suicide risk high for young people. 2009. URL: http://www.who.int/mediacentre/multimedia/podcasts/2009/suicide_prevention_20090915/en/ [accessed 2013-12-22] [WebCite Cache ID 6M3MEvYtJ]
2. Internet World Stats. World Internet Users and Population Statistics. 2012. URL: http://www.Internetworldstats.com/stats.htm [accessed 2013-12-23] [WebCite Cache ID 6M4KjrF5l]
3. Mehlum L. The Internet, Suicide, and Suicide Prevention. Crisis: The Journal of Crisis Intervention and Suicide Prevention 2006 Sep 1;21(4):186-188. [doi: 10.1027//0227-5910.21.4.186]
4. Alao AO, Soderberg M, Pohl EL, Alao AL. Cybersuicide: review of the role of the Internet on suicide. Cyberpsychol Behav 2006 Aug;9(4):489-493. [doi: 10.1089/cpb.2006.9.489] [Medline: 16901252]
5. Tam J, Tang WS, Fernando DJ. The Internet and suicide: A double-edged tool. Eur J Intern Med 2007 Oct;18(6):453-455. [doi: 10.1016/j.ejim.2007.04.009] [Medline: 17822655]
6. Biddle L, Donovan J, Hawton K, Kapur N, Gunnell D. Suicide and the Internet. BMJ 2008 Apr 12;336(7648):800-802 [FREE Full text] [doi: 10.1136/bmj.39525.442674.AD] [Medline: 18403541]
7. Recupero PR, Harms SE, Noble JM. Googling Suicide. J Clin Psychiatry 2008 Jun 15;69(6):878-888. [doi: 10.4088/JCP.v69n0601]
8. Durkee T, Hadlaczky G, Westerlund M, Carli V. Internet pathways in suicidality: a review of the evidence. Int J Environ Res Public Health 2011 Oct;8(10):3938-3952 [FREE Full text] [doi: 10.3390/ijerph8103938] [Medline: 22073021]
9. Collings S, Niederkrotenthaler T. Suicide prevention and emergent media: surfing the opportunity. Crisis 2012 Jan 1;33(1):1-4 [FREE Full text] [doi: 10.1027/0227-5910/a000141] [Medline: 22357378]
10. Gould MS. Suicide and the media. Ann NY Acad Sci 2001;932(1):200-224. [doi: 10.1111/j.1749-6632.2001.tb05807.x]
11. Niederkrotenthaler T, Sonneck G. Assessing the impact of media guidelines for reporting on suicides in Austria: interrupted time series analysis. Aust N Z J Psychiatry 2007 May;41(5):419-428. [doi: 10.1080/00048670701266680]
12. Insel BJ, Gould MS. Impact of modeling on adolescent suicidal behavior. Psychiatr Clin North Am 2008 Jun;31(2):293-316. [doi: 10.1016/j.psc.2008.01.007] [Medline: 18439450]
13. Niederkrotenthaler T, Voracek M, Herberth A, Till B, Strauss M, Etzersdorfer E, et al. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. Br J Psychiatry 2010 Sep;197(3):234-243 [FREE Full text] [doi: 10.1192/bjp.bp.109.074633] [Medline: 20807970]
14. Sisask M, Värnik A. Media roles in suicide prevention: a systematic review. Int J Environ Res Public Health 2012 Jan;9(1):123-138 [FREE Full text] [doi: 10.3390/ijerph9010123] [Medline: 22470283]
15. Lee DT, Chan KP, Yip PS. Charcoal burning is also popular for suicide pacts made on the Internet. BMJ 2005 Mar 12;330(7491):602 [FREE Full text] [doi: 10.1136/bmj.330.7491.602-b] [Medline: 15761009]
16. Becker K, Mayer M, Nagenborg M, El-Faddagh M, Schmidt MH. Parasuicide online: Can suicide websites trigger suicidal behaviour in predisposed adolescents? Nord J Psychiatry 2004;58(2):111-114. [doi: 10.1080/08039480410005602] [Medline: 15204216]
17. Rajagopal S. Suicide pacts and the Internet. BMJ 2004 Dec 4;329(7478):1298-1299 [FREE Full text] [doi: 10.1136/bmj.329.7478.1298] [Medline: 15576715]
18. Crump A. Suicide in Japan. The Lancet 2006 Apr 14;367(9517):1143-1143. [doi: 10.1016/S0140-6736(06)68502-4]
19. Naito A. Internet Suicide in Japan: Implications for Child and Adolescent Mental Health. Clinical Child Psychology and Psychiatry 2007 Oct 01;12(4):583-597. [doi: 10.1177/1359104507080990]

20. Becker K, Schmidt MH. When Kids Seek Help On-Line: Internet Chat Rooms and Suicide. Reclaiming Children and Youth: The Journal of Strength-based Interventions 2005;13(4):229-230.
21. Luxton DD, June JD, Fairall JM. Social media and suicide: a public health perspective. Am J Public Health 2012 May;102 Suppl 2:S195-S200. [doi: 10.2105/AJPH.2011.300608] [Medline: 22401525]
22. Boyce N. Pilots of the future: suicide prevention and the Internet. The Lancet 2010 Dec;376(9756):1889-1890. [doi: 10.1016/S0140-6736(10)62199-X]
23. Barak A. Emotional support and suicide prevention through the Internet: A field project report. Computers in Human Behavior 2007 Mar;23(2):971-984. [doi: 10.1016/j.chb.2005.08.001]
24. Hemelrijk E, van Ballegooijen W, Donker T, van Straten A, Kerkhof A. Internet-based screening for suicidal ideation in common mental disorders. Crisis 2012;33(4):215-221. [doi: 10.1027/0227-5910/a000142] [Medline: 22713975]
25. Gilat I, Shahar G. Emotional first aid for a suicide crisis: comparison between Telephonic hotline and Internet. Psychiatry 2007;70(1):12-18. [doi: 10.1521/psyc.2007.70.1.12] [Medline: 17492908]
26. Gould MS, Munfakh JL, Lubell K, Kleinman M, Parker S. Seeking help from the Internet during adolescence. J Am Acad Child Adolesc Psychiatry 2002 Oct;41(10):1182-1189. [doi: 10.1097/00004583-200210000-00007] [Medline: 12364839]
27. Mohd Daud T. [Unpublished dissertation for the degree of Master of Medicine(Psychiatry)]. In: Help seeking behaviour using the Internet: a study among sixteen year old Form Four secondary school students from high risk schools in Kuala Lumpur. Kuala Lumpur: Universiti Kebangsaan Malaysia; 2005.
28. Westerlund M, Hadlaczky G, Wasserman D. The representation of suicide on the Internet: implications for clinicians. J Med Internet Res 2012;14(5):e122 [FREE Full text] [doi: 10.2196/jmir.1979] [Medline: 23010086]
29. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the Internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
30. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
31. Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. 2004 URL: http://www.ihe.ca/documents/HTA-FR13.pdf [accessed 2013-12-23] [WebCite Cache ID 6M4O9E3NK]
32. Krysinska K, De Leo D. Telecommunication and Suicide Prevention: Hopes and Challenges for the New Century. OMEGA: The Journal of Death and Dying 2007 Sep 1;55(3):237-253. [doi: 10.2190/OM.55.3.e]
33. Miller K. The potential role of the Internet in suicide prevention. Counselling, Psychotherapy, and Health 2009;5(1):109-130.
34. van Spijker BA. Reducing the burden of suicidal thoughts through online self-help. VU University: Amsterdam; 2012.
35. Christensen H, Farrer L, Batterham PJ, Mackinnon A, Griffiths KM, Donker T. The effect of a Web-based depression intervention on suicide ideation: secondary outcome from a randomised controlled trial in a helpline. BMJ Open 2013;3(6):e002886 [FREE Full text] [doi: 10.1136/bmjopen-2013-002886] [Medline: 23811172]
36. van Spijker BA, Majo MC, Smit F, van Straten A, Kerkhof AJ. Reducing suicidal ideation: cost-effectiveness analysis of a randomized controlled trial of unguided Web-based self-help. J Med Internet Res 2012;14(5):e141 [FREE Full text] [doi: 10.2196/jmir.1966] [Medline: 23103835]
37. Watts S, Newby JM, Mewton L, Andrews G. A clinical audit of changes in suicide ideas with Internet treatment for depression. BMJ Open 2012;2(5):e001558 [FREE Full text] [doi: 10.1136/bmjopen-2012-001558] [Medline: 22983787]
38. Haas A, Koestner B, Rosenberg J, Moore D, Garlow SJ, Sedway J, et al. An interactive Web-based method of outreach to college students at risk for suicide. J Am Coll Health 2008;57(1):15-22. [doi: 10.3200/JACH.57.1.15-22] [Medline: 18682341]
39. Moutier C, Norcross W, Jong P, Norman M, Kirby B, McGuire T, et al. The suicide prevention and depression awareness program at the University of California, San Diego School of Medicine. Acad Med 2012 Mar;87(3):320-326. [doi: 10.1097/ACM.0b013e31824451ad] [Medline: 22373625]
40. Feigelman W, Gorman BS, Beal KC, Jordan JR. Internet support groups for suicide survivors: a new mode for gaining bereavement assistance. Omega (Westport) 2008;57(3):217-243. [Medline: 18837172]
41. Szumilas M, Kutcher S. Teen suicide information on the Internet: a systematic analysis of quality. Can J Psychiatry 2009 Sep;54(9):596-604. [Medline: 19751548]
42. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. British Journal of Guidance & Counselling 2010 May;38(2):191-204. [doi: 10.1080/03069881003600991]
43. Lester D. The Use of the Internet for Counseling the Suicidal Individual: Possibilities and Drawbacks. OMEGA--Journal of Death and Dying 2009 Jan 14;58(3):233-250. [doi: 10.2190/OM.58.3.e]
44. Gilat I, Shahar G. Suicide prevention by online support groups: an action theory-based model of emotional first aid. Arch Suicide Res 2009;13(1):52-63. [doi: 10.1080/13811110802572148] [Medline: 19123109]
45. Manning J, Vandeusen K. Suicide prevention in the dot com era: technological aspects of a university suicide prevention program. J Am Coll Health 2011;59(5):431-433. [doi: 10.1080/07448480903540507] [Medline: 21500064]

Abbreviations

CBT: cognitive behavioral therapy
CINAHL: Cumulative Index to Nursing and Allied Health Literature
iCBT: Internet cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 23.09.13; peer-reviewed by T Niederkrotenthaler, H Christensen, C Sinclair; comments to author 30.10.13; revised version received 23.12.13; accepted 09.01.14; published 28.01.14.

Please cite as:
Lai MH, Maniam T, Chan LF, Ravindran AV
Caught in the Web: A Review of Web-Based Suicide Prevention
J Med Internet Res 2014;16(1):e30
URL: http://www.jmir.org/2014/1/e30/
doi:10.2196/jmir.2973
PMID:24472876

Original Paper

Importance of Internet Surveillance in Public Health Emergency Control and Prevention: Evidence From a Digital Epidemiologic Study During Avian Influenza A H7N9 Outbreaks

Hua Gu[1*], MPH; Bin Chen[1*], MSc; Honghong Zhu[2], MD, PhD; Tao Jiang[1], BSc; Xinyi Wang[1], MSc; Lei Chen[1], BSc; Zhenggang Jiang[1], MSc; Dawei Zheng[1], BSc; Jianmin Jiang[1], MD, PhD

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
[2]Department of Public Health, College of Health and Human Services, Western Kentucky University, Bowling Green, KY, United States
[*]these authors contributed equally

Corresponding Author:
Jianmin Jiang, MD, PhD
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Rd
Binjiang District
Hangzhou, 310051
China
Phone: 86 571 87115009
Fax: 86 571 87115009
Email: jmjiang@cdc.zj.cn

Abstract

Background: Outbreaks of human infection with a new avian influenza A H7N9 virus occurred in China in the spring of 2013. Control and prevention of a new human infectious disease outbreak can be strongly affected by public reaction and social impact through the Internet and social media.

Objective: This study aimed to investigate the potential roles of Internet surveillance in control and prevention of the human H7N9 outbreaks.

Methods: Official data for the human H7N9 outbreaks were collected via the China National Health and Family Planning Committee website from March 31 to April 24, 2013. We obtained daily posted and forwarded number of blogs for the keyword “H7N9” from Sina microblog website and a daily Baidu Attention Index (BAI) from Baidu website, which reflected public attention to the outbreak. Rumors identified and confirmed by the authorities were collected from Baidu search engine.

Results: Both daily posted and forwarded number and BAI for keyword H7N9 increased quickly during the first 3 days of the outbreaks and remained at a high level for 5 days. The total daily posted and forwarded number for H7N9 on Sina microblog peaked at 850,000 on April 3, from zero blogs before March 31, increasing to 97,726 on April 1 and to 370,607 on April 2, and remaining above 500,000 from April 5-8 before declining to 208,524 on April 12. The total daily BAI showed a similar pattern of change to the total daily posted and forwarded number over time from March 31 to April 12. When the outbreak locations spread, especially into other areas of the same province/city and the capital, Beijing, daily posted and forwarded number and BAI increased again to a peak at 368,500 and 116,911, respectively. The median daily BAI during the studied 25 days was significantly higher among the 7 provinces/cities with reported human H7N9 cases than the 2 provinces without any cases ($P<.001$). So were the median daily posted and forwarded number and daily BAI in each province/city except Anhui province. We retrieved a total of 32 confirmed rumors spread across 19 provinces/cities in China. In all, 84% (27/32) of rumors were disseminated and transmitted by social media.

Conclusions: The first 3 days of an epidemic is a critical period for the authorities to take appropriate action through Internet surveillance to prevent and control the epidemic, including preparation of personnel, technology, and other resources; information release; collection of public opinion and reaction; and clarification, prevention, and control of rumors. Internet surveillance can be used as an efficient and economical tool to prevent and control public health emergencies, such as H7N9 outbreaks.

(J Med Internet Res 2014;16(1):e20) doi:10.2196/jmir.2911

KEYWORDS

influenza A virus, H7N9 subtype; Internet; surveillance; disease outbreak

Introduction

Human infections with new avian influenza A (H7N9) virus have been reported in Shanghai, Anhui, Zhejiang, and Jiangsu provinces in China beginning March 31, 2013 [1]. A total of 108 cases confirmed with DNA sequencing and 23 deaths of H7N9 infections have been reported until April 24, 2013. This is the first time that human infections with H7N9 virus have been found and reported in the world [1]. Cases were mainly distributed in provinces in the southeast of China. Almost 64% cases had exposure to animals or visited a live poultry market [2]. Clinical symptoms and signs were of sudden onset, including respiratory symptoms in the early stage, high fever (≥ 38°C) and cough, dyspnea after 5 to 7 days onset, and then severe progressive pneumonia. Some cases could rapidly develop acute respiratory distress syndrome and die [2,3]. This outbreak has received great public attention from all over the world. Its progress has been continuously reported by various media in China.

The Internet has been developed and popularized quickly in China in the past decade. According to the China Internet Development Statistical Report, there were 538 million Internet users (39.9% of the whole population) and 388 million cell phone Internet users by June 30, 2012 [4]. Internet users in China spend, on average, nearly 20 hours per week online [4]. They now have access to news and information through more channels including microblogs and social networking websites. In addition, they can expand news coverage by sharing and forwarding information online. Now there are millions of social media users and the number is increasing [4]. Sina microblog has the largest number of users in China, approximately 368 million users at the end of 2012. Web search engines are an important tool for Internet users to acquire information. In China, there are approximately 429 million users using search engine websites, among which Baidu website has the largest group of users [4]. In 2011, 77.2% Internet users preferred Baidu website as their first search engine [5]. The power of Internet users in China is becoming stronger, especially in public opinion leadership with increases in participation for big events. Reactions from Internet users can roughly represent public reactions to an event. Developed Internet and social media provide an excellent chance to understand public reactions to and clearly communicate about public health emergencies, including preparation, prevention, and control of emergent infectious disease outbreaks [6-8].

In recent years, China has experienced increasing outbreaks of emerging infectious diseases. The severe acute respiratory syndrome (SARS) outbreaks, epidemics, and pandemics in 2003 and influenza A H1N1 outbreaks in 2009 have led to strong repercussions among the public and media. The public and media reactions during an outbreak period are a double-edged sword. Moderate reaction may arouse individuals’ awareness of disease control and prevention, whereas overreaction could play a negative role in the population. During the period of SARS pandemics, there were network media information chaos, rumors, and public panic, including overbuying of medical-related products and the soaring prices of these products, which destabilized the society and caused many problems against disease control and prevention [9]. Control and prevention of emerging infectious disease outbreaks require public participation. How Internet users react to an emergent disease outbreak and how we can lead them to a proper response are extremely important for control and prevention of the disease in the population. During the H7N9 outbreak, we designed a study to investigate the potential roles of Internet surveillance on reactions of Internet users to the H7N9 outbreaks and provide evidence for government, health authorities, and the public to efficiently control and prevent public health emergency problems, such as H7N9, in the future.

Methods

Study Design and Population

Overview

We designed and conducted an exploratory digital epidemiologic study of public reaction to the outbreaks of human infection with H7N9 virus from March 31 to April 24, 2013. Our study population was Internet users in China during the study period of Avian Influenza A H7N9 outbreaks.

Case Sources

The first human case with H7N9 virus infection was reported by the China National Health and Family Planning Committee (CNHFPC) on March 31, 2013. We collected and analyzed data from the first 25 days of the outbreak (March 31 to April 24, 2013) as our study period because the government stopped daily case reports after April 24, 2013. All the reported cases and deaths were obtained from the website of the CNHFPC [10]. All cases were confirmed by DNA sequencing.

Internet Surveillance of Public Reactions

Since the H7N9 outbreak, we initiated Internet surveillance of Internet users’ reactions to it. Crowdsourcing [11] was used as an approach to collect data from millions of Internet users in Sina microblog and the Baidu website. Using the keyword “H7N9,” we collected the daily tally of relevant blogs posted and forwarded (daily posted and forwarded number) in Sina microblog from March 31 to April 24, 2013. We also obtained the daily posted and forwarded number for H7N9 in 9 provinces/cities, of which 7 provinces/cities (Shanghai, Anhui, Jiangsu, Zhejiang, Beijing, Henan, and Shandong) had reported H7N9 cases and 2 provinces (Hubei and Shaanxi) had no cases. The daily posted and forwarded number can be used as an estimate for the degree of attention to an event from the public. The daily posted and forwarded numbers in the 2 provinces without cases were used as a base/reference. These numbers were officially released by Sina microblog website [12].

The daily Baidu Attention Index (BAI) was used as another data source. The daily BAI computes the weighted sum of searching frequency for a keyword based on its daily search volume on the Baidu website [13]. The daily BAI for the keyword H7N9 was collected from the BAI webpage from March 31 to April 24, 2013 [13]. Meanwhile, the daily BAI for the keyword H7N9 from the 7 provinces that reported human H7N9 cases were also collected during the study period. The daily BAI data from Hubei and Shaanxi provinces in the Midwest of China, which had no human H7N9 cases reported, were collected as a base/reference to compare the differences of public reactions with the provinces with H7N9 cases.

In addition, we used the keywords "H7N9" and "rumor" to search for rumors of H7N9 that were officially verified and confirmed by the local authorities on the Baidu website. A total of 32 rumors were collected during the study period. We then categorized the rumors based on their characteristics. All the information obtained online was in simplified Chinese language and released publicly by the websites, but no personal identification information, such as name or email address, was collected. This study was approved by the Institutional Review Board in the Zhejiang Provincial Centers for Disease Control and Prevention.

Statistical Analysis

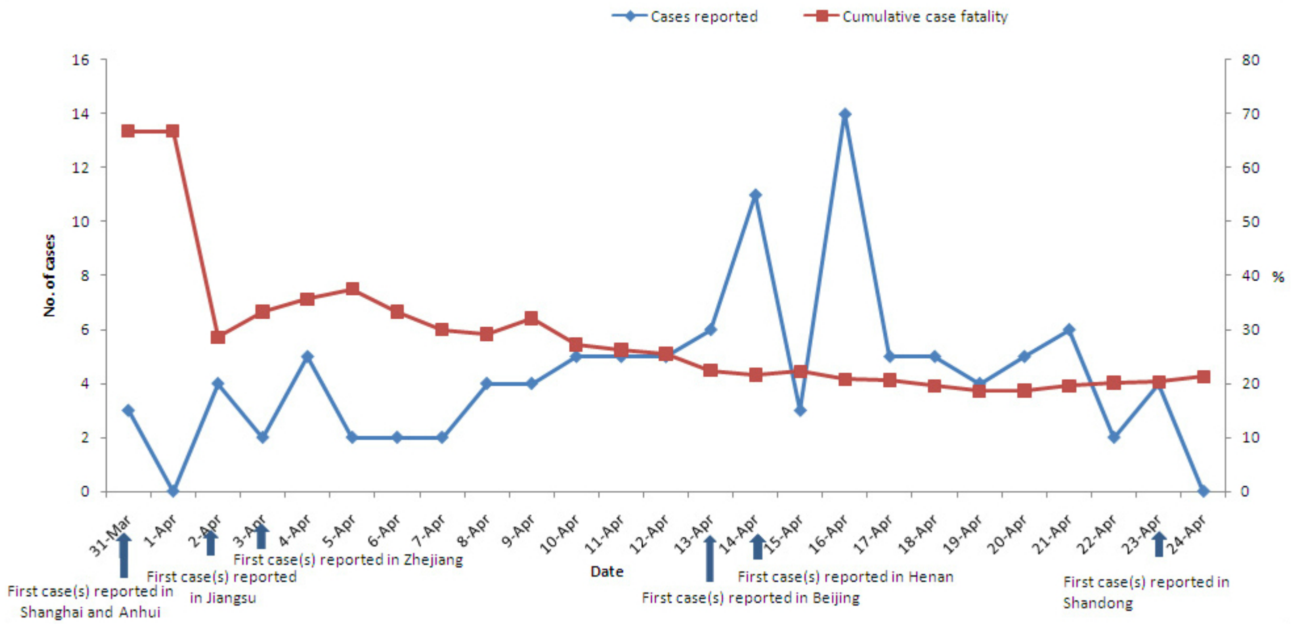
We graphed the curves of H7N9 epidemic and severity by number of daily cases, first case(s) reported in each province/city, and cumulative case fatality rate over time. The daily posted and forwarded number and BAI were graphed by date, number of daily cases, and first case(s) in each province/city, respectively, to explore public reaction changing to the H7N9 epidemic situation over time. Median (P_{50}) and range were used to describe the distributions of daily posted and forwarded number and BAI indexes by province/city. Spearman's rank correlation coefficient was used to check the relationship of epidemic trends with the daily posted and forwarded number and BAI, respectively, if the distribution of the variables is monotonic. Kruskal-Wallis test was used to test the differences in daily posted and forwarded number and BAI between provinces/cities with cases and without cases. Rumors of H7N9 outbreaks were qualitatively classified by date, province, rumor maker, media, and main content.

Results

H7N9 Epidemic Trend

On March 31, 2013, the CNHFPC reported the first human infections with the H7N9 virus, including 2 cases in Shanghai and 1 in Anhui province [10]. Subsequently, cases were successively reported in Jiangsu, Zhejiang, Beijing, Henan, and other provinces. Until April 24, 2013, a total of 108 cases and 23 deaths (cumulative case fatality rate = 21.3%) of H7N9 infections were reported in 7 provinces/cities in China [10]. The cumulative case fatality rate was 33% (6/18) during the first week and then gradually declined to 21.3% (23/108) because of timely prevention and treatment. By April 12, 43 cases had been reported in 4 provinces in Eastern China and then the epidemic expanded to Beijing on April 13 and Henan province on April 14. The last case during the study period was reported in Shandong province on April 23, 2013 (Figure 1).

Figure 1. The human cases of avian influenza A H7N9 virus infection reported with the first case(s) by province/city and the cumulative fatality rate in China from March 31 to April 24, 2013.



Daily Posted and Forwarded Number for H7N9 on Sina Microblog

The total daily posted and forwarded number for the keyword H7N9 on Sina microblog peaked at 850,000 on April 3, 2013, from zero before March 31, increasing to 97,726 on April 1, and to 370,607 on April 2, and remaining above 500,000 from April 5 to 8 before declining to 208,524 on April 12. When the first case was reported in Capital Beijing on April 12 and the first case in Henan province on April 14, the total daily posted and forwarded number returned to a relatively high level of 368,500 on April 13 and 339,822 on April 14, declined to 262,667 on April 15 and 235,131 on April 16, and increased again to 273,752 on April 17 and 370,995 on April 18. The total daily posted and forwarded number then decreased and remained at a level of 100,000 (Figure 2). The median of the total daily posted and forwarded number was 289,325 (range 76,482-853,027). Spearman’s rank correlation coefficient analyses showed that the daily posted and forwarded number was positively associated with cumulative case fatality rate (Spearman’s rank correlation coefficient = 0.60, P=.002), that is to say, higher cumulative case fatality rate higher daily posted and forwarded number at the early stage of epidemic and lower cumulative case fatality rate lower daily posted and forwarded number later (Figure 3).

Since the first human H7N9 case reported on March 31, the observed daily posted and forwarded number in each province/city showed a similar trend to the total daily posted and forwarded number. However, the daily posted and forwarded number in Beijing on April 14 and Henan on April 15 rose to another peak when the local area first reported cases (the daily posted and forwarded number peaked at 163,512 in Beijing on April 13 and peaked at 72,512 in Henan province on April 14 as shown in Figure 4).

The observed median daily posted and forwarded number during the 25 days studied was 36,256 (range 1830-232,368) in Shanghai, 4944 (217-31,312) in Anhui, 21,424 (1521-143,376) in Jiangsu, 24,720 (897-131,840) in Zhejiang, 57,680 (7867-166,448) in Beijing, 11,536 (316-56,032) in Henan, 28,016 (513-56,032) in Shandong, 9888 (327-47,792) in Hubei, and 924 (229-26,368) in Shaanxi. The mean rank value of the daily posted and forwarded number in each province/city with human H7N9 cases reported except Anhui was significantly higher (P<.001) than that in the 2 provinces without cases (Table 1).

Table 1. The observed median daily posted and forwarded number and Baidu Attention Index (BAI) for H7N9 by province/city on Sina microblog and Baidu website from March 31 to April 24, 2013.

Province/city	Daily posted and forwarded number[a]		Daily BAI[b]	
	Median	Mean rank value	Median	Mean rank value
Shanghai	36,256	150.26	4227	144.16
Anhui	4944	70.68	1723	62.16
Jiangsu	21,424	133.96	6592	175.68
Zhejiang	24,720	134.04	6432	174.66
Beijing	57,680	163.18	4790	148.60
Henan	11,536	97.08	2731	108.82
Shandong	28,016	127.54	3065	112.96
Hubei[c]	9888	88.92	1420	51.06
Shaanxi[c]	924	51.34	1301	38.90

[a] Kruskal-Wallis test of daily posted and forwarded number: χ^2_8=67.5 (P<.001).

[b] Kruskal-Wallis test of daily BAI: χ^2_8=129.2 (P<.001).

[c] Provinces without cases reported as base/reference.

Figure 2. The reported human H7N9 cases and the daily posted and forwarded number of H7N9 discussion trends on Sina microblog in China from March 31 to April 24, 2013.

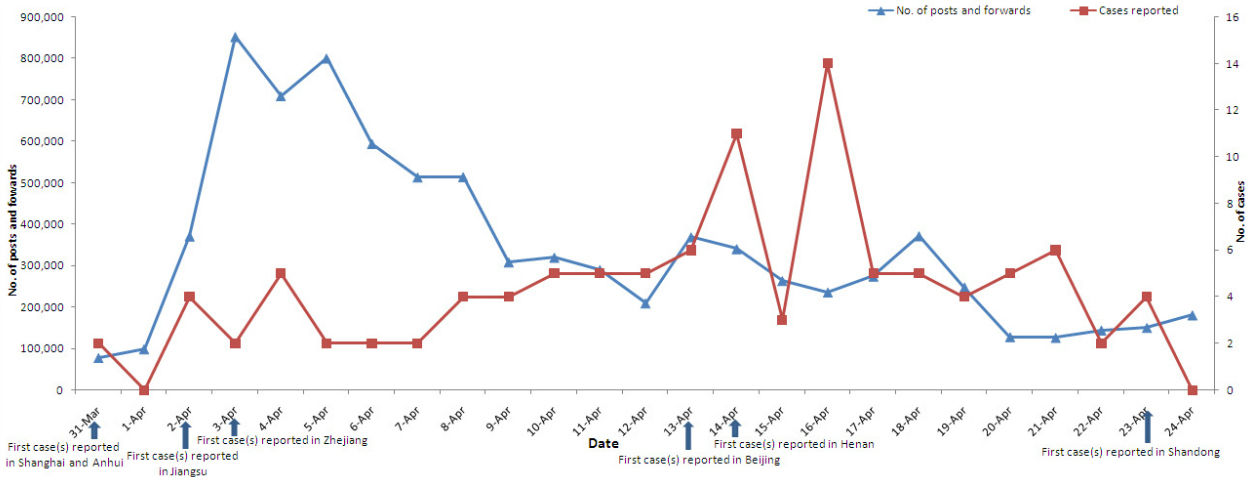


Figure 3. The daily posted and forwarded number of H7N9 discussion trends on Sina microblog positively associated with cumulative case fatality rate of human H7N9 infection in China from March 31 to April 24, 2013.

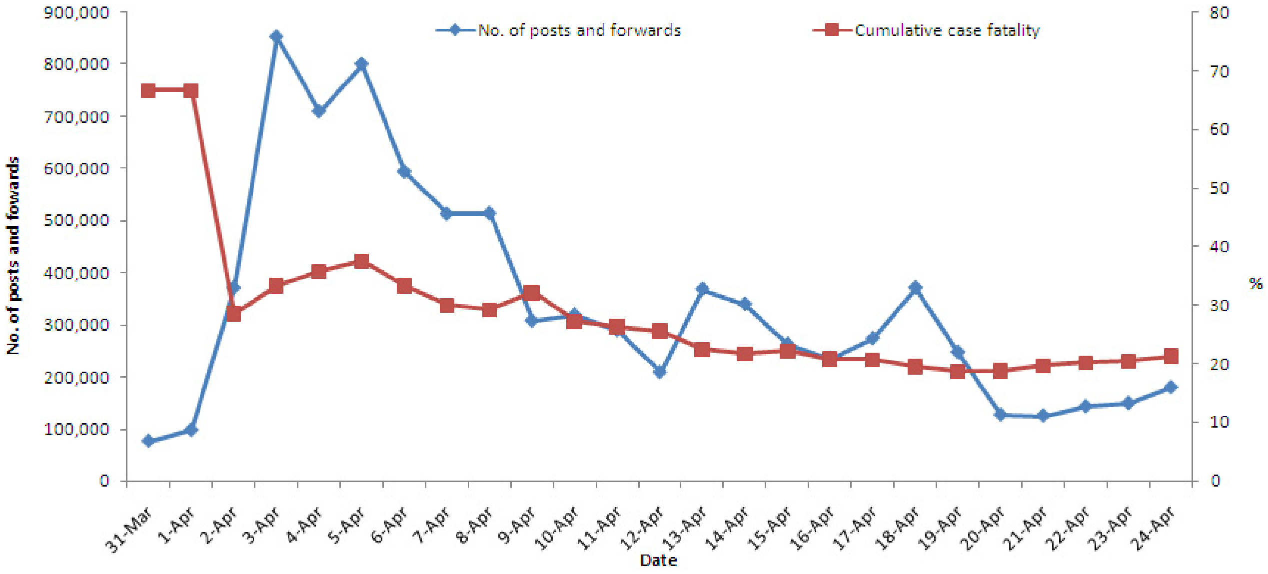
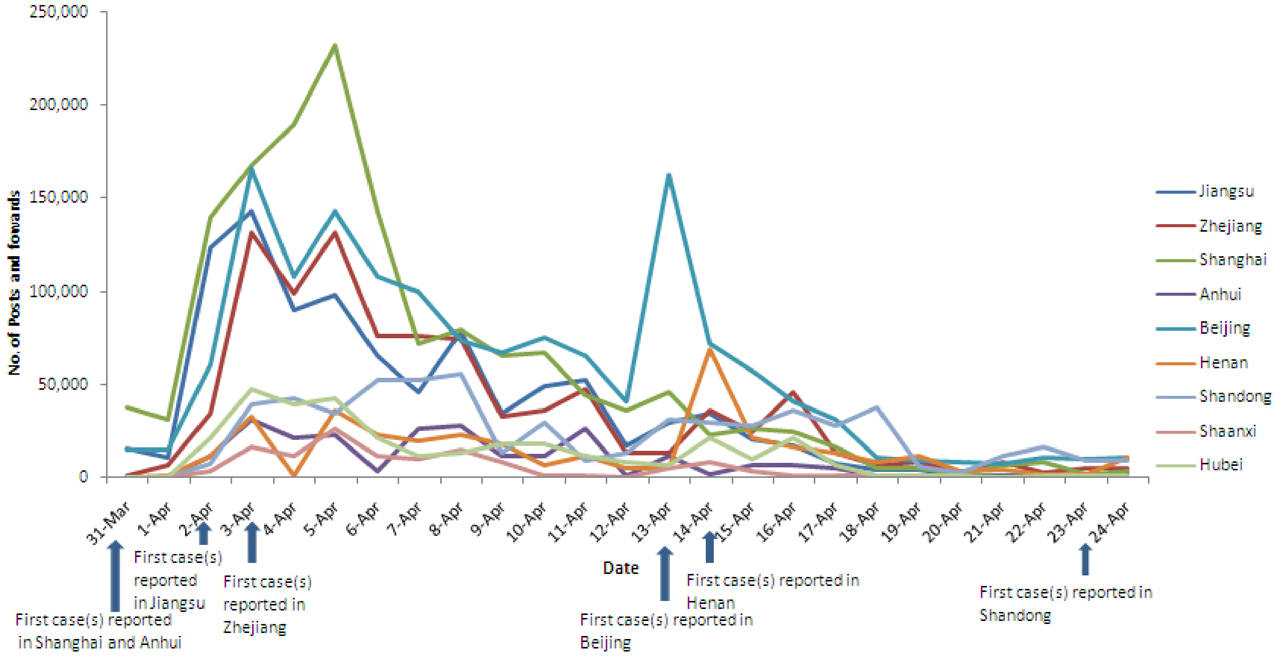


Figure 4. The daily posted and forwarded number for H7N9 on Sina microblog by province/city from March 31 to April 24, 2013.



First case(s) reported in Shanghai and Anhui

First case(s) reported in Jiangsu

First case(s) reported in Zhejiang

First case(s) reported in Beijing

First case(s) reported in Henan

First case(s) reported in Shandong

Daily Baidu Attention Index for H7N9

The daily BAI for keyword H7N9 increased sharply in the first 3 days of the H7N9 outbreak, with a peak at 126,825 on April 3. The daily BAI then declined with fluctuations, remaining at a high level above 100,000 from April 4 - 8. Following the cases reported in Beijing and Henan, the daily BAI increased again and peaked at 116,991 on April 15. Afterwards, it steadily decreased to a lower level of approximately 40,000. The median daily BAI was 79,687 (range 18,007-79,687) (Figure 5). Spearman's rank correlation coefficient analyses showed that the daily BAI was positively associated with cumulative case fatality rate (Spearman's rank correlation coefficient = 0.43, *P*=.04), that is to say, higher cumulative case fatality rate higher daily BAI at the early stage of epidemic and lower cumulative case fatality rate lower daily BAI later (Figure 6).

The daily BAI for H7N9 in each province/city increased substantially after the first human cases were reported regardless of where human cases were reported. The trend of daily BAI change in each province/city was similar to the total daily BAI. Moreover, there was also an obvious increase of the daily BAI in Beijing and Henan from April 14-15 because of the first case report in Beijing on April 13 and in Henan on April 14 and in Shandong province when the government first announced a case on April 23.

Overall, the median daily BAI was 4227 (range 0-9014) in Shanghai, 1723 (0-2765) in Anhui, 6592 (0-11,914) in Jiangsu, 6432 (0-15,011) in Zhejiang, 4790 (0-9488) in Beijing, 2731

(0-7592) in Henan, 3065 (0-6763) in Shandong, 1420 (0-2707) in Hubei, and 1301 (0-2198) in Shaanxi. The mean rank values of daily BAI in the provinces with reported cases were significantly higher than that in the provinces without cases (*P*<.001; Table 1).

Figure 5. The reported human H7N9 cases and the daily Baidu Attention Index (BAI) of H7N9 discussion trends on Baidu website in China from March 31 to April 24, 2013.

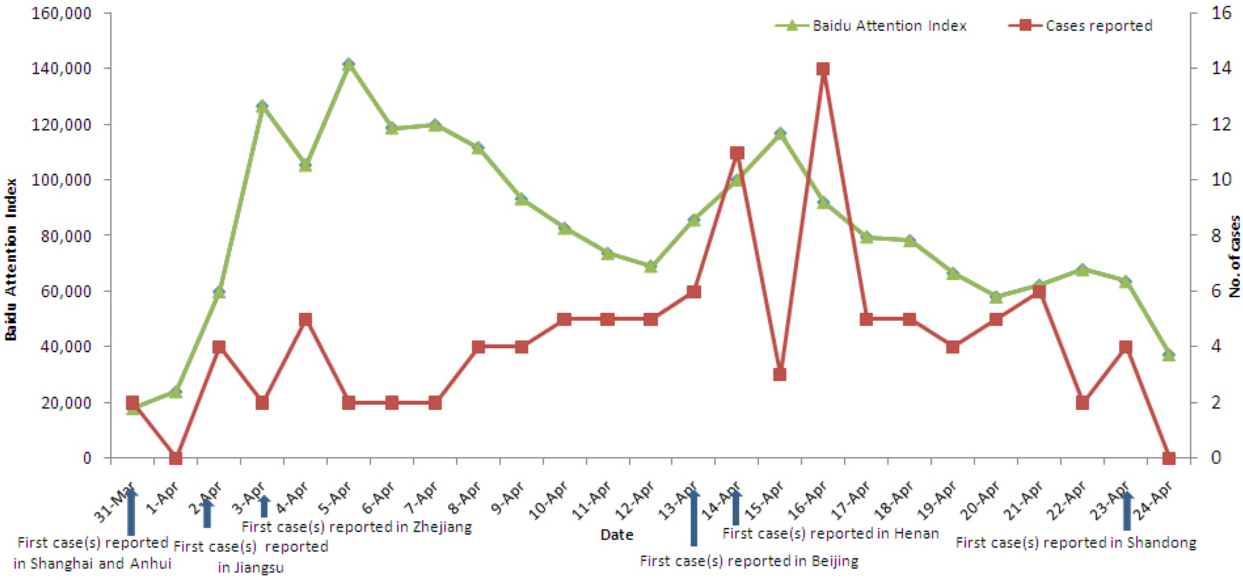


Figure 6. The daily Baidu Attention Index (BAI) of H7N9 discussion trends on Baidu website positively associated with cumulative case fatality rate of human H7N9 infection in China from March 31 to April 24, 2013.

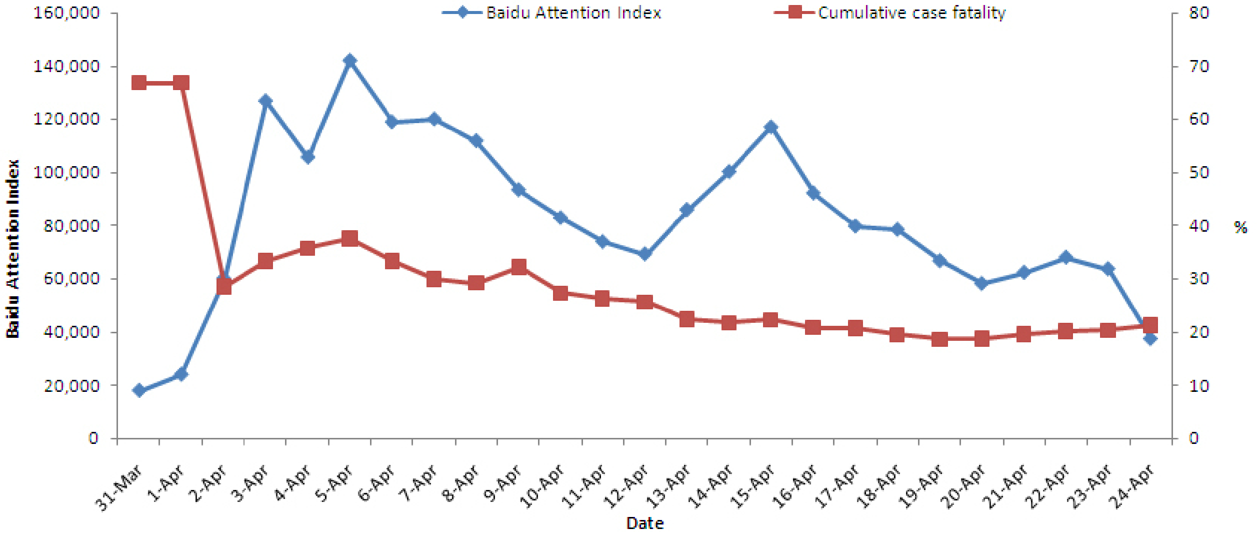
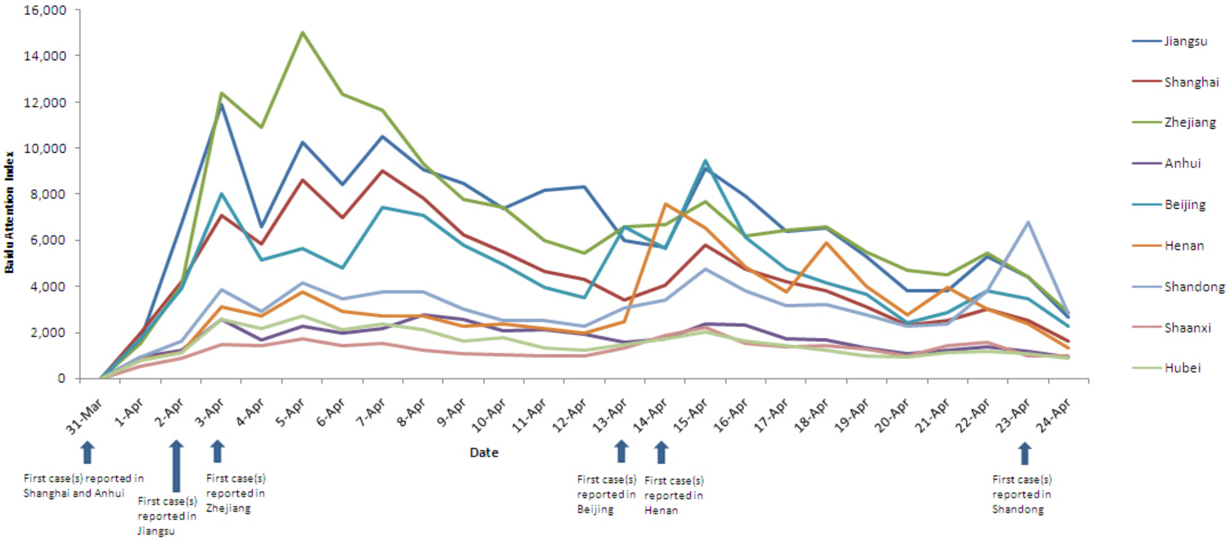


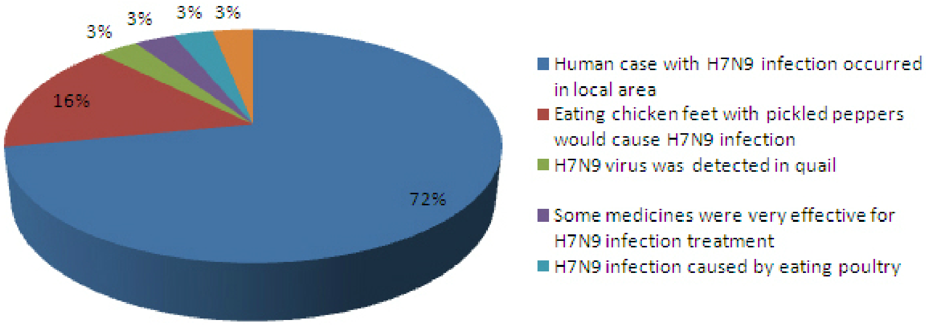
Figure 7. Daily Baidu Attention Index (BAI) by province/city in China from March 31 to April 24, 2013.



Rumors of H7N9 Outbreaks

All 32 rumors were disseminated by 58 people and distributed in 19 provinces in China. In all, 84% (27/32) rumors were disseminated through Sina microblog, QQ (the most popular Chinese communication software), and Wechat (an instant mobile communication software), and 16% (5/32) were spread via bulletin board system (BBS), cell phone message, and telephone calls. Of these 27 rumors, 72% (23/32) disseminated information that human infections with H7N9 virus occurred in the local areas, and 16% declared that eating chicken feet with pickled peppers (a popular Chinese food) would cause H7N9 infection (Figure 8). Among 15 rumor investigations that released the characteristics of the rumor makers, 5 were students; 5 were migrant workers; and 5 were farmers and company employees. Four rumor investigations announced the aims for making and disseminating rumors were to obtain more attention or get more visits to their individual forum or microblog. Some rumors had very strong transmissibility. The rumor that “there was a suspected human case with H7N9 virus infection in a local school” disseminated by a student from Gansu province, Western China, was forwarded by 34,000 people. The rumor that “eating chicken feet with pickled peppers would cause H7N9 infection” was spread across 5 provinces in China.

Figure 8. A total of 32 confirmed rumors for H7N9 outbreaks classified by main content during the H7N9 outbreaks in China from March 31 to April 24, 2013.



Discussion

Principal Findings

Our study used Internet surveillance to investigate the public reaction to the H7N9 outbreaks over time during the first 25 days of the H7N9 epidemic in China in the spring of 2013. The indexes of public reaction, daily posted and forwarded number and BAI, to H7N9 outbreaks were significantly higher in 7 provinces/cities with human H7N9 cases than those in 2 provinces without cases. Both daily posted and forwarded number and BAI were positively associated with the cumulative case fatality rate of human H7N9 infection. Our findings indicate that the first 3 days of an epidemic is a critical period for the authorities to take appropriate responsible action through Internet surveillance to control and prevent the epidemic, including preparation of personnel, technology, and other resources; information release; collection of public opinion and reaction; and clarification, prevention, and control of rumors. Internet surveillance can be used as an efficient and economical tool to control and prevent public health emergencies, such as H7N9 outbreaks.

This is the first study that investigated the importance of infoveillance [7,14-17] methods in monitoring public reaction to the H7N9 avian influenza outbreak on Sina microblog and Baidu search engines and to have collected rumors through the Internet in China. When an emerging infectious disease outbreak occurs, it is important for risk communicators in the government to clearly know the public concern [18]. In the past, to obtain information on the public, community, and media reactions or concern, most researchers used questionnaires and telephone interviews after the outbreak [19-23]. In contrast, network monitoring can be timelier to know public reactions to the outbreak because it is a real-time surveillance during the emergencies and also it is very economical without consuming too many resources [6,7,13,16,24,25]. During this human H7N9 avian influenza outbreak, we obtained the information by Internet monitoring and provided timely policy suggestions to the governments and related shareholders who play a role in disease outbreak control and prevention.

Emerging infectious diseases usually have the features of unknown pathogens and high mortality, no vaccine or effective treatment available, and quick and wide dissemination that trigger high public and media attention [18,26]. As soon as the government announced the outbreak, the public would have concern about the event and be eager to find relevant information to obtain a certain sense of safety [22,25]. Our study suggested that the first 3 days of the outbreak was the period when Internet attentions rose rapidly. At this point, the government needs to timely and transparently release the information, including personal protective methods of the disease, which can ease public tension quickly [18]. During this human H7N9 outbreak, when the first human case was announced, the authorities timely released case numbers and disease prevention knowledge online, which may be one of the key reasons for the early decline of public attention. However, the high attention to this outbreak lasted 3 - 5 days, which is a good time for the Department of Emergency Control and Prevention to disseminate health education knowledge and skills through network and social media.

Some studies have shown that the public concern with the H1N1 influenza infection was associated with the number of reported cases and hospitalizations by the governments [22,23]. However, other studies reported that although the number of flu cases rose, the tweets about the H1N1 influenza showed a downward trend [16]. This fact reflected that public concern or interests for the outbreak might be weakened along with the progress of the epidemic as information saturation sets in [16,27]. Our study showed public reaction was positively associated with the cumulative case fatality rate but not daily number of cases. When case fatality rate was high, the public attention was high. The public attention decreased with a decrease in case fatality rate. This indicates that the public attention decreases when the threat from the epidemic disease decreases. Our study showed a decreasing trend of Internet user's attention during the middle and later period. However, when there were new cases reported in a new province/city, the daily posted and forwarded number and BAI increased significantly, especially in the local province/city. This phenomenon indicates that the extension of epidemic area triggered more public attention than the increase of reported cases in the original outbreak areas. It is necessary to monitor the public reaction and take specific responses at that moment.

Rumors generate and occur early during the period of public health emergency [28-30]. The rumors of human H7N9 outbreaks collected in this study covered most provinces in China, and also spread quickly. Most of the rumors claimed that H7N9 avian influenza cases occurred in the local areas. The uncertainty of emerging infectious diseases often generates the rumors, which also suggests the importance of early information release [19,25,31-32]. In this new age of social media, microblogs are easy to generate and spread rumors, yet they could also be used to clarify wrong information and stop the dissemination.

Limitations

There were some limitations in this study. The daily posted and forwarded number and BAI were collected from the Internet and could not be standardized because of lack of exact numbers of daily Internet users. Thus, the mean rank of daily posted and forwarded number and BAI only provide relative comparisons between provinces/cities with and without H7N9 cases, but there is not much meaning to comparisons among individual provinces/cities. Our study was limited to the first 25 days of the outbreak because the government stopped daily case reports, instead using weekly case reports; therefore, this might not thoroughly reflect public reaction after the first 25 days. Our study primarily focused on Chinese websites, which limits the representativeness of the public reaction from Internet users on English and other language websites. Our study is an exploratory study that was designed to use a quick way to obtain public reaction during the study period of a public health emergency. The qualitative research on rumors was simple and did not use frameworks to analyze because of the diversity of the information released on the Internet. The results of our study could be applicable to public health emergency response, but may not be directly applied to public health practice yet. Further studies with advanced methods or study designs are needed.

Conclusions

This exploratory digital epidemiologic study provides evidence that Internet surveillance is a rapid and efficient way to evaluate public reactions, which will help prepare the government, health authorities, and the public to respond to public health emergencies. Our findings showed the public reaction increased quickly during the early stage of the outbreak. The first 3 days could be a "golden 3 days" for the government and public health authorities to release information in time and make the information transparent and open. The high public attention would last for a week; thus, this week could be a critical period for health education. Expansion of the disease into other areas has more impact on public attention than the number of cases reported in the same outbreak area. This implies that extra methods should be taken when the epidemic area is expanded. Rumors are always hard to avoid, but using media, such as releasing information via an official microblog, could be effective to quickly clarify rumors and prevent their further spread.

Acknowledgments

This study was supported by Zhejiang Leading Team of Science and Technology Innovation (2011R50021). We would like to thank the members in the Department of Infectious Disease in the Zhejiang Centers for Disease Control and Prevention (CDC) for their advice.

Conflicts of Interest

None declared.

References

1. Gao R, Cao B, Hu Y, Feng Z, Wang D, Hu W, et al. Human infection with a novel avian-origin influenza A (H7N9) virus. N Engl J Med 2013 May 16;368(20):1888-1897. [doi: 10.1056/NEJMoa1304459] [Medline: 23577628]
2. Li Q, Zhou L, Zhou M, Chen Z, Li F, Wu H, et al. Preliminary Report: Epidemiology of the Avian Influenza A (H7N9) Outbreak in China. N Engl J Med 2013 Apr 24:1-11 Epub ahead of print(forthcoming)(forthcoming). [doi: 10.1056/NEJMoa1304617] [Medline: 23614499]
3. Wu S, Wu F, He J. Emerging risk of H7N9 influenza in China. The Lancet 2013 May 4;381(9877):1539-1540. [doi: 10.1016/S0140-6736(13)60767-9] [Medline: 23602315]
4. China Internet Network Information Center. Beijing: China Internet Network Information Center; 2013. Statistical Report on Internet Development in China [in Chinese] URL: http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201301/P020130801546406723463.pdf [accessed 2013-10-08] [WebCite Cache ID 6KBt5fXnI]
5. China Internet Network Information Center. Beijing: China Internet Network Information Center; 2012. Chinese search engine market research report [in Chinese] URL: http://www.cnnic.cn/gywm/xwzx/rdxw/2011nrd/201207/W020111223371374614784.pdf [accessed 2013-10-08] [WebCite Cache ID 6KBtE5alf]
6. 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]
7. Henrich N, Holmes B. What the public was saying about the H1N1 vaccine: perceptions and issues discussed in on-line comments during the 2009 H1N1 pandemic. PLoS One 2011;6(4):e18479 [FREE Full text] [doi: 10.1371/journal.pone.0018479] [Medline: 21533161]
8. Xu XJ, Hu XY. The review on public health media surveillance and risk research [in Chinese]. Public Health and Preventive Medicine [in Chinese] 2011;22(5):72-77.
9. Fang N. The Management of Emergencies: Rethinking of SARS [in Chinese]. Beijing: China Social Science Press; Apr 01, 2005.
10. National Health and Family Planning Commission of People’s Republic of China. 2013. The daily report on avian influenza H7N9 [in Chinese] URL: http://www.moh.gov.cn/mohwsyjbgs/h7n9/list.shtml [accessed 2013-10-08] [WebCite Cache ID 6KBtQNbB5]
11. Goodchild MF, Glennon JA. Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth 2010 Sep;3(3):231-241. [doi: 10.1080/17538941003759255]
12. Sina Website. 2013. The Description of Hot Word in Sina micro blog [in Chinese] URL: http://data.weibo.com/index/hotword [accessed 2013-10-08] [WebCite Cache ID 6KBtVuqe3]
13. Baidu Website. 2013. The Help of Baidu Index [in Chinese] URL: http://www.baidu.com/search/index_help.html [accessed 2013-10-08] [WebCite Cache ID 6KBtbxzgd]
14. 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]
15. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
16. 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]
17. Tang T. Research on method of monitoring Net-mediated public sentiment based on analysis of search engine logs [in Chinese]. Journal of Intelligence 2012;31(8):27-30.
18. Outbreak communication: best practices for communicating with the public during an outbreak. Geneva: World Health Organization; 2005. URL: http://www.who.int/csr/resources/publications/WHO_CDS_2005_32web.pdf [accessed 2013-10-08] [WebCite Cache ID 6KBu5uYfx]
19. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of

three consecutive online surveys. BMC Public Health 2011;11:2 [FREE Full text] [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]

20. Lau JT, Griffiths S, Choi KC, Tsui HY. Widespread public misconception in the early phase of the H1N1 influenza epidemic. J Infect 2009 Aug;59(2):122-127. [doi: 10.1016/j.jinf.2009.06.004] [Medline: 19592114]
21. Lau JT, Lau M, Kim JH, Tsui HY. Impacts of media coverage on the community stress level in Hong Kong after the tsunami on 26 December 2004. J Epidemiol Community Health 2006 Aug;60(8):675-682 [FREE Full text] [doi: 10.1136/jech.2005.041897] [Medline: 16840756]
22. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul;339:b2651 [FREE Full text] [Medline: 19574308]
23. Rubin GJ, Potts HW, Michie S. The impact of communications about swine flu (influenza A H1N1v) on public responses to the outbreak: results from 36 national telephone surveys in the UK. Health Technol Assess 2010 Jul;14(34):183-266 [FREE Full text] [doi: 10.3310/hta14340-03] [Medline: 20630124]
24. Huang CM, Chan E, Hyder AA. Web 2.0 and internet social networking: a new tool for disaster management?--lessons from Taiwan. BMC Med Inform Decis Mak 2010;10:57 [FREE Full text] [doi: 10.1186/1472-6947-10-57] [Medline: 20925944]
25. Kittler AF, Hobbs J, Volk LA, Kreps GL, Bates DW. The Internet as a vehicle to communicate health information during a public health emergency: a survey analysis involving the anthrax scare of 2001. J Med Internet Res 2004 Mar 3;6(1):e8 [FREE Full text] [doi: 10.2196/jmir.6.1.e8] [Medline: 15111274]
26. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Inectious Diseases/SARS Community Outreach Team. Fear and stigma: the epidemic within the SARS outbreak. Emerg Infect Dis 2004 Feb;10(2):358-363 [FREE Full text] [doi: 10.3201/eid1002.030750] [Medline: 15030713]
27. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
28. Samaan G, Patel M, Olowokure B, Roces MC, Oshitani H, World Health Organization Outbreak Response Team. Rumor surveillance and avian influenza H5N1. Emerg Infect Dis 2005 Mar;11(3):463-466 [FREE Full text] [doi: 10.3201/eid1103.040657] [Medline: 15757567]
29. Tai Z, Sun T. The rumouring of SARS during the 2003 epidemic in China. Sociol Health Illn 2011 Jul;33(5):677-693. [doi: 10.1111/j.1467-9566.2011.01329.x] [Medline: 21592140]
30. Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 2006 Dec;63(12):3113-3123. [doi: 10.1016/j.socscimed.2006.08.004] [Medline: 16978751]
31. Duncan B. How the media reported the first days of the pandemic (H1N1) 2009: results of EU-wide media analysis. Euro Surveill 2009 Jul 30;14(30):19286 [FREE Full text] [Medline: 19643058]
32. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. N Engl J Med 2011 Jul 28;365(4):289-291. [doi: 10.1056/NEJMp1103591] [Medline: 21793742]

Abbreviations

BAI: Baidu Attention Index
SARS: severe acute respiratory syndrome
CNHFPC: China National Health and Family Planning Committee

Edited by G Eysenbach; submitted 28.08.13; peer-reviewed by C Freifeld, T Timpka, O Leal Neto; comments to author 19.09.13; revised version received 29.10.13; accepted 21.12.13; published 17.01.14.

Please cite as:
Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, Jiang Z, Zheng D, Jiang J
Importance of Internet Surveillance in Public Health Emergency Control and Prevention: Evidence From a Digital Epidemiologic Study During Avian Influenza A H7N9 Outbreaks
J Med Internet Res 2014;16(1):e20
URL: http://www.jmir.org/2014/1/e20/
doi:10.2196/jmir.2911
PMID:24440770

Original Paper

Guess Who's Not Coming to Dinner? Evaluating Online Restaurant Reservations for Disease Surveillance

Elaine O Nsoesie[1,2,3], PhD; David L Buckeridge[4,5,6], MD, PhD; John S Brownstein[1,2,4], PhD

[1] Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA, United States
[2] Department of Pediatrics, Harvard Medical School, Boston, MA, United States
[3] Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, MA, United States
[4] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada
[5] Surveillance Lab, McGill Clinical and Health Informatics, McGill University, Montreal, QC, Canada
[6] Agence de la santé et des services sociaux de Montréal, Direction de santé publique, Montreal, QC, Canada

Corresponding Author:
Elaine O Nsoesie, PhD
Children's Hospital Informatics Program, Boston Children's Hospital
1 Autumn St #14
Boston, MA, 02215
United States
Phone: 1 857 218 5108
Email: onelaine@vt.edu

Abstract

Background: Alternative data sources are used increasingly to augment traditional public health surveillance systems. Examples include over-the-counter medication sales and school absenteeism.

Objective: We sought to determine if an increase in restaurant table availabilities was associated with an increase in disease incidence, specifically influenza-like illness (ILI).

Methods: Restaurant table availability was monitored using OpenTable, an online restaurant table reservation site. A daily search was performed for restaurants with available tables for 2 at the hour and at half past the hour for 22 distinct times: between 11:00 am-3:30 pm for lunch and between 6:00-11:30 PM for dinner. In the United States, we examined table availability for restaurants in Boston, Atlanta, Baltimore, and Miami. For Mexico, we studied table availabilities in Cancun, Mexico City, Puebla, Monterrey, and Guadalajara. Time series of restaurant use was compared with Google Flu Trends and ILI at the state and national levels for the United States and Mexico using the cross-correlation function.

Results: Differences in restaurant use were observed across sampling times and regions. We also noted similarities in time series trends between data on influenza activity and restaurant use. In some settings, significant correlations greater than 70% were noted between data on restaurant use and ILI trends.

Conclusions: This study introduces and demonstrates the potential value of restaurant use data for event surveillance.

(J Med Internet Res 2014;16(1):e22) doi:10.2196/jmir.2998

KEYWORDS

population surveillance; restaurants; epidemics; outbreaks

Introduction

Global adoption of the Internet and mobile phone technologies has proven useful for gathering and disseminating data. Various novel data streams using these technologies have been explored as tools for augmenting traditional public health disease surveillance systems. These novel systems typically aim to improve the detection and monitoring of outbreaks in addition to disseminating information to the public and to public health professionals. Examples include search query volume [1], and digital surveillance systems harnessing news reports and social media, such as HealthMap [2,3] and Global Public Health Intelligence Network (GPHIN) [4]. Other innovative surveillance systems have explored the use of over-the-counter medication sales [5], telephone triage records [6], and school absenteeism [7]. The usefulness of these alternative data sources has been evaluated in several studies, especially for monitoring seasonal and pandemic influenza (eg, Besculides et al [8], Vergu et al [9], Yih et al [10], and Bernardo et al [11]). Another example includes the use of digital surveillance systems to mine early reports of an outbreak of acute respiratory infections, which later evolved into a pandemic in 2009 [12]. Similarly, during the recent 2013 H7N9 influenza outbreak in China, social media sites, such as Twitter and Sina Weibo (a Chinese social network site similar to Twitter), provided near real-time information on disease activity [13].

Similar to school absenteeism, over-the-counter medication sales, and volume of telephone triage service data, utilization of online restaurant reservation sites could also serve as a tool for event surveillance. Studies have noted that the percentage of meals consumed outside the home in the United States has increased [14-16]. Therefore, monitoring changes in restaurant use could possibly serve as a leading indicator of disruption resulting from social unrest, including a public health event. In particular, a decrease in restaurant use could serve as an early indicator of a disease-related event. In this study, we evaluated whether a rise in restaurant table availability was associated with an increase in influenza-like illness (ILI).

Methods

Data

All data on restaurant use were obtained from OpenTable [17], an online platform where individuals can make table reservations at restaurants with availabilities at different times of the day. The site caters to restaurants in various cities in several different countries with more than 28,000 restaurants in the database at the time of this writing. The number of registered restaurants is available for each city/region and varies over time as new restaurants join and existing ones either close or cancel their registration.

Each day from September 4, 2012 to April 30, 2013, at set times around lunch and dinner, we conducted a search to determine the number of restaurants with tables available for 2 people. To accommodate differences in regions and eating habits, we defined the lunch period as between 11:00 am-3:30 pm and dinner as between 6-11:30 pm. According to OpenTable policy, customers can cancel reservations up to 30 minutes before the reserved time. Therefore, we searched for restaurants with table availabilities 15 minutes before the times of interest. So, for reservations at 2:00 pm, we would search for available tables every day at 1:45 pm. In addition, we searched for restaurants with table availabilities at the hour and at half past the hour. This resulted in 20 distinct search times each day for each of the 10 study regions in the United States and Mexico. In the United States, table availability was examined for restaurants in Boston (Massachusetts), Atlanta (Georgia), Baltimore (Maryland), and Miami (Florida). For Mexico, we monitored table availabilities in Cancun (Quintana Roo), Mexico City (Distrito Federal), Puebla (Puebla), Monterrey (Nuevo Léon), Guadalajara (Jalisco), and the whole of Mexico. Since data were collected every day at the specified times, our observations formed time series curves of availabilities. Monitoring 10 regions at 20 search times resulted in 200 distinct time series.

Comparison to Data on Influenza Activity

By using data from the recent 2012-2013 severe influenza season, we tested the hypothesis that an increase in influenza activity was associated with a rise in restaurants with table availabilities. Since the number of restaurants in the system varied over time, we focused on the proportion of restaurants with available tables. For each region, the proportion of restaurants with available tables was defined as the number of restaurants with availabilities at time *t* divided by the total number of restaurants on OpenTable at time *t*. First, we examined the data to better understand trends in table availability during the baseline period. The baseline period was from September to October 2012 because influenza season typically runs from November to April in the northern hemisphere [18]. Observations on restaurant use during the baseline period could suggest best times for surveillance. Next, we calculated the average weekly proportion of restaurants with table availability at each sampling time and compared these data to weekly estimates of ILI. Cross-correlations between time series could be affected by bias because of temporal autocorrelation [19]. Bias in this study could be due to low-frequency patterns resulting from fewer numbers of restaurants open at particular hours of the day. Therefore, we applied prewhitening by fitting an autoregressive integrated moving average (ARIMA) model to availabilities, and then filtering the ILI values using the fitted model. The ARIMA model can be described as follows:

$$y_t=c+\Phi_1 y_{t-1}+\ldots+\Phi_p y_{t-p}-\theta_1 z_{t-1}-\ldots-\theta_p z_{t-p}+z_t$$

where c is a constant, y_t is the observation at time t, y_{t-p} are lagged values of the series, and z_t is a white noise process. Correlations were then examined between the residuals of the availabilities model and the filtered ILI values using the cross-correlation function (CCF). Data representing ILI activity was obtained from state surveillance systems [20-23], Google Flu Trends [24] and the Pan American Health Organization (PAHO) [25]. Google Flu Trends data was available at the city level for cities in the United States and at the province level for Mexico. We calculated correlations between city-level Google Flu Trends and availabilities data for cities in the United States. Due to the unavailability of data at the city level, we estimated correlations between Google Flu Trends state-level data and availabilities for the various cities in Mexico. PAHO’s estimated percent positive for influenza data was only available at the country level for Mexico. Weekly percent ILI (% ILI), resulting from physician visits was also available for all states in the United States. Additionally, for illustrative purposes, local polynomial regression fitting (LOESS) was used in smoothing curves presented in the Results. Smoothing was performed to capture the overall trend of the curves for comparison purposes. See Cleveland [26] for additional information on the LOESS model. Bonferroni adjustment was also applied to account for multiple comparisons as needed. The analysis was performed in R (The R Foundation for Statistical Computing, Wien, Austria).

Results

Data Summary: Baseline Period

Table 1. Mean proportion of restaurants with available tables at lunch (11:00 am-3:30 pm) and dinner (6:00-11:30 pm) times for cities in the United States and Mexico.

Sampling time	Regions, mean (SD)									
	United States				Mexico					
	Atlanta	Miami	Boston	Baltimore	Mexico City	Cancun	Guadalajara	Monterrey	Puebla	Mexico
11:00 am	0.598 (0.032)	0.668 (0.033)	0.595 (0.021)	0.575 (0.029)						
11:30 am	0.594 (0.032)	0.666 (0.033)	0.595 (0.020)	0.572 (0.028)						
12:00 pm	0.592 (0.035)	0.665 (0.035)	0.592 (0.021)	0.572 (0.028)	0.876 (0.054)	0.638 (0.090)[a]	0.896 (0.065)	0.841 (0.067)	0.982 (0.088)	0.777 (0.050)
12:30 pm	0.590 (0.033)	0.665 (0.035)	0.591 (0.022)[a]	0.573 (0.029)	0.878 (0.055)	0.638 (0.089)[a]	0.9 (0.065)	0.849 (0.068)	0.982 (0.085)	0.78 (0.050)
1:00 pm	0.589 (0.036)	0.664 (0.034)[a]	0.595 (0.022)	0.573 (0.029)	0.876 (0.054)	0.638 (0.089)[a]	0.903 (0.064)	0.855 (0.070)	0.981 (0.088)	0.78 (0.049)
1:30 pm	0.614 (0.034)	0.677 (0.032)	0.628 (0.023)	0.626 (0.029)	0.876 (0.054)	0.641 (0.089)	0.904 (0.066)	0.869 (0.072)	0.984 (0.087)	0.781 (0.048)
2:00 pm	0.604 (0.035)	0.684 (0.034)	0.624 (0.024)	0.622 (0.035)	0.876 (0.054)	0.641 (0.089)	0.896 (0.070)	0.87 (0.067)	0.983 (0.092)	0.781 (0.047)
2:30 pm	0.729 (0.045)	0.758 (0.041)	0.771 (0.034)	0.836 (0.040)	0.867 (0.054)	0.689 (0.092)	0.881 (0.070)	0.888 (0.072)	0.992 (0.092)	0.779 (0.046)
3:00 pm	0.906 (0.049)	0.765 (0.048)	0.887 (0.035)	0.877 (0.042)	0.855 (0.063)	0.693 (0.092)	0.859 (0.074)	0.891 (0.069)	0.99 (0.094)	0.771 (0.046)
3:30 pm	0.899 (0.051)	0.863 (0.041)	0.889 (0.038)	0.894 (0.043)	0.829 (0.075)	0.821 (0.121)	0.836 (0.078)	0.893 (0.071)	0.988 (0.099)	0.763 (0.050)
6:00 pm	0.884 (0.051)	0.876 (0.046)	0.846 (0.042)	0.877 (0.046)	0.833 (0.056)	0.89 (0.130)	0.858 (0.074)	0.912 (0.074)	0.99 (0.100)	0.773 (0.049)
6:30 pm	0.880 (0.051)	0.871 (0.047)	0.847 (0.040)	0.875 (0.044)	0.815 (0.056)	0.891 (0.132)	0.835 (0.075)	0.892 (0.079)	0.97 (0.107)	0.757 (0.048)
7:00 pm	0.884 (0.053)	0.869 (0.048)	0.860 (0.041)	0.879 (0.045)	0.807 (0.058)	0.891 (0.133)	0.829 (0.080)	0.877 (0.079)	0.961 (0.106)	0.749 (0.052)
7:30 pm	0.888 (0.052)	0.871 (0.048)	0.874 (0.043)	0.882 (0.045)	0.782 (0.056)	0.89 (0.131)	0.813 (0.076)	0.847 (0.080)	0.911 (0.097)	0.725 (0.049)
8:00 pm	0.887 (0.054)	0.873 (0.047)	0.874 (0.044)	0.879 (0.045)	0.775 (0.055)	0.891 (0.131)	0.813 (0.076)	0.839 (0.078)	0.911 (0.098)	0.719 (0.049)
8:30 pm	0.882 (0.051)	0.869 (0.047)	0.871 (0.043)	0.872 (0.045)	0.748 (0.056)	0.89 (0.131)	0.799 (0.078)	0.827 (0.084)	0.89 (0.091)	0.699 (0.048)
9:00 pm	0.863 (0.050)	0.864 (0.047)	0.858 (0.043)	0.845 (0.056)	0.74 (0.055)	0.889 (0.13)	0.786 (0.082)	0.799 (0.092)	0.891 (0.088)	0.69 (0.048)
9:30 pm	0.818 (0.066)	0.838 (0.052)	0.835 (0.045)	0.802 (0.076)	0.694 (0.065)	0.884 (0.131)	0.736 (0.095)	0.759 (0.093)	0.881 (0.089)	0.654 (0.050)
10:00 pm	0.682 (0.163)	0.779 (0.089)	0.721 (0.127)	0.656 (0.170)	0.674 (0.077)	0.883 (0.134)	0.688 (0.095)	0.691 (0.096)	0.865 (0.089)	0.634 (0.058)
10:30 pm	0.539 (0.229)[a]	0.677 (0.143)	0.598 (0.217)	0.520 (0.233)[a]	0.599 (0.077)	0.814 (0.148)	0.619 (0.078)	0.635 (0.098)	0.85 (0.097)	0.571 (0.065)
11:00 pm					0.546 (0.08)	0.774 (0.178)	0.563 (0.067)	0.582 (0.071)	0.776 (0.102)	0.525 (0.067)
11:30 pm					0.501 (0.117)[a]	0.717 (0.183)	0.527 (0.068)[a]	0.551 (0.091)[a]	0.705 (0.189)[a]	0.483 (0.097)[a]

[a] Sampling times with lowest proportion of availabilities.

Table availabilities varied by mealtime (lunch and dinner) and by city. Times with lowest availabilities in some cities could represent preferred dining times. Trends observed during preferred dining times should also be most affected by seasonal deviations in dining. Hereafter, we refer to times with lowest availabilities as “most preferred” and times with highest availabilities as “least preferred.” The most- and least-preferred dining times varied by region. Use of *t* tests suggested significant differences (*P*<.001) in availabilities between most- and least-preferred times across all regions. Additionally, mean availabilities at the most- and least-preferred times were also significantly different at lunch and dinner for most regions (*P*<.05), except for lunch in Monterrey (*P*=.23) and Puebla (*P*=.99). Times with the lowest mean proportion of restaurants with available tables at lunch were 2:00 pm (mean 0.608, SD 0.005) for Atlanta, Miami at noon (mean 0.673, SD 0.016), Boston at noon (mean 0.594, SD 0.009), Baltimore at noon (mean 0.589, SD 0.007), Mexico City at 3:30 pm (mean 0.757, SD 0.044), Cancun at 12:30 pm (mean 0.709, SD 0.048), Guadalajara at 3:30 pm (mean 0.818, SD 0.06), Monterrey at noon (mean 0.824, SD 0.055), Puebla at noon (mean 0.963, SD 0.035), and Mexico at 3:30 pm (mean 0.753, SD 0.035). All regions had the lowest proportion of restaurants with available tables for dinner between 10:30-11:30 pm, which could be because of fewer restaurants being open later in the night, especially for US cities. The mean proportion of restaurants with table availability for the entire study period (September 2012-April 2013) is summarized in Table 1.

Comparison to Data on Influenza Activity

We calculated 300 correlations between data representing ILI activity and restaurant use. Significant correlations (*P*<.05) between the restaurant table availability data for each region and percent positive for influenza from PAHO, percent ILI from states in the United States, and Google Flu Trends are summarized in Table 2. Correlations presented in Table 2 were calculated from prewhitened time series data. We present correlations at weekly lags of 0 and 1 because significant correlations at lag 0 suggest that increase in ILI were associated with immediate increase in availabilities. On the other hand, significant correlations at a 1-week lag indicate that increases in availabilities were followed by a noted increase in ILI. Understanding the correlation at different lags can inform the use of restaurant utilization data for modeling and forecasting future ILI activity. We present results only at lag 1 because the data are at a weekly resolution and studies have suggested that the mean duration of influenza symptoms is presumably less than 7 days [27].

As noted, all regions had the lowest proportion of restaurants with available tables for dinner between 10:30-11:30 pm during the baseline period. As seen in Table 2, these were also the times with the most significant correlations for Miami, Mexico City, and Mexico. Note that most of the dinnertime correlations for Atlanta and Baltimore were observed earlier in the evening compared to Miami. Overall, the highest correlations (>70%) were recorded for Atlanta at 12:30 pm and 3:00 pm at lag 0, and Baltimore at 7 pm at 1-week lag. More significant correlations were also noted between restaurant table availability and Google Flu Trend data compared to PAHO percent positive for influenza and state ILI data (see Table 2). Significant correlations were recorded between PAHO data and restaurant table availabilities for Mexico at 4 set times. As seen in Table 2, a few correlations were also observed between restaurant table availabilities in US cities and state ILI percentages. State ILI trends are not always identical to trends at the city level, which could explain the lack of significant correlations. No significant correlations were noted for Cancun, Guadalajara, Puebla, and Boston.

We present a sample plot showing a case in which the trend in restaurant table availability appears similar to the trend in estimated ILI activity for Miami in Figure 1. Note the dips in the curves before the peak observed during the weeks of Thanksgiving and Christmas holidays. The drop observed after the peak occurred during the week of Valentine’s Day. As seen in Figure 1, the overall trend in the data for availabilities at 10 pm for Miami was similar to that observed with Google Flu Trends, with the peak observed later. This could suggest that an increase in availabilities was observed after a rise in ILI-related queries. Further, 9 graphs representing trends observed during sampling times with the highest correlations for Baltimore, Atlanta, and Mexico are shown in Figure 2. Note the similarities and differences between the curves. In some cases, such as G-I in Figure 2, several of the peaks and troughs in the influenza data are captured by the data on restaurant utilization.

Table 2. Significant cross-correlations (*P*<.05) for data source and time for cities in the United States and Mexico.

Country/city	Data source								
	Google Flu Trends			State ILI			PAHO		
	Time	*r* %		Time	*r* %		Time	*r* %	
		Lag 0	Lag1		Lag 0	Lag 1		Lag 0	Lag 1
United States									
Miami	10:30 pm	55.06	58.19	10:30 pm	56.77				
	10:00 pm	53.04							
	9:30 pm		54.56						
	9:00 pm	55.73	65.85						
Baltimore	7:30 pm		63.60	3:30 pm	63.67				
	8:30 pm		49.93	7:00 pm		70.27			
	8:00 pm		62.75						
Atlanta	3:30 pm	69.74	46.54	12:30 pm	74.27				
	3:00 pm	70.15							
	6:30 pm	67.25							
	6:00 pm	59.85							
	7:30 pm	55.81	46.60						
	7:00 pm	69.01							
	8:30 pm	66.12	46.40						
	8:00 pm	60.35							
Mexico									
Mexico City	11:30 pm		36.43						
Monterrey	12:30 pm	51.19	47.43						
Mexico							10:30 pm		39.83
							10:00 pm		41.57
							11:00 pm		41.76
							11:30 pm		39.33

Figure 1. The trend in the proportion of restaurants with table availability at 10:00 PM for Miami compared to Google Flu Trend data for Miami. (A) Restaurant table availability and Google Flu Trend curve; (B) curve with LOESS smoothing to capture the overall trend in availabilities.

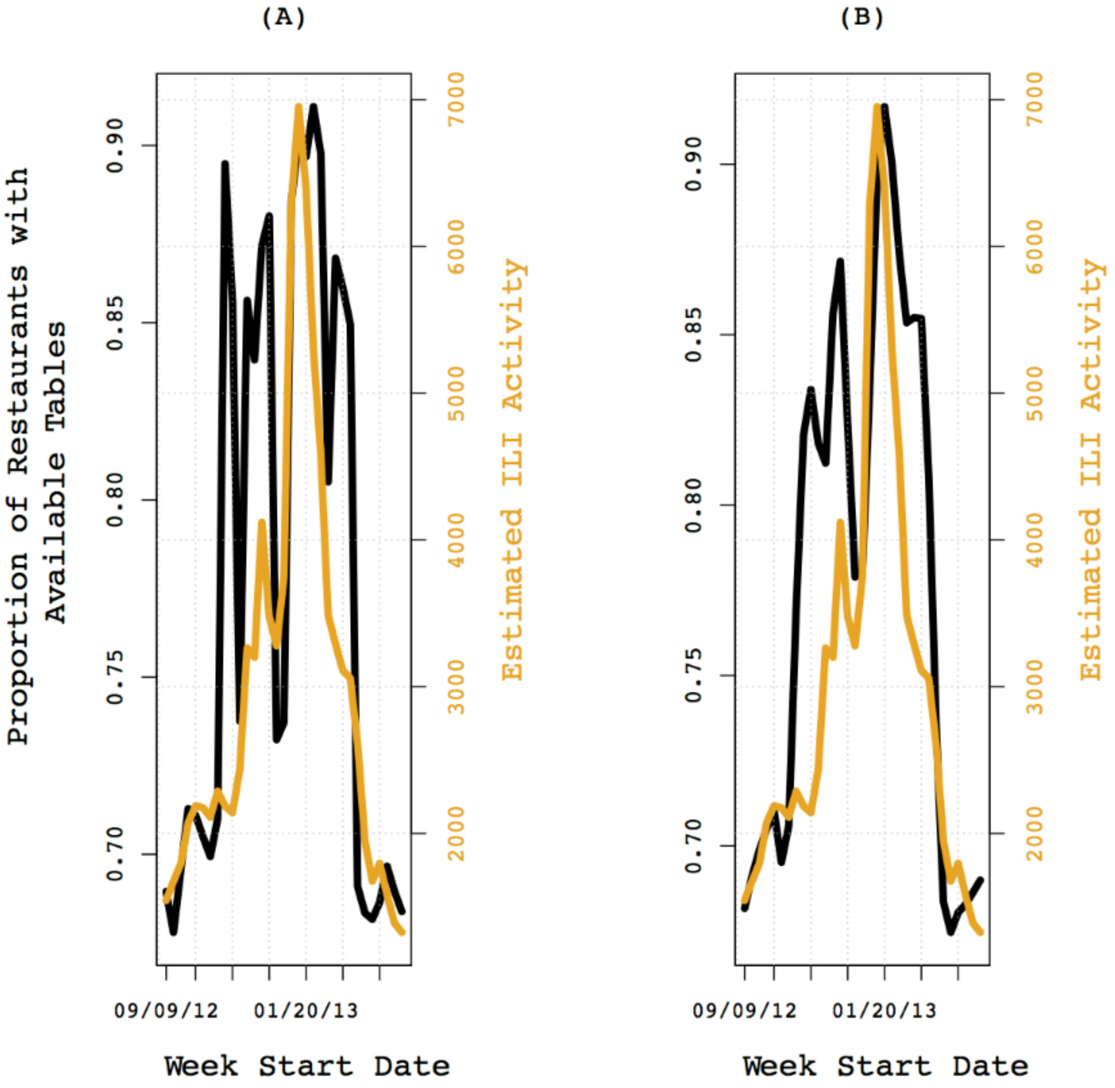
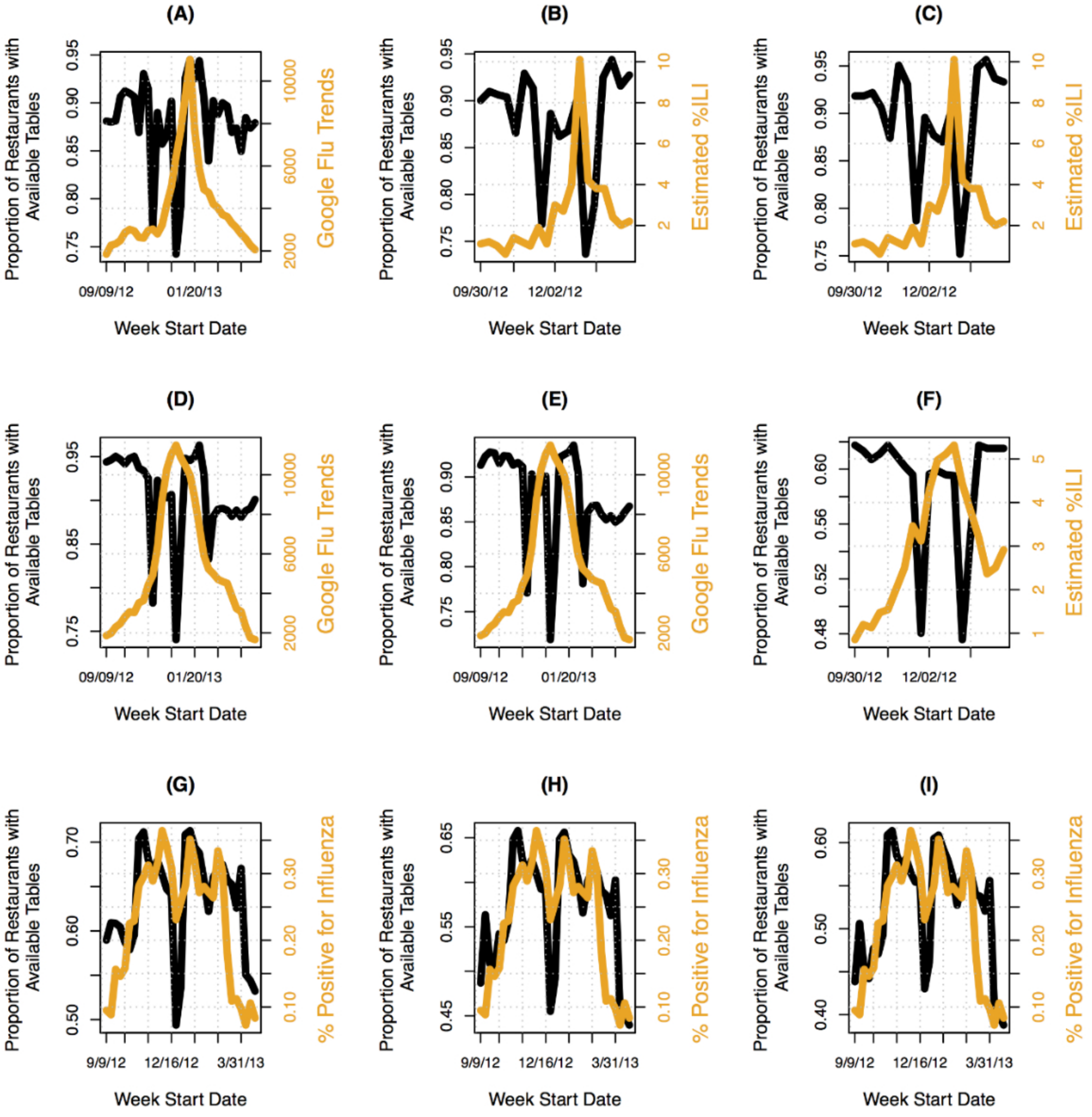


Figure 2. The trend in the proportion of restaurants with table availability compared to trend in various estimates of influenza activity for different regions. (A) Google Flu Trend and restaurant table availability for Baltimore at 7:30 PM; (B, C) estimated % influenza-like illness (ILI) for Maryland and restaurant table availability for Baltimore at 7:00 PM and 3:30 PM, respectively; (D, E) Google Flu Trend and restaurant table availability for Atlanta at 3:00 PM and 6:30 PM, respectively; (F) estimated % ILI for Georgia and restaurant table availability for Atlanta at 12:30 PM; (G-I) Pan American Health Organization (PAHO) % positive for influenza and restaurant table availability for Mexico at 10:00 PM,10:30 PM, and 11:00 PM, respectively.



Discussion

Principal Findings

In this paper, we introduced an easily accessible Internet-based data stream—online restaurant reservations—and demonstrated its potential value for event surveillance. More specifically, we observed significant correlations between restaurant table availability and Google Flu Trends, and influenza activity data at the city and country level in the United States and Mexico. In most cases, associations between restaurant use and measures of influenza activity were stronger when all data were at the same geographic resolution. For instance, correlations between restaurant use in Miami and Google Flu Trends data for Miami were stronger than correlations between ILI data for Florida and restaurant use data for Miami. This tendency is explained at least in part by the known variation in ILI trends when measured at different geographical resolutions (eg, city, state, country) [28,29]. We also observed the highest correlations (>70%) for Atlanta at 12:30 pm and 3:00 pm, and for Baltimore at 7 pm. Dinner times with significant correlations were observed later in the evening for regions in the south compared to others. These differences in observations across regions could potentially be explained by differences in dining habits and demographic differences, which could also affect both dining habits and influenza spread [30,31]. Other potential factors include social disruptions, socioeconomic factors, natural disasters, foodborne illnesses, etc. In addition to modeling the data for providing estimates of ILI before the release of official reports, in future studies we would also investigate any occurrences of social unrest and natural disasters, which might have affected the trend in the time series.

Limitations

Although possibly useful, there are several limitations inherent to the study and the data source impeding a full exploration of this system’s potential. Limitations include differentiating between seasonal changes and changes potentially resulting from a disease-related event. In addition, data on reservation cancellations would be most suitable. However, these data are currently unavailable. This issue could be easily remediated by creating a partnership with restaurants such that occurrences of and reasons for cancellations are recorded through a survey system. Furthermore, only samples of restaurants in each region are listed on OpenTable.

Conclusions

Despite these limitations, this preliminary analysis suggests that monitoring trends in restaurant table availabilities and cancellations could be useful for detecting social disruption, including disease-related events. Moreover, unlike school absenteeism, over-the-counter medication sales, and volume of telephone triage service data, which are traditionally difficult to access, reservation use data can easily be obtained from reservation sites. The global penetration of the Internet also suggests that such data sources could be easily harvested in the future. These novel data sources could serve as a stepping-stone to prompt further investigation of disease events if warranted. Observations made using this data can be further investigated by comparing trends to other alternative sources for disease surveillance, especially in situations where official reports on disease activity are delayed. Additionally, this data source can be fused with more traditional data streams for epidemic intelligence using ensemble modeling approaches.

Acknowledgments

We thank Sumiko Mekaru for suggestions regarding data analysis. This work is partially supported by a research grant the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

Authors' Contributions

All authors conceived and designed the study. Elaine O Nsoesie analyzed the data. All authors drafted, edited, and approved the final version of the paper. The sponsors had no role in the design, analysis, and writing of this manuscript.

Conflicts of Interest

None declared.

References

1. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
2. HealthMap. URL: http://healthmap.org/en/ [accessed 2013-11-27] [WebCite Cache ID 6LS0ph6rj]
3. 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]
4. Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. Can J Public Health 2006 Jan;97(1):42-44. [Medline: 16512327]

5. Magruder SF. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. Johns Hopkins University APL Technical Digest 2003;24:349-353.
6. Espino JU, Hogan WR, Wagner MM. Telephone triage: a timely data source for surveillance of influenza-like diseases. AMIA Annu Symp Proc 2003:215-219 [FREE Full text] [Medline: 14728165]
7. Peterson D, Andrews JS, Levy BS, Mitchell B. An effective school-based influenza surveillance system. Public Health Rep 1979;94(1):88-92 [FREE Full text] [Medline: 419298]
8. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City. BMC Public Health 2005;5:105 [FREE Full text] [doi: 10.1186/1471-2458-5-105] [Medline: 16212669]
9. Vergu E, Grais RF, Sarter H, Fagot JP, Lambert B, Valleron AJ, et al. Medication sales and syndromic surveillance, France. Emerg Infect Dis 2006 Mar;12(3):416-421 [FREE Full text] [doi: 10.3201/eid1205.050573] [Medline: 16704778]
10. Yih WK, Teates KS, Abrams A, Kleinman K, Kulldorff M, Pinner R, et al. Telephone triage service data for detection of influenza-like illness. PLoS One 2009;4(4):e5260 [FREE Full text] [doi: 10.1371/journal.pone.0005260] [Medline: 19381342]
11. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
12. Brownstein JS, Freifeld CC, Madoff LC. Influenza A (H1N1) virus, 2009--online monitoring. N Engl J Med 2009 May 21;360(21):2156 [FREE Full text] [doi: 10.1056/NEJMp0904012] [Medline: 19423868]
13. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
14. Poti JM, Popkin BM. Trends in energy intake among US children by eating location and food source, 1977-2006. J Am Diet Assoc 2011 Aug;111(8):1156-1164 [FREE Full text] [doi: 10.1016/j.jada.2011.05.007] [Medline: 21802561]
15. Nielsen SJ, Siega-Riz AM, Popkin BM. Trends in food locations and sources among adolescents and young adults. Prev Med 2002 Aug;35(2):107-113. [Medline: 12200094]
16. Kant AK, Graubard BI. Eating out in America, 1987-2000: trends and nutritional correlates. Prev Med 2004 Feb;38(2):243-249. [Medline: 14715218]
17. OpenTable. URL: http://www.opentable.com/ [accessed 2013-11-27] [WebCite Cache ID 6LS2EbL6P]
18. Truscott J, Fraser C, Cauchemez S, Meeyai A, Hinsley W, Donnelly CA, et al. Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. J R Soc Interface 2012 Feb 7;9(67):304-312 [FREE Full text] [doi: 10.1098/rsif.2011.0309] [Medline: 21715400]
19. Bloom RM, Buckeridge DL, Cheng KE. Finding leading indicators for disease outbreaks: filtering, cross-correlation, and caveats. J Am Med Inform Assoc 2007;14(1):76-85 [FREE Full text] [doi: 10.1197/jamia.M2178] [Medline: 17068353]
20. Massachusetts Department of Public Health Weekly Influenza Update. 2013 May 24. URL: http://ma-publichealth.typepad.com/files/weekly_flu_report_5-24-13.pdf [accessed 2013-11-29] [WebCite Cache ID 6LUrxiI3L]
21. Influenza in Maryland 2012-2013 Season Report. URL: http://phpa.dhmh.maryland.gov/influenza/fluwatch/Past%20Flu%20Season%20Summaries/FINAL%20FLU%20REPORT%202012_13_9SEP13_Final.pdf [accessed 2013-11-27] [WebCite Cache ID 6LS1IVWH4]
22. Weekly Influenza Surveillance in Georgia, 2012-2013. URL: http://health.state.ga.us/epi/flu/fluupd11.asp [accessed 2013-11-27] [WebCite Cache ID 6LS64bdkX]
23. Florida Flu Review 2012-2013 season. URL: http://www.floridahealth.gov/diseases-and-conditions/influenza/_documents/week17.pdf [accessed 2013-11-27] [WebCite Cache ID 6LS1tcLsa]
24. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
25. Pan American Health Organization. URL: http://ais.paho.org/phip/viz/ed_flu.asp [accessed 2013-11-27] [WebCite Cache ID 6LS1zwYrD]
26. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 1979 Dec;74(368):829-836. [doi: 10.1080/01621459.1979.10481038]
27. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, et al. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am J Epidemiol 2008 Apr 1;167(7):775-785 [FREE Full text] [doi: 10.1093/aje/kwm375] [Medline: 18230677]
28. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 2006 Apr 21;312(5772):447-451 [FREE Full text] [doi: 10.1126/science.1125237] [Medline: 16574822]
29. Brownstein JS, Wolfe CJ, Mandl KD. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. PLoS Med 2006 Sep;3(10):e401 [FREE Full text] [doi: 10.1371/journal.pmed.0030401] [Medline: 16968115]
30. Opatowski L, Fraser C, Griffin J, de Silva E, Van Kerkhove MD, Lyons EJ, et al. Transmission characteristics of the 2009 H1N1 influenza pandemic: comparison of 8 Southern hemisphere countries. PLoS Pathog 2011 Sep;7(9):e1002225 [FREE Full text] [doi: 10.1371/journal.ppat.1002225] [Medline: 21909272]

31. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proc Biol Sci 2010 Feb 22;277(1681):557-565 [FREE Full text] [doi: 10.1098/rspb.2009.1605] [Medline: 19864279]

Abbreviations

ARIMA: autoregressive integrated moving average
CCF: cross-correlation function
IARPA: Intelligence Advanced Research Projects Activity
ILI: influenza-like illness
PAHO: Pan American Health Organization

Edited by G Eysenbach; submitted 30.09.13; peer-reviewed by Z Huang, J Schwind; comments to author 30.10.13; revised version received 02.12.13; accepted 30.12.13; published 22.01.14.

Please cite as:
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
URL: http://www.jmir.org/2014/1/e22/
doi:10.2196/jmir.2998
PMID:24451921

Original Paper

Older Adult Experience of Online Diagnosis: Results From a Scenario-Based Think-Aloud Protocol

Tana M Luger[1*], MPH, PhD; Thomas K Houston[1,2*], MD, MPH; Jerry Suls[3*], PhD

[1]eHealth Quality Enhancement Research Initiative, Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veteran's Hospital, Bedford, MA, United States

[2]Health Informatics and Implementation Science, Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[3]Behavioral Research Program, National Cancer Institute, Bethesda, MD, United States

[*]all authors contributed equally

Corresponding Author:
Tana M Luger, MPH, PhD
eHealth Quality Enhancement Research Initiative
Center for Healthcare Organization and Implementation Research
Edith Nourse Rogers Memorial Veteran's Hospital
Building 70 (152)
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 781 687 2642
Fax: 1 781 687 3106
Email: Tana.Luger2@va.gov

Abstract

Background: Searching for online information to interpret symptoms is an increasingly prevalent activity among patients, even among older adults. As older adults typically have complex health care needs, their risk of misinterpreting symptoms via online self-diagnosis may be greater. However, limited research has been conducted with older adults in the areas of symptom interpretation and human-computer interaction.

Objective: The intent of the study was to describe the processes that a sample of older adults may use to diagnose symptoms online as well as the processes that predict accurate diagnosis.

Methods: We conducted a series of “think-aloud” protocols with 79 adults aged 50 years or older. Participants received one of two vignettes that depicted symptoms of illness. Participants talked out loud about their thoughts and actions while attempting to diagnose the symptoms with and without the help of common Internet tools (Google and WebMD’s Symptom Checker). Think-aloud content was categorized using an adapted Q-sort and general inductive approach. We then compared the think-aloud content of participants who were accurate in their diagnosis with those who were not.

Results: Nineteen descriptive codes were identified from the think-aloud content. The codes touched upon Web navigation, attempts to organize and evaluate online health information, and strategies to diagnose symptoms. Participants most frequently relied on a strategy where they reviewed and then rejected the online diagnoses if they contained additional symptoms than those that were depicted in the vignette. Finally, participants who were inaccurate in their diagnosis reported being confused by the diagnosis task, lacking confidence in their diagnosis, and using their past experiences with illness to guide diagnosis more frequently than those participants who accurately diagnosed the symptoms.

Conclusions: Older adult participants tended to rely on matching strategies to interpret symptoms, but many still utilized existing medical knowledge and previous illness experiences as a guide for diagnosis. Many participants also had difficulty navigating the Internet tools, which suggests an increased need for navigation aids in Web design. Furthermore, participants who were inaccurate in their diagnosis had more difficulty with the Internet tools and confusion with the task than those who were accurate. Future work in this area may want to utilize additional study design such as eye-tracking to further understand the coordination between Web navigation, online symptom information processing, and diagnostic strategies.

(J Med Internet Res 2014;16(1):e16) doi:10.2196/jmir.2924

KEYWORDS
information seeking behavior; Internet; age factors

Introduction

Older Adult Health Information-Seeking and Online Self-Diagnosis

Older Internet users tend to be of higher socioeconomic status, higher education levels, and be young-old (eg, 65-70 years) rather than old-old (85+ years) [1,2]. Among older adults with Internet access (estimated to be 85% of the US population aged 50-64 years and 58% of the population aged 65 years and older), more than three-fourths seek online health information [3]. Many older adult users view the Internet as an “invaluable resource” of information that can replace the library [4,5], especially for health-related topics. In fact, some older adults claim to use the Internet to prepare for physician’s visits or better understand the information offered by their health providers [4].

Searching for information to interpret one’s own physical symptoms, or “online self-diagnosis”, is increasingly prevalent, with 35% of US adults having attempted to diagnose their own symptoms online [6]. In addition, 29% of online older adults aged 50-64 years and 13% of online older adults aged 65 years or older used the Internet to diagnose personal symptoms. Researchers and physicians have been apprehensive about patient online self-diagnosis [7-10] as online health information has been found to be of varying quality [11] and patients typically have limited health literacy or understanding [12,13], both of which could lead patients to inaccurate symptom interpretations. In addition, as older adults typically have complex health care needs, including co-morbid conditions, their risk of misinterpreting physical symptoms via online self-diagnosis may be greater. Given the safety issues that it raises and the limited research that has investigated this phenomenon, it is important to better understand the ways that older adults diagnose symptoms online.

Models of Symptom Interpretation

Symptom interpretation and diagnostic decision-making research has naturally focused on health care professionals; however, most patients do not have the same breadth of expert knowledge to apply when interpreting their own symptoms. According to Leventhal’s common-sense model of self-regulation, a layperson interprets physical symptoms by accessing memories of past experiences with illness and general knowledge about a health concept [14-17]. Patients use this information to create a “representation” of their symptoms, which can then guide action. Similarly, work by Pennebaker [18] demonstrates that patients interpret their symptoms selectively. In other words, patients focus on physical sensations and external information, which confirm their beliefs about health and illness. This kind of reasoning stands in stark contrast to expert physicians who are thought to use “pattern recognition” or “if-then” rules to make inferences about clinical cases [19,20] and medical students who are thought to use hypothetico-deductive reasoning [21], forming an initial hypothesis based on symptoms and then collecting additional data in order to confirm or disprove their hypothesis.

Although this body of work can be used to begin to understand how patients might interpret or diagnose physical symptoms, it does not account for the additional cognitive and perceptual processes that are required to conduct and interpret an Internet search, especially for older adults [22,23]. Further, older adults may have unique needs and strategies for symptom diagnosis as a result of their unique health care needs. For these reasons, we sought to uncover the cognitive processes that older adults might use to diagnose symptoms online as well as the processes that predict accurate diagnosis.

Participants in this study engaged in a “think-aloud” procedure—talking about their thoughts and actions —while attempting to diagnose the symptoms depicted in a vignette. The think-aloud was adopted to obtain a detailed description of older adult experience with online symptom diagnosis. We utilized an adapted Q-sort [24] and general inductive approach [25] to categorize participants’ think-aloud content as well as relate the coded content to the previously outlined models of symptom interpretation. Thus, we also sought to examine whether participants relied more on past experiences with illness and prior medical knowledge (consistent with the common-sense model of self-regulation) or hypothetico-deductive reasoning to interpret physical symptoms online.

Methods

Study Design, Setting, and Sample

This is a human-computer interaction study that included a series of think-aloud protocols [26] conducted with 79 older adults aged 50 years or older. Participants were recruited from a registry of older adults from the counties surrounding the University of Iowa (predominantly Johnson County, Iowa). Participants were included in the study if they were (1) at least 50 years or older, (2) a community resident (ie, not living in a nursing home), (3) able to travel to the research laboratory for in-person data collection, (4) owned a computer at home, (5) did not have a previous diagnosis of dementia or cognitive impairment, and (6) did not show cognitive impairment or confusion on the Short Portable Mental Status Questionnaire (SPMSQ) [27] (ie, score ≤7), which was administered as a brief screen over the phone. Participants were told that the study was an investigation of how people search for health information. Those who participated received a US$10 community gift card in appreciation and parking vouchers for their time at the research laboratory. The study was approved by the University of Iowa Human Subjects Institutional Review Board.

Procedure

All older adult participants received one of two illness vignettes (see Materials) and were asked to diagnose the symptoms in the vignette using one of two common Internet tools (Google search engine and WebMD’s Symptom Checker). Participants were randomly assigned to a vignette and an Internet tool in order to attempt to mitigate participant differences in computer skills and previous illness experience across the manipulated study variables.

Participants were given explicit instructions (adapted from Ericsson and Simon’s protocols) about the think-aloud before beginning the task. Participants were told to approach the think-aloud “basically like you’re talking to yourself, but loud enough for other people to hear” and that the goal was to “think-aloud as continuously as possible”. Participants were also told that the exercise would end when “you’ve come upon a diagnosis that you are satisfied with”.

The think-aloud procedure was first demonstrated by the experimenter (TML), and then the participant was given the opportunity to practice thinking out loud. When the participant felt comfortable with the procedure, he or she was given one of two illness vignettes to read and diagnose on his or her own, without any electronic aid. If the participant remained silent for five seconds, he or she was reminded to “please keep talking”. Participants were asked to choose one specific diagnosis (a specific illness or condition) in order to complete the task. No other prompting or questioning came from the experimenter regarding the diagnosis. The participant was audio-recorded during the think-aloud to allow for later transcription and analysis.

Participants then diagnosed the same symptoms, while thinking aloud, with the aid of one of two Internet tools. If the participant appeared confused or frustrated with the Internet tool for more than five seconds, the experimenter provided computer support in the form of describing the interface in more detail or describing what Web actions were available to the participant. Participants were limited to 30 minutes of search time. Most completed the task between 15 and 20 minutes. Finally, participants completed quantitative questionnaires, including demographics and computer skills (see Materials). The session typically took between one hour and two hours, depending on the participant’s interest in the tasks and speed answering the questionnaires.

Materials

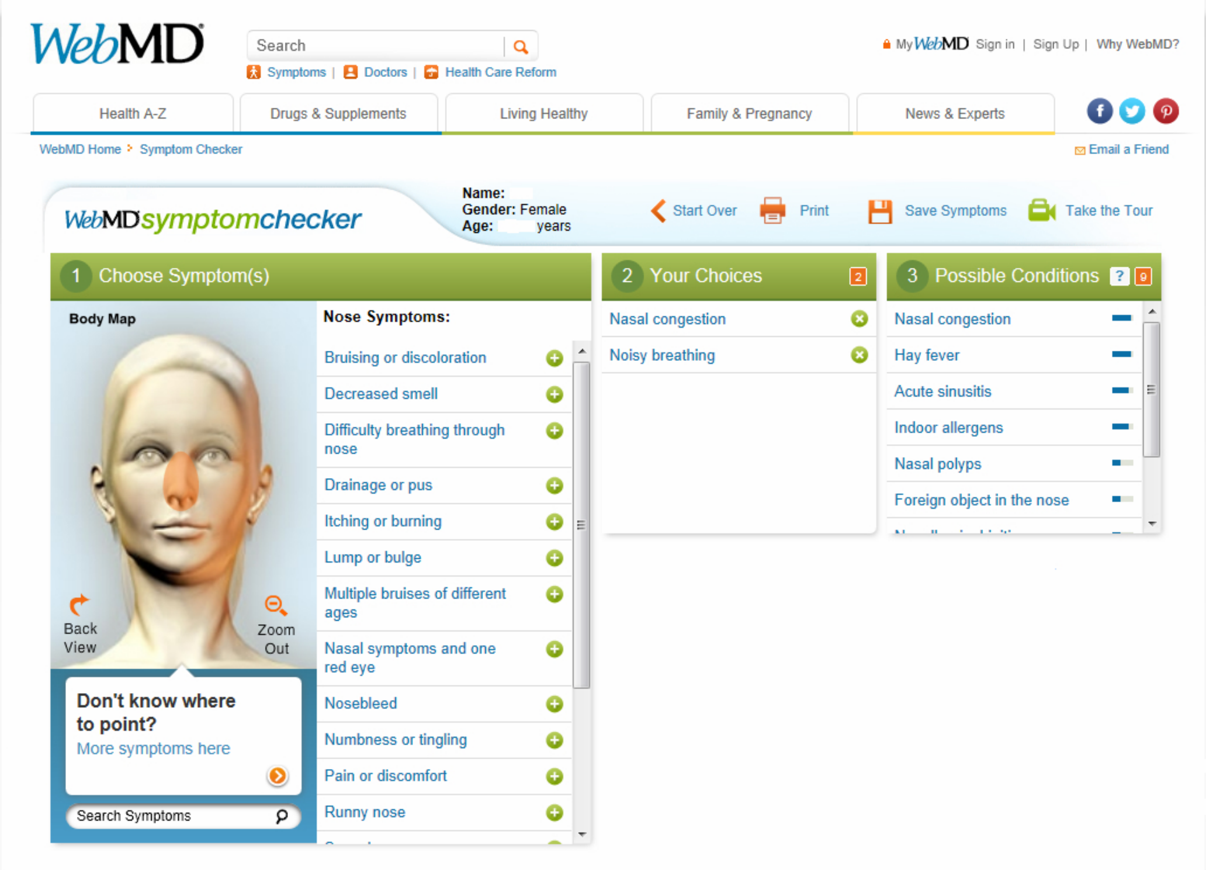
Illness Vignettes

Two vignettes were developed for the current study (see Multimedia Appendix 1). The vignettes depicted the symptoms of an acute health condition: mononucleosis or scarlet fever. These conditions were selected as they are rare in older adults, but still relatively common in the general population. This was to ensure that few participants would have recent experiences with the illness that could influence their diagnostic process. Participants were instructed to read the vignette as if they were experiencing the symptoms themselves. Vignettes were drafted from symptom information found at Mayo Clinic’s website [28] as well as the National Institute of Allergy and Infectious Diseases website [29]. Information was combined from multiple sites so that a Google search would not point directly to the site from which the information was drawn. Ten graduate students piloted both vignettes using both Google and WebMD’s Symptom Checker. Seven out of the ten students obtained the correct diagnosis for both vignettes.

Internet Tools

The common Internet tools of Google’s search engine [30] and WebMD’s Symptom Checker [31] were employed. Google is the number one visited website in the world [32] and provides users with webpages related to search queries by ranking the relative usefulness of Internet sites [33]. WebMD’s Symptom Checker is a consumer decision aid for the purpose of self-diagnosis (see Figure 1). The application features an avatar (or pictorial representation) of the human body. In order to diagnose, the user clicks on the area of the body where his or her symptoms are located and inputs descriptors of the symptoms such as “pain”, “tenderness”, or “warm to touch”. The application then asks tailored questions based on the location of symptoms, the descriptors, and the user’s response to each previous question. After gleaning enough information, the application will present a list of potential diagnoses. The user can click on a diagnosis to get more information about its symptoms and severity as well as recommendations for care.

Figure 1. WebMD symptom checker.



Demographics and Health

Age, gender, ethnicity, education, and income were collected via a self-report survey. Recent health history was measured using a retrospective symptom checklist [34]. Participants indicated the frequency (0 = not at all; 6 = extremely/much) with which they had experienced each of 15 symptoms (eg, headache, dizziness) in the past 21 days. To measure chronic health history, participants completed a checklist of common chronic conditions [35] (eg, diabetes, pneumonia). Participants indicated whether or not they had ever experienced each condition.

Computer Skills and Familiarity

Participants reported the number of hours per week that they used a home computer as well as the number of years that they had owned a home computer in order to gauge general computer familiarity. Participants were also verbally asked whether or not they had previous experience with the Internet tool to which they were assigned (Google or WebMD's Symptom Checker). Participant responses were documented in the think-aloud audio-recording and subsequent written transcript as a yes/no response.

Accuracy

Participants were required to choose one specific diagnosis (ie, a specific illness or condition) in order to complete the task. Participant diagnoses were documented in the think-aloud audio-recording and subsequent written transcript. Participants were deemed to have made an accurate diagnosis if they chose the same illness as the one depicted in the illness vignette to which they were assigned (ie, mononucleosis or scarlet fever).

Analysis: Think-Aloud Content Coding

In order to assess the design of the current study and establish an initial code list, pilot think-aloud data was collected from 15 participants aged 18 years or older at the University of Iowa Hospitals and Clinics. Audio recordings of pilot participant think-alouds (both without and using Internet tools) were transcribed into verbatim text. Each independent clause (or segment) was off-set on its own line for ease of analysis. Utilizing an adapted Q-sort method [2], the experimenter and a group of research assistants (n=9) independently read the transcript segments to identify and label categories of content/meaning. The team then met to discuss the content categories, to combine similar categories, and further refine the code list. This final version of the list was used to analyze transcripts for the current study.

In the current study, a team of 13 additional research assistants (who had not participated in the pilot work) coded segmented transcripts using the previously compiled code list. The team was instructed to label the segmented lines of the transcript with the codes that they believed were depicted. The team was advised that not every line needed to be coded and that some lines may depict more than one code. Each segmented transcript was coded by two research assistants independently in order to assess inter-rater reliability. The experimenter served as the arbitrator if there was a discrepancy in coding. Finally, the codes were examined by the authors utilizing a general inductive approach [3] to identify higher-level descriptive themes.

Results

Participant Characteristics

Participant characteristics can be found in Table 1. Participants were predominantly Caucasian (98%, 77/79) with an overall mean age of 63.97 years (SD 7.68). Most participants were highly educated with all having achieved some college, while 52% (41/79) had earned a post-graduate degree. In addition, most participants earned between US$50,000 and $75,000 per year (35%, 28/79). Participants were healthy, experiencing a mean of 3.11 (SD 0.55) physical symptoms in the past three weeks (out of 14 total symptoms) and a mean of 2.58 (SD 1.59) health conditions in their lifetime (out of 17 total conditions).

In terms of computer experience, the average participant had owned a home computer for almost 20 years (mean 18.17 years, SD 8.14). In addition, most participants used their home computer, on average, almost 20 hours per week (mean 18.77 years, SD 13.33). Finally, most participants had previously had experience with the Internet tool to which they were assigned (63%, 45/72).

Table 1. Participant characteristics (n=79).

Characteristics		n (%) / mean (SD)
Age, years		63.97 (7.23)
Gender		
	Male	31 (39.24%)
	Female	48 (60.76%)
Income (USD)		
	Less than $15,000 per year	3 (3.80%)
	$15,000-25,000 per year	6 (7.59%)
	$25,000-50,000 per year	21 (26.58%)
	$50,000-75,000 per year	28 (35.44%)
	$75,000 or more per year	21 (26.58%)
Education		
	Some high school	---
	High school graduate	---
	Some college	10 (12.66%)
	Associate’s degree	7 (8.86%)
	Bachelor’s degree	21 (26.58%)
	Post-graduate degree	41 (51.90%)
Number of recent physical symptoms		3.12 (2.37)
Number of lifetime health conditions		2.58 (1.59)
Years of computer ownership		18.17 (8.14)
Hours of home computer use per week		18.77 (13.33)
Familiar with Internet Tool (n=72) [a]		
	Yes	45 (62.50%)
	No	27 (37.50%)

[a] Seven participants failed to respond to this interview question and were not included in analyses regarding familiarity.

Accuracy of Diagnosis

The characteristics of accurate and inaccurate participants can be found in Table 2. Less than half of participants came to an accurate diagnosis for the illness vignette symptoms during their search (41%, 32/79). Participants who accurately diagnosed the symptoms were similar in gender, yearly income, education, years of home computer ownership, and familiarity with the Internet tool to those participants who were inaccurate in their diagnosis (see Table 2). Participants who accurately diagnosed the symptoms appeared to be slightly younger (mean 61.72 years, SD 6.17) than those who were inaccurate (mean 65.51 years, SD 7.54). In addition, those who were accurate reported an average of 3.54 (SD 2.53) recent physical symptoms and 3 (SD 1.34) lifetime health conditions as compared to the 2.83 (SD 2.23) recent symptoms and 2.28 (SD 1.69) lifetime health conditions that those who were inaccurate reported. Finally, participants who were accurate in their diagnosis used their home computers for 22.94 (SD 16.68) hours per week as compared to the 15.93 (SD 9.66) hours per week of those who were inaccurate.

Table 2. Participant and study characteristics by accuracy of diagnosis.

Characteristics		Accurate diagnosis (n=32) n (%) / mean (SD)	Inaccurate diagnosis (n=47) n (%) / mean (SD)
Search Method			
	Google	16 (50.00%)	25 (53.25%)
	WebMD Symptom Checker	16 (50.00%)	22 (46.75%)
Illness Vignette			
	Mononucleosis	19 (59.38%)	18 (38.30%)
	Scarlet Fever	13 (40.63%)	29 (61.70%)
Age		61.72 (6.17)	65.51 (7.54)
Gender			
	Male	10 (31.25%)	21 (44.68%)
	Female	22 (68.75%)	26 (55.32%)
Income			
	Less than $15,000 per year	3 (9.38%)	--
	$15-25,000 per year	3 (9.38%)	3 (6.38%)
	$25-50,000 per year	5 (15.63%)	16 (34.04%)
	$50-75,000 per year	14 (43.75%)	14 (29.80%)
	$75,000 or more per year	7 (21.90%)	14 (29.79%)
Education			
	Some high school	---	---
	High school graduate	---	---
	Some college	3 (9.38%)	7 (14.89%)
	Associate’s degree	5 (15.63%)	2 (4.26%)
	Bachelor’s degree	10 (31.25%)	11 (23.40%)
	Post-graduate degree	14 (43.75%)	27 (57.45%)
Number of recent physical symptoms		3.54 (2.53)	2.83 (2.23)
Number of lifetime health conditions		3.01 (1.34)	2.28 (1.69)
Years of computer ownership		19.66 (9.22)	17.16 (7.24)
Hours of home computer use per week		22.94 (16.68)	15.93 (9.66)
Familiar with Internet Tool (n=72) [a]			
	Yes	16 (50.00%)	29 (61.70%)
	No	12 (37.50%)	15 (31.90%)

[a] Seven participants failed to respond to this interview question and were not included in analyses regarding familiarity.

Descriptive Findings From Think-Aloud Content

Overview

Descriptive findings from the 19 codes are presented below and in Tables 3 and 4. From the think-aloud content, we identified three major areas related to online symptom diagnosis: (1) Internet tool navigation, (2) symptom information processing, and (3) diagnostic strategies.

Table 3. Think-aloud content codes and participant endorsement (n=79).

Code		Description	Percentage of participants expressing code, n (%)
Navigation			
	Web orientation	Comments about the layout or features of the website	70 (88.61%)
	Web navigation	Direct actions taken on the computer	78 (98.73%)
	Internet problem	Trouble or issue with the computer application	64 (81.01%)
Symptom information processing			
	Reading	Reading directly from the vignette or Web screen	78 (98.73%)
	Paraphrasing	Stating information found in the vignette or Web screen	79 (100.00%)
	Judgment of relevancy	Deciding whether to use information or not	72 (91.14%)
	Credibility	Discussing the source of information or trust in information	22 (27.85%)
	Confusion	Questions or statements that reflect confusion about content	44 (55.70%)
	Discussing unknowns	Talking about information that is unknown or uncertain	68 (86.08%)
	Lack of confidence	Uncertainty in a diagnosis or not knowing enough to make specific diagnosis	30 (37.97%)
Diagnostic strategy			
	Action plan	Stating an action that could be taken to achieve the goal of diagnosing	74 (93.67%)
	Hypothesis	Making a guess about what the diagnosis could be	78 (98.73%)
	Symptom	Selecting a specific symptom from the vignette on which to focus and search for	76 (96.20%)
	Confirmation	Matching the symptoms in the story with information about a particular diagnosis	58 (73.42%)
	Negation	A difference between the symptoms in the story and a particular diagnosis (mismatch)	72 (91.14%)
	Previous experience	Relating the symptoms or diagnosis to personal experiences	35 (44.30%)
	Previous knowledge	Relating the symptoms or diagnosis to medical information previously known	55 (69.62%)
	Cause	A potential cause of the illness (eg, a virus or germ)	41 (51.90%)
	Suggested action	Discussing potential actions for the symptoms	40 (50.63%)

Table 4. Think-aloud content codes by accuracy of diagnosis.

Code		Accurate diagnosis Participants expressing code (n=32)	Inaccurate diagnosis Participants expressing code (n=47)
Navigation			
	Web orientation	29 (90.63%)	45 (95.74%)
	Web navigation	32 (100.00%)	46 (97.87%)
	Internet problem	24 (75.00%)	40 (85.11%)
Symptom information processing			
	Reading	32 (100.00%)	46 (97.87%)
	Paraphrasing	32 (100.00%)	47 (100.00%)
	Cause	16 (50.00%)	25 (53.19%)
	Judgment of relevancy	27 (84.38%)	45 (95.74%)
	Credibility	9 (28.13%)	13 (27.66%)
	Confusion	16 (50.00%)	28 (59.57%)
	Discussing unknowns	28 (87.50%)	40 (85.11%)
	Suggested action	12 (37.50%)	28 (59.57%)
	Lack of confidence	9 (28.13%)	21 (44.68%)
Diagnostic strategy			
	Action plan	29 (90.63%)	45 (95.74%)
	Hypothesis	32 (100.00%)	46 (97.87%)
	Symptom	30 (93.75%)	46 (97.87%)
	Confirmation	24 (75.00%)	34 (72.34%)
	Negation	29 (90.63%)	43 (91.49%)
	Previous experience	13 (40.63%)	22 (46.81%)
	Previous knowledge	23 (71.88%)	32 (68.09%)

Internet Tool Navigation

Analysis of the think-aloud content showed that participants frequently commented on issues surrounding the use and navigation of the Internet tools. For example, participants seemed highly focused on the actions that they were taking on the computer (eg, “Type that in and hit Enter”); “web navigation” was the second most frequently identified code (14.34%, 1472/10,262) with almost all participants commenting on navigation (99%, 78/79). In addition, many participants also made comments about the layout or features of the website that they were visiting (eg, “Well, here’s a tool from the Mayo Clinic”) to orient themselves to visited websites. Finally, many participants mentioned difficulty with the computer programs, either not knowing how to navigate them or not knowing how to troubleshoot after an error message (81% of participants, 64/79; eg, “Oh, where, where did Question B go? I don’t know where Question B is. What happened there? Umm, am I at the top of Question B?”).

Symptom Information Processing

Participants made comments that indicated attempts to organize or evaluate the symptom and illness information encountered online. For example, participants analyzed the usefulness of the online information (eg, “Well, darn, that’s not gonna help”), followed by stating what information they were lacking (eg, “Um, but I don’t know how old this particular person is”). However, only a quarter of participants commented on the source or credibility of the online information (eg, “And the page I’m looking at, MedicineNet.com, that looks very reliable”). Diagnosing the vignette also appeared to be a difficult task for participants as about half of participants demonstrated confusion about how to attempt to diagnose the vignette symptoms (eg, “I’m kind of at a loss where to go now”). In addition, some participants also seemed hesitant to make a diagnosis and demonstrated that they were not confident in the diagnosis that they had settled upon (eg, “I can’t diagnose this by myself”).

Diagnostic Strategy

Participants utilized a number of strategies to attempt to diagnose the symptoms. For example, many participants planned the steps that they would take to diagnose before implementing those actions (eg, “So I guess what I will do is, uh, try to think of something that Google will be interested in trying to answer”). Many participants also focused on only one symptom at a time, inputting each symptom into the Internet tool separately, rather than attempting to diagnose the entire collection of symptoms at once (eg, “High fever…I’m going to put this in quotes”). Some participants considered the cause of the vignette symptoms (eg, “Could be bacterial, could be viral”) or suggested some sort of action that should be taken in response to the symptoms such as going to the doctor or asking for antibiotics. Participants most frequently relied on a strategy where they reviewed and then rejected potential diagnoses that contained additional symptoms than those that were depicted in the vignette (eg, “No, this person is not short of breath”). This strategy was used more than confirming potential diagnoses by comparing whether the illness/condition information contained symptoms that, in fact, matched the vignette symptoms (eg, “It fits some of it. High fever and lymph nodes”). About two-thirds of participants utilized lay/existing medical knowledge to diagnose the symptoms, even while using the computer simultaneously (eg, “Ummm, colon polyps, that’s not symptomatic”). In addition, approximately half of participants described memories of previous experiences with the symptoms and illness to aid in diagnosis (44%, 35/79; eg, “Been there, done that, um, so I had it as a kid”). However, these strategies were identified less frequently than the matching strategies discussed above.

Accuracy and Think-Aloud Content

We then compared the think-aloud content between participants who accurately diagnosed the symptoms and those who did not (see Table 4). Participants who inaccurately diagnosed the symptoms seemed to express more difficulty with the diagnosis task. For example, 60% (28/47) of inaccurate participants mentioned confusion about the task as compared to only half of accurate participants (50%, 16/32). Similarly, 45% (21/47) of inaccurate participants lacked confidence in their diagnosis as compared to 28% (9/32) of accurate participants. In addition, participants who were inaccurate appeared to have more difficulty navigating the computer: 85% (40/47) of inaccurate participants mentioned having an Internet problem or difficulty with the Internet tool as compared to 75% (24/32) of accurate participants.

In terms of diagnostic strategy, there was little difference between the proportion of inaccurate and accurate participants who utilized a confirmation strategy (inaccurate: 72%, 34/47 vs accurate: 75%, 24/32) or a negation strategy (inaccurate: 92%, 43/47 vs accurate: 91%, 29/32). However, inaccurate participants seemed to rely more on their previous experiences with illness (47%, 22/47) than accurate participants (41%, 13/32). In addition, inaccurate participants seemed less likely to utilize previous medical knowledge (68%, 32/47) than accurate participants (72%, 23/32).

Discussion

Processes of Online Self-Diagnosis

In this think-aloud protocol, the process of older adult online self-diagnosis was explored. Older adult participants frequently commented on navigating the websites visited. Participants also organized information by considering what else they would need to know or whether information encountered was useful. Most participants tended to diagnose physical symptoms through a matching process, utilizing information encountered online.

Our results most aligned with hypothetico-deductive reasoning strategies where participants utilized the additional health information available online to confirm or reject various illnesses/conditions. However, some participants also relied on existing medical knowledge to diagnose the symptoms, noting potential causes of symptoms and treatment-seeking actions, as well as recalled previous personal experiences with the symptoms, which informed their diagnosis. These findings are similar to the common-sense model.

Interestingly, more participants who were inaccurate in their online symptom diagnosis mentioned previous experiences with illness than those who were accurate. According to the common-sense model, symptom interpretation is typically undertaken with the help of heuristics or automatic rules [19,36]. Use of heuristics allows laypeople to interpret symptoms more quickly and with less cognitive effort. Heuristics although helpful, often lead to erroneous conclusions [37]. Because the symptoms of the illness vignettes were relatively common, it is possible that inaccurate participants were misled by memories of previous experiences with illness that showed similar, yet distinct patterns of symptoms. Misdiagnosing common symptoms through use of heuristics is similar to the “pattern rule” of the common-sense model where diffuse symptoms are more susceptible to interpretation errors [36]. In contrast, participants who were accurate in their diagnosis may have been relying on more effortful comparison of vignette symptoms and online symptom information. This more effortful matching, even if initially guided by past illness experience, may have been better informed and less reliant on heuristics as these participants were deploying more cognitive resources. Thus, online self-diagnosis may disrupt the reliance on heuristics for symptom interpretation as typically described by the common-sense model. Future work may want to examine what factors predict more effortful processing of symptoms and whether online self-diagnosis can encourage such processing.

Participants who were inaccurate in their online symptom diagnosis also had more difficulty with both the task and the Internet tools. In addition, these participants reported using their home computer for fewer hours per week than accurate participants. Thus, it may be that basic computer skills are predictive of the ability to obtain an accurate online diagnosis. However, inaccurate participants reported being more familiar with the Internet tool to which they were assigned, which would typically suggest experience and more skill. In addition, Sharit and colleagues [24] found that Internet knowledge was related to performance on an information-seeking task although not sufficient to explain performance. Future work may want to obtain measures of performance (eg, speed of diagnosis, cognitive abilities) and computer skills to clarify predictors of accuracy of online symptom diagnosis.

Of note, few participants focused on the source or credibility of the information that they were reviewing, consistent with previous research findings on young and middle-aged adults [13,38]. This may be due to a focus on other activities like Web navigation, which was frequently commented on by participants. Other researchers [39] have used eye-tracking technology to investigate where older adults focus attention during an online health search. This may help to elucidate why credibility was ignored or not verbalized in our sample.

Implications for Web Design

These findings point to the need for changes and/or updates to current popular health websites. For example, because older adult participants appeared to focus on navigating websites and Web applications, developers may want to update webpages or Web tools with clear navigation aids that guide users as to how the page is structured, how the user can backtrack (ie, return to an earlier viewed page), and how the search bar can best be used. Furthermore, webpages with visual summaries of information, such as tables or figures, may help to decrease effort, allowing users to focus their energies on other aspects of information search.

As few older adult participants noted the credibility of the information source, this seems an important target area. Though past attempts have been made to create user tools for ascertaining the quality and credibility of online information (eg, check lists, website “branding”) [40,41], more effort needs to be made to advise lay searchers of these tools and encourage their use. This would help to ensure that users are gaining access to accurate information via credible sources.

Limitations

A think-aloud investigation of older adult online health information-seeking produced themes that related to layperson diagnostic strategies, symptom information processing, and especially Web navigation. However, there are factors that limit the generalization of the findings. First, the sample of older adults was predominantly Caucasian, highly educated, and of a comfortable income. In addition, all participants had access to a computer and the Internet at home, suggesting that they already possessed basic computer and Internet skills. While this does limit the generalizability of our findings, previous surveys [6,42] confirm that the majority of adults who search for health information online tend to be of similar backgrounds as our participants (eg, Caucasian, educated, and of higher income). In addition, three-fourths of participants were not familiar with the Internet tool to which they were assigned, and thus, were experiencing the tool for the first time. Nevertheless, a study of older adults in different socioeconomic and geographic locations may demonstrate more variability in the strategies used to find online health information to diagnose symptoms. Furthermore, the study was performed on a university campus that is home to a comprehensive medical center. Thus, our sample of older adults has consistent access to medical care and so may not typically need to search for diagnosis. This may not be true for rural older adults who might lack an easily accessed source of care. Additional investigation as to the type of older adult who engages in online diagnosis may prove beneficial.

Conclusions

This exploratory study investigated the process of older adult online symptom diagnosis. Few studies have systematically examined this recent phenomenon, especially among older adults. Our findings suggest that, in our sample, older adults tend to rely on hypothetico-deductive matching to diagnose physical symptoms but still may utilize existing medical knowledge and illness experiences to guide diagnosis. This may be because navigating websites and Web tools is a cognitively complex task, providing older adults few resources to sort through the extensive amount of health information online. Thus, additional Web development is necessary to make online search more efficient and accurate for older adult users. In addition, we found that few older adult participants mentioned the credibility of the information that they were viewing. Increased dissemination of previously produced Web tools would be beneficial to ensuring that older adults can access the most appropriate information.

Given the popularity of online self-diagnosis, this study represents the first of its kind in attempting to describe the process that an older patient takes for symptom interpretation. In addition, we focused on a population that is less frequently represented in human-computer interaction studies. While our study provides an initial picture of how some older adults might attempt online self-diagnosis, future work will want to utilize additional study design such as eye-tracking in order to further understand the complex coordination between Web navigation, online symptom information processing, and patient diagnostic strategies.

Acknowledgments

Dr Luger would like to acknowledge the University of Iowa Department of Psychology and the College of Nursing, which provided funding for the apparatus used in the study. She would also like to thank Drs Timothy Hogan and Gemmae Fix for their review of subsequent versions of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Illness vignettes.

[PDF File (Adobe PDF File), 23KB - jmir_v16i1e16_app1.pdf]

References

1. Campbell RJ. Internet-based health information seeking among low-income, minority seniors living in urban residential centers. Home Health Care Management & Practice 2008 Jul 08;21(3):195-202. [doi: 10.1177/1084822308322648]

2. Macias W, McMillan S. The return of the house call: the role of Internet-based interactivity in bringing health information home to older adults. Health Commun 2008 Jan;23(1):34-44. [doi: 10.1080/10410230701805174] [Medline: 18443991]
3. Zickuhr K, Madden M. Older adults and Internet use. 2012 URL: http://www.pewInternet.org/~/media/Files/Reports/2012/PIP_Older_adults_and_Internet_use.pdf [accessed 2013-08-29] [WebCite Cache ID 6JEfuSXhl]
4. McMillan SJ, Avery EJ, Macias W. From have nots to watch dogs. Information, Communication & Society 2008 Aug;11(5):675-697. [doi: 10.1080/13691180802126745]
5. Hilt ML, Lipschultz JH. Elderly Americans and the Internet: E-mail, TV news, information and entertainment websites. Educational Gerontology 2004 Jan;30(1):57-72. [doi: 10.1080/03601270490249166]
6. Fox S, Duggan M. 2013. Health Online 2013 URL: http://www.pewInternet.org/~/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2013-08-29] [WebCite Cache ID 6JEfgLrwR]
7. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with Internet-based health information? J Med Internet Res 2006 Sep;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]
8. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
9. 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]
10. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
11. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
12. Cashen MS, Dykes P, Gerber B. eHealth technology and Internet resources: barriers for vulnerable populations. J Cardiovasc Nurs 2004;19(3):209-14; quiz 215. [doi: 10.1097/00005082-200405000-00010] [Medline: 15191264]
13. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Viera A, Crotty K, et al. Health literacy interventions and outcomes: an updated systematic review. Evid Rep Technol Assess (Full Rep) 2011 Mar(199):1-941. [Medline: 23126607]
14. Leventhal H, Safer MA, Panagis DM. The impact of communications on the self-regulation of health beliefs, decisions, and behavior. Health Educ Q 1983;10(1):3-29. [doi: 10.1177/109019818301000101] [Medline: 6629788]
15. Leventhal H, Cameron L. Behavioral theories and the problem of compliance. Patient Education and Counseling 1987 Oct;10(2):117-138. [doi: 10.1016/0738-3991(87)90093-0]
16. Leventhal H, Leventhal EA, Contrada RJ. Self-regulation, health, and behavior: A perceptual-cognitive approach. Psychology & Health 1998 Jul;13(4):717-733. [doi: 10.1080/08870449808407425]
17. Cameron LD, Leventhal H. The self-regulation of health and illness behaviour. London: Routledge; 2003.
18. Pennebaker JW. The psychology of physical symptoms. New York: Springer-Verlag; 1982.
19. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cognitive Science 1986;10(1):91-116. [doi: 10.1207/s15516709cog1001_4]
20. Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. Med Educ 2003 Aug;37(8):695-703. [doi: 10.1046/j.1365-2923.2003.01577.x]
21. Elstein A, Schulman L, Sprafka S, Allal L. Medical problem solving: an analysis of clinical reasoning. Cambridge, Mass: Harvard University Press; 1978.
22. Sharit J, Hernández MA, Czaja SJ, Pirolli P. Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the Web. ACM Trans Comput Hum Interact 2008 May;15(1):3 [FREE Full text] [doi: 10.1145/1352782.1352785] [Medline: 20011130]
23. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
24. McKeown B, Thomas D. Q methodology. Newbury Park, Calif: Sage Publications; 1988.
25. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. American Journal of Evaluation 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
26. Ericsson KA, Simon HA. Protocol analysis: verbal reports as data. Cambridge, Mass: MIT Press; 1993.
27. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975 Oct;23(10):433-441. [Medline: 1159263]
28. Mayo Clinic. URL: http://www.mayoclinic.com/ [accessed 2013-08-29] [WebCite Cache ID 6JEhKIxW3]
29. National Institute of Allergy and Infectious Diseases. URL: http://www.niaid.nih.gov/Pages/default.aspx [accessed 2013-08-29] [WebCite Cache ID 6JEiC1AbB]
30. Google. URL: http://www.google.com [accessed 2013-08-29] [WebCite Cache ID 6JDqpHXj4]
31. WebMD Symptom Checker. URL: http://symptoms.webmd.com/ [accessed 2013-08-29] [WebCite Cache ID 6JDqxmTmd]
32. Alexa.com. Google.com Site Info. URL: http://www.alexa.com/siteinfo/google.com [accessed 2013-08-29] [WebCite Cache ID 6JDqVLYLD]

33. Google. About the Company - Google URL: http://www.google.com/intl/en/about/company/ [accessed 2013-08-29] [WebCite Cache ID 6JDqlvWbK]
34. Larsen RJ. Neuroticism and selective encoding and recall of symptoms: evidence from a combined concurrent-retrospective study. J Pers Soc Psychol 1992 Mar;62(3):480-488. [doi: 10.1037/0022-3514.62.3.480] [Medline: 1560338]
35. Pecoraro RE, Inui TS, Chen MS, Plorde DK, Heller JL. Validity and reliability of a self-administered health history questionnaire. Public Health Rep 1979 Jun;94(3):231-238 [FREE Full text] [Medline: 451132]
36. Scott SE, Walter FM, Webster A, Sutton S, Emery J. The model of pathways to treatment: conceptualization and integration with existing theory. Br J Health Psychol 2013 Feb;18(1):45-65. [doi: 10.1111/j.2044-8287.2012.02077.x] [Medline: 22536840]
37. Kahneman D, Tversky A. Subjective probability: a judgment of representativeness. Cognitive Psychology 1972;3(3):430-454. [doi: 10.1007/978-94-010-2288-0_3]
38. Eastin MS. Credibility assessments of online health information: The effects of source expertise and knowledge of content. Journal of Computer-Mediated Communication 2001;6(4). [doi: 10.1111/j.1083-6101.2001.tb00126.x]
39. Kules B, Xie B. Older adults searching for health information in MedlinePlus - an exploratory study of faceted online search interfaces. Proc. Am. Soc. Info. Sci. Tech 2012 Jan 11;48(1):1-10. [doi: 10.1002/meet.2011.14504801137]
40. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
41. Fritch JW. Heuristics, tools, and systems for evaluating Internet information: helping users assess a tangled Web. Online Information Review 2003;27(5):321-327. [doi: 10.1108/14684520310502270]
42. Fox S. 2011. Health Topics URL: http://www.pewInternet.org/~/media//Files/Reports/2011/PIP_HealthTopics.pdf [accessed 2013-08-29] [WebCite Cache ID 6JEh9KlAp]

Edited by G Eysenbach; submitted 29.08.13; peer-reviewed by S Knight, M Masoni; comments to author 30.10.13; revised version received 14.11.13; accepted 15.11.13; published 16.01.14.

Please cite as:
Luger TM, Houston TK, Suls J
Older Adult Experience of Online Diagnosis: Results From a Scenario-Based Think-Aloud Protocol
J Med Internet Res 2014;16(1):e16
URL: http://www.jmir.org/2014/1/e16/
doi:10.2196/jmir.2924
PMID:24434479

Original Paper

Evaluating a Web-Based Health Risk Assessment With Tailored Feedback: What Does an Expert Focus Group Yield Compared to a Web-Based End-User Survey?

Sandra Vosbergen[1], MSc; Guy R Mahieu[1], BSc; Eva K Laan[2], MSc; Roderik A Kraaijenhagen[3], MD, PhD; Monique WM Jaspers[1,4], PhD; Niels Peek[1], PhD

[1]Academic Medical Center, Department of Medical Informatics, Amsterdam, Netherlands

[2]Academic Medical Center, Department of Public Health, Amsterdam, Netherlands

[3]NIPED Research Foundation, Amsterdam, Netherlands

[4]Academic Medical Center, Center for Human Factors Engineering of Health Information Technology, Amsterdam, Netherlands

Corresponding Author:
Niels Peek, PhD
Academic Medical Center
Department of Medical Informatics
Room J1b-110.1
PO Box 226600
Amsterdam, 1100DD
Netherlands
Phone: 31 20 5667872
Fax: 31 20 6919840
Email: n.b.peek@amc.uva.nl

Abstract

Background: Increasingly, Web-based health applications are developed for the prevention and management of chronic diseases. However, their reach and utilization is often disappointing. Qualitative evaluations post-implementation can be used to inform the optimization process and ultimately enhance their adoption. In current practice, such evaluations are mainly performed with end-user surveys. However, a review approach by experts in a focus group may be easier to administer and might provide similar results.

Objective: The aim of this study was to assess whether industrial design engineers in a focus group would address the same issues as end users in a Web-based survey when evaluating a commercial Web-based health risk assessment (HRA) with tailored feedback.

Methods: Seven Dutch companies used the HRA as part of their corporate health management strategy. Employees using the HRA (N=2289) and 10 independent industrial designers were invited to participate in the study. The HRA consisted of four components: (1) an electronic health questionnaire, (2) biometric measurements, (3) laboratory evaluation, and (4) individually tailored feedback generated by decision support software. After participating in the HRA as end users, both end users and designers evaluated the program. End users completed an evaluation questionnaire that included a free-text field. Designers participated in a focus group discussion. Constructs from user satisfaction and technology acceptance theories were used to categorize and compare the remarks from both evaluations.

Results: We assessed and qualitatively analyzed 294 remarks of 189 end users and 337 remarks of 6 industrial designers, pertaining to 295 issues in total. Of those, 137 issues were addressed in the end-user survey and 148 issues in the designer focus group. Only 7.3% (10/137) of the issues addressed in the survey were also addressed in the focus group. End users made more remarks about the usefulness of the HRA and prior expectations that were not met. Designers made more remarks about how the information was presented to end users, quality of the feedback provided by the HRA, recommendations on the marketing and on how to create more unity in the design of the HRA, and on how to improve the HRA based on these issues.

Conclusions: End-user surveys should not be substituted for expert focus groups. Issues identified by end users in the survey and designers in the focus group differed considerably, and the focus group produced a lot of new issues. The issues addressed in the focus group often focused on different aspects of user satisfaction and technology acceptance than those addressed by the survey participants; when they did focus on the same aspects, then the nature of issues differed considerably in content.

(J Med Internet Res 2014;16(1):e1) doi:10.2196/jmir.2517

KEYWORDS

health risk assessment; health information systems; qualitative research; evaluation; end users; professional review; designers; optimization

Introduction

Over the last decade, there has been increasing interest in Web-based health applications for the prevention and management of chronic diseases. However, significant problems have been reported with the reach and utilization of these health applications [1,2]. According to Wixom and Todd [3], certain factors influence the use of information technology. For example, users of information technology must find it relatively advantageous, easy to use, and compatible with their beliefs and attitudes [3]. To increase the adoption of an innovation, we need insight into end users' real-life experiences and the potential issues to be solved. Various formative evaluation methods to collect this kind of information are available.

A previous evaluation of a Web-based health risk assessment (HRA) with tailored feedback, which will be the focus of this study, showed that only 33.7% of the 6790 employees invited to participate in the HRA did so [4,5]. To optimize this HRA, two evaluations were performed independently. After participation in the HRA as end users, user feedback was collected with a Web-based survey, and expert feedback was collected in a focus group with industrial design engineers. These evaluation techniques are often deployed in practice after a system's implementation, as they are relatively efficient and easy to administer. However, until now little research has been done to support the claim that end-user surveys are the best option to get insight into experiences and issues influencing adoption of health innovations. Our main question was whether an expert focus group would bring up similar issues to those addressed in the end-user survey. The results of the Web-based survey were published previously in [5]. If a focus group with industrial design engineers would come up with similar issues as revealed by the Web-based survey, a strong point would be made for collecting feedback through this method in a pilot implementation. If these issues were solved in a system's redesign before its full implementation, end users may perceive the innovation as more valuable, ultimately increasing its adoption rate.

Studies in the field of usability have already shown that, along with end users, usability experts can contribute significantly to the development process of applications because they can give a detailed analysis of potential problems [6] and identify their causes [7]. Research in the fields of usability and text evaluation has generally shown that the types of problems revealed in end users' and experts' evaluations [7-11] and by different evaluation strategies [8,12-16] only slightly overlap. However, these studies mostly focused on revealing usability or text understanding issues before implementation of a product, in a laboratory setting, or a study setting. Usability is only one aspect that influences the acceptance of a system, as it mainly focuses on system function [17]; whereas, it is the total user experience that influences the final adoption of an innovation through users' perception of the value and service quality [18]. To enhance the adoption of innovations, we need broad-based information about user experiences that influence their satisfaction with and acceptance of its design post-implementation.

In this study, we compared two commonly used methods to inform the optimization process of a system after its implementation: end-user surveys and expert focus groups. We assessed whether those two methods would produce similar feedback on issues affecting user satisfaction and technology acceptance. Surveys among large groups of end users are generally performed to provide information on end users' overall satisfaction with a health care innovation and its effect on their health-related behavior [19-21]. These surveys generally include an open-text field to provide additional space for users to describe their experiences. Industrial design engineers, instead of Web design evaluators, were invited as focus group members because their expertise lies in applying approaches focused on improving user experiences and exploring—and finding solutions to—problems in a broad spectrum of design aspects ranging from esthetics and ergonomics to human-product interaction, user needs research, and usability [22]. Therefore, they could provide a broad view of issues regarding the complete HRA user experience, including those parts not directly associated with the Web-based part of the HRA. Focus groups were chosen because industrial design engineers are familiar with this technique as most of their work is performed in teams. Also, focus groups are a well-established technique in market research for the design and optimization of new innovations (eg, [23]), as well as for human factor research and usability evaluation [24].

As a common denominator to lump the results of the evaluations of industrial design engineers and end users together, we used constructs from user satisfaction and technology acceptance theories [3]. These theories describe various domains that potentially influence the end user's attitude towards a system such as information quality, service quality, organizational quality, system quality, end users' outcome expectations, ease of use, and usefulness of a product.

The aim of this study was to assess whether industrial design engineers in a focus group would address the same issues as end users in a Web-based survey when evaluating a commercial Web-based HRA with tailored feedback. Additionally, we aimed to gain insight into the kinds of issues addressed in the focus group and end-user survey, using constructs of user satisfaction and technology acceptance. This study is a comparison of the yield of two evaluation techniques used as they commonly are in practice, not as a traditional experiment.

Methods

Description of the Web-Based Health Risk Assessment

The focus of this study is on the evaluation of a Web-based HRA program. HRAs collect health-related data that are used to evaluate an individual's health status and health risks, typically screening for risk factors of chronic diseases and identifying related health behaviors [25]. They can be used to assess average health risks in a group or population or to provide tailored feedback on health risks to individuals in an effort to help them reduce their risks. In the latter case, the primary goal is to prevent high-incidence chronic diseases such as coronary heart disease, cancer, and diabetes. HRAs used in the workplace have shown promising results in the prevention of chronic diseases [25-28]. Nowadays, HRAs are often disseminated via Web-based technology. Web-based HRAs can reach more people, thus facilitating the impact of HRAs on public health [29].

The Web-based HRA with tailored feedback evaluated in this study consisted of four components completed in the following sequence: (1) a Web-based electronic health questionnaire, (2) biometric evaluation, (3) laboratory evaluation, and (4) an individually tailored feedback report based on the results of the first three components provided in a Web-based environment. The aim of the health questionnaire was to capture data on a participant's health and lifestyle. It consisted of questions concerning sociodemographic variables, family and personal medical history, health complaints, psychological functioning, lifestyle behavior, and perceived health perception. The biometric data that were collected consisted of a participant's weight, height, waist circumference, and blood pressure measurements. Blood and urine samples were collected by certified health professionals at predetermined sites and analyzed at a lab. All procedures and components of the HRA were exactly the same for both end users and industrial design engineers, except for the way in which participants' data on weight, height, waist circumference, and blood pressure was collected. The end users' biometric data was collected at the predetermined sites by certified health professionals at the same time as the blood and urine samples. Industrial design engineers collected this data themselves at home using a toolbox provided by the HRA.

The tailored feedback report was automatically generated by a computerized decision support system (CDSS) after the participant had completed all HRA components in full. The report described the individual's overall health risk (expressed by a compass metaphor and a four-color system), results on health risks in five categories (behavioral, psychological, physical, personal medical history/family risk, and work-related) distinguished by a color system per category, an explanation of the risks and potential benefits of taking preventive action, and actionable advice for improving a participant's health. When the participant's health was "seriously off-track" (the color red), a referral to a health professional was provided for further medical evaluation and treatment. A more extensive description of the HRA, including examples and screenshots, can be found elsewhere [5].

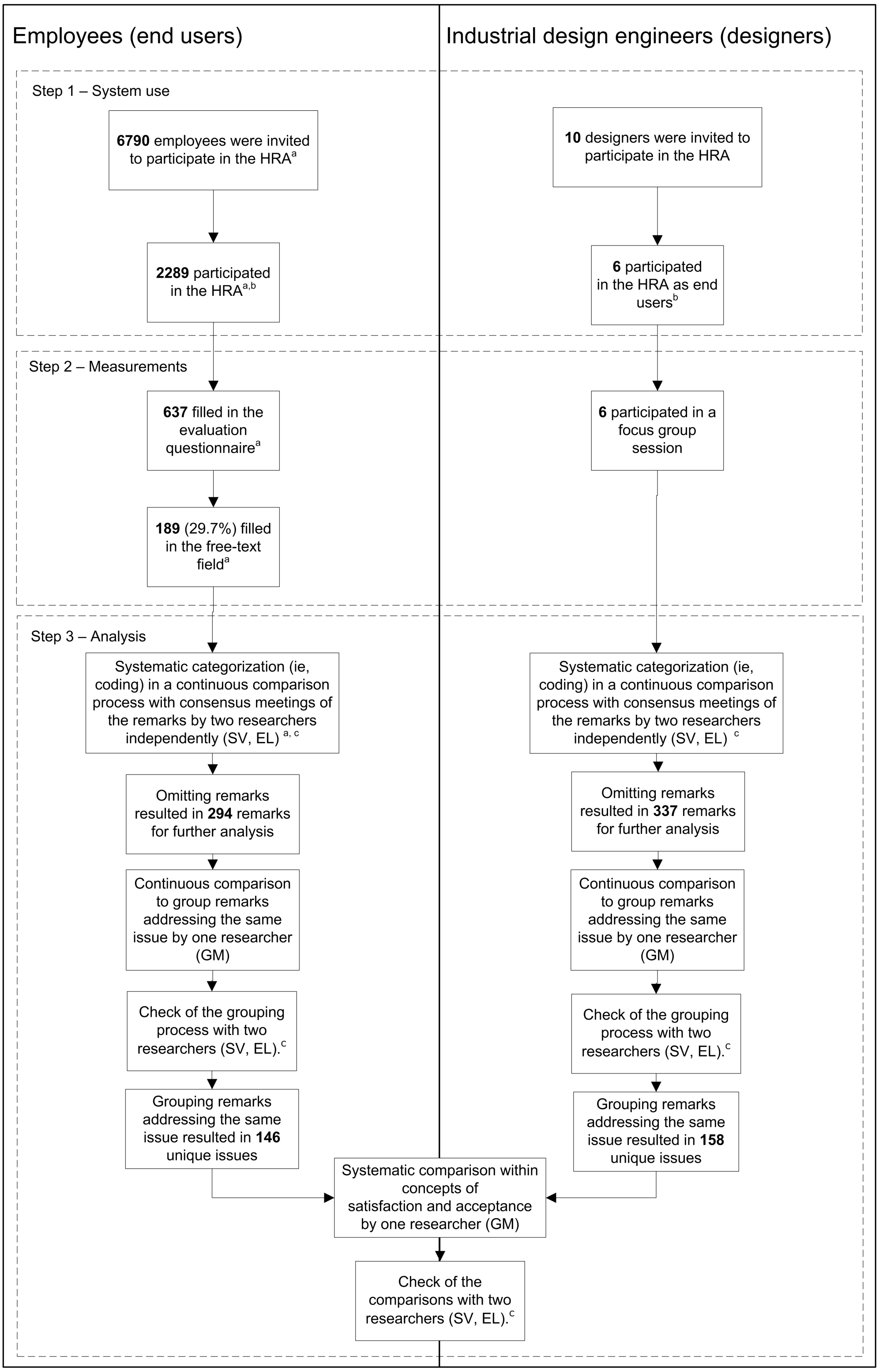
Design and Study Population

The study design is depicted in Figure 1. For the end user evaluation, the HRA was implemented in seven Dutch companies as part of their corporate health management strategy between 2007 and 2008. Employers initiated the HRA by sending invitations to their employees. A single reminder was sent 2 weeks later. These invitations consisted of a description of the HRA and a notification that participation was voluntary, paid for by the employer, all personal data would be treated confidentially, and no results would be shared with their employer or any other party. A Web-based survey was subsequently used to evaluate their experiences with the HRA. Participating employees are referred to as "end users".

Ten industrial design engineers were recruited via the researchers' own network to acquire the recommended group size of 6-10 people in focus group research [30]. All invited industrial design engineers had obtained a bachelor's and master's degree from the Faculty of Industrial Design Engineering at the Delft University of Technology (DUT) or were finishing such a master's degree. The bachelor and master programs at the Faculty of Industrial Design Engineering are a mixture of theory concerning various aspects of design and applied design projects in collaboration with the business community. The designers were invited to participate in the same HRA for evaluative purposes in 2009. The invitation included a description of the HRA, the goal of the research, and explained that they were first expected to take part in the HRA as an end user and then jointly evaluate the HRA in a focus group session afterwards. It also explained privacy aspects and that participation was free of charge. We refer to the industrial design engineers as "designers".

Before conducting the actual evaluation, all survey participants and focus group members participated in the Web-based HRA with tailored feedback as end users. To complete the HRA, only one use scenario is possible. It is therefore unlikely that the two evaluator groups differed in the time spent completing the HRA. No further instructions were given on what aspects of the program to evaluate or focus on during the evaluation because we wanted to evaluate the HRA in the broadest sense and aimed to see what issues both evaluation techniques would yield naturally. End users and designers did not have any experience with this type of program, as the HRA program evaluated in this study is the first in its kind in the Netherlands.

Figure 1. Study design and participant flow.



[a] These steps were part of the study by Vosbergen and Laan [5].

[b] Except for how biometric measurements were collected, the HRA and its procedure was exactly the same for both groups.

[c] If consensus was not reached another researcher was consulted and resolved the disagreement

Measurements

End Users

Six weeks after having participated in the HRA, end users received an evaluation questionnaire by email. This two-part questionnaire consisted of a quantitative and a qualitative section. The first part assessed end users' satisfaction with the HRA and the initiation of health behavior change after participation. In this section, respondents were asked to appraise the different components of the HRA and the overall satisfaction with the HRA on an ordinal ranking scale. Results concerning end users' satisfaction with the HRA are reported elsewhere [5]. Participants were also asked whether they initiated health behavior changes after participation and which behaviors they changed. These results are reported in [4].

The current study focuses on the free-text comments in the second part of the questionnaire. This free-text field was preceded by the following text: "It's possible that some things were not contained in the above questions, or that you weren't able to express these things as you would have liked. If this is the case, please enter them below." The outcomes of the detailed analysis of the remarks that were made by end users in this free-text field can be found in [5].

Designers

After completing the HRA as an end user, designers were asked to participate in a focus group discussion. The focus group meeting was led by a moderator and introduced as follows: "I actually want to start with your general impression of the HRA itself, for example how the process works, its strong and weak points. If you already have improvements in mind, you can also mention them." The entire focus group session was videotaped for evaluative purposes after receiving the participants' consent and was subsequently transcribed verbatim.

Data Analysis

Qualitative data analysis was aimed at comparing issues addressed in the end-user survey and in the designer focus group. To achieve sufficient reliability, our main goal was to code the remarks made in both evaluations exactly the same. To do so, an iterative coding, grouping and comparison process was applied.

A codebook was constructed to analyze all textual remarks made by end users and designers. This codebook was based on the domains and topics described in Table 1 of the article by Wixom and Todd (p. 88 [3]), which integrates constructs from user satisfaction and technology acceptance theories. These theories were developed to understand and evaluate factors explaining users' perceptions about information systems to assess actual usage of these systems. We decided to use all topics and domains, rather than the framework of Wixom and Todd [3], to evaluate the HRA in the broadest sense. The codebook was used to analyze all remarks made by end users and designers alike. The final codebook can be found in Multimedia Appendix 1.

The end users' remarks had already been categorized using the codebook for the study by Vosbergen and Laan et al [5]. In this section, we describe only the methods we applied for qualitative analysis of the transcripts of the focus group session with the designers, since remarks of both the end user and designer group were analyzed in exactly the same way. At the start of the analysis, the codebook contained all domains and topics described in Table 1 of Wixom and Todd [3]. During analysis, interpretations of these domains and topics were specified to the HRA. Two researchers (SV and EL) independently categorized remarks according to the following topic schemes: (1) the component of the HRA addressed, (2) the domain and topic from the codebook, and (3) whether the remark was positive, negative, neutral, or a recommendation. Based on these topic schemes, the remarks in the transcripts were divided into individual remarks during coding. For example, if a designer's remark was about both the electronic health questionnaire and the feedback report, it was divided into two separate remarks. If the remaining part of the transcript covered another topic from the codebook, the remarks in the transcript were divided again based on these topics.

End user and designer remarks were categorized in an iterative process, in which independent coding by the two researchers was alternated with consensus meetings. These meetings were aimed at resolving discrepancies so that remarks of end users and designers would be coded exactly the same. During these meetings, consensus was sought by comparing and discussing the individual codes assigned to every text excerpt until all discrepancies were solved and the categorizations and interpretation of the codebook were exactly the same. If consensus could not be reached, a third researcher (NP for end users, GM for designers) resolved the disagreement. After every change in interpretation of the codebook or in categorization of the remarks, all previously categorized data was checked for consistency with the decisions made during the preceding meeting. There were no remarks that did not pertain to any of the domains of the original codebook. If a remark did not pertain to one of the topics within the original codebook, a topic from another domain was considered or an appropriate new topic was added to the codebook. Adding a (new) topic to the codebook was always done in mutual agreement during a consensus meeting. The topics that were copied or added can be found in Multimedia Appendix 1.

Different remarks may or may not address the same issue. To identify the set of unique issues that were addressed, remarks that were syntactically different but semantically similar, meaning that they addressed the same issue, were grouped together. This was done in a bottom-up fashion. Initially, all remarks were viewed as singleton groups (ie, containing a single remark). Subsequently, groups of remarks addressing the same issue were joined. This was done recursively until all groups addressed different issues. For example, all remarks about printing the HRA's feedback report (eg, "Too bad the results can't be printed out so you can read through them at a time and place of your choosing", "Can't print out recommendations for healthier eating, a real pity...") were placed in one group. Remarks that had been categorized with different domains or topics were never merged into a single group, unless the coding was found to be inconsistent during the grouping process (see below). Remarks that had been categorized with the same domain and topic could be merged into one group, but only if they addressed the same issue. The grouping was performed separately for remarks made by end users and remarks made by designers.

Remarks about the procedures for collecting biometric data measurements were omitted from the dataset prior to analysis because they could not be compared due to differences between end users and designers. The grouping of equivalent remarks was performed by one researcher (GM). As remarks that had been categorized with different domains or topics were assumed never to be equivalent, combinations were first sought within topics. As a reliability check on the categorization of the data, combinations were then sought within domains, and finally across domains (ie, within the topics from user satisfaction and technology acceptance theories). If an inconsistency was found, then discrepancies were resolved in a consensus meeting with three researchers (GM, EL, SV). After this process was completed, 2 researchers (EL and SV) independently assessed the resulting groups of remarks, followed by a consensus meeting to resolve discrepancies. For each of the final groups, one remark was chosen that best portrayed the issue addressed by the group. This representative remark is given to illustrate the qualitative results. Finally, the issues identified in the end-user survey and designer focus group were compared qualitatively on equivalency in the same manner as described previously for the grouping process.

Results

Study Population

Figure 1 depicts the flow of the participants through the study and the number of resulting remarks and issues for both groups. End user response to the evaluation questionnaire was 27.8% (637/2289 participants in the HRA); 29.6% (189/637) of those end users made one or more remarks in the free-text field. Participants were mainly male (386/637, 60.6%) and an average age of 46.49 years (SD 8.76); 36.1% of these participants were aged 50 or older. 48.2% (307/637) of participants had a high education level, 30.0% (191/637) an intermediate education level, and 21.8% (139/637) had a low education level. Except for age, there were no significant differences in sociodemographics between respondents who made remarks in the free-text field compared to those who did not. Respondents who made remarks were nearly 2 years older (mean 47.8, SD 8.1 years) than those who did not make remarks (mean 45.9, SD 9.0 years). More information on end users' backgrounds and responses to the evaluation questionnaire is available in Vosbergen and Laan et al [5].

Furthermore, a total of 6/10 recruited designers participated in the focus group. Of those, 5 obtained their master's degree in the 12 months before the start of the study and the other was in the final phase of obtaining a master's degree. One designer obtained a degree in Strategic Product Design (corporate strategy and opportunities to develop sound product development portfolios), 4 obtained a degree in Design for Interaction (the way in which people interact with products; designing products appropriate to users' needs and expectations), and one was finishing the master curriculum Integrated Product Design (systematic approach of product design; designing innovative products and service combinations). Four of the designers had gained experience with either product design, product evaluation research, or sales and project management in a company during their studies.

Four of the six designers completed all steps of the HRA within 1 month. The other 2 were unable to complete the HRA before the focus group session due to time limitations (n=1) and fear of the needles used to collect the blood sample (n=1). Both designers, however, did complete all steps in the HRA; however, they did not receive their own risk profile and recommendations but received a fictitious health risk profile and fictitious health recommendations instead to be able to complete the evaluation. The designers' focus group session lasted around 3.5 hours.

Issues Identified in the End-User Survey and Designer Focus Group

Methods

After omitting the remarks about the approach used to collect biometric data, categorization of the free-text field and transcripts resulted in 294 remarks made by end users and 337 remarks made by designers. Grouping together similar remarks revealed 137 unique issues addressed by end users only, 148 unique issues addressed by designers only, and 10 unique issues addressed by both. Representative remarks of the issues that were addressed by both evaluators are found in Multimedia Appendix 2.

Table 1 lists the results of the systematic comparison of the issues identified in the end-user survey and in the designer focus group. In both evaluations all domains were relevant, but some topics were addressed in only one of the evaluations or were not relevant at all (see Multimedia Appendix 1). There were also topics that were addressed in both evaluations, but comparison of the associated issues showed that the nature of these issues were often dissimilar. Below, we first provide a high-level overview of the similarities and differences in issues identified in the evaluations for each of the user satisfaction and technology acceptance domains. Then we provide a more detailed description of these similarities and differences, using representative remarks to illustrate the findings.

Table 1. Numbers of unique issues addressed in the end-user survey and in the designer focus group, categorized using constructs of user satisfaction and technology acceptance.

Domain and topic	Total number of issues addressed by end users or designers, n	Issues addressed by end users, n	Issues addressed by designers, n	Issues identified by both groups, n	Percentage of issues in this domain made by both groups, % (n)
Ease of use	18	10	10	2	11.1 (2/18)
Easy to use	11	9	4	2	
User-friendly	7	1	6	0	
Information quality	77	27	54	4	5.2 (4/77)
Accuracy	7	7	1	1	
Completeness	18	7	12	1	
Format	39	9	32	2	
Language	3	2	1	0	
Precision	7	1	6	0	
Volume	3	1	2	0	
Organizational factors	54	14	41	1	1.8 (1/54)
Communication	28	2	26	0	
Organizational competition	5	5	0	0	
Documentation	2	0	2	0	
Error recovery	5	4	1	0	
Management	8	0	8	0	
Data security	5	2	4	1	
Time	1	1	0	0	
Outcome expectations	30	26	5	1	3.3 (1/30)
Accuracy	3	3	0	0	
Confidence in the system	8	8	1	1	
Feeling of control	3	3	0	0	
Expectations	13	11	2	0	
Health effects	3	1	2	0	
Service quality	54	34	22	2	3.7 (2/54)
Attitude	1	0	1	0	
Communication with program staff	5	4	1	0	
Means of input for the HRA	23	10	14	1	
Processing of change requests	4	4	0	0	
Relationship with program staff	2	2	1	1	
Response time	1	0	1	0	
Schedule of products or services	11	8	3	0	
Staff support	1	1	0	0	
Technical competence of program staff	6	5	1	0	
System quality	27	10	17	0	0.0 (0/27)

Domain and topic	Total number of issues addressed by end users or designers, n	Issues addressed by end users, n	Issues addressed by designers, n	Issues identified by both groups, n	Percentage of issues in this domain made by both groups, % (n)
Accessibility	3	2	1	0	
Efficiency	13	2	11	0	
Errors	2	1	1	0	
Flexibility	3	2	1	0	
Language	3	0	3	0	
Tailoring	2	2	0	0	
Timeliness	1	1	0	0	
Usefulness	35	26	9	0	0.0 (0/35)
Relevancy	8	5	3	0	
Usefulness	27	21	6	0	
Percentage of issues in all domains made by both groups					3.4 (10/295)

Overview of Similarities and Differences per Domain

Issues identified by designers concerning “ease of use” mostly focused on the user friendliness of the system and possible improvements, while those of end users were merely focused on practical issues that were encountered (eg, not being able to print the feedback). Twice as many issues were addressed by designers compared to end users concerning “quality of information”. The majority of these issues addressed by designers focused on information completeness and on the format of information throughout the HRA. End user issues regarding these topics focused on accuracy of the information provided by the HRA. Within the “organizational factors” domain, most of the issues brought up by designers concerned availability of accurate information about the HRA before actually using it. They discussed various aspects of how to motivate potential participants to use the program (eg, Table 2, remark d5). Conversely, within the “outcome expectations” domain, five times more issues were addressed by end users. These remarks mainly concerned the fit of the system to their prior expectations and perceived reliability of the HRA’s feedback. In the “service quality” domain, both groups identified a similar amount of issues (34 for end users, 22 for designers); however, only two issues were addressed by both groups. A majority of the issues identified by designers categorized under this domain focused on methods and data collection tools used to deliver the different components of the HRA and how these related to usefulness of the system (eg, Table 2, remark d6). End users also remarked on the methods and data collection tools used, but only one of those issues overlapped with those of designers. Furthermore, the issues identified by end users covered more topics within this domain. Within the “system quality” domain, a few issues were addressed by both groups for every topic. The eleven designer issues that concerned various aspects of the system’s efficiency domain, however, stand out within this domain. Finally, end users brought up at least 2.5 times more issues in the “usefulness” domain. They focused mainly on the extent to which the HRA actually helped them solve their health problems (eg, Table 2, remark e8).

Table 2. Examples of end-users' remarks and designers' remarks to illustrate the differences in yield of the two evaluation methods.

End-user remark	Designer remark
Remark e1 After receiving the results, I didn't really understand the feedback. I got the advice to eat healthier. I actually already started eating healthier food some time ago (and I have indicated this in the questionnaire). Still I received this advice, but I wouldn't know what else to do. (Information quality/Completeness)	Remark d1 Just explain that there is one way to complete things. People should just…you just have to guide them, because that's the most useful. It's also the most efficient. (Information quality/Completeness)
Remark e2 My general practitioner was really unsatisfied with how the HRA works, there is no explanation given about what has been tested precisely et cetera. Due to these results, there have been, according to my general practitioner, needless blood tests via STAR. (Information quality/Accuracy)	Remark d2 What you can also do, you have those five or six subjects in the feedback report, you can also have a compass per health subject. Because then you'll see a compass at the left top, this compass does this (mimics a compass pointer): for this subject you go wrong and for this one you go well. (Information quality/Format)
Remark e3 The feedback report of the HRA got across "fiercer" on me than it was in fact. It is a good realization and certainly a good provocation to take action. Nevertheless, I had preferred a few things to be expressed more subtle. (Information quality/Format)	Remark d3 Make it a little bit more personal, instead of relating it to some kind of standard. Because now you get: this is healthy, this is you, you fail in this, or you fail on that. But just say something like: This is personal, this is what you are now, and this is what you could do, this, that, et cetera. Instead of, relating it to the mass, you are wrong here and there and there. (Information quality/Format)
Remark e4 The feedback report is clear, but I am wondering whether one can conclude from these limited tests how healthy I live and what my physical condition is. (Outcome expectations/Confidence)	Remark d4 If I were to do it, I would just like my house style to be consistent. (Organizational factors/Communication)
Remark e5 I realize you want the phrasing of the questions to be as clear as possible. In a number of cases, the answers are oversimplified. The actual situation is sometimes far removed from the possible answers and consequently the results also give a different (more negative) picture. (Outcome expectations/Feeling of control)	Remark d5 You have to get someone excited, so you can say that it is a gift to the people. […] And that you, for example, mention somewhere: this package costs X euro, but the government and the employer believe it is important that…Just a story that they know that it is no garbage, but that it is actually really valuable and that they get it for free. (Organizational factors/Communication)
Remark e6 The examination does not have any relation to my work activities. Work-related problems/complaints are insufficiently covered because of this. (Outcome expectations/expectations).	Remark d6 That's when I thought, I will first fill in the questionnaire before I set to work with measuring my blood pressure. Or was it that I first waited for the lab box, I don't know anymore. I first want to have everything, you know, to get an overview of what I have to do. (Service quality/Means of input for the HRA)
Remark e7 During a face-to-face talk you could have given a lot more information and clarified things, and also have had a more thorough physical examination. (Service quality/Means of input for the HRA)	Remark d7 With the card you can activate your account and you also have to use the card to perform your measurements (has the card in his hands). Everything you do is stored on your card and if you log in with the card, your data will automatically be stored on the website, on your account. And the parameters, the outcomes of the blood tests are eventually also stored on your card and with this card you can go to the general practitioner and he explains to you what to do. (Service quality/Means of input for the HRA)
Remark e8 I don't think everybody needs to have an HRA. It upsets people more than anything else and doesn't give any guarantees at all. It's useful for (hereditary) diseases in the family. The question remains as to whether this should be done through the employer. (Usefulness/Relevance)	Remark d8 You can immediately register at the Internet. You immediately receive a login, so you can immediately fill in the questionnaire and only if you have completed the questionnaire you receive the home measurement tools. (System quality/Efficient)
Remark e9 Loss of time, drawing conclusions based on length, weight and a few simple Internet-based questions. (Usefulness/Usefulness)	Remark d9 It is ideal to just, as an employee, so to speak not with your Christmas box, but…So you get it from your employer, you think: 'Hey that employer has exerted himself, it is a subsidized thing.' I really think it is a good cause. (Usefulness/Relevance)
Remark e10 A polyp has been removed from my intestines on two different occasions. According to the specialist, one of these would certainly have become malignant. (Usefulness/Usefulness)	

Similarities

Issues that were found to be similar for both evaluator groups mainly focused on the feedback report provided by the HRA and on various aspects of using the Web-based part of the system. Both groups had some comments on the presentation and communication of the feedback report. They remarked that risk categories and recommendations mentioned in the feedback report could be clearer if a more extensive explanation on how to interpret the results would have been given: "What is required is that you get a printout of all the recommendations along with the referrals to a website or individual. Along the lines of, what is good, what isn't, and what does it actually mean?" [Designer #4, Information quality/Completeness]

Some also indicated that the feedback was not communicated in a way that optimally motivated them to change their lifestyle behavior: "Because all communication is written (in other words, in the system), it's easy to disregard any potential recommendations" [End user, female, age 56, Information quality/Format] and "No, but the results also didn't really provide the motivation to go into them at length" [Designer #6, Information quality/Format].

Some respondents of both groups also felt that the feedback was not inferred from both the answers to the questions and the biometric data, but merely represented their answers to the questionnaire: "Results are based on the answers that are filled in, and not on blood and urine tests" [End user, female, age 31, Outcome expectations/Confidence in the system]. Some end users added that if the HRA was more transparent about the origin of the results (eg, the data and evidence that led to these particular results), this could mitigate these problems. This suggestion was not made by designers.

Differences

Table 2 provides some examples of remarks to illustrate differences in yield of the two evaluation methods; a more comprehensive overview can be found in Multimedia Appendix 3. Within the "information quality" domain, designers mainly discussed how health risks and recommendations were displayed. The designers had extensive discussions on various design choices, like the choice of the compass metaphor for displaying the overall health risk of an individual. They found that if a metaphor is used, it should be used throughout the entire system and not just for displaying a single aspect (eg, remark d2). End users also commented on the use of the metaphor, but because they found the feedback displayed through use of the compass too alarming (eg, remark e3). Furthermore, designers also had discussions on ways to clarify the feedback report and present it in a way that might better motivate users to change their health behavior (eg, remark d3).

Both end users and designers addressed issues concerning the information completeness. End users remarked that they felt that the clarification of the values of the blood tests performed, of what to do with the feedback received (eg, remark e1), and the clarification of the tests performed to the general practitioner was unsatisfactory (eg, remark e2). In contrast to end users, designers' discussions mainly focused on the lack of guidance while completing the various HRA components (eg, health questionnaire, biometric measurements). They felt it was not clear enough in which order these components should be completed. They proposed explaining the single path for completing the entire program in a clear-cut way (remark d1).

The sequence of HRA components was further discussed in terms of efficiency (System quality/Efficiency). Designers indicated that because it was not possible to continue with another HRA component without completing the previous component and because the order of components was not clear, the process was not efficient and would therefore be a barrier for participants to complete the HRA (eg, remark d8). The designers also noted problems with the presentation of the health risks and feedback and discussed potential improvements extensively (eg, remark d3). The end users also identified issues concerning the presentation but added that the feedback report could have contained more practical advice (eg, remark e3).

Within the "organizational factors" domain, designers focused on the "look and feel" of the entire HRA from a marketing perspective, including both the Web-based aspects as well as pamphlets and other materials provided with the HRA. They were both impressed by some aspects of the HRA and disappointed by others. For example, they were impressed by the use of colors but disappointed by the use of stock photos within the Web-based environment. They also discussed the style of the various HRA components and recommended the use of a more consistent house style throughout the components (eg, remark d4).

End users, on the other hand, identified many more issues in the "outcome expectations" domain, which focused particularly on the ways in which the HRA met their expectations (eg, remark e6) and the confidence they had in the system (eg, remark e4). They felt they were unable to enter all the information they deemed necessary to generate the HRA feedback. Some mentioned that the questionnaire was not sufficiently tailored to their personal situations. End users were concerned that this affected the reliability of the provided health risk profile and feedback (eg, remark e5). Consequently, some end users had the desire to consult a health professional and have the opportunity to provide the aforementioned details during a conversation.

Remarks about the need for a clinical follow-up appointment also appeared in the "service quality" domain. In this domain, both end users and designers identified issues about the methods and data collection tools used in the HRA. Some end users wanted a follow-up appointment to discuss the results with a clinician and expected there would be more physical tests in the HRA (Service quality/Means of input for the HRA). In general, these end users preferred a more personal approach and the opportunity to clarify particular aspects of their health (eg, remark e7).

In contrast, designers focused mainly on exploring the opportunities for the future of HRAs with new means of performing the process of assessing one's health risks and giving feedback. New input methods ranged from the possibilities of having all data and feedback stored on one medical chip card to a lab inside a bus that allows end users to do all biometric tests close to home (eg, remark d7).

In the "usefulness" domain, a clear difference was seen in the type of issues identified by both groups concerning the relevance of the HRA. On one hand, designers were positive and thought that providing the HRA to this population was a good way of addressing health prevention options (eg, remark d9). On the other hand, some end users doubted the relevance for various personal reasons, and some also doubted its usefulness for the population as a whole (eg, remark e8).

Issues about the HRA's usefulness were mainly identified by end users; they formulated those with either with positive or skeptical remarks. For example, positive remarks made by end users showed that they were satisfied because the HRA had brought a serious health problem to light (eg, remark e10) or because they were happy to have their good health confirmed. Skeptical participants indicated they felt the HRA was of limited value because the feedback presented only what they could have come up with themselves (eg, remark e9).

Discussion

Key Contributions

In this study, we assessed the output of a professional review by designers in a focus group compared to that of an end-user survey of a Web-based HRA with tailored feedback. Designers in their focus group identified 10 out of the 137 issues (7%) addressed by end users in their survey and brought to light many new issues.

To our knowledge, no previous studies have investigated whether designers in focus groups bring up similar issues as end-user surveys after participating in the same intervention. Despite designers' full participation in the HRA as end users, our results show tendencies similar to studies in the fields of usability evaluation and text evaluation that compare the output of expert versus end user evaluations and the different methodologies applied [8,11,13]. The 7% overlap in the issues addressed in the evaluations of our study is, however, even lower than the 10-50% overlap found in these studies [8,11]. In their review of document evaluation methods, de Jong and Schellens [8] concluded that in expert evaluations many new issues are reported and that for expert evaluators the feedback of real readers often contains many surprising insights. The results of our study are in line with this conclusion.

The use of qualitative methods in this study provided insight into experiences of end users and designers. By analyzing their remarks using constructs of widely used theories on user satisfaction and technology acceptance, we were not only able to find that issues addressed by designers were dissimilar to those expressed in the end-user survey, but also to analyze differences in the nature of remarks made by both groups and pinpoint precisely which aspects of the intervention these differences pertain to. The analysis of these differences showed that although in both evaluations all high-level domains from the field of user satisfaction and technology acceptance were addressed, several lower-level topics were addressed in only one of the evaluations. When lower-level topics were addressed in both evaluations, often the nature of the associated remarks was different.

The differences in issues addressed show that end-user surveys should not be substituted for professional review in a focus group, despite the fact that both groups took on roles as HRA participants. On one hand, end-user surveys in the evaluation of HRAs give insight into immediate, practical opportunities for improvement, its perceived usefulness, and into whether the HRA meets or could be optimized to meet their expectations and enhance its adoption. These findings are in contrast with those of Lai [31] and of Lathan et al [32]. Lai found that end users were focused on system access and navigation, while we did not. Lathan et al found that end users were interested in their ability to use the system efficiently and effectively [32]. These differences in study results might be explained by the different foci of research (eg, research on system usability versus research on system acceptance and satisfaction) and suggest that the kind of evaluation method chosen greatly influences the kinds of issues an evaluation yields. Smilowitz et al concluded that nonexpert evaluations in real-life settings may uncover unforeseen problems introduced when the program is used [33]. Similarly, we observed various problems that affected end users' perception of the system that were not discussed by designers. For example, the fact that confidence in the system is of major importance for the decision to implement health behavior recommendations could be deduced from answers by end users but was not discussed by designers. This and similar examples highlight the value of performing evaluations with end users in real-life settings to tackle unforeseen problems and that those problems cannot be found with designer focus groups.

On the other hand, designer focus groups yield insight into various aspects of proper design, and issues about the HRA's user friendliness, efficiency, the display and completeness of the provided information, and organizational factors potentially impacting its adoption. Their discussion mainly focused on how the HRA can be improved in a broad variety of aspects (ie, idea generation) such as possibilities for (re)design, marketing and cost reduction, and its dissemination, while they presumed the usefulness of the innovation in general (ie, it is a good cause, it should be presented as a present). The fact that the designers emphasized the efficiency of the HRA contrasts with findings from Lathan's study, which suggested that end users are more interested in their ability to use the system efficiently and effectively [32]. Our results are similar to those of Lai, who reported that experts tend to focus on information design [31].

In summary, the key difference between end-user surveys and designer focus groups is as follows. End users report on immediate experiences with problems and not their causes. In contrast, designers in focus groups discuss how to solve problems that have been encountered and formulate additional requirements for improving the HRA's design. This difference was illustrated by how remarks by end users and designers were formulated. For example, end users formulated their remarks more like "I thought the feedback was…I would prefer…", while designers formulated the same issue as end users as "the feedback should be presented so that…"

The implication of our study is that HRA evaluation can be performed only with end-user surveys, or with a combination of both end-user surveys and designer focus groups, and the choice should depend on the focus of the evaluation. Among the most significant advantages of evaluating HRAs by means of end-user surveys are their potentially wide reach, their low cost, that they provide the opportunity to quickly determine the perspectives of end users, and that they can be used periodically for comparisons over time. However, they often have low response rates and associated risks of bias and do not allow participants to elaborate on the issues addressed. Adding professional reviews with focus groups to the evaluation is a choice that depends on the information needed for optimization of the innovation and the available resources. This method is easy to administer and is a relatively quick and low-cost method for obtaining information about potential attitudes and experiences of participants. The data analysis, however, is often time consuming and the quality of the data collection is in part dependent on the skills of the research team performing the focus group. Combining both end-user surveys and designer focus groups in HRA evaluations might enhance its adoption and design, as both evaluations resulted in a different focus within a theoretical framework that has previously been shown to be useful in understanding system impact and actual usage behavior [3,34]. Whether this actually enhances adoption must be proven in future studies.

Limitations

Various points for improving the HRA were identified from the issues addressed by end users and designers. However, from our study we are unable to determine whether the issues addressed in the designer focus group are valid nor how important they are. To explore this, these issues could be presented to end users to assess whether these issues actually influence utility and acceptance of the HRA. The added value of using one or both of the evaluation methods could be explored by implementing the suggested improvements in the HRA and studying the rates of adoption. It may also be useful to investigate whether some form of cross-fertilization is possible or whether the performance of designer focus groups can be enhanced by training.

The conditions of the two evaluation methods that were compared differed in various aspects from each other. The methods differed in the composition of the evaluator group, evaluation technique, and in the instructions given to the two groups in the evaluation. Therefore we are unable to comment on the causes of the differences found in the yield of both evaluations. Additionally, varying aspects of the evaluation techniques applied might provide different results. However, the goal was to compare the yield of the two methods in the way they are commonly used in practice to inform the optimization of new health technology innovations post-implementation. Our goal was not to compare them in a traditional experiment that aims to control these types of variations through methodology.

Although the designers in the focus group were specialized in different areas of design engineering, the study sample was essentially monodisciplinary. A group with varying disciplinary backgrounds might have delivered more comprehensive answers and shown different results. However, a disadvantage of multidisciplinary groups is that they might not speak the same language. Further research into evaluating HRAs post-implementation with a more multidisciplinary group might, however, prove valuable.

Another shortcoming of our study is the fact that the response rate to the free-text field was low when compared to other satisfaction surveys [35,36]. The explanation is likely that filling in the free-text field requires additional time and effort. Respondents had already filled in the structured part of the evaluation questionnaire, and they may have felt that this already covered all of their remarks on the HRA. Also, end users in the survey could have interpreted the question preceding the open-text field in two different ways, that is, as an invitation to comment on the HRA in general or on the survey itself. However, qualitative analysis of the end user remarks showed that most end users interpreted the question as an invitation to comment on the HRA in general; only 2.9% of the remarks were about the survey itself. Nevertheless, we cannot ignore that specific groups of end users left the free-text field blank.

Finally, categorization of the designers' remarks was performed after all end user responses had already been categorized, rather than doing this simultaneously. However, this should not have influenced the interpretation of their remarks, as we kept going back to the data of both end users and designers in our iterative process if changes were made in our categorization or framework. Subsequently, 1 researcher rather than 2 researchers grouped together identical issues and systematically compared the issues identified by end users and designers to limit the workload. However, 2 researchers independently assessed these final issues and made changes where necessary. These 2 researchers validated the comparison of issues revealed by the end-user survey and designer focus group.

Conclusions

Most post-implementation evaluations of health promotion programs, including eHealth applications, are performed with end users. In this study, professional review by a focus group of industrial designers of a Web-based HRA proved that this type of evaluation mainly yields new issues when compared to the feedback provided by end users in a Web-based survey. The end-user survey gave insight into the extent to which expectations and needs of users were met and suggested how the HRA could be improved to enhance its adoption in practice. Designers in the focus group gave more constructive criticism and provided recommendations to improve the HRA design and its marketing. They focused not only on potential problems but also made suggestions for (re)design, marketing, costs, and opportunities for future HRAs. We recommend that end-user surveys not be substituted for professional review in a focus group. Instead a combination of both methods in the evaluation of HRAs and other health promotion programs may prove more advantageous.

Acknowledgments

This study was funded by The Netherlands Organisation for Health Research and Development (ZonMw) through an unrestricted research grant: MyCARDSS, self-regulation services for cardiovascular risk management (#300020008). We would like to thank all of the employees at the study worksite and the industrial designers for taking part.

Authors' Contributions

NP and MWM were the study's principal investigators and contributed to interpreting the data. SV designed and wrote the original proposal, with support from NP. SV also carried out the data collection. SV and EL performed the coding of all data and assessed the grouping and comparison process. GM performed the data analysis. SV drafted the manuscript, based on GM's research report. The manuscript was reviewed by the other authors. All authors read and approved the final manuscript.

Conflicts of Interest

Dr Roderik A Kraaijenhagen is medical director and co-owner of NDDO Institute for Prevention and Early Diagnostics (NIPED). The program that was studied was developed by this company, which currently markets the program in the Netherlands. Eva K Laan was employed part-time at NIPED (4 hours a week) and received a salary during this study.

Multimedia Appendix 1

Characteristics and interpretations of the codebook used to structure the data.

[PDF File (Adobe PDF File), 70KB - jmir_v16i1e1_app1.pdf]

Multimedia Appendix 2

All equivalent issues identified by end users and designers about the Web-based HRA with tailored feedback.

[PDF File (Adobe PDF File), 48KB - jmir_v16i1e1_app2.pdf]

Multimedia Appendix 3

Illustrative remarks within domains to illustrate the differences in yield of the two evaluation methods.

[PDF File (Adobe PDF File), 46KB - jmir_v16i1e1_app3.pdf]

References

1. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
2. Marshall AL. Challenges and opportunities for promoting physical activity in the workplace. J Sci Med Sport 2004 Apr;7(1 Suppl):60-66. [Medline: 15214603]
3. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. Information Systems Research 2005 Mar;16(1):85-102. [doi: 10.1287/isre.1050.0042]
4. Colkesen EB, Niessen MA, Peek N, Vosbergen S, Kraaijenhagen RA, van Kalken CK, et al. Initiation of health-behaviour change among employees participating in a web-based health risk assessment with tailored feedback. J Occup Med Toxicol 2011;6:5 [FREE Full text] [doi: 10.1186/1745-6673-6-5] [Medline: 21388524]
5. Vosbergen S, Laan EK, Colkesen EB, Niessen MA, Kraaijenhagen RA, Essink-Bot ML, et al. Evaluation of end-user satisfaction among employees participating in a web-based health risk assessment with tailored feedback. J Med Internet Res 2012;14(5):e140 [FREE Full text] [doi: 10.2196/jmir.2067] [Medline: 23111097]
6. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
7. Doubleday A, Ryan M, Springlett M, Sutcliff A. A comparison of usability techniques for evaluating design. In: Designing interactive systems: processes, practices, methods and techniques. New York: Association for Computing Machinery; 1997:101-110.
8. de Jong M, Schellens P. Toward a document evaluation methodology: what does research tell us about the validity and reliability of evaluation methods? IEEE Trans Profess Commun 2000 Sep;43(3):242-260. [doi: 10.1109/47.867941]
9. Fu L, Salvendy G, Turley L. Effectiveness of user testing and heuristic evaluation as a function of performance classification. Behaviour & Information Technology 2002 Jan;21(2):137-143. [doi: 10.1080/02699050110113688]
10. Lentz L, De Jong M. The evaluation of text quality: expert-focused and reader-focused methods compared. IEEE Trans Profess Commun 1997 Sep;40(3):224-234. [doi: 10.1109/47.649557]
11. Tan W, Liu D, Bishu R. Web evaluation: Heuristic evaluation vs. user testing. International Journal of Industrial Ergonomics 2009 Jul;39(4):621-627. [doi: 10.1016/j.ergon.2008.02.012]
12. Law EL, Hvannberg ET. Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation. In: Proceedings of the third Nordic conference on Human-computer interaction. New York: Association for Computing Machinery (ACM) Press; 2004 Oct 23 Presented at: Human-computer Interaction; 23-27 Oct. 2004; Tampere, Finland p. 241-250.
13. Molich R, Ede MR, Kaasgaard K, Karyukin B. Comparative usability evaluation. Behaviour & Information Technology 2004 Jan;23(1):65-74. [doi: 10.1080/0144929032000173951]

14. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems. New York: ACM Press; 1990 Presented at: SIGCHI conference on Human factors in computing systems; April 01 - 05, 1990; Seattle, WA, USA p. 249-256. [doi: 10.1145/97243.97281]
15. Sienot M. Pretesting Web sites: A comparison between the plus-minus method and the think-aloud method for the World Wide Web. Journal of Business and Technical Communication 1997 Oct 01;11(4):469-482. [doi: 10.1177/1050651997011004006]
16. de Jong M, Lentz L. Focus: Design and Evaluation of a Software Tool for Collecting Reader Feedback. Technical Communication Quarterly 2001 Oct;10(4):387-401. [doi: 10.1207/s15427625tcq1004_2]
17. Friedman C, Wyatt J. Evaluation methods in biomedical informatics. New York: Springer; 2006:1-47.
18. Petre M, Minocha S, Roberts D. Usability beyond the website: An empirically-grounded e-commerce evaluation instrument for the total customer experience. Behaviour & Information Technology 2006 Mar;25(2):189-203. [doi: 10.1080/01449290500331198]
19. Kaplan B. Evaluating informatics applications--clinical decision support systems literature review. Int J Med Inform 2001 Nov;64(1):15-37. [Medline: 11673100]
20. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899]
21. van Velsen L, van der Geest T, Klaassen R, Steehouder M. User-centered evaluation of adaptive and adaptable systems: a literature review. Knowl Eng Rev 2008 Sep 3;23(03):-. [doi: 10.1017/S0269888908001379]
22. de Noblet J. Industrial Design: Reflection of a Century - 19th To 21st Century. Paris: Flammarion; 1993.
23. Savage PA, Millen DR, Bayerl JP. Designing features for display-based systems. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society; 1995 Oct 01 Presented at: Human Factors and Ergonomics Society Annual Meeting; 9-13 Oct. 1995; San Diego, CA p. 182-186. [doi: 10.1177/154193129503900302]
24. Jordan PW. An introduction to usability. London: Taylor & Francis; 1998.
25. Soler RE, Leeks KD, Razi S, Hopkins DP, Griffith M, Aten A, Task Force on Community Preventive Services. A systematic review of selected interventions for worksite health promotion. The assessment of health risks with feedback. Am J Prev Med 2010 Feb;38(2 Suppl):S237-S262. [doi: 10.1016/j.amepre.2009.10.030] [Medline: 20117610]
26. Carnethon M, Whitsel LP, Franklin BA, Kris-Etherton P, Milani R, Pratt CA, American Heart Association Advocacy Coordinating Committee, Council on EpidemiologyPrevention, Council on the Kidney in Cardiovascular Disease, Council on Nutrition, Physical ActivityMetabolism. Worksite wellness programs for cardiovascular disease prevention: a policy statement from the American Heart Association. Circulation 2009 Oct 27;120(17):1725-1741 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192653] [Medline: 19794121]
27. Goetzel RZ, Ozminkowski RJ. The health and cost benefits of work site health-promotion programs. Annu Rev Public Health 2008;29:303-323. [doi: 10.1146/annurev.publhealth.29.020907.090930] [Medline: 18173386]
28. Niessen MA, Kraaijenhagen RA, Dijkgraaf MG, Van Pelt D, Van Kalken CK, Peek N. Impact of a Web-based worksite health promotion program on absenteeism. J Occup Environ Med 2012 Apr;54(4):404-408. [doi: 10.1097/JOM.0b013e31824d2e43] [Medline: 22441493]
29. Cowdery JE, Suggs LS, Parker S. Application of a Web-based tailored health risk assessment in a work-site population. Health Promot Pract 2007 Jan;8(1):88-95. [doi: 10.1177/1524839905278892] [Medline: 16980572]
30. Morgan DL. Focus groups as qualitative research. Thousand Oaks: SAGE; 1997.
31. Lai TY. Iterative refinement of a tailored system for self-care management of depressive symptoms in people living with HIV/AIDS through heuristic evaluation and end user testing. Int J Med Inform 2007 Oct;76 Suppl 2:S317-S324 [FREE Full text] [doi: 10.1016/j.ijmedinf.2007.05.007] [Medline: 17616431]
32. Lathan CE, Sebrechts MM, Newman DJ, Doarn CR. Heuristic evaluation of a web-based interface for internet telemedicine. Telemed J 1999;5(2):177-185. [doi: 10.1089/107830299312140] [Medline: 10908430]
33. Smilowitz ED, Darnell MJ, Benson AE. Are we overlooking some usability testing methods? A comparison of lab, beta, and forum tests. Behaviour & Information Technology 1994 Jan;13(1-2):183-190. [doi: 10.1080/01449299408914597]
34. Abugabah A, Sanzogni L. The Impact of Information Systems on User Performance: A critical review and theoretical model. : World Academy of Science, Engineering and Technology (WASET); 2009 Sep Presented at: International Conference on Information Systems (ICIS); Sept. 2009; Amsterdam p. 809-819.
35. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. Int J Qual Health Care 1998 Aug;10(4):311-317 [FREE Full text] [Medline: 9835247]
36. Jones R, Pitt N. Health surveys in the workplace: comparison of postal, email and World Wide Web methods. Occup Med (Lond) 1999 Nov;49(8):556-558 [FREE Full text] [Medline: 10658310]

Abbreviations

CDSS: computerized decision support system
DUT: Delft University of Technology

HRA: health risk assessment

Edited by G Eysenbach; submitted 08.01.13; peer-reviewed by T van der Geest, PY Yen, PC Masella; comments to author 12.05.13; revised version received 10.07.13; accepted 12.08.13; published 02.01.14.

Please cite as:
Vosbergen S, Mahieu GR, Laan EK, Kraaijenhagen RA, Jaspers MWM, Peek N
Evaluating a Web-Based Health Risk Assessment With Tailored Feedback: What Does an Expert Focus Group Yield Compared to a Web-Based End-User Survey?
J Med Internet Res 2014;16(1):e1
URL: http://www.jmir.org/2014/1/e1/
doi:10.2196/jmir.2517
PMID:24384408

Short Paper

How Accurate is Web-Based Self-Reported Height, Weight, and Body Mass Index in Young Adults?

Kirrilly Pursey[1*], BND; Tracy L Burrows[1*], BND, PhD; Peter Stanwell[2*], PhD; Clare E Collins[1*], BSC, Dip Nutr&Diet, PhD

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

[2] Faculty of Health and Medicine, Priority Research Centre for Translational Neuroscience and Mental Health, University of Newcastle, Callaghan, Australia

[*] all authors contributed equally

Corresponding Author:
Tracy L Burrows, BND, PhD
Faculty of Health and Medicine
Priority Research Centre for Physical Activity and Nutrition
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 0249215514
Fax: 61 0249217053
Email: Tracy.Burrows@newcastle.edu.au

Abstract

Background: Web-based approaches are an effective and convenient medium to deliver eHealth interventions. However, few studies have attempted to evaluate the accuracy of online self-reported weight, and only one has assessed the accuracy of online self-reported height and body mass index (BMI).

Objective: This study aimed to validate online self-reported height, weight, and calculated BMI against objectively measured data in young Australian adults.

Methods: Participants aged 18-35 years were recruited via advertisements on social media sites and reported their current height and weight as part of an online survey. They then subsequently had the same measures objectively assessed by a trained researcher.

Results: Self-reported height was significantly overestimated by a mean of 1.36 cm (SD 1.93; *P*<.001), while self-reported weight was significantly underestimated by –0.55 kg (SD 2.03; *P*<.001). Calculated BMI was also underestimated by –0.56 kg/m^2 (SD 0.08; *P*<.001). The discrepancy in reporting resulted in the misclassification of the BMI category of three participants. Measured and self-reported data were strongly positively correlated (height: *r*=.98, weight: *r*=.99, BMI: *r*=.99; *P*<.001). When accuracy was evaluated by BMI category and gender, weight remained significantly underreported by females (*P*=.002) and overweight/obese participants (*P*=.02).

Conclusions: There was moderate to high agreement between self-reported and measured anthropometric data. Findings suggest that online self-reported height and weight can be a valid method of collecting anthropometric data.

(J Med Internet Res 2014;16(1):e4) doi:10.2196/jmir.2909

KEYWORDS

Internet; height; weight; body mass index; self-report

Introduction

Web-based approaches are becoming an increasingly popular and effective medium to collect epidemiological data and deliver eHealth interventions [1,2]. Web-based delivery is more cost effective than face-to-face interaction [3], can improve access to services for those in rural and remote locations, and allows provider contact with a large number of people simultaneously [4]. Online data collection and delivery of programs is also convenient with materials accessible at any time online, allowing for participation at times that are more opportune or outside regular hours [5].

To be effective, data that are self-reported via eHealth studies need to be reliable and accurate so that a participant's health status can be assessed and progress can be monitored. Discrepancies between measured and self-reported anthropometric data can lead to a misclassification of weight status and can thus affect assessment of participant health. Therefore the validation of self-reported Web-based data is essential.

Previous research indicates that a variety of factors including gender, age, and body mass index (BMI) can affect the accuracy of paper-based and interview-based self-reported anthropometric data [6,7]. There is a tendency for height to be overestimated and weight and BMI to be underestimated [8-12]. This leads to a subsequent misclassification of BMI category as a result of misreported anthropometric data [7,8,13], which is significant given that BMI is a commonly used indicator of health status in epidemiological research. Given that these differences exist, it is likely that similar differences may exist between online self-report and measured data; however to date, the latter has not been well explored.

Self-reported data are subject to influence by factors including social desirability and mode of data collection, leading to estimation bias of anthropometric data [14]. Social norms to conform to a certain body ideal can affect reporting of anthropometric data [15], with one study reporting the classification of more individuals as obese in face-to-face interviews compared to via telephone interviews [16]. Mail-in surveys are associated with more accurate reporting of anthropometric data because participants are not as likely to be affected by the social pressures associated with data collection via interview. Similar to mail-in surveys, it may be assumed that the anonymity of Web-based data collection may result in more accurate self-reporting compared to face-to-face and telephone interviews. However, in a study conducted by Lassalle et al [17], reporting bias of Web-based self-reported anthropometric data was similar to that observed in face-to-face interviews. Therefore, the level of reporting bias associated with Web-based self-reporting of anthropometric data must be studied to determine its accuracy and appropriateness as a method of data collection.

To the authors' knowledge, very few studies have assessed online self-reported data. One study that recruited adult participants (N=2513) found significant underreporting of online self-reported weight by –0.49 kg and overreporting of online self-reported height by 0.56 cm. This resulted in the significant underreporting of BMI in the study ($P<.05$). In addition, Bonn et al found significant underreporting of weight (mean difference -1.2 kg, SD 2.6), although they validated online self-reported weight alone [18] without validating online self-reported height or BMI. Similarly, a study conducted by Harvey-Berino [3] found significant underreporting of weight (mean difference -0.86 kg). However, this study was set in the context of a weight loss intervention, which could have potentially made them more aware of their current weight, that is, they were less likely to have misreported their data, although that could not be assessed. Additionally, the study sample population were all in the overweight/obese category (mean BMI 35.6 m/kg^2, SD 6.5, range 25-50).

No studies have attempted to validate online self-reported height, weight, and BMI data in a young adult population. To the best of our knowledge, the current study is the first to evaluate accuracy of Web-based self-reported height compared to measured height in the young adult population in all weight categories. This has allowed for the calculation and comparison of BMI from self-reported and measured data in the youth population. The aim of this study was to validate self-reported height, weight, and calculated BMI data via an online survey compared to objectively measured data in young Australian adults.

Methods

Participants

Males and females living in New South Wales, Australia, aged 18-35 years were recruited from March-May 2013 via a media release coordinated by the University of Newcastle, Australia, including advertisements on the university website, online blog, and "virtual snowballing" using social media sites including Facebook. Participants were excluded if they were currently pregnant or not currently living in Australia. This study was conducted as part of a 164-item online food addiction survey that took approximately 30 minutes to complete. The survey included questions investigating perceptions of food addiction, demographic details, anthropometric data, and current dietary habits. As part of the survey, respondents were asked to self-report their current height and weight via the online tool, SurveyMonkey [19]. Demographic information was collected including gender, age, education, and postal code. Socioeconomic status was determined using the Socio-Economic Indexes for Areas (SEIFA) deciles, whereby postal areas receive a score of 1 to 10, with the lowest 10% of areas given the score of 1 and the highest 10% of areas are given the score of 10. Self-reported BMI was calculated from online self-reported height and weight using the standard equation, weight/height^2 (kg/m^2).

Upon completion of the survey, participants were invited to attend a voluntary anthropometric measurement session on campus at the University of Newcastle and were contacted via email to select their preferred time via the online scheduling link, Doodle [20]. Within 1 month of completing the survey, body measurements were taken by a trained assessor using a standardized protocol. Participant height was measured to 0.1 cm by the BSM370 Stadiometer, and verbal instructions were provided according to the stretch stature method [21]. Weight, fat mass, and fat free mass were measured to 0.1 kg using the InBody720 bioelectrical impedance analyzer with shoes and heavy clothing removed. BMI was calculated using the same equation used for the calculation of self-reported BMI, and participants were subsequently classified as underweight (<18.49 kg/m^2), healthy weight (18.0-24.99 kg/m^2), overweight (25-29.99 kg/m^2) or obese (>30 kg/m^2) using the World Health Organization cut points [22]. At the end of the session, participants were provided with personalized feedback of their results relative to normative standards. Written informed consent was obtained from all participants prior to measurement. This study was approved by the University of Newcastle Human Research Ethics committee.

Statistics

Participant characteristics were checked for normality and analyzed descriptively, with mean (standard deviation) reported. Paired *t* tests were used to evaluate differences between self-reported and measured data. Pearson correlation was used to examine the strength of linear relationships between self-reported and measured data. To further investigate the relationship between variables, a multiple regression model using age, gender, and BMI was used. The degree of agreement between self-reported and measured data was also assessed using Bland-Altman plots [23]. Cohen’s *d* was used to compare effect sizes across different measures [24] and allowed for a more direct comparison of intervention effects on each outcome variable. These were calculated using the mean difference and the pooled standard deviation of the group (d=M_1-M_2 / σ_{pooled}). Respondents were grouped and analyzed by age (18-25 years or >25-35 years), gender, and BMI category (healthy weight or overweight/obese) to determine differences between self-reported and measured data both within groups and across groups. Due to the small sample sizes, overweight and obese participants were grouped and analyzed together and underweight participants were excluded from analysis. Statistics were computed using Stata V12. Significance level was set at .05.

Results

A total of 504 participants completed the broader food addiction survey, with n=117 in the validation study (23.2%). Participant characteristics are described in Table 1. Participants were predominantly female (79.5%, 93/117) with mean age 23.74 years (SD 3.92, range 18-35) and were from a range of socioeconomic backgrounds (35% SEIFA 5-6 deciles). The most commonly reported highest level of education achieved by the participants was a high school certificate (46.2%, 54/117) followed by a university degree (29.1%, 34/117). Mean BMI calculated from measured data was 24.18 kg/m^2 (SD 5.62, range 16.3-53) with the majority of participants classified as healthy weight (73.5%, 86/117). Three participants were classified as underweight, 16 as overweight, and 12 as obese. BMI calculated using self-reported data did not change BMI classification significantly with 5 participants classified as underweight, 87 as healthy weight, 13 as overweight, and 12 as obese. There were no significant differences between study participants and nonparticipants (n=367) of the larger survey with respect to demographic variables and self-reported height and weight (*P*>.05).

Table 1. Baseline data of adults participating in the Web-based food addiction study.

Characteristics	Male	Female	Total
Participants (n)	24	93	117
Age (years), mean (SD)	24.54 (3.57)	23.45 (4.54)	23.74 (3.92)
SEIFA [a] deciles, n (%)			
1-2 (lowest)	0 (0.0)	4 (4.3)	4 (3.4)
3-4	5 (20.8)	23 (24.7)	28 (23.9)
5-6	6 (25.0)	35 (37.6)	41 (35.0)
7-8	5 (20.8)	15 (16.1)	20 (17.1)
9-10 (highest)	8 (33.3)	16 (17.2)	24 (20.5)
BMI [b] category (kg/m^2), n (%)			
Underweight	0 (0.0)	3 (3.2)	3 (2.6)
Healthy weight	19 (79.2)	67 (72.0)	86 (73.5)
Overweight	4 (16.7)	12 (12.9)	16 (13.7)
Obese	1 (4.2)	11 (11.8)	12 (10.2)
Fat mass (kg), mean (SD)	11.20 (8.87)	20.33 (13.28)[c]	18.43 (13.00)
Fat free mass (kg), mean (SD)	65.20 (9.02)	46.18 (6.34)[d]	50.10 (10.34)
Body fat (%), mean (SD)	13.89 (6.85)	28.58 (9.38)[d]	25.48 (10.63)

[a] SEIFA: socioeconomic index for areas

[b] BMI: body mass index; weight (kg) / height (m)^2

[c] *P*=.002

[d] *P*<.001

Differences in self-reported and measured data as well as effect sizes are reported in Tables 2 and 3. Mean self-reported height, 169.35 cm (SD 8.86) was significantly higher than measured height (mean 167.99 cm, SD 8.37; mean difference 1.36 cm, SD 1.93, *P*<.001). Self-reported weight (mean 67.93 kg, SD 17.39) was significantly lower than measured weight (mean 68.48 kg, SD 17.59; mean difference –0.55 kg, SD 2.03, *P*=.004). As a result of the discrepancies between self-reported and measured height and weight, BMI calculated from self-reported height and weight was significantly lower than measured BMI (mean –0.56 kg/m^2, SD 0.08, *P*<.001). Self-reported height and weight and calculated BMI were highly correlated with the corresponding measured data (height: *r*=.98, weight: *r*=.99, BMI: *r*=.99; *P*<.001). Figures 1-3 display the Bland Altman plots for the average versus mean difference in self-reported and actual measurements. The limits of agreement (LOA) were wide for each variable of height, weight, and BMI. At the group level, the majority of values fell within the LOA (2SD) indicating a fairly good level of agreement. In descending order, the mean difference (LOA) when compared to measured data for each variable was –0.55 kg (–4.61, 3.51) for weight, 0.56 kg/m^2 (–1.26, 2.37) for BMI, and 1.37 cm (–2.49, 5.22) for height. Analysis using Cohen's *d* showed there was no or little effect (Cohen's *d*≤.50) on all variables (height, weight, BMI); range was 0.01-0.30. Despite the lower number of male participants, effect sizes were generally higher for males than females.

When grouped by BMI category, self-reported and measured weight did not differ significantly in healthy weight participants (*P*=.07) but remained significantly underreported in overweight/obese participants (*P*=.02). Discrepancies between self-reported and measured weight and BMI were significant between those overweight/obese compared to healthy weight (*P*=.02 and *P*=.03 respectively). Self-reported weight was significantly underreported by females (*P*=.002) but not by males (*P*=.71). Differences between self-reported and measured height were significant for males and females (*P*=.02 and *P*=.03 respectively). When grouped by age category, individuals aged 18-25 years were found to significantly underreport weight (*P*=.02), but not in individuals >25 years (*P*=.06). Height and BMI remained significantly misreported for all groups when grouped by BMI, age, and gender. When controlling for variables including gender, age, and BMI, the relationship between self-reported and measured data for each outcome measure remained highly significant (*P*<.001). However, these additional explanatory variables did not increase the strength of the associations.

Table 2. Differences between Web-based self-reported and measured height (cm) and weight (kg) in adults (n=117) grouped by BMI, age, and gender.

	Self-reported height	Measured height	Difference	*P* [a]	*d* [b]	Self-reported weight	Measured weight	Difference	*P* [a]	*d* [b]
	Mean (SD)	Mean (SD)	Mean (SD)			Mean (SD)	Mean (SD)	Mean (SD)		
All	169.35 (8.86)	167.99 (8.37)	1.36 (1.93)	<.001	0.16	67.93 (17.39)	68.48 (17.59)	0.55 (2.03)	.004	0.03
Gender										
Male (n=24)	180.48 (7.52)	178.69 (6.94)	1.79 (1.71)	<.001	0.25	76.13 (13.81)	76.31 (14.72)	–0.18 (2.38)	.71	0.01
Female (n=93)	166.50 (6.56)	165.20 (6.21)	1.29 (2.02)	<.001	0.20	65.82 (17.70)	66.48 (17.80)	–0.60 (1.91)	.002	0.04
P value between groups				.23					.32	
Age (years)										
18-25 (n=87)	169.11 (8.82)	167.76 (8.36)	1.35 (1.99)	<.001	0.15	65.74 (16.02)	66.22 (16.00)	–0.50 (2.00)	.028	0.03
>25-35 (n=30)	170.05 (9.08)	168.63 (8.50)	1.42 (1.75)	<.001	0.17	74.30 (19.82)	75.04 (20.45)	–0.74 (2.06)	.058	0.03
P value between groups				.88					.55	
BMI[c] (kg/m^2)										
Healthy weight (n=86)	169.04 (9.10)	167.63 (8.56)	1.41 (2.05)	<.001	0.16	61.81 (7.71)	62.12 (7.51)	–0.31 (1.63)	.08	0.04
Overweight/Obese (n=28)	170.89 (8.30)	169.59 (8.00)	1.30 (1.63)	<.001	0.16	89.16 (21.63)	90.52 (21.40)	–1.36 (2.97)	.02	0.06
P value between groups				.80					.02	

[a] *P* value

[b] Cohen's *d*

[c] BMI: body mass index; weight (kg) / height (m)^2

Table 3. Differences between Web-based self-reported and measured BMI (kg/m^2) in adults (n=117) grouped by BMI, age, and gender.

	Self-reported BMI, mean (SD)	Measured BMI, mean (SD)	Difference, mean (SD)	Cohen's *d*
All	23.63 (5.60)	24.18 (5.62)	–0.56 (0.08)[a]	0.11
Gender				
Male (n=24)	23.30 (3.38)	23.80 (3.66)	0.51 (0.97)[b]	0.14
Female (n=93)	23.75 (6.18)	24.31 (6.06)	0.57 (0.89)[a]	0.10
P value between groups			.78	
Age (years)				
18-25 (n=87)	22.94 (5.21)	23.46 (5.17)	0.52 (0.91)[a]	0.12
>25-35 (n=30)	25.63 (6.29)	26.28 (6.40)	0.65 (0.89)[a]	0.11
P value between groups			.51	
BMI[c] (kg/m^2)				
Healthy weight (n=86)	21.60 (1.74)	22.06 (1.58)	–0.46 (0.77)[a]	0.30
Overweight/Obese (n=28)	30.57 (7.56)	31.46 (7.26)	–0.89 (1.22)[a]	0.12
P value between groups			.03	

[a] *P*<.001

[b] *P*=.02

[c] BMI: body mass index; weight (kg) / height (m)^2

Figure 1. Level of agreement between self-reported and measured height (cm). Solid line represents the mean difference and dotted line represents the limits of agreement (LOA).

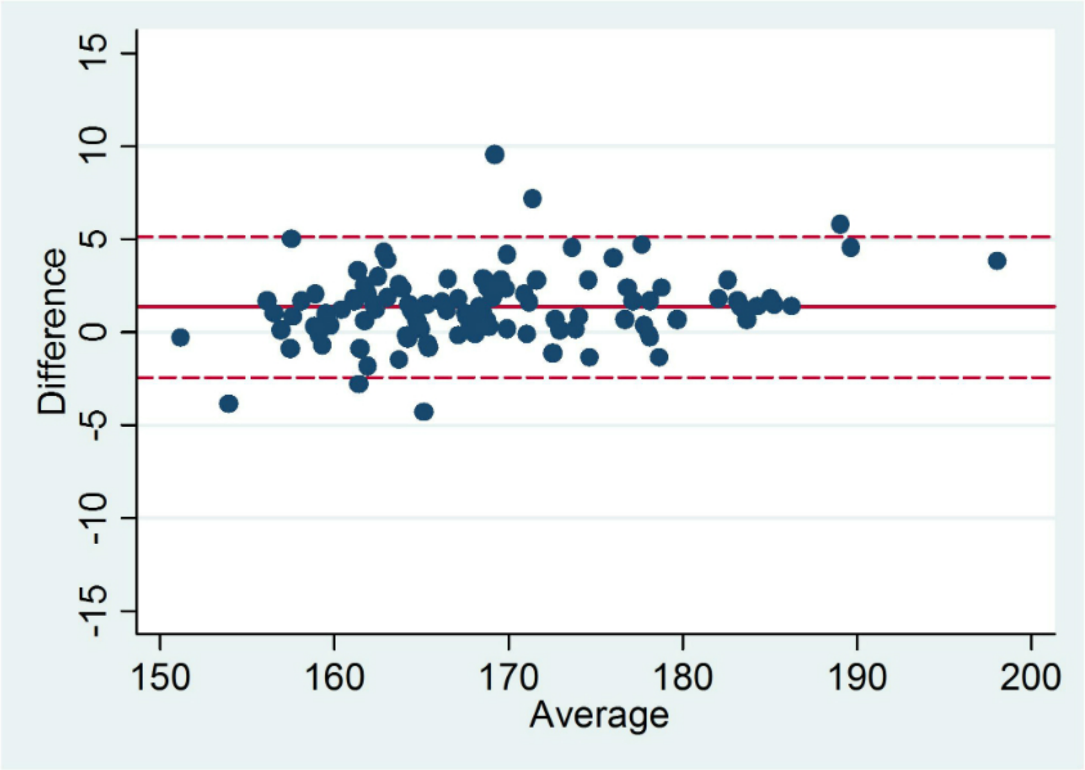


Figure 2. Level of agreement between self-reported and measured weight (kg). Solid line represents the mean difference and dotted line represents the limits of agreement (LOA).

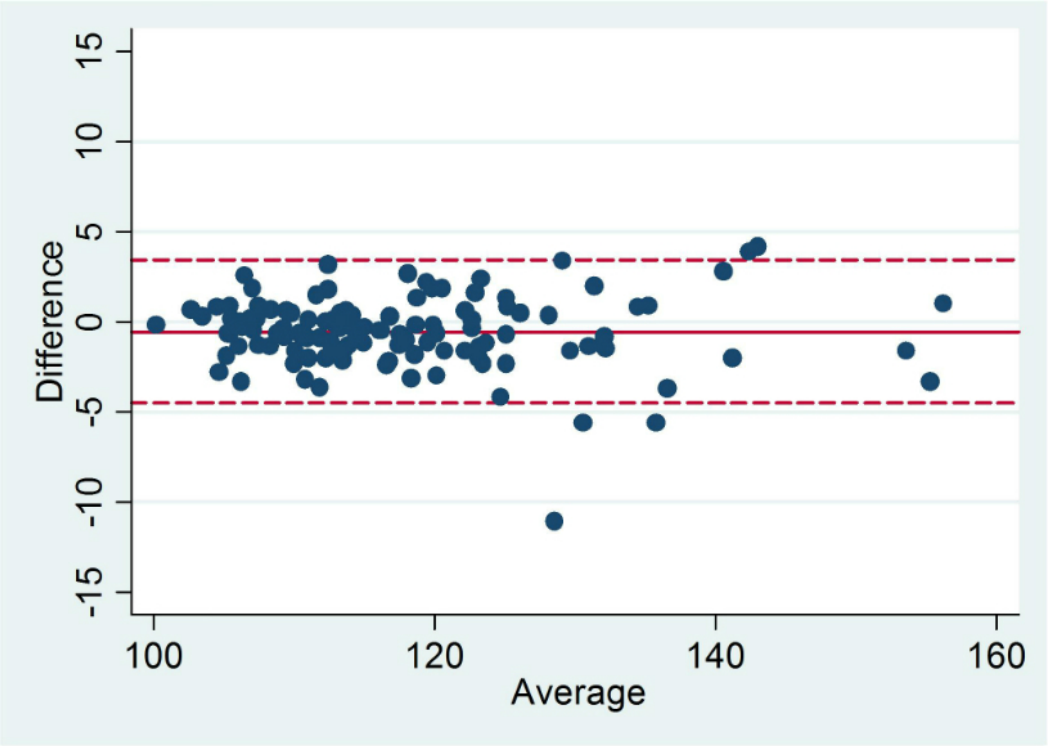
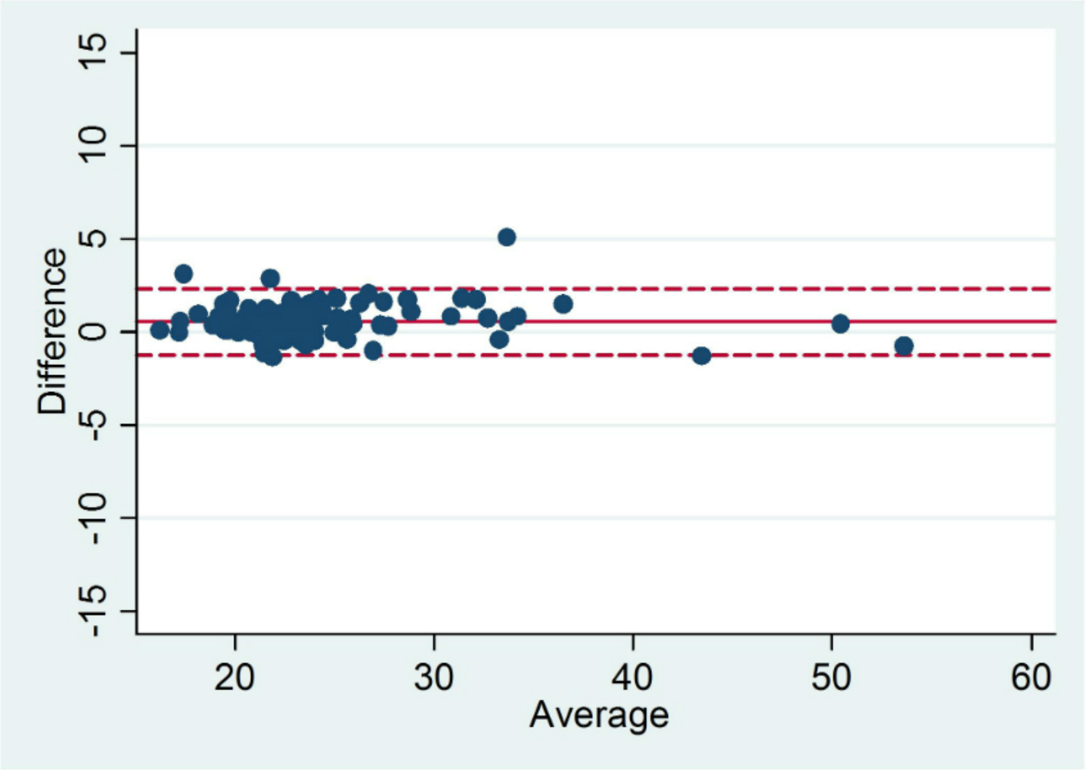


Figure 3. Level of agreement between BMI calculated from self-reported and measured data (kg/m^2). Solid line represents the mean difference and dotted line represents the limits of agreement (LOA).



Discussion

Principal Findings

To our knowledge, this is the first study to evaluate the accuracy of online self-reported height and weight in a young adult population. Compared to objectively measured data, online self-reported height was significantly overestimated while weight was significantly underestimated. As a consequence of the differences in self-reported and measured height and weight, self-reported BMI was significantly underestimated by participants; this underestimation of BMI changed the classification of BMI category of three participants. Measured and self-reported height, weight, and BMI were all strongly positively correlated with moderate levels of agreement. When grouped by BMI, age, and gender, self-reported weight remained significantly underreported by females, overweight/obese participants, and individuals <25 years. Effect sizes in this study (Cohen's *d*) for subgroups were considered small and likely to reflect the overall small sample size.

There was fairly good agreement between self-reported and measured data in the current study. Thus, online self-reported height and weight can be accepted as a satisfactory method of data collection in Web-based weight interventions, which is in agreement with international studies [3,17,18]. The discrepancies between self-reported and measured weight in the current study were smaller than those reported by two previous online studies validating weight alone [3,18] but were greater than those reported by a third larger study that also assessed height and BMI [17]. In addition, misreporting of anthropometric data in the current study related to BMI classification [17] and gender bias [18] is consistent with previous research in the area. These discrepancies in online self-reported data highlight that the same medium should be used to collect data for repeated measures within trials.

Strengths and Limitations

The generalizability of the current study may be limited by the recruitment of a convenience sample of predominantly female participants who were interested in a survey about food addiction. Those who volunteered to be measured were participants in an online food addiction survey where the only incentive was providing personalized feedback regarding height, weight, and body composition. Thus it is possible that these individuals may be more motivated than the general population, and the smaller magnitude of differences in the current study compared to previous paper-based self-reported studies [9-12] could be evidence that measurements may be affected by volunteer bias. However, the participants that were measured were representative of the larger online survey sample.

Another limitation of the study is the time lapse of 1 month between self-report and measurement. Logistical issues related to time taken to recruit participants and accessing measurement facilities resulted in a longer period of time between self-report and measurement than previously conducted studies. This time-lag between self-report and measurement could potentially be enough time for weight to have changed. This is particularly important in college-aged participants whose weight has been shown to fluctuate rapidly [25] or could be enough time for weight to change if an individual was participating in a weight loss program. It is possible that the instructions given to the participants in self-reporting data, the clothing worn by participants at time of measurement, and the use of different measuring equipment by participants compared to the calibrated equipment used by the trained assessor could have introduced measurement bias [8]. However, we would expect to see systematic and larger differences between self-reported and measured data than the results obtained if differences were due to measurement error.

Strengths of the study include the use of measured height to validate calculated BMI and the inclusion of adults from all weight categories to allow for comparison of self-reported anthropometric data based on weight status. The current study is important as it includes individuals from a range of weight status categories, including healthy weight.

Conclusions

Self-reported height was significantly overestimated and self-reported weight significantly underestimated by Australian adults aged 18-35 years. However, there was fairly good agreement between self-reported and measured data, and these were strongly positively correlated. When grouped by BMI category and demographic data, self-reported weight remained significantly underreported by individuals classified as overweight/obese, females, and individuals <25 years only. These findings suggest that online self-reported height and weight can be a valid method of collecting anthropometric data and calculating BMI. Future studies with larger sample sizes and repeated measures over time in eHealth research contexts are required.

Conflicts of Interest

None declared.

References

1. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of Web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
2. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ. A 12-week commercial Web-based weight-loss program for overweight and obese adults: Randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):306-321. [doi: 10.2196/jmir.1980]

3. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a Web-based obesity treatment program. Telemed J E Health 2011 Nov;17(9):696-699 [FREE Full text] [doi: 10.1089/tmj.2011.0032] [Medline: 21882997]
4. Mehring M, Haag M, Linde K, Wagenpfeil S, Frensch F, Blome J. Effects of a general practice guided Web-based weight reduction program - results of a cluster-randomized controlled trial. BMC Fam Pract Jun;14. 10.1186/1471-2296- 2013;14:76. [doi: 10.1186/1471-2296-14-76]
5. Hansen AW, Gronbaek M, Helge JW, Severin M, Curtis T, Tolstrup JS. Effect of a Web-based intervention to promote physical activity and improve health among physically inactive adults: A population-based randomized controlled trial. J Med Internet Res 2012;14(5):e145. [doi: 10.2196/jmir.2109]
6. Gunnare NA, Silliman K, Morris MN. Accuracy of self-reported weight and role of gender, body mass index, weight satisfaction, weighing behavior, and physical activity among rural college students. Body Image 2013 Jun;10(3):406-410. [doi: 10.1016/j.bodyim.2013.01.006] [Medline: 23419637]
7. Larsen JK, Ouwens M, Engels RC, Eisinga R, van Strien T. Validity of self-reported weight and height and predictors of weight bias in female college students. Appetite 2008;50(2-3):386-389. [doi: 10.1016/j.appet.2007.09.002] [Medline: 18006115]
8. Gorber SC, Tremblay M, Moher D, Gorber B. A comparison of direct vs self-report measures for assessing height, weight and body mass index: A systematic review. Obes Rev 2007;8(4):307-326. [doi: 10.1111/j.1467-789X.2007.00347.x]
9. Wang Z, Patterson CM, Hills AP. A comparison of self-reported and measured height, weight and BMI in Australian adolescents. Aust NZ J Publ Heal 2002 Oct;26(5):473-478. [doi: 10.1111/j.1467-842X.2002.tb00350.x]
10. Danubio ME, Miranda G, Vinciguerra MG, Vecchi E, Rufo F. Comparison of self-reported and measured height and weight: Implications for obesity research among young adults. Econ Hum Biol 2008;6(1):181-190. [doi: 10.1016/j.ehb.2007.04.002]
11. Krul AJ, Daanen HA, Choi H. Self-reported and measured weight, height and body mass index (BMI) in Italy, the Netherlands and North America. Eur J Public Health 2011 Aug;21(4):414-419 [FREE Full text] [doi: 10.1093/eurpub/ckp228] [Medline: 20089678]
12. Burton NW, Brown W, Dobson A. Accuracy of body mass index estimated from self-reported height and weight in mid-aged Australian women. Aust N Z J Public Health 2010 Dec;34(6):620-623. [doi: 10.1111/j.1753-6405.2010.00618.x] [Medline: 21134066]
13. Grossschadl F, Haditsch B, Stronegger WJ. Validity of self-reported weight and height in Austrian adults: sociodemographic determinants and consequences for the classification of BMI categories. Public Health Nutr 2012;34(6):20-27. [doi: 10.1017/S1368980011001911]
14. Larson MR. Social desirability and self-reported weight and height. Int J Obes Relat Metab Disord 2000 May;24(5):663-665. [Medline: 10849592]
15. Brestoff JR, Perry IJ, Van den Broeck J. Challenging the role of social norms regarding body weight as an explanation for weight, height, and BMI misreporting biases: development and application of a new approach to examining misreporting and misclassification bias in surveys. BMC Public Health 2011;11:331. [doi: 10.1186/1471-2458-11-331]
16. Béland Y, St-Pierre M. Mode effects in the Canadian community health survey: A comparison of CATI and CAPI. In: Advances in Telephone Survey Methodology. New York: John Wiley & Sons, Inc; 2007:297-314.
17. Lassale C, Peneau S, Touvier M, Julia C, Galan P, Hercberg S. Validity of Web-based self-reported weight and height: results of the Nutrinet-Sante study. J Med Internet Res 2013;15(8):e152. [doi: 10.2196/jmir.2575]
18. 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]
19. Survey Monkey. 2013. URL: http://www.surveymonkey.com/ [accessed 2013-08-22] [WebCite Cache ID 6J3JFyaWJ]
20. Doodle online scheduling. 2013. URL: http://doodle.com/?locale=en [accessed 2013-08-22] [WebCite Cache ID 6J3JTtz45]
21. Stewart A, Marfell-Jones M, Olds T, de Ridder H. International standards for anthropometric assessment. In: International Society for the Advancement of Kinanthropometry. Lower Hutt, New Zealand: ISAK; 2011:125.
22. World Health Organization. BMI classification. URL: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html [accessed 2013-08-22] [WebCite Cache ID 6J3JWlEm4]
23. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]
24. Cohen J. Statistical power analysis for behavioural sciences. 2nd ed. Hillsdale, New Jersey, USA: Lawrence Earlbaum and Associates; 1988.
25. Gow RW, Trace SE, Mazzeo SE. Preventing weight gain in first year college students: An online intervention to prevent the "freshman fifteen". Eating Behaviors 2010;11(1):33-39. [doi: 10.1016/j.eatbeh.2009.08.005]

Abbreviations

BMI: body mass index
LOA: limits of agreement
SEIFA: Socio-Economic Indexes for Areas

Edited by G Eysenbach; submitted 26.08.13; peer-reviewed by S Bonn, J Kim, A Sarkin; comments to author 19.09.13; revised version received 24.10.13; accepted 10.11.13; published 07.01.14.

Please cite as:
Pursey K, Burrows TL, Stanwell P, Collins CE
How Accurate is Web-Based Self-Reported Height, Weight, and Body Mass Index in Young Adults?
J Med Internet Res 2014;16(1):e4
URL: http://www.jmir.org/2014/1/e4/
doi:10.2196/jmir.2909
PMID:24398335

Original Paper

Internet and Mobile Technology Use Among Urban African American Parents: Survey Study of a Clinical Population

Stephanie J Mitchell[1*], PhD; Leandra Godoy[1], PhD; Kanya Shabazz[1], BS; Ivor B Horn[2*], MD, MPH

[1]Children's National Medical Center, Center for Translational Science, George Washington University School of Medicine, Washington, DC, United States

[2]Children's National Medical Center, Division of General Pediatrics/Center for Translational Science, George Washington University School of Medicine, Washington, DC, United States

*these authors contributed equally

Corresponding Author:
Ivor B Horn, MD, MPH
Children's National Medical Center
Division of General Pediatrics/Center for Translational Science
George Washington University School of Medicine
111 Michigan Ave, NW
Washington, DC, 20010
United States
Phone: 1 202 476 6908
Fax: 1 202 476 3245
Email: ihorn@childrensnational.org

Abstract

Background: There is considerable potential for mobile technologies to empower pediatric patients and families by improving their communication with health professionals. National surveys suggest minority parents frequently communicate via mobile technology, but it is uncertain how amenable they are to receiving health care information in this format. Although the low cost and far reach characteristics of mobile health (mHealth) technology makes it advantageous for communication with minority parents, data on acceptance are needed.

Objective: The objective of the study was to determine utilization of mobile and Internet technology by African American parents in an urban, underserved population, and to assess their interest in receiving health information via text messaging or other technologies (eg, social media and the Internet).

Methods: A survey was administered to parents of children aged 1-12 years covered by public insurance receiving care at 3 pediatric primary care centers in Washington, DC.

Results: The African American sample (N=302) was composed of primarily single (75.8%, 229/302) mothers. Almost half had more than a high school education (47.7%, 144/302) and incomes above US $25,000 per year (43.0%, 130/302). Most (97.0%, 293/302) reported owning a cell phone, of which 91.1% (275/302) used it to text and 78.5% (237/302) used it to access the Internet. Most had service plans with unlimited text and data, but 26.5% (80/302) experienced service interruptions in the previous year. Home Internet access was more prevalent among those with higher income (86.2%, 112/130), but it was still relatively pervasive among lower income families (66.9%, 83/124). In adjusted logistic regression models, African American mothers with income greater than US $25,000 annually were 4 times as likely to own a tablet computer than their lower income counterparts. Of the participants, 80.8% (244/302) used social networking, primarily Facebook, and 74.2% (224/302) were interested in joining a social networking group about a health topic concerning their child. Although relatively few African American mothers (17.9%, 54/302) shared health information via texting, there was strong interest in receiving health information via mobile phones (87.4%, 264/302). There was no significant difference in Internet/mobile device use or interest in using these outlets to send/receive information about their children’s health between parents of healthy children and parents of children with chronic health conditions.

Conclusions: Urban African American parents are active users of the Internet and mobile technology for social interactions, but they are less likely to use it for accessing or communicating health information. However, most parents expressed an interest in receiving health information or utilizing social networking to learn more about health topics. Mobile technology and social networks may be an underutilized method of providing health information to underserved minority populations.

(J Med Internet Res 2014;16(1):e9) doi:10.2196/jmir.2673

KEYWORDS

health communication; mobile phone; social networking, African Americans; pediatrics

Introduction

As with all forms of Web-based technology, advances in mobile health (mHealth) and health communication technology are rapidly increasing. According to the National Institutes of Health Consensus Group, mHealth is defined as the use of mobile and wireless devices to improve health outcomes, health care services, and health research [1]. Health communication technology is used specifically for communication between patients or caregivers and health care providers about health [2], and it is typically a means of streamlining the delivery of health information and services [3]. Some of the many advantages of mHealth communication technologies include the relatively low cost and the ease with which apps and text messages can be widely distributed because of the popularity of mobile technology.

Data from the Pew Research Center's Internet & American Life Project indicate that technology use has expanded rapidly over the past decade and is now pervasive [4,5]. As of spring 2013, approximately 85% of American adults used the Internet, 56% owned a smartphone, and 63% went online wirelessly with a laptop or cell phone [4-6]. Furthermore, 72% of Internet users reported looking online for health information; 31% of cell phone owners and 52% of smartphone owners say they have used their phone to look up health or medical information [7]. Despite overall increases in access to and use of the Internet and mHealth technology, differences in access and use vary by race/ethnicity, age, income, and education [4,8]. With regard to race and ethnicity, current (2013) Internet usage rates for non-Hispanic black adults (85%) are comparable with those of non-Hispanic white adults (86%), with usage rates for Hispanic adults lagging slightly behind (76%) [5]. However, underrepresented minority (URM) populations, including African Americans and English-speaking Hispanic persons, are more likely to own a mobile phone and to use mobile technology than their white counterparts [9]. Moreover, Hispanic persons and African Americans are more likely than white cell phone users to look for health information on their phones [10].

At the same time, URMs experience higher rates of childhood chronic diseases. The highest prevalence of asthma in the United States is among non-Hispanic black children who are almost 7 times more likely to die of asthma than white children [11]. URM children are also at high risk for obesity and type 2 diabetes [12]. Because of its popularity among URM populations, its relatively low cost, and its demonstrated efficacy in facilitating parent-provider communication and improving health behaviors, mHealth communication technology is uniquely well-suited for addressing pediatric health disparities. mHealth technology has been successfully used to improve patient-provider communication [2,13], chronic disease self-management [14], and preventive health behaviors [15]. The efficacy of mHealth technologies is, in part, because of their convenience as a method of health information exchange [16] and they are effective for modifying health behavior because behavioral cues (eg, reminders) can be sent/received asynchronously when and where they are most relevant or appropriate [17-18].

One arena in which mHealth communication technology can have important public health effects is in pediatric settings, where clear and frequent communication between parents and providers is important for managing children's chronic diseases (eg, asthma, diabetes). Parents (or other caregivers) serve as the intermediary between pediatric patients and providers; they determine when health care is sought and are primarily responsible for relaying information about their child's health to providers [2]. Thus, facilitating parent-provider communication through mobile technology is a promising method for addressing disparities in childhood chronic diseases.

Although there is great potential for mHealth to reduce pediatric health disparities, more information is needed to understand how URM parents/patients use communication technology, particularly in reference to their children's health-related issues. Most of the research among African American parents has focused on their use of and access to Web-based health information. Not only do the reported rates of Internet use vary by population sampled, but these data are out of date considering how rapidly and pervasively Internet-capable mobile devices are being disseminated [8,19-20]. More specific information is needed on URM parents' usage patterns and receptivity to receiving health information via mobile technology. To our knowledge, only 1 study has been conducted on urban parents' use of mobile technology within a clinical pediatric setting [21]. This study affirmed the widespread popularity and frequent use of mobile technology (eg, 75% of respondents used some form of digital technology daily), including searching for health information (eg, 58% searched for medical info at least once in the preceding week) among families seeking care from urban pediatric primary care centers. However, information was not gathered in several areas that would be necessary to understand the way in which mobile technology could be used to address health disparities, such as participants' race/ethnicity, the impact of child health status on use, service interruptions, functions of cell phone use, and specific modes in which parents may be interested in accessing health information.

The current study reports on the use of mobile phones, the Internet and wireless devices, and social media among a sample of urban minority parents of children attending pediatric community health clinics. To help guide the development and marketing of mHealth communication technologies for URM parents, we examine how patterns of access, use, and preferences relating to mHealth technology vary by socioeconomic characteristics (ie, education, income). We additionally explore the extent to which parents of children with chronic diseases differ from other parents in their use of technologies due to the particular potential of mHealth in addressing childhood chronic conditions.

Methods

Participants

A clinical sample of parents whose children received primary care services at 1 of 3 pediatric primary care practices in the Washington, DC metropolitan area between July 2011 and December 2012 were recruited to participate in the study. A clinical sample, as opposed to a sample drawn from the general population of minority parents of young children, was chosen to understand how providers may better engage and communicate with parents through technology. English-speaking parents were eligible if they were at least 18 years old, their child's legal guardian, had a child aged between 12 months and 12 years, and they could identify a primary care provider for the child. For this analysis, the sample (N=314) was constrained to self-identified African American parents (n=302).

Procedures

Trained research assistants invited parents/guardians to participate while they were in the pediatric clinic waiting area, and screened interested parents for eligibility. Upon meeting inclusion criteria and providing informed consent, parents/guardians completed the 5-10 minute paper survey. The Children's National Health System Institutional Review Board approved all study procedures.

Measures

A demographics questionnaire was administered to collect information on parent age, race/ethnicity (self-reported by the parents of the children from a list including white/Caucasian, black/African American, Asian/Pacific Islander, American Indian/Alaskan, Latino/Hispanic, or other), marital status, relationship to child, household composition, annual income, education level, occupation, and child's health status (ie, diagnosis of chronic illness). Parents indicated the level of education they had completed from the following choices: some grade school, completed junior high school, high school graduate or graduate equivalency degree (GED), some college/2-year college, college graduate, and postgraduate study. Based on the distribution, responses were recoded into less than a high school degree/GED or more than a high school degree/GED. Similarly, parents' selection from 6 categories of annual household income (< US $10,000, US $10,000-US $24,999, US $25,000-US $49,999, US $50,000-US $74,999, US $75,000-US $99,999, and ≥US $100,000) were recoded into < US $25,000 and ≥ US $25,000.

The survey measure of mobile and Internet technology utilization was an adaptation of the Pew Internet & American Life Project, a telephone survey of 2252 Americans older than 18 years [4]. The Pew survey included 23 multiple-choice questions covering topics such as ownership of electronic devices, Internet access and usage, mobile phone plans, and utilization of mobile phone features, such as mobile Internet and email, text messages, video messages, etc. Our adapted measure followed a similar format with 27 additional questions, including 14 regarding mobile phone use, 9 regarding Internet usage, and 4 regarding social networking use.

Statistical Analyses

Frequencies were generated to characterize cell phone, Internet, and social networking use as well as parents' interest in receiving child health information through these channels. Next, chi-square tests were conducted comparing frequencies/proportions of each technology use outcome across groups defined by education (> high school vs ≤high school), annual income (>US $25,000 vs ≤US $25,000), and health status of children (chronic illness vs no chronic illness).

Any socioeconomic (eg, education, income) differences that were significant according to chi-square tests were further examined using a logistic regression adjusting for the other predictor variables. Statistical significance was set at P=.01 because of the number of tests conducted and the higher chances of type I (false positive) errors.

Results

Sample Demographics

See Table 1 for participant characteristics. This sample included primarily single (75.8%, 229/302) mothers (84.1%, 254/302) with an average of 2.5 (SD 1.6) children. Participants ranged in age from 18 to 59 years (mean 31.5, SD 8.0), and household sizes ranged from 1 to 10 persons (mean 4.2, SD 1.7). Close to half had more than a high school education (47.7%, 144/302) and incomes greater than US $25,000 per year (43.0%, 130/302). Almost one-third had a child with asthma (31.1%, 94/302); 53.6% (162/302) had children without any chronic illness.

Table 1. Demographic characteristics of African American participants (N=302).

Demographics	Mean (SD)	Range	n (%)
Parent age (years)	31.5 (8.0)	18-59	
Relationship to child			
Biological mother			254 (84.1)
Biological father			26 (8.4)
Legal guardian			21 (7.0)
Marital status			
Married			51 (16.9)
Single			229 (75.8)
Divorced/separated			18 (6.0)
Number of children	2.5 (1.6)	1-10	
Household total	4.2 (1.7)	1-10	
Children	2.4 (1.4)	0-7	
Adults	1.9 (0.9)	1-7	
Highest education			
Less than high school			18 (6.0)
High school/GED			138 (45.7)
Some college			110 (36.4)
Bachelor degree or more			34 (11.3)
Income (US $)			
<10,000			58 (19.2)
10,000-24,999			66 (21.9)
25,000-49,999			87 (28.8)
50,000-74,999			25 (8.3)
75,000-99,999			11 (3.6)
>100,000			7 (2.3)
Don’t know			40 (13.2)
Child chronic health problems			
None			162 (53.6)
Asthma			94 (31.1)
Diabetes			2 (0.7)
Obesity			2 (0.7)
Other			14 (4.6)
Multiple conditions			14 (7.3)

Cell Phone Use

When asked “what do you use your cell phone for,” most participants selected functions in addition to voice calls: 91.1% (275/302) text, 70.9% (214/302) send/receive email, and 78.5% (237/302) access the Internet (Table 2). Most had service plans with unlimited texts (86.4%, 261/302) and data (76.5%, 231/302), but 26.5% (80/302) experienced service interruptions (of unknown duration) in the previous year. When asked to indicate how many texts they sent/received on an average day, 18.2% (55/302) said they sent/received more than 50. Although most (66.2%, 200/302) sent messages about important personal matters, only 17.9% (54/302) reported sharing health information via text.

Table 2. Participant cell phone use by education and income (N=302).

Cell phone use	Overall n (%)	Parent education n (%)		Household income (US $) n (%)	
		≤High school (n=156)	>High school (n=144)	<25,000 (n=124)	≥25,000 (n=130)
Do you have a cell phone? (yes)[a]	293 (97.0)	94.2 (147)	144 (100)	118 (95.2)	127 (97.7)
What do you use your cell phone for?					
Calls[a]	288 (95.4)	143 (91.7)	144 (100)	116 (93.5)	127 (97.7)
Texts[a]	275 (91.1)	135 (86.5)	139 (96.5)	112 (90.3)	120 (92.3)
Emails[a]	214 (70.9)	102 (65.4)	112 (77.8)	83 (66.9)	99 (76.2)
Internet	237 (78.5)	116 (74.4)	120 (83.3)	98 (79.0)	103 (79.2)
Service plan					
My cell phone plan includes text messaging with...unlimited texts	261 (86.4)	128 (89.5)	131 (91.0)	106 (90.6)	116 (91.3)
My plan includes data with...unlimited data	231 (76.5)	113 (77.9)	116 (81.7)	94 (79.7)	104 (83.2)
In the past year, has your cell phone service been interrupted? (yes)	80 (26.5)	43 (29.5)	36 (25.0)	39 (33.1)	36 (28.3)
Who do you text?					
Family[a]	273 (90.4)	132 (84.6)	139 (96.5)	111 (89.5)	122 (93.8)
Friends	278 (92.1)	139 (89.1)	137 (95.1)	113 (91.1)	122 (93.8)
Coworkers[a,b]	135 (44.7)	53 (34.0)	82 (56.9)	46 (37.1)	81 (62.3)
On an average day, how many text messages do you send or receive on your cell phone?					
≤10[a]	55 (18.2)	21 (14.6)	33 (23.6)	19 (16.4)	29 (23.6)
11-25[a]	87 (28.8)	35 (24.3)	51 (36.4)	34 (29.3)	45 (36.6)
26-50[a]	63 (20.9)	37 (25.7)	51 (18.6)	28 (24.1)	24 (19.5)
>50	55 (18.2)	30 (20.8)	25 (17.9)	26 (22.4)	20 (16.3)
When you text do you...?					
Just say hello and chat[a,b]	248 (82.1)	119 (76.3)	127 (88.2)	101 (81.5)	11 (85.4)
Do things related to work	127 (42.1)	52 (33.3)	75 (52.1)	49 (39.5)	71 (54.6)
Send multiple messages to discuss important personal matters[a]	200 (66.2)	93 (59.6)	106 (73.6)	78 (62.9)	97 (74.6)
Share information about your health	54 (17.9)	29 (18.6)	25 (17.4)	30 (24.2)	16 (12.3)

[a] Chi-square tests indicate a significant (P<.01) difference between parents with ≤ high school vs > high school education.

[b] Chi-square tests indicate a significant (P<.01) difference between parents with annual household incomes < US $25,000 vs ≥ US $25,000.

There were several significant socioeconomic differences in terms of cell phone use. Chi-square tests (Table 2) indicated that parents with more than a high school education were more likely to own a cell phone (χ^2_1=8.7, P=.003) and use nonvoice phone features (eg, text messaging; χ^2_1 = 9.7, P=.002). Multivariate logistic regression (Table 3) controlling for household income also showed that parents with more than a high school education were 6 times more likely to use their cell phones to send text messages compared to parents with less education (P=.002).

Chi-square tests suggested that those with higher incomes (>US $25,000) were more likely to use cell phones for work purposes (eg, to text coworkers: χ^2_1 =17.2, P<.001; send messages related to work: $\chi2_1$ =6.0 P=.02) than those with lower incomes (Table 2). The income-related difference in likelihood of texting coworkers remained significant in the multivariate model (Table 3) controlling for parent education (P=.001). There were no significant differences in cell phone use by child health status (results not shown).

Table 3. Odds of reporting various cell phone, Internet, and social network usage for parents with more than a high school education or a household income above US $25,000.[a]

Technology use	>High school education		Household income ≥ US $25,000	
	OR	95% CI	OR	95% CI
Cell phone use				
Own cell phone	0	0	1.04	0.25-4.40
Uses these functions:				
Calls	0	0	1.43	0.36-5.70
Texts	6.23[b]	1.96-19.83	0.76	0.30-1.95
Emails	1.78	0.99-3.21	1.28	0.71-2.30
Sends texts to these people:				
Family	5.12[b]	1.60-16.38	1.13	0.43-2.97
Coworkers	1.58	0.92-2.69	2.41[b]	1.41-4.11
Send/receive >50 texts/day on average	0.89[b]	0.45-1.75	0.71	0.36-1.39
When you text do you...?				
Just say hello and chat	2.26	1.09-4.66	1.04	0.51-2.10
Do things related to work	1.53	0.87-2.69	1.54	0.87-2.71
Send multiple messages to discuss important personal matters	1.47	0.84-2.58	1.49	0.85-2.61
Share information about your health	0.92	0.47-1.82	0.45	0.22-0.90
Internet usage				
Do you own any of the following items?				
None	0.34[b]	0.15-0.76	0.37[b]	0.17-0.82
Desktop computer	1.61	0.94-2.75	2.44[b]	1.43-4.17
Laptop or notebook	1.70	0.99-2.92	2.11[b]	1.22-3.62
iPod or MP3 player	1.60	0.94-2.73	2.18[b]	1.28-3.72
Game console	1.02	0.60-1.74	2.10[b]	1.23-3.57
Tablet computer	1.12	0.45-2.76	4.14[b]	1.44-11.89
Have Internet access at home	1.68	0.89-3.17	2.65[b]	1.38-5.09
Ever used Internet for...				
Send/read email	5.24[b]	2.08-13.17	0.74	0.33-1.66
Get news online	3.29[b]	1.75-6.18	1.23	0.67-2.27
Get health info	1.98	1.16-3.37	0.88	0.52-1.51
Social networking (which sites)				
MySpace	1.03	0.41-2.58	0.23[b]	0.08-0.67
LinkedIn	5.10	1.09-23.79	2.91	0.78-10.91

[a] Models include both greater than high school education and household income above US $25,000 so that ORs reflect independent influence of each predictor while adjusting for the other.

[b] OR is significant (*P*<.01).

Internet Use

In this sample, most parents owned either a desktop (43.7%, 132/302) or laptop computer (55.3%, 167/302), and approximately three-quarters (75.5%, 228/302) reported having Internet access at home (see Table 4). Most parents (69.5%, 210/302) reported using the Internet to get news or visit social networking sites (69.9%, 211/302), whereas only 53.0%

(160/302) reported using the Internet to get health information. Nevertheless, more than 87% (264/302) reported willingness or interest in receiving health info online through email or in texts.

Chi-square analyses (Table 4) indicated that parents with more education were more likely to own a computer (χ^2_1 = 18.6, *P*<.001) and have Internet access at home (χ^2_1 = 10.4, *P*=.001). In addition, those with more education were more likely to use the Internet to send/read email (χ^2_1 = 22.9, *P*<.001), get news (χ^2_1 = 23.5, *P*<.001), and get health information (χ^2_1 = 9.7, *P*=.002) than those with less than a high school education. When controlling for household income (Table 3), parents with more education were more likely to own an Internet-capable device (*P*=.009) and to use the Internet for email and news (*P*<.001). Those with more than a high school education were more likely to seek health information online than those with less education, but this did not reach statistical significance (*P*=.03).

Chi-square tests (Table 4) and multivariate logistic regressions (Table 3) revealed that parents with higher incomes were more likely to own all types of Internet-capable devices (χ^2_1 = 12.1, *P*=.001) and twice as likely to have Internet access at home (*P*=.002). There were no significant differences in Internet use or interest in online health information by child health status (results not shown).

Table 4. Participant Internet use by education and income (N=302).

Internet use	Overall n (%)	Parent education n (%)		Household income (US$) n (%)	
		≤High school (n=156)	>High school (n=144)	<25,000 (n=124)	≥25,000 (n=130)
Do you own any of the following items?					
None[a,b]	52 (17.2)	41 (26.3)	11 (7.6)	29 (23.4)	10 (7.7)
Desktop computer[a,b]	132 (43.7)	53 (34.0)	78 (54.2)	41 (33.1)	76 (58.5)
Laptop or notebook[a,b]	167 (55.3)	70 (44.9)	96 (66.7)	57 (46.0)	93 (71.5)
iPod or MP3 player[a,b]	114 (37.7)	46 (29.5)	68 (47.2)	37 (29.8)	67 (51.5)
Game console[b]	150 (49.7)	71 (45.5)	79 (54.9)	53 (42.7)	80 (61.5)
Tablet computer[b]	26 (8.6)	9 (5.8)	17 (11.8)	5 (4.0)	20 (15.4)
Do you have Internet access at home (other than a cell phone)? (yes)[a,b]	228 (75.5)	106 (67.9)	120 (83.3)	83 (66.9)	
Have you ever used the Internet to do any of the following?					
Send/read email[a]	254 (84.1)	116 (74.4)	136 (94.4)	107 (86.3)	115 (88.5)
Get news online[a]	210 (69.5)	90 (57.7)	119 (82.6)	85 (68.5)	103 (79.2)
Online chat	155 (51.3)	79 (50.6)	76 (52.8)	69 (55.6)	64 (49.2)
Social networking site	211 (69.9)	102 (65.4)	108 (75.0)	89 (71.8)	93 (71.5)
Write or read blogs	161 (53.3)	78 (50.0)	82 (56.9)	61 (49.2)	74 (56.9)
Use Twitter	102 (33.8)	50 (32.1)	51 (35.4)	41 (33.1)	46 (35.4)
Watch video sharing site	191 (63.2)	92 (59.0)	98 (68.1)	81 (65.3)	83 (63.8)
Search engine	209 (69.2)	101 (64.7)	107 (74.3)	91 (73.4)	89 (68.5)
Get health info[a]	160 (53.0)	68 (43.6)	91 (63.2)	67 (54.0)	73 (56.2)
Video games	121 (40.1)	59 (37.8)	61 (42.4)	56 (45.2)	51 (39.2)
Would you be willing to receive email or text messages to get health information? (yes)	264 (87.4)	141 (91.0)	121 (84.0)	112 (90.3)	109 (83.8)
Would you be Interested in receiving health information on the Internet via email or online? (yes)	255 (84.4)	133 (85.8)	120 (83.9)	109 (88.6)	109 (83.8)
Would you be interested in keeping track of your child's health online? (yes)	279 (92.4)	141 (91.0)	136 (94.4)	115 (92.7)	120 (92.3)

[a] Chi-square tests indicate a significant (*P*<.01) difference between parents with ≤high school vs >high school education.

[b] Chi-square tests indicate a significant (*P*<.01) difference between parents with annual household incomes <US $25,000 vs ≥US $25,000.

Social Networking

More than 80.8% (244/302) of parents in our sample used social networking, primarily Facebook (78.8%, 238/302), and close to half (46.4%, 140/302) accessed these sites daily (see Table 5). Almost three-quarters (74.2%, 224/302) were interested in joining a social networking group about a health topic concerning their child.

There were very few education or income differences in social networking activity among African American mothers. Chi-square tests ($\chi^2_1 = 9.8$, *P*=.002) and multivariate logistic regressions (Table 3; *P*=.005) indicated that those with lower incomes were more likely to use MySpace. Also, chi-square tests (Table 5) showed parents with more education ($\chi^2_1 = 11.4$, *P*=.001) and higher household incomes ($\chi^2_1 = 7.0$, *P*=.008) were more likely to use LinkedIn compared to their counterparts, but multivariate results suggest the effects of education and income were not independent (Table 3). There were no significant differences in social networking by child health status (results not shown).

Table 5. Participant social networking by education and income (N=302).

Social networking use	Overall n (%)	Parent education n (%)		Household income (US $) n (%)	
		≤High school (n=156)	>High school (n=144)	<25,000 n=124)	≥25,000 (n=130)
Ever use social networking	244 (80.8)	130 (83.3)	112 (77.8)	104 (83.9)	99 (76.2)
Which social networking sites do you use?					
Facebook	238 (78.8)	126 (80.8)	110 (76.4)	102 (82.3)	96 (73.8)
MySpaceb	24 (7.9)	14 (9.0)	10 (6.9)	18 (14.5)	5 (3.8)
LinkedIn[a,b]	16 (5.3)	2 (1.3)	14 (9.7)	3 (2.4)	13 (10.0)
Twitter	90 (29.8)	47 (30.1)	41 (28.5)	41 (33.1)	35 (26.9)
How often do you visit social networking sites?					
Every day	140 (46.4)	74 (57.4)	65 (58.0)	55 (52.9)	61 (61.6)
2-3 days/week	50 (16.6)	25 (19.4)	25 (22.3)	24 (23.1)	16 (16.2)
Weekly	33 (10.9)	18 (14.0)	14 (12.5)	17 (16.3)	12 (12.1)
Monthly	20 (6.6)	12 (9.3)	8 (7.1)	8 (7.7)	10 (10.1)
Would you join a social networking group about a health topic concerning your child? (yes)	224 (74.2)	114 (74.5)	108 (78.8)	98 (83.1)	94 (74.0)

[a] Chi-square tests indicate a significant (*P*<.01) difference between parents with ≤high school vs >high school education.

[b] Chi-square tests indicate a significant (*P*<.01) difference between parents with annual household incomes <US $25,000 vs ≥US $25,000.

Discussion

This study aimed to guide the development of mHealth communication technologies designed for URM pediatric patients, particularly technologies that could address chronic conditions for which there are racial/ethnic disparities. We examined patterns of mHealth communication technology access, use, and preferences among a socioeconomically diverse sample of African American parents of underserved children in urban community health centers.

Our data confirm that African American adults, primarily mothers in this case, are avid users of mobile phones. Nearly everyone in our sample (97%) reported owning a cell phone, which is similar to recent findings by the Pew Research Center on African American cell phone ownership (93%) [9]. Approximately 90% of all participants used their cell phones to text, and approximately 80% used them to access the Internet. These rates are even higher than those for non-Hispanic black adults reported in Pew's May 2012 national survey (80% and 60% for texting and Internet access, respectively) [22]. The higher rates in our sample may be due to participants' residing in a metropolitan area [23], but they also may be related to the participants' specific age and status as parents which has an associated need to be easily accessible. The health communication implication of their pervasive use of mobile phones for texting and Internet access is that these channels may be a more acceptable means than voice calls for health care providers to communicate with parents/caregivers of pediatric patients. Although relatively few African American mothers (18%) shared health information via text, there was strong interest in receiving health information via mobile phones (87%). Thus, findings support both high levels of nonvoice mobile phone use and high levels of interest in receiving health information via this channel.

A caveat in the ubiquity of mobile phone use is that over 25% of African American parents in our sample experienced cell phone service interruption in the prior year. This means that for a notable proportion of African American mothers, mobile phones are not a totally reliable form of communication. Nevertheless, there is probably not a more consistent means of contacting these parents/caregivers. We know of no other studies that have examined the frequency and pervasiveness of cell phone service interruptions. Therefore, it would be useful to collect more information about service interruptions and turnover in phone numbers when designing mHeath interventions for this population.

Our data also indicate that most urban African American mothers can access the Internet at home (76%). Although home Internet access in our sample was more prevalent among those with higher income (86%), it was still relatively pervasive among lower income families (67%). Our findings regarding home Internet access were similar to those of DeMartini and colleagues [21], who found that 80% of their sample of urban parents had home Internet access. Also, our finding that lower income parents were less likely to report home Internet access parallels the Pew Center data, although national rates for adults are lower (46% of those with annual incomes less than US $30,000 reported home broadband access) [24].

In addition to discrepancies in home Internet access, there were also significant socioeconomic discrepancies in mobile device ownership. Although there were no significant differences in cell phone ownership, mothers with higher incomes were at least twice as likely to own various other mobile devices (eg, laptop, tablet computer) than their lower income counterparts, which parallels the Pew Center's 2013 survey data [25]. This income disparity is important for developers to account for when considering the platforms and user interface of apps and other mHealth technologies.

We found that email and websites, including social networks, are other venues with large potential for communicating with URM parents about children's health management because 84% reported email use and 70% reported social networking use (46% reported daily use). Our findings regarding the frequency of email and social networking use parallel data on all adults from the Pew Center (88% of Internet users send or receive email and 67% use a social networking site) [26] and from DeMartini and colleagues (91% of parents reported having an email account) [21]. Moreover, 84% of our sample reported interest in receiving health information on the Internet via email or online, and more than half are already using the Internet to get health information. Despite potential privacy concerns, 74% of mothers were interested in joining a social networking group about a health topic concerning their children. DeMartini and colleagues similarly found among their sample of urban parents, strong interest in the use of mHealth data with more than 70% of parents reporting that they would use digitally supplied health information.

We found no significant difference in Internet/mobile device use or interest in using these outlets to send/receive information about their children's health between parents of healthy children and parents of children with a chronic health condition. We know of no other study that has examined this issue, and more research is needed to confirm our findings. Yet, our results suggest that mHealth could be harnessed to reduce racial disparities across both chronic diseases and more general child health measures (eg, vaccination receipt)[27].

This survey of urban African American mothers of pediatric primary care patients is an important first step in understanding acceptability of mHealth communication technology in this population, but there is more specific and nuanced information that will be necessary for designing technologies targeting them. For example, this survey did not ask about smartphone ownership or operating systems used (eg, Android, Apple), cell phone service providers, utilization of different types of apps, or desire/need for mobile products for managing children's chronic conditions. Furthermore this was a convenience sample that may not be representative of URM parents who do not seek primary care for their children (ie, rely on emergency departments) and does not reflect the usage and preferences of non-English speakers. Future studies of mobile technology use in pediatric settings should measure the usage of patients themselves (ie, children) and assess eHealth literacy to understand the capacity for content uptake by patients/caregivers.

Overall, this study confirmed that African American parents, even those with lower incomes, are frequent users of various mobile technologies. Although few are currently using such modalities for managing their children's health, there was pervasive interest in doing so. Designers and researchers should note that mobile use varies by parent education and income and may need to account for service interruptions. Nonetheless, African American parents' access and interest in using mHealth make it a promising platform for reducing pediatric health disparities.

Acknowledgments

This project was supported by Award Number UL1RR031988 /UL1TR000075 from the NIH National Center for Advancing Translational Sciences. Research reported in this publication was also supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award Number P20MD000198. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. US Department of Health and Human Services, Health Information Technology and Quality Improvement. 2013. Video report: What is mHealth? URL: http://www.hrsa.gov/healthit/mhealth.html [accessed 2013-07-16] [WebCite Cache ID 6I9xJclEO]
2. Gentles SJ, Lokker C, McKibbon KA. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. J Med Internet Res 2010;12(2):e22 [FREE Full text] [doi: 10.2196/jmir.1390] [Medline: 20562092]
3. HealthyPeople.gov. 2012. Health Communication and Health Information Technology URL: http://www.healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=18 [accessed 2013-04-04] [WebCite Cache ID 6FdJDBQab]
4. Zickuhr K, Smith A. Digital Differences. Washington, DC: Pew Internet & American Life Project; 2012 Apr 13. URL: http://www.pewinternet.org/Reports/2012/Digital-differences.aspx [accessed 2013-04-11] [WebCite Cache ID 6FnmhB24L]
5. Pew Internet & American Life Project. 2013. Trend data (adults): Demographics of Internet users URL: http://pewinternet.org/Trend-Data-(Adults)/Whos-Online.aspx [accessed 2013-07-11] [WebCite Cache ID 6I2NbjLy0]
6. Smith A. Smartphone ownership - 2013 update. Washington, DC: Pew Internet & American Life Project; 2013 Jun 5. URL: http://www.pewinternet.org/~/media/Files/Reports/2013/PIP_Smartphone_adoption_2013.pdf, [accessed 2013-11-27] [WebCite Cache ID 6LRnXrEvA]
7. Fox S. Pew Internet: Health. Washington, DC: Pew Internet & American Life Project; 2013 Jul 01. URL: http://www.pewinternet.org/Commentary/2011/November/Pew-Internet-Health.aspx [accessed 2013-07-12] [WebCite Cache ID 6I2ppmgxa]
8. Kind T, Huang ZJ, Farr D, Pomerantz KL. Internet and computer access and use for health information in an underserved community. Ambul Pediatr 2005;5(2):117-121. [doi: 10.1367/A04-107R.1] [Medline: 15780014]
9. Brenner J. Pew Internet: Mobile. Washington, DC: Pew Internet & American Life Project; 2013 Jun 06. URL: http://pewinternet.org/Commentary/2012/February/Pew-Internet-Mobile.aspx [accessed 2013-07-12] [WebCite Cache ID 6I2qCoXDw]
10. Fox S, Duggan M. Mobile Health 2012. Washington, DC: Pew Internet & American Life Project; 2012 Nov 08. URL: http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2013-07-11] [WebCite Cache ID 6I2CA4UPh]
11. Akinbami LJ, Moorman JE, Garbe PL, Sondik EJ. Status of childhood asthma in the United States, 1980-2007. Pediatrics 2009 Mar;123 Suppl 3:S131-S145 [FREE Full text] [doi: 10.1542/peds.2008-2233C]
12. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. JAMA 2012 Feb 1;307(5):483-490. [doi: 10.1001/jama.2012.40] [Medline: 22253364]
13. Ngo-Metzger Q, Hayes GR, Yunan Chen, Cygan R, Garfield CF. Improving communication between patients and providers using health information technology and other quality improvement strategies: focus on low-income children. Med Care Res Rev 2010 Oct;67(5 Suppl):246S-267S. [doi: 10.1177/1077558710375431] [Medline: 20675346]
14. Neville R, Greene A, McLeod J, Tracey A, Tracy A, Surie J. Mobile phone text messaging can help young people manage asthma. BMJ 2002 Sep 14;325(7364):600 [FREE Full text] [Medline: 12228151]
15. Kharbanda EO, Stockwell MS, Fox HW, Andres R, Lara M, Rickert VI. Text message reminders to promote human papillomavirus vaccination. Vaccine 2011 Mar 21;29(14):2537-2541. [doi: 10.1016/j.vaccine.2011.01.065] [Medline: 21300094]
16. Ruppel EK, Rains SA. Information sources and the health information-seeking process: An application and extension of channel complementarity theory. Communication Monographs 2012 Sep;79(3):385-405. [doi: 10.1080/03637751.2012.697627]
17. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
18. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
19. Cohall AT, Cohall R, Dye B, Dini S, Vaughan RD. Parents of urban adolescents in Harlem, New York, and the Internet: a cross-sectional survey on preferred resources for health information. J Med Internet Res 2004 Dec 3;6(4):e43 [FREE Full text] [doi: 10.2196/jmir.6.4.e43] [Medline: 15631967]
20. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
21. Demartini TL, Beck AF, Klein MD, Kahn RS. Access to digital technology among families coming to urban pediatric primary care clinics. Pediatrics 2013 Jul;132(1):e142-e148. [doi: 10.1542/peds.2013-0594] [Medline: 23753100]
22. Duggan M, Rainie L. Cell Phone Activities 2012. Washington, DC: Pew Internet & American Life Project; 2012 Nov 25. URL: http://pewinternet.org/Reports/2012/Cell-Activities.aspx [accessed 2013-07-12] [WebCite Cache ID 6I2rBkWFH]
23. Fox S. Rural e-patients face access challenges. Washington, DC: Pew Internet & American Life Project; 2012 Nov 26. URL: http://www.pewinternet.org/Commentary/2012/November/Rural-epatients-face-access-challenges.aspx [accessed 2012-04-11] [WebCite Cache ID 6Fnmk4SQz]

24. Brenner J, Rainie L. Pew Internet: Broadband. Washington, DC: Pew Internet & American Life Project; 2012 May 01. URL: http://pewinternet.org/Commentary/2012/May/Pew-Internet-Broadband.aspx [accessed 2013-07-11] [WebCite Cache ID 6I2rI3IHr]
25. Zickuhr K. Tablet Ownership 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jun 10. URL: http://pewinternet.org/Reports/2013/Tablet-Ownership-2013/Findings.aspx [accessed 2013-07-11] [WebCite Cache ID 6I2rQfraT]
26. Pew Internet & Life Project. 2012 Dec. Trend data (Adults) URL: http://pewinternet.org/Trend-Data-(Adults)/Online-Activites-Total.aspx [accessed 2013-07-12] [WebCite Cache ID 6I2rcVKcl]
27. Stockwell MS, Kharbanda EO, Martinez RA, Lara M, Vawdrey D, Natarajan K, et al. Text4Health: impact of text message reminder-recalls for pediatric and adolescent immunizations. Am J Public Health 2012 Feb;102(2):e15-e21. [doi: 10.2105/AJPH.2011.300331] [Medline: 22390457]

Abbreviations

GED: graduate equivalency degree
mHealth: mobile health
URM: underrepresented minority

Edited by G Eysenbach; submitted 16.04.13; peer-reviewed by A Moorhead, N Warshay; comments to author 15.06.13; revised version received 16.07.13; accepted 22.07.13; published 13.01.14.

Please cite as:
Mitchell SJ, Godoy L, Shabazz K, Horn IB
Internet and Mobile Technology Use Among Urban African American Parents: Survey Study of a Clinical Population
J Med Internet Res 2014;16(1):e9
URL: http://www.jmir.org/2014/1/e9/
doi:10.2196/jmir.2673
PMID:24418967

Original Paper

Biological Calibration for Web-Based Hearing Tests: Evaluation of the Methods

Marcin Masalski[1,2], MD, PhDEng; Tomasz Grysiński[2], PhDEng; Tomasz Kręcicki[1], MD, PhD

[1]Department and Clinic of Otolaryngology, Head and Neck Surgery, Wroclaw Medical University, Wrocław, Poland

[2]Institute of Biomedical Engineering and Instrumentation, Wroclaw University of Technology, Wrocław, Poland

Corresponding Author:
Marcin Masalski, MD, PhDEng
Department and Clinic of Otolaryngology, Head and Neck Surgery
Wroclaw Medical University
Wybrzeże L Pasteura 1
Wrocław, 50-367
Poland
Phone: 48 71 734 37 00
Fax: 48 71 733 12 09
Email: marcin.masalski@pwr.wroc.pl

Abstract

Background: Online hearing tests conducted in home settings on a personal computer (PC) require prior calibration. Biological calibration consists of approximating the reference sound level via the hearing threshold of a person with normal hearing.

Objective: The objective of this study was to identify the error of the proposed methods of biological calibration, their duration, and the subjective difficulty in conducting these tests via PC.

Methods: Seven methods have been proposed for measuring the calibration coefficients. All measurements were performed in reference to the hearing threshold of a normal-hearing person. Three methods were proposed for determining the reference sound level on the basis of these calibration coefficients. Methods were compared for the estimated error, duration, and difficulty of the calibration. Web-based self-assessed measurements of the calibration coefficients were carried out in 3 series: (1) at a otolaryngology clinic, (2) at the participant's home, and (3) again at the clinic. Additionally, in series 1 and 3, pure-tone audiometry was conducted and series 3 was followed by an offline questionnaire concerning the difficulty of the calibration. Participants were recruited offline from coworkers of the Department and Clinic of Otolaryngology, Wroclaw Medical University, Poland.

Results: All 25 participants, aged 22-35 years (median 27) completed all tests and filled in the questionnaire. The smallest standard deviation of the calibration coefficient in the test-retest measurement was obtained at the level of 3.87 dB (95% CI 3.52-4.29) for the modulated signal presented in accordance with the rules of Bekesy's audiometry. The method is characterized by moderate duration time and a relatively simple procedure. The simplest and shortest method was the method of self-adjustment of the sound volume to the barely audible level. In the test-retest measurement, the deviation of this method equaled 4.97 dB (95% CI 4.53-5.51). Among methods determining the reference sound level, the levels determined independently for each frequency revealed the smallest error. The estimated standard deviations of the difference in the hearing threshold between the examination conducted on a biologically calibrated PC and pure-tone audiometry varied from 7.27 dB (95% CI 6.71-7.93) to 10.38 dB (95% CI 9.11-12.03), depending on the calibration method.

Conclusions: In this study, an analysis of biological calibration was performed and the presented results included calibration error, calibration time, and calibration difficulty. These values determine potential applications of Web-based hearing tests conducted in home settings and are decisive factors when selecting the calibration method. If there are no substantial time limitations, it is advisable to use Bekesy method and determine the reference sound level independently at each frequency because this approach is characterized by the lowest error.

(J Med Internet Res 2014;16(1):e11) doi:10.2196/jmir.2798

KEYWORDS

pure-tone audiometry; computer-assisted instruction; self-examination

Introduction

Sound systems of modern home electronic equipment, such as a personal computer (PC), tablet, or smartphone, offer opportunities to conduct hearing examinations at low cost and on a large scale [1-4]. The population of people who are computer literate is aging and their hearing sensitivity is declining. Therefore, the number of individuals potentially interested in this type of testing is increasing. Additionally, research shows that the use of the Internet is higher in the hearing-impaired population in comparison to similar age groups in the general population [5,6].

Hearing tests conducted remotely in home settings on PCs can be divided into 2 groups depending on the necessity of conducting prior calibration. The examinations which do not require prior calibration are usually screening tests represented by speech-in-noise tests [1,7-9]. The speech-in-noise test involves the evaluation of speech intelligibility in relation to signal-to-noise ratio; therefore, the knowledge of the absolute sound level is not required. The speech-in-noise test contributes to increased identification of hearing loss [1] and is more useful in screening tests than a short questionnaire [7]. Additionally, sensitivity of the test can be improved after applying low-band noise [9].

However, most hearing tests, including the basic examination in the form of pure-tone audiometry, require prior calibration of the system, and its omission leads to significant measurement errors [10]. Calibration consists of determining the reference sound level. For the purposes of the hearing test conducted in in-home conditions, it can be performed in a number of ways. The calibration of a PC system can be carried out in a laboratory setting beforehand and then later used for home-based examinations [11]. Another solution is to prepare software that will cooperate with an audio set whose parameters are known, consisting of a sound card and headphones [12]. In this case, to conduct a home-based examination requires purchasing a particular set. Both the previously mentioned solutions limit accessibility of the hearing test because they require efforts that are unjustified in the case of a single hearing test. In light of this, biological calibration seems a sensible solution, consisting of approximation of the reference sound level by the hearing threshold of a person with normal hearing. Usually the reference sound level is assumed at 0 decibel hearing level (dB HL).

Honeth et al [3] used biological calibration based on evaluation of the hearing threshold of a person with normal hearing at the following frequencies: 500 Hz, 1 kHz, 2 kHz, 6 kHz, and 8 kHz. The task of the reference person was to set the volume marker at the level at which the sound was barely audible. In this way, the reference sound level was determined individually for each frequency. The test results were compared with pure-tone audiometry and exhibited the greatest error at 2 and 4 kHz, corresponding to 5.6 dB (SD 8.29) and 5.1 dB (SD 6.9), respectively. In all, 89% of the tests were conducted on the same computer and with the same reference person. Masalski and Kręcicki [4] also used biological calibration based on evaluation of the hearing threshold of a reference person by using a volume marker. Calibration was conducted for 1 kHz only, and the values of 0 dB HL at other frequencies were calculated on the basis of the A-weight filter. Self-examinations conducted by the participants on their home computers calibrated by their normal-hearing family members showed a mean error of the hearing threshold compared to pure-tone audiometry at the level of -1.35 dB (SD 10.66).

The error analysis of the pure-tone audiometry conducted on a PC calibrated by the biological method showed significant influence of the calibration error [4]. The standard deviation of the calibration error at 1 kHz was 6.19 dB, whereas the measurement was additionally burdened with an estimation error of 0 dB HL conducted on the basis of the A-weight filter at other frequencies. The highest estimation error was at 250 Hz at the level of 7.28 dB. Nevertheless, sensitivity and specificity values calculated for the detection of noise-induced hearing loss, compared with pure-tone audiometry, were found to be reasonable (ie, at the level of sensitivity 0.89, 95% CI 0.74-1.0 and specificity 0.89, 95% CI 0.76-1.0). Similar sensitivity and specificity were obtained by Honeth et al [3] (sensitivity 0.75, 95% CI 0.51-0.90 and specificity 0.96, 95% CI 0.96-0.99).

The application of pure-tone audiometry based on biological calibration depends significantly on the measurement error. Because of the much larger error of biological calibration than tolerance required by the standards (ie, ±3 dB in the frequency range 125 Hz to 5 kHz [13]), the home test cannot be an alternative to classical pure-tone audiometry. However, it may be applied as a screening test as well as in other situations suggested in other studies [3,4], such as self-monitoring of hearing for some disorders (eg, fluctuating hearing loss, tinnitus, sudden deafness, otosclerosis, Ménière's disease), during treatment with ototoxic drugs, in large-scale epidemiological studies, in cases of limited access to specialist equipment (eg, at the general practitioner's office or in countries with low economic status), and also as a telemedical examination combined with a questionnaire to determine the direction of further treatment. However, before verification of these applications it is advisable to optimize biological calibration [4].

This paper presents 7 methods of measuring the calibration coefficients. All measurements were performed in reference to the hearing threshold of a normal-hearing person. For each method, the measurement error was determined, as well as the timeframe for its calibration and the difficulty level. Next, 3 methods were proposed for determining the reference sound level on the basis of these calibration coefficients and an error analysis was conducted for each.

Methods

Overview

The proposed methods of biological calibration consist in measuring the calibration coefficient that describes the threshold sound level of the reference person. Seven calibration methods were proposed: (1) calibration using an amplitude-modulated signal, (2) calibration using 2 sounds differing by 5 dB, (3) calibration using 2 sounds differing by 2 dB, (4) the ascending method with a 5-dB step, (5) the ascending method with a 2-dB step, (6) calibration based on Bekesy audiometry using the continuous signal, and (7) calibration based on Bekesy audiometry using an amplitude-modulated signal. In methods 1-5, the assessment was conducted for the following frequencies: 125 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz. In methods 6 and 7, the frequency was changed in a continuous way from 62.5 Hz to 16 kHz. The sound signal was presented bilaterally.

Amplitude-Modulated Signal Method

In the calibration with amplitude-modulated signal (method 1), the presented signal was amplitude-modulated using rectangular envelope with frequency of 1 Hz and modulation depth of 100%. The task of the reference person was to set the volume marker in such a way that the generated sound be barely audible. The step of the volume marker was 1 dB.

Dual Tone Methods

During calibration with 2 sounds differing in intensity (methods 2 and 3), 2 tone signals with a given frequency and a duration of 1 s were presented in turns. The task of the reference person was to set the volume marker in such a way that the louder of the 2 sounds was still audible, and the quieter inaudible. In method 2, the signals differed by 5 dB, whereas in method 3, the difference was 2 dB. The step of the volume marker was 1 dB.

Ascending Methods

The ascending method (methods 4 and 5) was based on the ascending algorithm used for the assessment of the hearing threshold in pure-tone audiometry [14]. A signal with a given frequency was presented for a random duration from 2 s to 7 s. The task of the reference person was to press a button on hearing a sound and release it when it was no longer audible. The button should be pressed up to 2 s from the start of playing the sound and released up to 2 s after it stopped. The level of the tone was reduced in 10-dB steps until no further response occurred, and then it was increased in 5-dB steps until the participant responded. The calibration coefficient was defined as the lowest level at which responses occurred in at least half of the series of ascending trials with a minimum of 2 responses required at that level. No more than 5 previously conducted ascending trials were taken into account. Method 4 used 5-dB and 10-dB steps. Method 5 used a 4-dB step down and a 2-dB step up.

Bekesy Methods

During calibrations based on Bekesy audiometry, the frequency of a presented signal was increased at the speed of 1 octave/60 s, simultaneously with the change of its intensity. The task of the reference person was to press a button on hearing the signal and keep it pressed for as long as the sound was audible. The intensity of the sound was reduced at a speed of 2 dB/s when the sound was audible, and increased at the same speed when the sound was inaudible. The value of the calibration coefficient was determined as the mean of the values of the sound intensity at which a change in the status of the button occurred, after rejecting the outliers on the basis of the Grubbs' test [15]. Coefficients were determined for the frequencies of 125 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz by calculating the mean of the range ±0.5 octave. Method 6 used a continuous signal, whereas method 7 used a signal modulated in amplitude by a sinusoidal envelope with frequency of 2 Hz and modulation depth of 100%.

All 7 methods were implemented in Java technology in the form of applets embedded in a Web browser. The calibration coefficients expressing the sound intensity in decibels, together with the duration of examinations, were recorded in the database. On completion of all the tests, the participant filled in an offline questionnaire (Multimedia Appendix 1) on the difficulty of the tests by assigning each method a value from 0 (the easiest method) to 10 (the hardest method).

Reference Sound Level Methods

In addition to the 7 methods of measuring the calibration coefficients, 3 methods were proposed for determining the reference sound level: (1) the reference sound level determined independently for each frequency depending on the value of calibration coefficient measured at this frequency, (2) the reference sound level estimated by a model fitted to calibration coefficients determined at all frequencies, and (3) as (2) except the model was fitted to a single coefficient determined at the frequency characterized by the smallest measurement error.

Participants were recruited offline from coworkers of Department and Clinic of Otolaryngology, Wroclaw Medical University, Poland, using face-to-face prompting from September 2012 to March 2013. The eligibility criteria were age younger than 35 years, lack of previous hearing problems, owning headphones and a PC at home and basic skills to operate it, and the willingness to participate in the research. Each participant performed calibration using all 7 methods 3 times. In series 1, the study was carried out in a sound booth with the use of notebook Dell Vostro 1310 with Microsoft Windows 7 operational system and Technics RP-F290 headphones; in series 2, each person was asked to perform calibration on their own home computer using their own headphones in the quietest conditions possible, preferably late in the evening or at night to minimize background noise level and to create conditions close to those in the sound booth; and series 3 was the repetition of examinations from series 1. Because of the relatively long duration of the series, the participants were informed about the option of taking a break when they felt tired, and most of the participants took advantage of this. In series 1 and 3, pure-tone audiometry was performed with the use of a clinical audiometer Interacoustic AD229e and TDH-39 headphones calibrated in accordance with ISO 389-1:1998. The hearing threshold was determined by using the ascending method in accordance with ISO 8253-1:2010. Additionally, based on the pure-tone audiometry, the bilateral hearing threshold was calculated by choosing for each frequency the threshold of the ear that heard better at this particular frequency.

A test-retest analysis of calibration coefficients was conducted, as well as 1-way ANOVA for measurement duration and difficulty. Calibration errors were determined by means of variance estimation. Statistical analyses were performed on the basis of confidence intervals that were estimated in the same way. Estimation of the variance was conducted based on measurement variances and their confidence intervals calculated from the variance and the sample size [16].

Results

Test-Retest Analysis

The 25 participants (11 men, 14 women), aged between 22 and 35 years (median 27), who took part in the study completed all the examinations and filled in the questionnaire. All participants were skilled in computer use. On the basis of series 1 and 3, a test-retest analysis was conducted. For each method, a mean difference, standard deviation of the difference and corresponding confidence intervals were calculated (Table 1). Mean values for the dual tone (2 dB), Bekesy (continual), and Bekesy (modulated) methods were significantly different than zero at the significance level of *P*=.05. At *P*=.01, this relation was insignificant for all the methods. The smallest standard deviation was obtained for the Bekesy (modulated) method.

Table 1. Mean difference and standard deviation of the difference with corresponding confidence intervals at *P*=.05 of hearing thresholds and calibration coefficients between series 1 and 3 calculated jointly for 8 frequencies (N=25).

Method	Difference (dB), mean (95% CI)	Difference (dB), SD (95% CI)
Hearing threshold		
Right ear	–0.38 (–1.13, 0.38)	5.40 (4.92, 5.99)
Left ear	0.13 (–0.62, 0.87)	5.34 (4.86, 5.92)
Both ears	–0.13 (–0.65, 0.40)	5.37 (5.02, 5.77)
Bilateral	–0.35 (–1.04, 0.34)	4.92 (4.48, 5.46)
Calibration coefficient		
Modulated signal	–0.09 (–0.78, 0.61)	4.97 (4.53, 5.51)
Dual tone (5 dB)	1.05 (0.00, 2.11)	7.54 (6.87, 8.37)
Dual tone (2 dB)	1.20 (0.06, 2.35)	8.18 (7.45, 9.07)
Ascending (5-dB step)	–0.15 (–1.00, 0.70)	6.05 (5.51, 6.71)
Ascending (2-dB step)	–0.10 (–0.80, 0.60)	5.00 (4.55, 5.54)
Bekesy (continual)	0.88 (0.19, 1.57)	4.92 (4.48, 5.46)
Bekesy (modulated)	0.63 (0.09, 1.17)	3.87 (3.52, 4.29)

Duration of Calibration

Durations of calibration in relation to the calibration methods are presented in Figure 1. The durations were significantly different (*P*<.001).

The shortest times were obtained for the modulated signal method and both dual tone methods (5 and 2 dB), which consisted in self-adjusting the volume marker. The mean duration of calibration based on the ascending method with the step of 5 dB was comparable to the duration of calibration using both Bekesy methods (continual and modulated). In the Bekesy methods, the outliers are those examinations that were paused momentarily.

Calibration’s Degree of Difficulty

The degree of difficulty of the methods were significantly different (*P*<.001). The easiest method of calibration was the modulated signal method consisting in self-adjusting the volume marker in such a way that the presented tone was barely audible. Subsequently, the easiest methods were based on Bekesy’s audiometry (Figure 2).

Figure 1. Calibration durations for all 7 calibration methods in series 1-3 (N=25). The horizontal line in each box represents the median, top and bottom box borders represent 75th and 25th percentiles, respectively; crosses represent outliers. MOD: modulated signal; 2TONE5: dual tone (5 dB); 2TONE2: dual tone (2 dB); ASC5: ascending (5-dB step); ACS2: ascending (2-dB step); BEK: Bekesy (continual); BEKM: Bekesy (modulated).

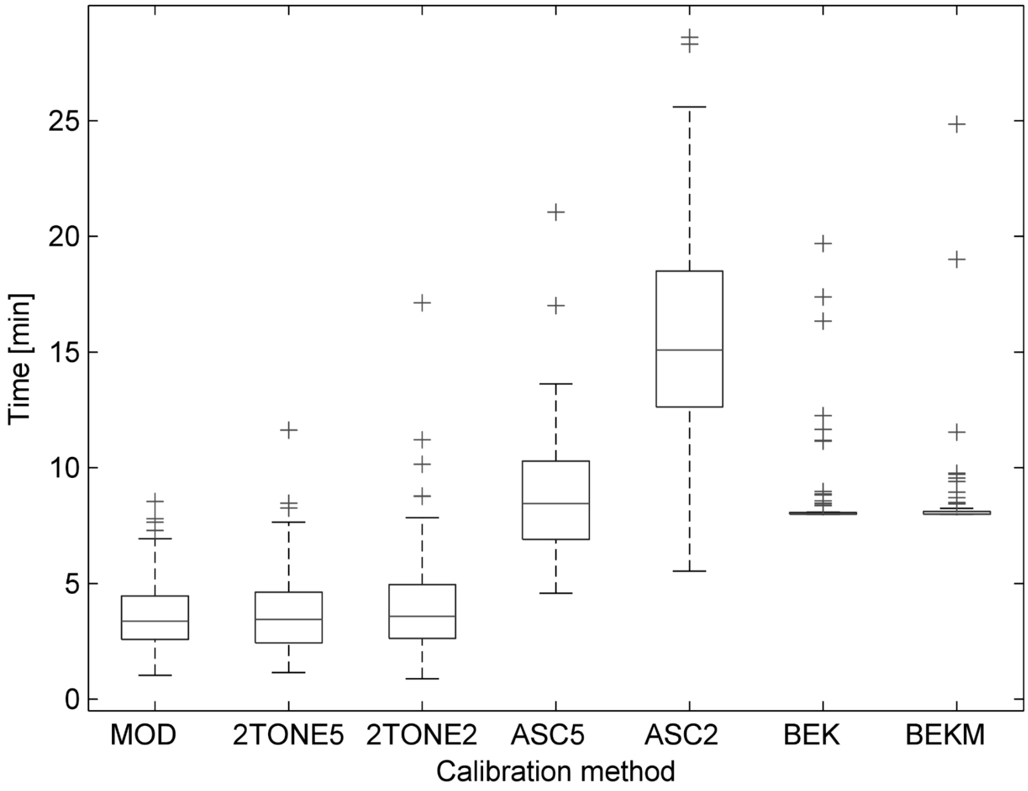
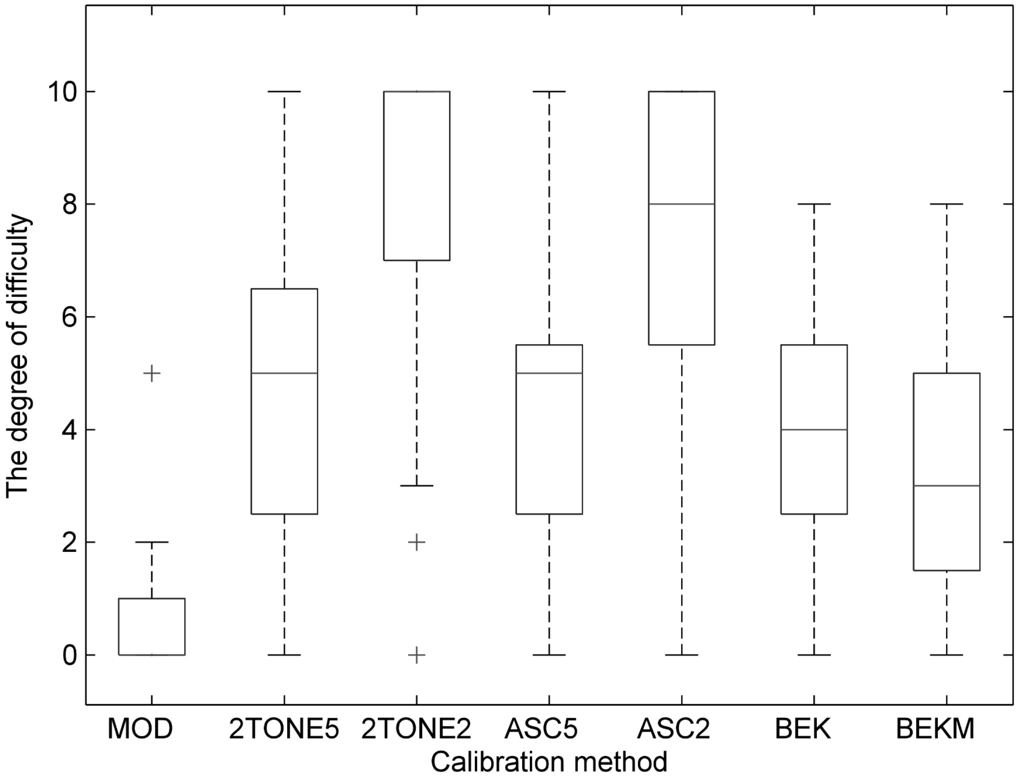


Figure 2. Difficulty ratings of the calibration methods evaluated by 25 participants (0=easiest; 10=hardest). The horizontal line in each box represents the median, top and bottom box borders represent 75th and 25th percentiles, respectively; crosses represent outliers. MOD: modulated signal; 2TONE5: dual tone (5 dB); 2TONE2: dual tone (2 dB); ASC5: ascending (5-dB step); ACS2: ascending (2-dB step); BEK: Bekesy (continual); BEKM: Bekesy (modulated).



Evaluation of the Frequency Response Model

There were 3 methods used to determine the reference sound level. Two were based on the frequency response model of a common sound card and headphones set. Therefore, comparison of methods requires prior evaluation of the model that was conducted using standard deviation of the residual. The mean standard deviation of the residual was calculated on the basis of the differences between the model and the coefficients in series 2 after taking into account the measurement error of coefficients, bilateral hearing threshold of the reference person, and measurement error of this threshold. Measurement error of calibration coefficients and the measurement error of bilateral hearing threshold were calculated from the test-retest differences between series 1 and 3. The standard deviation of the residual estimated in this way describes the difference between the actual coefficients and those calculated for the model fitted on their basis. This standard deviation is independent of the measurement method and the hearing threshold of the reference person. Detailed calculations are presented subsequently. A detailed list of all equations can be found in Multimedia Appendix 2.

Let us assume that C_i is the real value of the calibration coefficient at frequency f_i, and c_i denotes its value determined with some error. Moreover, let the model be given as a set of coefficients M_i estimated as follows:

M_i=mean(C)–FR(f_i) (1)

where mean(C) is the mean value of coefficients C_i and FR(f_i) is the frequency response of the model.

Let's assume that random variable X describes the desired difference between the model M_i and coefficients C_i and random variable Y denotes the determination error of the coefficient c_i, namely the difference C_i–c_i. Let's also define the random variable Z as the difference between the model M_i and determined coefficient c_i. In this way, we obtain 3 random variables X, Y, and Z, which take on the following values x_i, y_i, and z_i:

$x_i=M_i-C_i$ (2)

$y_i=C_i-c_i$ (3)

$z_i=M_i-c_i$ (4)

It is worth noting that random variables X and Y are independent. The standard error of the model fitted to the real calibration coefficients does not depend on the determination error of these coefficients. Therefore, bearing in mind that the variance of the sum of 2 independent random variables is the sum of their variances, the desired variance of the random variable X is:

variance(X)=variance(Z)–variance(Y) (5)

The mean value of the determined coefficient is close to the mean value of the real coefficient mean(c)≈mean(C). Then, on the basis of equation 1 we can calculate that $M_i \approx m_i$, where m_i is the model estimated on the basis of coefficients c_i:

m_i=mean(c)–FR(f_i) (6)

Bearing in mind that $M_i \approx m_i$, the variance of random variable Z may be calculated on the basis of the difference between coefficients c_i and the model m_i estimated on their basis, according to equation 7 (see Multimedia Appendix 2).

Coefficient c_i was determined by subtracting the bilateral hearing threshold from the measured calibration coefficient (equation 8). Therefore, standard deviation of the random variable Y expressing the standard error of coefficient c_i depends on the measurement error of calibration coefficient and the measurement error of the bilateral hearing threshold. Both measurement errors were calculated on the basis of the standard deviation of the differences in test-retest examination (equation 9 and Table 1).

c=(measured calibration coefficient)–(bilateral hearing threshold) (8)

variance(Y) = (calibration method test-retest difference SD)^2/2+(bilateral threshold test-retest difference SD)^2/2 (9)

which, on the basis of equations 5, 7, and 9, allows to estimate variance of the random variable X determining the model's error.

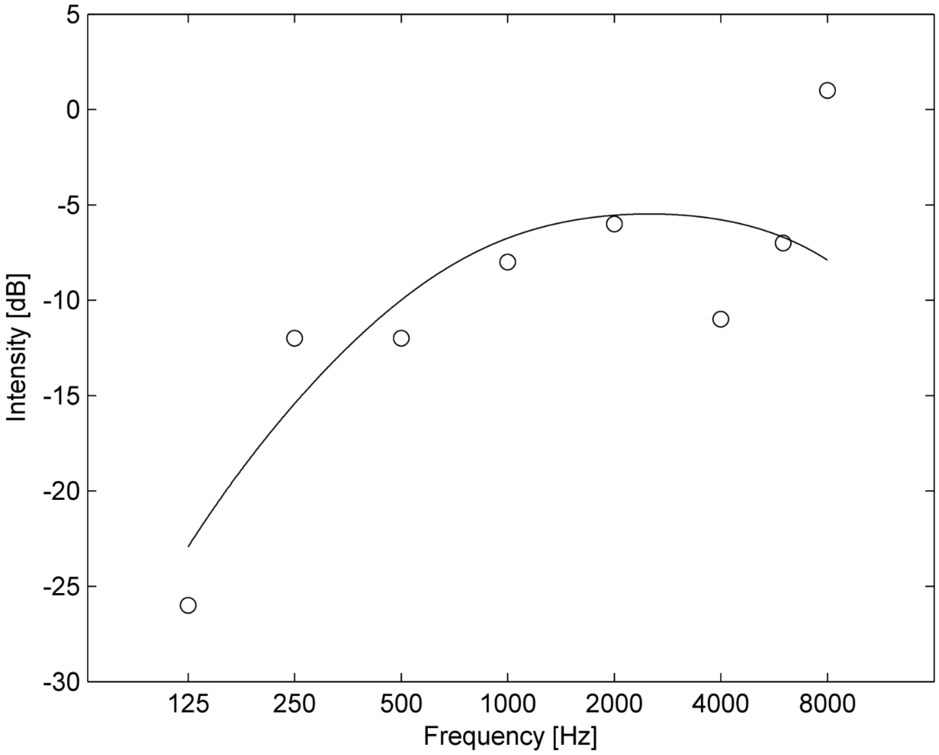
Following further calculations, a model based on an A-weight filter was assumed [17] (see equation 10 in Multimedia Appendix 2 and Figure 3).

For each calibration conducted in series 2, variance(Z) of the residual of the model was calculated (equation 7), and averaged for every calibration method. Next, for each calibration method, variance(Y) was computed on the basis of the standard deviation of the test-retest examination (equation 9 and Table 1). Finally, variance of residual of the model variance(X) was estimated independently for each calibration method (equation 5 and Table 2). The mean of standard deviations of residual of the model (model SD 6.57 dB, 95% CI 5.59-7.54) was used for further calculations.

Table 2. Standard deviation of residual of the model based on A-weight filter estimated by means of measurements at 8 frequencies carried out by 25 participants.

Calibration method	Model residual (dB), SD
Modulated signal	7.11
Dual tone (5 dB)	6.58
Dual tone (2 dB)	6.36
Ascending (5-dB step)	6.50
Ascending (2-dB step)	7.18
Bekesy (continual)	6.52
Bekesy (modulated)	5.69

Figure 3. The model of the frequency response fitted to a sample set of calibration coefficients.



The Reference Sound Level

The error of determining the reference sound level was estimated on the basis of intermediate values: the standard deviation of the bilateral hearing threshold in a population of people with normal hearing, the measurement error of calibration coefficients, and, in the case of methods based on the model, previously calculated error of the model expressed by the standard deviation of the residual. The standard deviation of the bilateral hearing threshold was determined from audiograms after eliminating the assessment error on the basis of the test-retest examination. Measurement error of calibration coefficients was also calculated from a test-retest examination.

The standard deviation of the bilateral hearing threshold measured by the means of pure-tone audiometry is affected by the population variability and measurement error. Knowing, that measurement error is equal to the standard deviation of the bilateral hearing threshold difference in test-retest examination (Table 1) divided by a square root of 2, the standard deviation of the real bilateral hearing threshold can be calculated from equation 11 (Table 3).

(measured bilateral threshold SD)^2=(real bilateral threshold SD)^2+(bilateral threshold test-retest difference SD)^2/2 (11)

Table 3. Standard deviation of the bilateral hearing threshold measured in series 1 by 25 participants, estimated measurement error and standard deviation of the real bilateral hearing threshold after eliminating the measurement error with corresponding confidence intervals at *P*=.05.

Frequency	Measured threshold (dB), SD (95% CI)	Measurement error (dB) SE (95% CI)	Real threshold (dB) SD (95% CI)
Bilateral hearing threshold			
At125 Hz	5.58 (4.35, 7.76)	3.48 (3.17,3.86)	4.35 (2.98, 6.73)
At 250 Hz	5.42 (4.23, 7.53)		4.15 (2.80, 6.47)
At 500 Hz	4.41 (3.44, 6.13)		2.70 (1.34, 4.76)
At 1 kHz	3.82 (2.98, 5.31)		1.57 (0.00, 3.65)
At 2 kHz	4.11 (3.21, 5.72)		2.19 (0.51, 4.22)
At 4 kHz	5.00 (3.90, 6.96)		3.59 (2.28, 5.79)
At 6 kHz	6.61 (5.16, 9.20)		5.62 (4.08, 8.35)
At 8 kHz	7.03 (5.49, 9.78)		6.11 (4.48, 8.98)
In range 125 Hz-8 kHz	5.36 (4.87, 5.95)		4.07 (3.70, 4.53)
Mean bilateral hearing threshold			
In range 125 Hz-8 kHz	2.62 (2.05, 3.65)	1.23 (1.12,1.37)	2.32 (1.71, 3.38)

The standard deviation of the bilateral hearing threshold difference in test-retest examination was calculated jointly for all frequencies due to lack of significant differences between frequencies in the 1-way ANOVA at the level of statistical significance *P*=.05.

Analogical computation were carried out for the mean value of bilateral hearing threshold, assuming the measurement error divided by a square root of 8 as the mean was for 8 frequencies (Table 3).

The error of the independent coefficients method (determining the reference sound level independently for each frequency on the basis of the calibration coefficient at this frequency) depends on the distribution of the bilateral hearing threshold in the population and the measurement error of the calibration coefficient. Measurement error of the calibration coefficient can be easily calculated from the standard deviation of the difference in the test-retest examination by dividing its value by the square root of 2. Therefore, the mean error of the independent coefficients method across all frequencies may be expressed in the following equation:

(independent coefficients SD)^2=(real bilateral threshold in the range 125 Hz-8 kHz SD)^2+(calibration method test-retest difference SD)^2/2 (12)

where bilateral threshold in the range 125 Hz-8 kHz SD is the standard deviation of the bilateral hearing threshold calculated jointly for all values reduced by the mean at relevant frequencies (Table 3).

The modeled coefficients method consists in estimation of the reference sound level on the basis of the model fitted to the mean value of 8 calibration coefficients determined at various frequencies. Therefore, its error is connected with distribution of the mean bilateral threshold, the error of determining the mean of 8 calibration coefficients, and the standard error of the model. Similarly, as for the independent coefficients method, the error of mean of 8 coefficients can be calculated from the standard deviation of the difference in test-retest examination by dividing its value by the square root of 2, to obtain the error for single coefficient, and by the square root of 8, to obtain the error for the mean. Thus:

(modeled coefficients SD)^2=(real mean bilateral threshold SD)^2+(calibration method test-retest difference SD)^2/16+(model SD)^2 (13)

Finally, the error of single frequency method consisting in estimating the reference sound level determined on the basis of the model fitted to 1 calibration coefficient at the frequency with the lowest standard deviation will be:

(single frequency SD)^2=(real bilateral threshold at 1 kHz SD)^2+(calibration method test-retest difference SD)^2/2+(model SD)^2 (14)

The standard errors of each method are presented in Table 4. For practical reasons, the differences in the hearing threshold between measurements on clinical audiometer and biologically calibrated PC were estimated (Table 5). These hearing thresholds were assumed to be obtained by means of ascending methods; therefore, the variances of the calibration methods were increased by the variance of test-retest examination for the ascending method. The variance calculated jointly for both ears was used (Table 1).

Table 4. The standard error of biological calibration with corresponding confidence intervals at *P*=.05 estimated on the basis of measurements carried out by 25 participants.

Method	Reference sound level (dB), SE (95% CI)		
	Independent coefficients	Modeled coefficients	Single coefficient
Modulated signal	5.38 (4.89, 5.98)	7.07 (5.95, 8.38)	7.61 (6.35, 9.24)
Dual tone (5 dB)	6.71 (6.10, 7.45)	7.21 (6.09, 8.53)	8.60 (7.32, 10.26)
Dual tone (2 dB)	7.07 (6.43, 7.85)	7.26 (6.13, 8.57)	8.89 (7.60, 10.55)
Ascending (5-dB step)	5.91 (5.37, 6.56)	7.12 (6.00, 8.44)	7.99 (6.73, 9.63)
Ascending (2-dB step)	5.39 (4.90, 5.99)	7.07 (5.95, 8.38)	7.62 (6.36, 9.25)
Bekesy (continual)	5.35 (4.87, 5.95)	7.07 (5.95, 8.38)	7.59 (6.33, 9.23)
Bekesy (modulated)	4.90 (4.46, 5.45)	7.03 (5.91, 8.34)	7.28 (6.02, 8.91)

Table 5. The standard deviation of the difference in the hearing threshold determined by means of the ascending method between measurements on clinical audiometer and the biologically calibrated personal computer, together with corresponding confidence intervals at *P*=.05 estimated on the basis of measurements carried out by 25 participants.

Method	Hearing threshold difference (dB), SD (95% CI)		
	Independent coefficients	Modeled coefficients	Single coefficient
Modulated signal	7.60 (7.01, 8.31)	8.88 (7.79, 10.17)	9.31 (8.09, 10.89)
Dual tone (5 dB)	8.59 (7.90, 9.42)	8.99 (7.89, 10.29)	10.14 (8.88, 11.77)
Dual tone (2 dB)	8.88 (8.16, 9.74)	9.03 (7.93, 10.33)	10.38 (9.11, 12.03)
Ascending (5-dB step)	7.98 (7.35, 8.74)	8.92 (7.83, 10.22)	9.63 (8.39, 11.23)
Ascending (2-dB step)	7.61 (7.01, 8.31)	8.88 (7.79, 10.17)	9.32 (8.10, 10.90)
Bekesy (continual)	7.58 (6.99, 8.29)	8.88 (7.78, 10.17)	9.30 (8.08, 10.88)
Bekesy (modulated)	7.27 (6.71, 7.93)	8.84 (7.75, 10.14)	9.05 (7.84, 10.61)

Discussion

Principal Findings

This paper presents methods of biological calibration of a PC for hearing examination by determining the reference sound level on the basis of the hearing threshold of the reference person. Seven methods of measuring calibration coefficients and 3 methods of determining reference sound level on the basis of these coefficients were proposed and analyzed. On the basis of 3 series of measurements conducted by 25 participants, the difference between classical pure-tone audiometry and audiometry based on biological calibration was estimated. The smallest standard deviation of the difference was obtained for the Bekesy (modulated) method with the independent coefficients method at the level of 7.27 dB (95% CI 6.71-7.93).

Comparison of Measurement Methods

The lowest standard deviation in test-retest examination at the level of 3.87 dB (95% CI 3.52-4.29) was obtained using the Bekesy (modulated) method, which entails assessment of the hearing threshold by means of the amplitude-modulated sound according to the rules of Bekesy's audiometry. This value is in-line with the standard deviation of the test-retest examination of Bekesy's audiometry [18]. The Bekesy (modulated) method is of moderate duration and is relatively easy to conduct. The modulated signal method, which consists in self-adjusting the volume of the amplitude-modulated sound to the barely audible level, turned out to be the easiest and the quickest method. In the test-retest examination, the standard deviation of this method was 4.97 dB (95% CI 4.53-5.51). The greatest error was found in the dual tone methods consisting in self-adjusting the volume of 2 generated sound signals differing slightly in intensity by a constant value in such a way that only the louder of the 2 sounds was audible.

Comparison of Sound Reference Level Determination Methods

The estimated error of determining reference sound level turned out to be the lowest for the independent coefficients method and higher for the modeled coefficients method (Table 4). This relation was statistically significant for the modulated signal, ascending (5-dB step), ascending (2-dB step), Bekesy (continual), and Bekesy (modulated) methods (*P*=.05). The highest error occurred for the single frequency method. However, when compared with the modeled coefficients method, statistical significance was achieved only for the dual tone (5 dB) method (*P*=.05).

The standard error of the modeled coefficients method was estimated for the model determined in the frequency range 125 Hz-8 kHz. When the range is limited to 250 Hz-8 kHz, the standard error of the model decreases from 6.57 dB (95% CI 5.59-7.54) to 5.98 dB (95% CI 4.45-7.50). This improves the modeled coefficients method, but the independent coefficients method is still more accurate. However, in this case the relation remained statistically significant only for the Bekesy (modulated) method (*P*=.05).

In the single frequency method, only 1 coefficient is needed to fit the model, which indicates calibration time is 8 times shorter at the cost of higher calibration error.

Comparison With Previous Work

Some of the presented calibration methods have been used in other studies. In Masalski and Kręcicki [4], calibration was carried out using the dual tone (5 dB) method with single frequency method. The standard deviation of the difference in the hearing threshold between PC-based test and pure-tone audiometry was 10.66 dB, which is in-line with the present study (SD 10.14 dB, 95% CI 8.88-11.77). In Bexelius et al [3], the pure-tone audiometry was compared with the test carried out on a PC calibrated by means of the modulated signal method with independent coefficients method. In all, 89% of the measurements were performed on the same PC and using the same reference person. The standard deviation was obtained at the level of 8.29 dB and 6.9 dB at frequencies 2 kHz and 4 kHz, respectively. These results are also consistent with the present study. The standard deviation for these modulated signal and independent calibration coefficient methods was estimated at the level of 7.60 dB (95% CI 7.01-8.31) (Table 5), whereas if we assume that the reference person is the same by setting the standard deviation of real bilateral threshold in the range 125 Hz-8 kHz to 0 in equation 12, we get 6.42 dB (95% CI 6.13-6.75).

Other Factors Affecting Accuracy

Calibration error strongly depends on the hearing threshold of the reference person. This applies especially to the independent coefficients and single frequency methods, in which the sound reference level at a single frequency is determined on the basis of a single measurement, contrary to modeled coefficients method, which uses mean hearing threshold. To verify the obtained results, the distribution of the hearing threshold of the participants was compared with literature data (Table 6). The standard deviation of the hearing threshold in this study is significantly smaller (*P*=.01) than the results presented in some studies [19-22], is in-line with one study [23], and is larger than other studies [24-27].

Table 6. Summary of standard deviations of the hearing threshold in decibels for participants with normal hearing in the literature.

Study	N	Hearing threshold (dB), SD								Mean SD
		125 Hz	250 Hz	500 Hz	1k Hz	2k Hz	4k Hz	6k Hz	8k Hz	
Taylor et al, 1967 [24]										
18-24 years	46	4.6	4.1	3.9	3.5	3.3	4.1	4.7	4.6	4.1
25-34 years	33	4.8	4.3	3.9	3.8	4.5	4.3	4.8	5.1	4.4
Robinson, Sutton, 1979 [19] [a]										
Men	1636	6.6	6.1	5.6	5.6	6.6	7.8	8.9	9.9	7.1
Women	1578	6.1	5.6	5.6	5.6	6.1	7.2	8.2	9.8	6.8
Arlinger, 1982 [21]	10			7.8	5.2	7.4	6.5	7.4	7.6	7.0
Arlinger, 1991 [25]	30	5.7	5.2	5.0	4.9	4.6	5.1	6.7	6.1	5.4
Lutman, Davis, 1994 [26]	241		4.2	4.4	4.2	4.6	6.9	7.8	7.9	5.7
Han, Poulsen, 1998 [27]	31		4.7	4.4	3.5	4.9	6.6	6.3	7.1	5.4
Johansson, 2002 [23] [b]										
Men	266	5.7	5.5	5.5	5.4	5.6	6.7	7.0	7.5	6.1
Women	337	5.5	5.2	4.8	4.5	5.2	7.9	7.4	6.1	5.8
Engdahl et al, 2005 [22] [c]										
Men	3587		7.6	5.7	5.3	7.0	8.9	9.5	8.0	7.4
Women	1840		6.3	5.8	5.3	5.7	7.0	9.1	7.6	6.7
Current study	25	5.9	6.1	5.0	4.2	4.9	5.6	8.0	9.1	6.1

[a] Estimated on the basis of the model for the age of 25 years, consistent with ISO 7029, 2000 [20].

[b] Estimated on the basis of centiles calculated from the model for the age of 25 years.

[c] Estimated on the basis of centiles in the age range 20-29 years.

The examinations in this study were conducted on young employees and interns of the Otolaryngology Clinic (ie, persons familiar with the subject of hearing examinations). This may have led to better calibration results and shorter duration of the examination than in a population of young people with good hearing without experience with hearing examinations.

In the calculations, it was assumed that the examinations conducted on home computers are not burdened with an error resulting from the presence of background noises other than the fan noise. This assumption was made because during home examinations and in those conducted in the sound booth the fan noise was the loudest and the most disturbing sound. Thus, the estimated calibration error takes into account the fan noise. However, in the case of other background noises, the error may turn out to be bigger.

Calibration methods presented in the paper were implemented as Java applets embedded in browsers. However, their application is not limited only to Web-based tests, but may also be used for offline determination of the reference sound level or on mobile devices. Moreover, in the case of tablets or smartphones, calibration error may turn out to be smaller because of the lack of fan noises.

When conducting examinations on a PC with the use of headphones with very high sensitivity instead of regular ones, interferences of the sound card or other electronic systems may affect the stimulus. During examination at home, such incidents occurred in 2 of 25 cases. As a result, it was impossible to perform the examination. After changing headphones from professional to regular ones, the examination was completed without any problems.

Calibration accuracy may be improved if it is conducted by 2 or more reference persons [3]. The greatest improvement may be expected in the case of the independent coefficients method, whose standard deviation should reduce proportionally to the square root of the number of persons conducting calibration. In the case of the modeled coefficients and single frequency methods, the improvement will be less visible because increasing the number of reference persons does not affect the model's error.

Another method of improving the accuracy is to introduce additional conditions to reject inaccurate calibrations. For example, calibration using the independent coefficients method may be rejected as the difference between coefficients exceeds the predetermined threshold [3]. In the case of the modeled coefficients method, the condition may be imposed on the difference between the value of the calibration coefficient and the model. In the single frequency method, it is possible to do an additional measurement and verify its value with the model. Moreover, for all methods based on Bekesy's audiometry, verification can be based on the difference between the intensities at which the sound starts to be audible and the intensities at which the sound ceases to be audible.

Recommendations

The final choice of the calibration method will depend on the desired accuracy of calibration and the time for its performance. If considerable accuracy is required, it is advisable to use the independent coefficients method, whereas when quick calibration is the priority, the single frequency method is preferable. The application of the modeled coefficients method is not justified because of higher calibration error than is in the independent coefficients method at the same duration.

Two of the 7 methods of measuring calibration coefficients seem worth noting: the modulated signal and Bekesy (modulated) methods. The choice of the better of the 2 is not obvious. The Bekesy (modulated) method is the most accurate at moderate duration, whereas the modulated signal method is the fastest at moderate accuracy. Additionally, the modulated signal method is the easiest, and the Bekesy (modulated) method is the second easiest. However, the methods differ significantly in the complexity of implementation with the Bekesy (modulated) method being more complex. On the other hand, in the case of Bekesy (modulated) method, the measurement can be easily verified on the basis of the differences between the intensities at which the stimulus starts or stops being audible.

Therefore, if there are no substantial time limitations, it is advisable to use Bekesy (modulated) method with independent coefficients method, which have the lowest error. When a simple and quick calibration is required, modulated signal method with single frequency method should be chosen.

Acknowledgments

The authors of this paper would like to thank the workers and interns of the Otolaryngology Clinic who agreed to take part in the examinations. The research described in this paper was carried out as part of a project (Kluczowy Stażysta no KSW/13/I/2011) cofinanced by the European Social Fund.

Conflicts of Interest

The first author (MM) is the owner of an Internet portal (e-audiologia.pl) that offers online hearing tests.

Multimedia Appendix 1

Translation of the questionnaire on difficulty of the calibration.

[PDF File (Adobe PDF File), 38KB - jmir_v16i1e11_app1.pdf]

Multimedia Appendix 2

All equations used for this paper: (1) the model, (2) the difference between the model and the real calibration coefficient, (3) the determination error of the calibration coefficient, (4) the difference between the model and the determined coefficient, (5) the variance of random variable X, (6) the model estimated on the basis of determined coefficients, (7) the variance of random variable

Z, (8) the value of determined coefficient, (9) the variance of random variable Y, (10) the frequency response of the model, (11) the standard deviation of the participants' bilateral hearing threshold measured with ascending method, (12) the standard error of the independent coefficients method, (13) the standard error of the modeled coefficients method, (14) the standard error of the single frequency method.

[PNG File, 693KB - jmir_v16i1e11_app2.png]

Multimedia Appendix 3

CONSORT-EHEALTH checklist V1.6.2 [28].

[PDF File (Adobe PDF File), 996KB - jmir_v16i1e11_app3.pdf]

References

1. Smits C, Merkus P, Houtgast T. How we do it: The Dutch functional hearing-screening tests by telephone and Internet. Clin Otolaryngol 2006 Oct;31(5):436-440. [doi: 10.1111/j.1749-4486.2006.01195.x] [Medline: 17014457]
2. Bexelius C, Honeth L, Ekman A, Eriksson M, Sandin S, Bagger-Sjöbäck D, et al. Evaluation of an Internet-based hearing test--comparison with established methods for detection of hearing loss. J Med Internet Res 2008;10(4):e32 [FREE Full text] [doi: 10.2196/jmir.1065] [Medline: 18940783]
3. Honeth L, Bexelius C, Eriksson M, Sandin S, Litton JE, Rosenhall U, et al. An Internet-based hearing test for simple audiometry in nonclinical settings: preliminary validation and proof of principle. Otol Neurotol 2010 Jul;31(5):708-714. [doi: 10.1097/MAO.0b013e3181de467a] [Medline: 20458255]
4. Masalski M, Kręcicki T. Self-test Web-based pure-tone audiometry: validity evaluation and measurement error analysis. J Med Internet Res 2013;15(4):e71 [FREE Full text] [doi: 10.2196/jmir.2222] [Medline: 23583917]
5. Thorén ES, Oberg M, Wänström G, Andersson G, Lunner T. Internet access and use in adults with hearing loss. J Med Internet Res 2013;15(5):e91 [FREE Full text] [doi: 10.2196/jmir.2221] [Medline: 23659867]
6. Henshaw H, Clark DP, Kang S, Ferguson MA. Computer skills and Internet use in adults aged 50-74 years: influence of hearing difficulties. J Med Internet Res 2012;14(4):e113 [FREE Full text] [doi: 10.2196/jmir.2036] [Medline: 22954484]
7. Smits C, Kramer SE, Houtgast T. Speech reception thresholds in noise and self-reported hearing disability in a general adult population. Ear Hear 2006 Oct;27(5):538-549. [doi: 10.1097/01.aud.0000233917.72551.cf] [Medline: 16957503]
8. Leensen MC, de Laat JA, Dreschler WA. Speech-in-noise screening tests by Internet, part 1: test evaluation for noise-induced hearing loss identification. Int J Audiol 2011 Nov;50(11):823-834. [doi: 10.3109/14992027.2011.595016] [Medline: 21988504]
9. Leensen MC, de Laat JA, Snik AF, Dreschler WA. Speech-in-noise screening tests by Internet, part 2: improving test sensitivity for noise-induced hearing loss. Int J Audiol 2011 Nov;50(11):835-848. [doi: 10.3109/14992027.2011.595017] [Medline: 21970351]
10. Kimball SH. Inquiry into online hearing test raises doubts about its validity. The Hearing Journal 2008;61(3):38-46. [doi: 10.1097/01.HJ.0000314717.42321.4a]
11. Choi JM, Lee HB, Park CS, Oh SH, Park KS. PC-based tele-audiometry. Telemed J E Health 2007 Oct;13(5):501-508. [doi: 10.1089/tmj.2007.0085] [Medline: 17999612]
12. Platforma Badań Zmysłów. Senses examination platform URL: http://platformabadanzmyslow.pl/en/about.html [accessed 2012-09-24] [WebCite Cache ID 6AvYBYhVy]
13. Jiang T. Important revisions of ANSI S3.6-1989:ANSI S3.6-1996 American National Standard Specification for Audiometers. Canadian Journal of Speech-Language Pathology and Audiology 1998;22(1):5-9 [FREE Full text]
14. British Society of Audiology. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. Berkshire, UK: British Society of Audiology; 2011 Sep 24. URL: http://www.thebsa.org.uk/docs/Guidelines/BSA_RP_PTA_FINAL_24Sept11.pdf [accessed 2012-05-27] [WebCite Cache ID 67yJElAP1]
15. Barnett V, Lewis T. Outliers in Statistical Data. Chichester, UK: Wiley; 1994.
16. Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. Cambridge, UK: Cambridge University Press; 2002.
17. International Electrotechnical Commission. International Standard IEC 61672-2:2003(E). Electroacoustics - Sound level meters - Part 2: Pattern evaluation tests. Geneva: IEC; 2003:1-79.
18. Erlandsson B, Håkanson H, Ivarsson A, Nilsson P. Comparison of the hearing threshold measured by manual pure-tone and by self-recording (Békésy) audiometry. Audiology 1979;18(5):414-429. [Medline: 496724]
19. Robinson DW, Sutton GJ. Age effect in hearing - a comparative analysis of published threshold data. Audiology 1979;18(4):320-334. [Medline: 475664]
20. International Organization for Standardization. International Standard ISO 7029:2000(E), Acoustics - Statistical distribution of hearing thresholds as a function of age. Geneva: ISO; 2000:1-9.
21. Arlinger S. Normal threshold of hearing at preferred frequencies. Scand Audiol 1982;11(4):285-286. [Medline: 7163775]

22. Engdahl B, Tambs K, Borchgrevink HM, Hoffman HJ. Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag Hearing Loss Study. Int J Audiol 2005 Apr;44(4):213-230. [Medline: 16011050]
23. Johansson MS, Arlinger SD. Hearing threshold levels for an otologically unscreened, non-occupationally noise-exposed population in Sweden. Int J Audiol 2002 Apr;41(3):180-194. [Medline: 12033636]
24. Taylor W, Pearson J, Mair A. Hearing thresholds of a non-noise-exposed population in Dundee. Br J Ind Med 1967 Apr;24(2):114-122 [FREE Full text] [Medline: 6023076]
25. Arlinger SD. Normal hearing threshold levels in the low-frequency range determined by an insert earphone. J Acoust Soc Am 1991 Nov;90(5):2411-2414. [Medline: 1774411]
26. Lutman ME, Davis AC. The distribution of hearing threshold levels in the general population aged 18-30 years. Audiology 1994 Dec;33(6):327-350. [Medline: 7741666]
27. Han LA, Poulsen T. Equivalent threshold sound pressure levels for Sennheiser HDA 200 earphone and Etymotic Research ER-2 insert earphone in the frequency range 125 Hz to 16 kHz. Scand Audiol 1998;27(2):105-112. [Medline: 9638829]
28. 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

dB: decibel
dB HL: decibel hearing level
PC: personal computer

Edited by G Eysenbach; submitted 27.06.13; peer-reviewed by DW Swanepoel, I Brooks; comments to author 04.08.13; revised version received 09.10.13; accepted 30.10.13; published 15.01.14.

Please cite as:
Masalski M, Grysiński T, Kręcicki T
Biological Calibration for Web-Based Hearing Tests: Evaluation of the Methods
J Med Internet Res 2014;16(1):e11
URL: http://www.jmir.org/2014/1/e11/
doi:10.2196/jmir.2798
PMID:24429353

Original Paper

The Psychometric Properties of CollaboRATE: A Fast and Frugal Patient-Reported Measure of the Shared Decision-Making Process

Paul James Barr[1], BSc, MSc Public Health, PhD; Rachel Thompson[1], BPsySC(Hons), PhD; Thom Walsh[1], MS, MSPT, PhD; Stuart W Grande[1], PhD; Elissa M Ozanne[2,3], PhD; Glyn Elwyn[1,3], BA, MB, BCh, MSc, PhD

[1]The Dartmouth Center for Health Care Delivery Science, Dartmouth College, Hanover, NH, United States
[2]Institute for Health Policy Studies, Department of Surgery, University of California, San Francisco,, San Francisco, CA, United States
[3]The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Glyn Elwyn, BA, MB, BCh, MSc, PhD
The Dartmouth Center for Health Care Delivery Science
Dartmouth College
37 Dewey Field Road
Hanover, NH, 03755
United States
Phone: 1 1 603 646 2295
Fax: 1 1 603 646 1269
Email: glynelwyn@gmail.com

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2015/2/e32/

Abstract

Background: Patient-centered health care is a central component of current health policy agendas. Shared decision making (SDM) is considered to be the pinnacle of patient engagement and methods to promote this are becoming commonplace. However, the measurement of SDM continues to prove challenging. Reviews have highlighted the need for a patient-reported measure of SDM that is practical, valid, and reliable to assist implementation efforts. In consultation with patients, we developed CollaboRATE, a 3-item measure of the SDM process.

Objective: There is a need for scalable patient-reported measure of the SDM process. In the current project, we assessed the psychometric properties of CollaboRATE.

Methods: A representative sample of the US population were recruited online and were randomly allocated to view 1 of 6 simulated doctor-patient encounters in January 2013. Three dimensions of SDM were manipulated in the encounters: (1) explanation of the health issue, (2) elicitation of patient preferences, and (3) integration of patient preferences. Participants then completed CollaboRATE (possible scores 0-100) in addition to 2 other patient-reported measures of SDM: the 9-item Shared Decision Decision Making Questionnaire (SDM-Q-9) and the Doctor Facilitation subscale of the Patient’s Perceived Involvement in Care Scale (PICS). A subsample of participants was resurveyed between 7 and 14 days after the initial survey. We assessed CollaboRATE’s discriminative, concurrent, and divergent validity, intrarater reliability, and sensitivity to change.

Results: The final sample consisted of 1341 participants. CollaboRATE demonstrated discriminative validity, with a significant increase in CollaboRATE score as the number of core dimensions of SDM increased from zero (mean score: 46.0, 95% CI 42.4-49.6) to 3 (mean score 85.8, 95% CI 83.2-88.4). CollaboRATE also demonstrated concurrent validity with other measures of SDM, excellent intrarater reliability, and sensitivity to change; however, divergent validity was not demonstrated.

Conclusions: The fast and frugal nature of CollaboRATE lends itself to routine clinical use. Further assessment of CollaboRATE in real-world settings is required.

(J Med Internet Res 2014;16(1):e2) doi:10.2196/jmir.3085

KEYWORDS

decision making; physician-patient relations; psychometrics/Instrumentation; patient participation; questionnaires; Internet

Introduction

Health care that is patient-centered and supports patient engagement has become an integral aspect of health policy [1-3]. Shared decision making (SDM) has been described as the pinnacle of patient-centered care [4], relevant to managing long-term conditions and situations where multiple treatment options exist. However, to date, implementation has been limited [2,5]. To encourage adoption, SDM has been included in the Patient Protection and Affordable Care Act, as a quality metric for new health care payment and service delivery models [2]. The challenge of developing a measure of the SDM process that is psychometrically sound and suitable for use in routine care forms a barrier to the realization of this plan [6,7] and impedes SDM implementation [8,9].

Measuring the SDM process using observational instruments is laborious, costly, and not conducive to rapid data feedback. Patient-reported measurement of the SDM process may be implemented more successfully. We found 5 such measures: the dyadic OPTION scale [10], the Facilitation of Patient Involvement in Care Scale [11], the Perceived Involvement in Care Scale (PICS) [12], the 9-item Shared Decision Making Questionnaire (SDM-Q-9) [13], and the modified Control Preferences Scale [14]. Four of these measures [10-13] contain 5 or more items, which introduces a patient burden that complicates their integration into usual care. All 5 measures also refer explicitly to a “decision” despite recognition that patients may not always realize that a decision has been made [15,16]. Three of the measures [10,12,13] refer to a single decision, limiting their applicability for health care encounters in which several decisions are made [17,18] and although the psychometric properties of some measures are promising [7], important qualities, such as discriminative validity and intrarater reliability, are often unreported.

A fast and frugal, valid and reliable, patient-reported measure of the SDM process that is applicable to a wide range of clinical settings—especially the primary care setting where varied and often unanticipated decisions are made—is needed. Encouraged by the success of short health measures in other fields [19-23], we developed a 3-item measure of the SDM process, CollaboRATE, in partnership with patients [24,25]. CollaboRATE represents a formative measurement model, assessing the extent to which each of 3 core shared decision-making tasks (or dimensions) are present in a clinical encounter: (1) explanation of the health issue, (2) elicitation of patient preferences, and (3) integration of patient preferences [24]. To date, we have completed the first of 3 planned stages in the development of CollaboRATE: (1) item development with target users, (2) psychometric performance in simulated encounters, and (3) psychometric properties in real clinical populations. In the first stage, we conducted a series of cognitive interviews, where we have shown CollaboRATE to be fast to complete, easy to understand, and to consist of items that are interpreted in the way intended [24]. Our aim in this study, the second stage of CollaboRATE development, was to assess the psychometric properties of CollaboRATE using simulated clinical encounters.

Methods

Participants

Participants were adults, 18 years of age or older, residing in the United States, and proficient in English. CollaboRATE was designed to be used in any health care encounter. As such, the target population for CollaboRATE is any person visiting a health provider. Therefore, recruitment quotas, based on the 2010 US Census, were imposed to ensure the sample approximated the US population in terms of gender, age, and educational attainment. Participants were recruited via Survey Sampling International (Shelton, CT), an online survey sampling company with experience in sampling participants for health care research. Survey Sampling International provides small incentives for participation; all respondents were entered into a quarterly draw for US $12,500. The Internet is now a well-established and recognized mode of recruiting participants into research allowing investigators to include hard to reach populations, such as ethnic minorities, with the potential to reduce measurement error, missing data, and respondent attrition. In the recent US Census (2011), 71.7% of Americans reported having access to the Internet at home [26]. The representativeness of data gathered from Internet panels has been shown to be comparable to that from probability-based general population samples [27].

Simulated Encounters

We created a series of simulated encounters using avatars with audio overlay, where a female patient consulted a male clinician about a prolapsed lumbar disk. Each encounter included zero, 1, 2, or 3 dimensions of SDM (Table 1). In total, 6 encounters were created. No encounters were created that included preference integration in the absence of preference elicitation because this was considered implausible. Encounters were scripted to represent realistic encounters, were spoken by volunteers with American accents, and overlaid on computer animations (Multimedia Appendices 1-6). Seven trained independent raters assessed the level of SDM in each encounter using 2 validated observational measures: the Observer OPTION measure [28] and the Rochester Participatory Decision-Making Scale (RPAD) [29]. As expected, observer ratings demonstrated a linear increase in the mean level of SDM as the number of dimensions increased (Figure 1).

Table 1. Number of core dimensions of shared decision making (SDM) included in each simulated encounter.

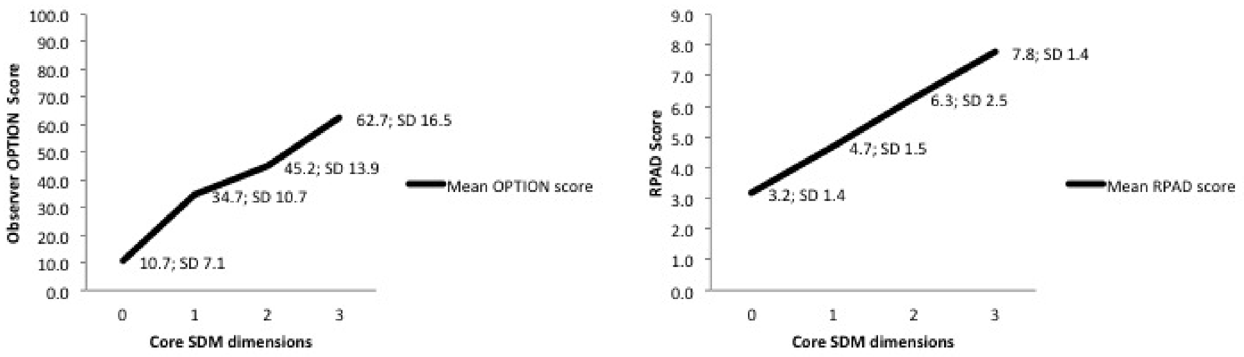
Encounter	Level of SDM	Dimensions of SDM	Explanation[a]	Preference elicitation[b]	Preference integration[c]	Length (min:s)
1	None	0	No	No	n/a	2:10
2	Low	1	Yes	No	n/a	6:05
3	Low	1	No	Yes	No	3:55
4	Medium	2	Yes	Yes	No	7:52
5	Medium	2	No	Yes	Yes	4:49
6	High	3	Yes	Yes	Yes	8:45

[a] Thorough explanation of health-related information to patient.

[b] Patients' health-related preferences, views, or opinions elicited.

[c] Patients' preferences integrated in decision making.

Figure 1. Independent observer ratings (n=7) of the simulated clinical encounters using the observer OPTION Scale and Rochester Participatory Decision-Making Scale.



Measures

CollaboRATE

CollaboRATE is a 3-item measure of the SDM process. Items included are:

1. How much effort was made to help you understand your health issues?
2. How much effort was made to listen to the things that matter most to you about your health issues?
3. How much effort was made to include what matters most to you in choosing what to do next?

Participants are instructed to reflect on a health care encounter and then asked to complete the CollaboRATE survey. We administered 2 different response scales to examine their psychometric properties separately. CollaboRATE-10 was a 10-point anchored scale, ranging from 0 (no effort was made) to 9 (every effort was made). CollaboRATE-5 was a 5-point Likert scale, with responses of 0 (no effort was made), 1 (a little effort was made), 2 (some effort was made), 3 (a lot of effort was made), and 4 (every effort was made). We also used 2 scoring methods to enable us to examine their psychometric properties separately. For CollaboRATE mean, we summed participant's scores on the 3 items and multiplied by 3.704, transforming to a scale from 0 to 100 (for CollaboRATE-10) and the sum of participant's scores on the 3 items on the original scale from 0 to 12 (for CollaboRATE-5). For CollaboRATE top score, we coded participants as 1 (yes) when they recorded the highest response on the scale for all 3 items and as 0 (no) in all other situations.

Other Patient-Reported Measures of the Shared Decision-Making Process

We administered the 9-item SDM-Q-9 [13]. Responses were on a 6-point Likert scale ranging from completely disagree to completely agree with total scores on the survey ranging from 0 to 100. We also administered the 5-item Doctor Facilitation subscale of the PICS (PICS-DFS) [12]. Responses were on binary scale (yes or no) and total scores on the survey ranged from 0 to 5.

Clinician Technical Skills

We asked, "How would you rate the technical skills (thoroughness, carefulness, competence) of the provider in the video?" [30] to measure clinician technical skills. Responses were coded as 1 (excellent) or 0 (very good, good, fair, or poor).

Participant Characteristics

We assessed participants' gender, age, educational attainment [31], ethnicity and race [32], and language(s) spoken at home [33] using standard measures. We assessed health care utilization, measured using a single item, "In the last 12 months, did you make any appointment to see a specialist" with a yes/no response option [34]. Self-reported health status was measured using 2 questions: "Do you have any long-standing illness or disability?” (yes/no response option) and, if yes, “Does this illness or disability limit your activities in any way?” (yes/no response option) [35]. Decision-making role preferences were measured using the Control Preferences Scale [36].

Procedure

Prospective participants were provided with a link to an online information sheet. On the information sheet, participants were informed of the purpose of the survey, the time needed to complete the survey (approximately 15 minutes), and ensured that all data would be stored securely, confidentially, and used only for the purpose of the research project. They were given the number and email of a member of the study team (PJB) to contact if they had any questions. Those who consented were able to enter the online survey system. The survey was created by the research team, piloted with academics and members of the public (n=10) to refine wording, and hosted in Qualtrics, a company specializing in online survey design and data capture. Participants completed items assessing their characteristics and then were randomly allocated by a survey software algorithm to view one of the simulated encounters. Participants were restricted to viewing the simulated encounter once. Potential participants were eligible for inclusion only if they viewed the whole encounter, took the minimum amount of time required to complete the survey questions (3 minutes for the initial survey and 45 seconds for the resurvey), and completed the survey within 1 hour of commencement. Participants were asked to imagine themselves in the position of the patient and then to assess the encounter by completing CollaboRATE, SDM-Q-9, and PICS-DFS. CollaboRATE was administered using both response scales; the presentation of response scales was counterbalanced to attenuate possible order effects. Participants were prevented from making multiple survey entries. This was achieved by preventing respondents with the same Internet Protocol (IP) address from taking the survey again.

A random subsample of participants from each of the encounters was resurveyed between 1 and 2 weeks after initial survey completion [37]. Some of these participants were shown the same encounter a second time, whereas others were shown a different encounter. Participants were asked to complete CollaboRATE using both sets of response scales, which were again counterbalanced. The Committee for the Protection of Human Subjects at the Dartmouth College Institutional Review Board (IRB) approved the study (CPHS #23687).

Statistical Analysis

The statistical analyses conducted to assess the psychometric properties of CollaboRATE are provided in Table 2. We conducted all analyses using both response scales of CollaboRATE (CollaboRATE-10 and CollaboRATE-5), and both scoring methods (CollaboRATE mean and CollaboRATE top score). As CollaboRATE represents a formative model of SDM, we did not assess internal consistency. Similarly, we did not assess floor or ceiling effects as because the artificial manipulation makes such an assessment invalid. Analyses were conducted using Stata 12 (StataCorp LP, College Station, TX, USA).

Table 2. Statistical analyses conducted to assess psychometric properties of CollaboRATE.

Psychometric property	Definition	Assessment[a]	CollaboRATE analyses	
			Mean	Top score
Discriminative validity	Ability of the measure to yield low scores when the construct under measurement is absent, and higher scores as the presence of the construct increases [37]	Between-dimension comparisons of CollaboRATE scores	ANOVA, planned comparisons (between-groups *t* test or Welch test)	Chi-square test
Concurrent validity	Presence of correlation between measures that claim to measure the same construct [37]	Relationship between CollaboRATE and the 2 other measures of SDM (SDM-Q-9 and PICS-DFS)	Pearson product moment correlation (*r*) [38]	Point-biserial correlation (rpb) [39]
Divergent validity	Absence of correlation between measures that claim to measure different constructs [37]	Relationship between CollaboRATE and the clinician technical skills question	Pearson product moment correlation (*r*)	Point-biserial correlation (rpb)
Intrarater reliability	Consistency of ratings of the same encounter, across 2 time points by the same rater [37]	Comparison of CollaboRATE scores on initial survey and resurvey for participants exposed to the same encounter	Intraclass correlation coefficients (ICC 2,2; 2-way mixed effects model of absolute agreement)	Cohen’s kappa coefficient [40]
Sensitivity to change	Ability of the measure to detect change in the specified construct, regardless of whether it is deemed meaningful to the decision maker [37,41]	Comparison of CollaboRATE scores on initial survey and resurvey for participants exposed to the “opposite” encounter on resurvey (e.g., low SDM on initial survey, high SDM on resurvey)	Paired *t* test	McNemar’s test

[a] SDM: shared decision making; SDM-Q-9: 9-item Shared Decision Making Questionnaire; PICS-DFS: 5-item Doctor Facilitation subscale of the Perceived Involvement in Care Scale.

Sample Size Calculation

To detect an estimated 15% difference in the proportion of participants with a top score on CollaboRATE between the encounter with 3 dimensions present (estimated 75% top score) and an encounter with 2 dimensions present (estimated 60% top score), with 90% power, 216 participants per encounter were required. We planned to resurvey 30 participants initially exposed to each of the encounters who would be exposed to the same encounter again (providing 95% power to detect a minimum intraclass correlation coefficient, ICC, of 0.65) and a further 30 participants initially exposed to the 2 extreme encounters (zero dimensions and 3 dimensions) who would be exposed to the opposite encounter.

Results

Participant Flow

A total of 2026 participants completed the initial survey. Before analysis, 685 (33.8%) were excluded for taking less than the minimum reasonable time to complete the survey, resulting in a total of 1341 eligible participants included. A total of 388 participants were approached for resurvey. Prior to analysis, 137 (35.3%) were excluded for taking less than the minimum reasonable time to complete the survey, resulting in a total of 251 eligible participants included in the resurvey.

Participant Characteristics

Characteristics of the participants were similar to that of the US population. Participants' characteristics across the dimensions were comparable (Table 3) although there were statistically significant differences in race (*P*=.04). 6.94% (93/1341) of participants did not report age or gender, but no differences in CollaboRATE scores were found between those that did and did not report age (*P*=.45) or gender (*P*=.76). The acceptability of CollaboRATE items was demonstrated by less than 1% (8/1341) of participants missing any of the items.

Discriminative Validity

The discriminative validity of CollaboRATE was demonstrated with significant increases in scores as progressively more of the dimensions were included in the encounters (Table 4). For all analyses, a significant overall association between CollaboRATE and number of dimensions was found (data available on request).

This was true for both response scales and both scoring methods. The discriminative validity of the 2 other measures of SDM was also demonstrated. No significant differences were observed in CollaboRATE between the 2 encounters that included 1 dimension of SDM, nor between the 2 encounters that included 2 dimensions of SDM, on either response scales or scoring method (analysis available upon request). A further description of CollaboRATE scores per item is presented in Table 5.

Table 3. Participant characteristics by group.

Sociodemographic and health care characteristics	Number of dimensions[a]				Total (n=1341)	US population[b]
	0 (n=270)	1 (n=443)	2 (n=425)	3 (n=203)		
Gender, n (%)						
Female	128 (50.0)	218 (53)	223 (57.2)	104 (54.5)	673 (53.9)	50.8%
Male	128 (50.0)	193 (47.0)	167 (42.8)	87 (45.6)	575 (46.1)	49.2%
Age (years), n (%)						
18-44	85 (39.5)	172 (50.3)	165 (50.5)	77 (46.7)	499 (47.6)	48.1%
45-64	79 (36.7)	110 (32.2)	102 (31.2)	55 (30.3)	346 (33.0)	34.7%
65+	51 (23.7)	60 (17.5)	60 (18.4)	33 (20.0)	204 (19.5)	17.2%
Educational attainment						
High school graduate or less	108 (40.3)	179 (40.4)	169 (40.0)	84 (41.6)	540 (40.4)	42.7%
Some college, no degree	56 (20.9)	103 (23.3)	100 (23.6)	41 (20.3)	300 (22.5)	16.7%
Associate’s or bachelor’s degree	79 (29.5)	120 (27.1)	126 (29.8)	55 (27.2)	380 (28.4)	29.5%
Master’s, professional, or doctoral degree	25 (9.3)	41 (9.3)	28 (6.6)	22 (10.9)	116 (8.7)	11.1%
Ethnicity						
Hispanic or Latino	21 (8.1)	28 (6.4)	32 (7.7)	17 (8.6)	98 (7.5)	16.3%
Not Hispanic or Latino	238 (91.9)	409 (93.6)	384 (92.3)	181 (91.4)	1212 (92.5)	83.7%
White alone	205 (86.1)	315 (77.0)	314 (81.8)	148 (81.8)	982 (81.0)	63.7%
Race						
One race	260 (97)	423 (96.1)	413 (97.6)	198 (98.5)	1299 (97.5)	97.1%
White	229 (85.5)	341 (77.5)	342 (80.9)	161 (80.1)	1073 (80.6)	72.4%
Black or African American	18 (6.7)	45 (10.2)	39 (9.2)	27 (13.4)	129 (9.7)	12.6%
American Indian and Alaska Native	4 (1.5)	4 (0.9)	6 (1.4)	0	14 (1.1)	0.9%
Asian	4 (1.5)	21 (4.8)	15 (3.6)	4 (2.0)	44 (3.3)	4.8%
Native Hawaiian and Other Pacific Islander	0	4 (0.9)	0	0	4 (0.3)	0.2%
Some other race	5 (1.9)	13 (3.0)	11 (2.6)	6 (3.0)	35 (2.6)	6.2%
Two or more races	8 (3.0)	17 (3.9)	10 (2.4)	3 (1.5)	33 (2.5)	2.9%
Language spoken at home						
English only	243 (91.4)	386 (89.4)	370 (88.5)	87 (42.9)	1177 (89.9)	80.4%
Language other than English	23 (8.7)	46 (10.7)	48 (11.5)	116 (57.1)	133 (10.2)	19.6%
Health care experiences and preferences						
Long-standing illness or disability						
Yes, and limits activities	55 (20.6)	117 (26.6)	122 (28.8)	62 (30.8)	356 (26.7)	–
Yes, and does not limit activities	26 (9.7)	38 (8.6)	37 (8.7)	23 (11.4)	124 (9.3)	–
No	186 (68.7)	285 (64.8)	264 (62.4)	116 (57.7)	851 (63.9)	–
Specialist appointment in last 12 months						
Yes	120 (44.6)	228 (51.6)	223 (52.5)	104 (51.2)	661 (49.4)	–
No	149 (55.4)	214 (48.4)	202 (47.5)	99 (48.8)	678 (50.6)	–
Decision-making role preferences						
Patient alone	37 (13.8)	74 (16.7)	64 (15.1)	35 (17.2)	210 (15.7)	–
Patient with provider input	117 (43.5)	167 (37.7)	159 (37.4)	69 (34.0)	512 (38.2)	–
Shared	84 (31.2)	162 (36.6)	164 (38.6)	84 (41.4)	494 (36.9)	–

Sociodemographic and health care characteristics	Number of dimensions[a]				Total (n=1341)	US population[b]
	0 (n=270)	1 (n=443)	2 (n=425)	3 (n=203)		
Provider with patient input	17 (6.3)	22 (5.0)	17 (4.0)	11 (5.4)	67 (5.0)	–
Provider alone	14 (5.2)	18 (4.1)	21 (4.9)	4 (2.0)	57 (4.2)	–

[a] Frequencies may not sum to the total due to missing data.

[b] Gender and age data were taken from the 2010 Census [42], educational attainment data correspond to the population aged ≥25 years and were taken from the Current Population Survey 2012 Annual Social and Economic Supplement [31], ethnicity and race data were taken from 2010 Census [43], and language data were taken from the 2006-2008 American Community Survey [33].

Table 4. Discriminative validity of CollaboRATE, the 9-item Shared Decision Making Questionnaire (SDM-Q-9), and the 5-item Doctor Facilitation subscale of the Perceived Involvement in Care Scale (PICS-DFS).

Discriminative validity	Number of dimensions				Contrasts between dimensions[a]									Valid[b]
	0 (n=270)	1 (n=443)	2 (n=425)	3 (n=203)	0 vs 1			1 vs 2			2 vs 3			
					t (df)	χ^2_1	*P*	*t (df)*	χ^2_1	*P*	*t (df)*	χ^2_1	*P*	
CollaboRATE-10, mean (SD)	46.0 (29.9)	69.6 (26.2)	82.0 (21.6)	85.8 (19.1)	–10.68 (505.1)		<.001	–7.58 (844.1)		<.001	–2.25 (447.7)		.01	Yes
CollaboRATE-5, mean (SD)	5.2 (3.4)	7.8 (3.1)	9.4 (2.6)	10.0 (2.3)	–10.15 (525.2)		<.001	–8.37 (845.0)		<.001	–2.66 (435.3)		.008	Yes
CollaboRATE-10 top score, n (%)	13 (4.9)	79 (17.9)	131 (31.1)	81 (39.9)		24.9	<.001		20.5	<.001		4.7	.03	Yes
CollaboRATE-5 top score, n (%)	16 (6.0)	76 (17.2)	136 (32.2)	85 (42.3)		18.4	<.001		26.0	<.001		6.1	.01	Yes
SDM-Q-9, mean (SD)	37.1 (27.9)	63.2 (23.4)	75.1 (19.8)	82.0 (16.0)	–12.85 (490.6)		<.001	–8.09 (852.9)		<.001	–4.69 (484.6)		<.001	Yes
PICS-DFS, mean (SD)	1.60 (1.9)	3.2 (1.7)	3.9 (1.2)	4.3 (0.9)	–11.64 (510.8)		<.001	–7.07 (798.2)		<.001	–4.64 (496.4)		<.001	Yes

[a] Two-sample *t* test with unequal variances for contrasts of means.

[b] Yes=psychometric property found in this sample; no=psychometric property not found in this sample.

Table 5. CollaboRATE scores by item.

CollaboRATE items	Number of dimensions			
	0 (n=270)	1 (n=443)	2 (n=425)	3 (n=203)
CollaboRATE-10, mean (SD)				
Item 1 (information)	4.50 (2.76)	6.63 (2.36)	7.48 (1.94)	7.70 (1.86)
Item 2 (preference)	4.19 (2.83)	6.25 (2.49)	7.36 (2.07)	7.69 (1.75)
Item 3 (integration)	3.77 (2.96)	5.94 (2.74)	7.32 (2.16)	7.80 (1.73)
CollaboRATE-5, mean (SD)				
Item 1 (information)	1.98 (1.10)	2.74 (1.02)	3.18 (0.85)	3.34 (0.84)
Item 2 (preference)	1.73 (1.21)	2.58 (1.31)	3.14 (0.95)	3.27 (0.83)
Item 3 (integration)	1.51 (1.33)	2.49 (1.25)	3.13 (1.02)	3.39 (0.82)
CollaboRATE-10 top score, n (%)				
Item 1 (information)	22 (8.2)	117 (26.5)	175 (41.4)	97 (47.8)
Item 2 (preference)	20 (7.5)	103 (23.4)	172 (40.6)	95 (46.8)
Item 3 (integration)	23 (8.6)	98 (22.1)	180 (42.6)	99 (48.8)
CollaboRATE-5 top score, n (%)				
Item 1 (information)	26 (9.7)	109 (24.6)	175 (41.3)	104 (51.5)
Item 2 (preference)	24 (8.9)	111 (25.2)	184 (43.3)	95 (47.0)
Item 3 (integration)	24 (9.0)	111 (25.1)	194 (45.8)	112 (55.2)

Concurrent Validity, Divergent Validity, and Intrarater Reliability

The concurrent validity of CollaboRATE was demonstrated with moderate to strong positive correlations between the 2 other measures of SDM for both response scales and both scoring methods (see Table 6). Divergent validity of CollaboRATE was not demonstrated, with moderate to strong positive correlations also observed with the clinician technical skills rating for both response scales and both scoring methods. Intrarater reliability of CollaboRATE mean scores was demonstrated for both response scales, with excellent intraclass correlations observed between Time 1, initial survey completion, and Time 2, resurvey, scores. Intrarater reliability of CollaboRATE top scores was also demonstrated for both response scales, with moderate agreement observed between Time 1 and Time 2 scores (Table 6).

Sensitivity to Change

Sensitivity to change of CollaboRATE was demonstrated with significant differences observed between scores for encounters with zero and 3 dimensions of SDM (within participants) for both response scales and both scoring methods (Table 7).

Table 6. Concurrent validity, divergent validity, and intrarater reliability of CollaboRATE.

Psychometric properties of CollaboRATE	Statistic	95% CI	*P*	Relationship	Valid/reliable[a]
Concurrent validity (with SDM-Q-9)					
CollaboRATE-10 mean	*r*=0.79	0.77, 0.81	<.001	Strong, positive	Yes
CollaboRATE-5 mean	*r*=0.80	0.78, 0.82	<.001	Strong, positive	Yes
CollaboRATE-10 top score	rpb=0.49	0.45, 0.53	<.001	Moderate, positive	Yes
CollaboRATE-5 top score	rpb=0.50	0.46, 0.54	<.001	Strong, positive	Yes
Concurrent validity (with PICS-DFS)					
CollaboRATE-10 mean	*r*=0.67	0.64, 0.70	<.001	Strong, positive	Yes
CollaboRATE-5 mean	*r*=0.68	0.65, 0.71	<.001	Strong, positive	Yes
CollaboRATE-10 top score	rpb=0.36	0.31, 0.41	<.001	Moderate, positive	Yes
CollaboRATE-5 top score	rpb=0.37	0.32, 0.42	<.001	Moderate, positive	Yes
Divergent validity (with clinician technical skills rating)					
CollaboRATE-10 mean	rpb=0.42	0.37, 0.46	<.001	Moderate, positive	No
CollaboRATE-5 mean	rpb=0.46	0.42, 0.51	<.001	Moderate, positive	No
CollaboRATE-10 top score	Agreement=83.4 % Kappa= 0.53	0.48, 0.59	<.001	Moderate	No
CollaboRATE-5 top score	Agreement=83.8 % kappa=0.55	0.50, 0.60	<.001	Moderate	No
Intrarater reliability (Time 1 to Time 2)					
CollaboRATE-10 mean	ICC (2,2)=0.86	0.82, 0.90	<.001	Excellent	Yes
CollaboRATE-5 mean	ICC (2,2)=0.82	0.76, 0.87	<.001	Excellent	Yes
CollaboRATE-10 top score	Agreement=84.7 % kappa=0.56	0.42, 0.70	<.001	Moderate	Yes
CollaboRATE-5 top score	Agreement=82.4 % kappa=0.58	0.44, 0.72	<.001	Moderate	Yes

[a] Yes=psychometric property found in this sample; no=psychometric property not found in this sample.

Table 7. Sensitivity to change of CollaboRATE.

CollaboRATE	Time 1 to Time 2 (n=29)					Time 1 to Time 2 (n=33)				
	Dimensions		Statistic			Dimensions		Statistic		
	0	3	t_{28}	χ^2_1	*P*	3	0	*t* (*df*)	χ^2_1	*P*
CollaboRATE-10 mean, mean (SD)	38.0 (29.0)	78.9 (28.0)	–6.75		<.001	82.2 (18.3)	66.3 (25.5)	3.58 (32)		<.001
CollaboRATE-5 mean,[a] mean (SD)	4.5 (3.2)	9.0 (3.8)	–5.87		<.001	9.7 (2.6)	7.4 (3.2)	4.73 (31)		<.001
CollaboRATE-10 top score, n (%)	1 (3.5)	12 (41.4)		11.0	.001	11 (33.3)	3 (9.1)		8.0	.008
CollaboRATE-5 top score, n (%)	1 (3.5)	13 (44.8)		12.0	<.001	13 (39.4)	5 (15.2)		6.4	.02

[a] 1 missing response for CollaboRATE-5 Time 1 (3 dimensions) to Time 2 (0 dimensions).

Discussion

Principal Findings

In simulated patient-clinician encounters, CollaboRATE, a patient-reported measure of the SDM process, demonstrated discriminative validity, concurrent validity, intrarater reliability, and sensitivity to change. Divergent validity was not demonstrated. Although further testing in real-world clinical care is needed, these results provide a solid foundation on which to consider this measure a fast and frugal measure of the SDM process.

CollaboRATE discriminated between all levels of SDM. It was particularly effective when discriminating between the absence and presence of any level of SDM. Although the discriminative ability of CollaboRATE was evident between moderate and high SDM encounters, the magnitude of differences was smaller. A greater number of recordings would be required to detect differences between moderate and high SDM in real-world settings.

CollaboRATE performed as well as the 2 most-commonly used patient-reported measures of SDM process. All 3 measures (CollaboRATE, SDM-Q-9, and the PICS-DFS) demonstrated excellent psychometric qualities, including discriminative validity (previously unreported for both SDM-Q-9 and the PICS-DFS). CollaboRATE scores remained consistent when retested over a 1- to 2-week period. CollaboRATE was also capable of detecting a change on resurvey in the level of SDM when participants viewed a clinical encounter with a different number of core dimensions.

There was little difference in the psychometric properties of CollaboRATE when a 10-point anchored scale or 5-point Likert response scale was used. Further discussion and testing with patients and clinicians, in real clinics is required to decide which is preferred. In addition, top score analysis was also conducted as part of our analysis and mirrored the psychometric properties of CollaboRATE when treated as a continuous outcome, with the exception of reduced intrarater reliability.

Strengths and Limitations of the Study Method

A strength of our method is that the use of simulated encounters, delivered via the Internet, allowed us to examine discriminative validity in ways that are not possible in clinical settings. The use of simulated medical encounters in this manner has been used successfully in previous studies and is deemed an important intermediary step to real-world testing [10,44-46]. It could be argued that assessing a measure under idealized circumstances is highly desirable because it is less time consuming, less intrusive for participants, and less costly. Moreover, if a measure cannot perform under ideal and controlled circumstances, it is unlikely to succeed in the mire of clinical practice. In addition, we have successfully demonstrated that the Internet can be used to successfully deliver and conduct this type of psychometric assessment in the field of SDM. We hope to encourage other measure development researchers to consider this approach in the future before testing in the clinical setting. Our choice of animated characters rather than real-life video recordings was to avoid potential rater bias that has been commonly reported with the latter [47]. The simulated encounters are freely available for use as teaching or research resources (Multimedia Appendices 1-6).

A limitation is potential confounding because of the differing durations of the simulated encounters, which increased as more dimensions of SDM were included. However, we argue that this is also likely reflective of how SDM might increase in the clinical setting. In addition, the varying length of times across each of the 6 scenarios could not be standardized without introducing more bias. We plan to assess the impact of consultation length on SDM in usual care. We were also limited to creating encounters that dealt with only 1 health issue. This may be reflective of specialist care, but it is less reflective of primary care. Our choice of clinician technical skills as measure of divergent validity appeared inadequate as none of the measures could meet this criteria using this question in the current sample. Finally, there was the potential of introducing selection bias, as approximately 35% of participants were excluded for not taking the minimal required time to view the encounter and complete the survey. However, we feel this risk was low as are use of quotas ensured a representative sample of the US population.

Results in Context

Our findings contribute further evidence that short patient-reported measures can produce valid and reliable results [19,22,48,49] and we believe that CollaboRATE addresses this gap in the field of SDM process measurement [7]. The psychometric qualities of CollaboRATE in the current study compare well with the reported psychometric properties of existing measures [6,7]. We demonstrated intrarater reliability, as has been shown for the Facilitation of Patient Involvement in Care Scale [11]. We also demonstrated CollaboRATE's discriminative validity and sensitivity to change. To our knowledge, this is the first time these aspects of validity have been demonstrated in a measure of the SDM process. Although short whole-encounter measures of the SDM process exist currently [11,14], we believe CollaboRATE is more understandable for patients because it avoids explicit reference to decisions made within the encounter [15,16].

Implications

To date, measures of the SDM process are not routinely implemented in clinical practice. We believe that CollaboRATE can assist in this effort because it is easy to understand and allows for uncomplicated analysis. Our previous work demonstrates that CollaboRATE is also easily administered to patients and has high face validity [24]. In addition there are practical benefits of short tools for both research (eg, reduced respondent burden), and policy (eg, ease of interpretation, implementation, and cost) [20,21]. The potential of CollaboRATE to assess SDM generically increases the potential scope of its use, whether patients seek help for long-term conditions or in situations where alternative treatments need to be compared. Although lengthier observer- and patient-reported measures of the SDM process can provide more detail about the consultation, CollaboRATE is better positioned to be used on a larger scale to produce valid and reliable measurement of the SDM process while also enabling faster feedback to clinics and clinicians. This patient-reported feedback can have positive effects on clinical practice [50] and patient participation in medical care associated with a range of positive health outcomes [51].

Conclusion

We have developed a fast and frugal measure of the SDM process that has sound psychometric properties when tested in a simulated setting. Stage 3 evaluation of CollaboRATE in real-world clinical settings, including its psychometric properties and feasibility, is now required.

Acknowledgments

We would like to acknowledge Greg McHugo, PhD; Shama Alam, MSc PH; Robin Paradis Montibello, MLIS; Arianna Blaine, MPH; and the independent raters for their support on this project. This project was supported through the Dartmouth Center for Health Care Delivery Science. No external sources of funding were used.

Conflicts of Interest

Glyn Elwyn has received funding from the Informed Medical Decisions Foundation, Boston, MA, USA, and provides ad hoc consulting to Emmi Solutions. No other authors have conflicts of interests to report.

Multimedia Appendix 1

Simulated doctor-patient encounter with no attempt at shared decision making.

[MP4 File (MP4 Video), 6MB - jmir_v16i1e2_app1.mp4]

Multimedia Appendix 2

Simulated doctor-patient encounter with a low level of shared decision making: detailed Information on health issue provided, no preference elicitation or integration.

[MP4 File (MP4 Video), 17MB - jmir_v16i1e2_app2.mp4]

Multimedia Appendix 3

Simulated doctor-patient encounter with a low level of shared decision making: preference elicitation, little information on health issue provided and no preference integration.

[MP4 File (MP4 Video), 11MB - jmir_v16i1e2_app3.mp4]

Multimedia Appendix 4

Simulated doctor-patient encounter with a moderate level of shared decision making: detailed Information on health issue provided and preferences elicited, no preference integration.

[MP4 File (MP4 Video), 22MB - jmir_v16i1e2_app4.mp4]

Multimedia Appendix 5

Simulated doctor-patient encounter with a moderate level of shared decision making: preferences elicited and integrated, little information on health issue provided.

[MP4 File (MP4 Video), 14MB - jmir_v16i1e2_app5.mp4]

Multimedia Appendix 6

Simulated doctor-patient encounter with a high level of shared decision making: detailed Information on health issue provided, preferences elicited and integrated.

[MP4 File (MP4 Video), 25MB - jmir_v16i1e2_app6.mp4]

References

1. Coulter A, Edwards A, Elwyn G, Thomson R. Implementing shared decision making in the UK. Z Evid Fortbild Qual Gesundhwes 2011 Jan;105(4):300-304. [doi: 10.1016/j.zefq.2011.04.014] [Medline: 21620325]
2. Frosch DL, Moulton BW, Wexler RM, Holmes-Rovner M, Volk RJ, Levin CA. Shared decision making in the United States: policy and implementation activity on multiple fronts. Z Evid Fortbild Qual Gesundhwes 2011;105(4):305-312. [doi: 10.1016/j.zefq.2011.04.004] [Medline: 21620326]
3. Härter M, van der Weijden T, Elwyn G. Policy and practice developments in the implementation of shared decision making: an international perspective. Z Evid Fortbild Qual Gesundhwes 2011 Jan;105(4):229-233. [doi: 10.1016/j.zefq.2011.04.018] [Medline: 21620313]
4. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]

5. Gravel K, Légaré F, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. Implement Sci 2006 Jan;1:16 [FREE Full text] [doi: 10.1186/1748-5908-1-16] [Medline: 16899124]
6. Simon D, Loh A, Härter M. Measuring (shared) decision-making--a review of psychometric instruments. Z Arztl Fortbild Qualitatssich 2007 May;101(4):259-267. [Medline: 17601182]
7. Scholl I, Koelewijn-van Loon M, Sepucha K, Elwyn G, Légaré F, Härter M, et al. Measurement of shared decision making - a review of instruments. Z Evid Fortbild Qual Gesundhwes 2011 Jan;105(4):313-324. [doi: 10.1016/j.zefq.2011.04.012] [Medline: 21620327]
8. Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013 Feb;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078] [Medline: 23381520]
9. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards A, Clay K, et al. “Many miles to go …”: a systematic review of the implementation of patient decision support interventions into routine clinical practice. BMC Med Inform Decis Mak 2013;13(Suppl 2):S14. [doi: 10.1186/1472-6947-13-S2-S14]
10. Melbourne E, Sinclair K, Durand MA, Légaré F, Elwyn G. Developing a dyadic OPTION scale to measure perceptions of shared decision making. Patient Educ Couns 2010 Feb;78(2):177-183. [doi: 10.1016/j.pec.2009.07.009] [Medline: 19647970]
11. Martin LR, DiMatteo MR, Lepper HS. Facilitation of patient involvement in care: development and validation of a scale. Behav Med 2001 Jan;27(3):111-120. [doi: 10.1080/08964280109595777] [Medline: 11985184]
12. Lerman CE, Brody DS, Caputo GC, Smith DG, Lazaro CG, Wolfson HG. Patients' Perceived Involvement in Care Scale: relationship to attitudes about illness and medical care. J Gen Intern Med 1990;5(1):29-33. [Medline: 2299426]
13. Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. Patient Educ Couns 2010 Jul;80(1):94-99. [doi: 10.1016/j.pec.2009.09.034] [Medline: 19879711]
14. Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA 1997 May 14;277(18):1485-1492. [Medline: 9145723]
15. Entwistle VA, Skea ZC, O'Donnell MT. Decisions about treatment: interpretations of two measures of control by women having a hysterectomy. Soc Sci Med 2001 Sep;53(6):721-732. [Medline: 11511048]
16. Entwistle VA, Watt IS, Gilhooly K, Bugge C, Haites N, Walker AE. Assessing patients' participation and quality of decision-making: insights from a study of routine practice in diverse settings. Patient Educ Couns 2004 Oct;55(1):105-113. [doi: 10.1016/j.pec.2003.08.005] [Medline: 15476997]
17. Hauer KE, Fernandez A, Teherani A, Boscardin CK, Saba GW. Assessment of medical students' shared decision-making in standardized patient encounters. J Gen Intern Med 2011 Apr;26(4):367-372 [FREE Full text] [doi: 10.1007/s11606-010-1567-7] [Medline: 21108048]
18. Weiss MC, Peters TJ. Measuring shared decision making in the consultation: a comparison of the OPTION and Informed Decision Making instruments. Patient Educ Couns 2008 Jan;70(1):79-86. [doi: 10.1016/j.pec.2007.09.001] [Medline: 17942270]
19. de Boer AG, van Lanschot JJ, Stalmeier PF, van Sandick JW, Hulscher JB, de Haes JC, et al. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? Qual Life Res 2004 Mar;13(2):311-320. [Medline: 15085903]
20. Bowling A. Just one question: If one question works, why ask several? J Epidemiol Community Health 2005 May;59(5):342-345 [FREE Full text] [doi: 10.1136/jech.2004.021204] [Medline: 15831678]
21. Littman AJ, White E, Satia JA, Bowen DJ, Kristal AR. Reliability and validity of 2 single-item measures of psychosocial stress. Epidemiology 2006 Jul;17(4):398-403. [doi: 10.1097/01.ede.0000219721.89552.51] [Medline: 16641618]
22. Lindberg P, Josephson M, Alfredsson L, Vingård E. Comparisons between five self-administered instruments predicting sick leaves in a 4-year follow-up. Int Arch Occup Environ Health 2009 Jan;82(2):227-234. [doi: 10.1007/s00420-008-0326-0] [Medline: 18408948]
23. DeSalvo KB, Jones TM, Peabody J, McDonald J, Fihn S, Fan V, et al. Health care expenditure prediction with a single item, self-rated health measure. Med Care 2009 Apr;47(4):440-447. [doi: 10.1097/MLR.0b013e318190b716] [Medline: 19238099]
24. Elwyn G, Barr PJ, Grande SW, Thompson R, Walsh T, Ozanne EM. Developing CollaboRATE: a fast and frugal patient-reported measure of shared decision making in clinical encounters. Patient Educ Couns 2013 Oct;93(1):102-107. [doi: 10.1016/j.pec.2013.05.009] [Medline: 23768763]
25. CollaboRATE. URL: http://www.collaboratescore.org/collaborate.html [accessed 2013-12-18] [WebCite Cache ID 6Ly0pLKdC]
26. File T. Computer and Internet use in the United States: Population characteristics. Washington, DC: US Census Bureau; 2013 May. URL: http://www.census.gov/hhes/computer/files/2011/p20-568.pdf [accessed 2013-11-04] [WebCite Cache ID 6Kt1Dibjn]
27. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. J Clin Epidemiol 2010 Nov;63(11):1169-1178 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.11.021] [Medline: 20688473]

28. Elwyn G, Hutchings H, Edwards A, Rapport F, Wensing M, Cheung WY, et al. The OPTION scale: measuring the extent that clinicians involve patients in decision-making tasks. Health Expect 2005 Mar;8(1):34-42. [doi: 10.1111/j.1369-7625.2004.00311.x] [Medline: 15713169]
29. Shields CG, Franks P, Fiscella K, Meldrum S, Epstein RM. Rochester Participatory Decision-Making Scale (RPAD): reliability and validity. Ann Fam Med 2005;3(5):436-442 [FREE Full text] [doi: 10.1370/afm.305] [Medline: 16189060]
30. Rubin HR, Gandek B, Rogers WH, Kosinski M, McHorney CA, Ware JE. Patients' ratings of outpatient visits in different practice settings. Results from the Medical Outcomes Study. JAMA 1993 Aug 18;270(7):835-840. [Medline: 8340982]
31. US Census Bureau. 2012. Annual social and economic supplement URL: http://www.census.gov/hhes/socdemo/education/data/cps/ [accessed 2013-11-04] [WebCite Cache ID 6Kt2r0UWz]
32. Office of Management and Budget. Federal Register Notice. Washington, DC: The White House; 1997 Oct 30. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity URL: http://www.whitehouse.gov/omb/fedreg_1997standards [accessed 2013-11-04] [WebCite Cache ID 6Kt3743VT]
33. Shin HB, Kominski RA. Language Use in the United States: 2007. Washington, DC: US Census Bureau; 2010 Apr. URL: http://www.census.gov/hhes/socdemo/language/data/acs/ACS-12.pdf [accessed 2013-11-04] [WebCite Cache ID 6Kt3LUgOz]
34. CAHPS Clinician & Group Surveys: Supplemental Items for the Adult Surveys 2.0. 2012 May 01. URL: https://cahps.ahrq.gov/surveys-guidance/docs/2357a_Adult_Supp_Eng_20.pdf [accessed 2013-12-20] [WebCite Cache ID 6M0uPqWq5]
35. Walker A, Maher J, Coulthard M, Goddard E, Thomas M. Living in Britain: Results from the 2000/01 General Household Survey. London: Office for National Statistics; 2001. URL: http://www.ons.gov.uk/ons/rel/ghs/general-household-survey/2001-edition/index.html [accessed 2013-12-20] [WebCite Cache ID 6M0vTJqfl]
36. Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. Can J Nurs Res 1997;29(3):21-43. [Medline: 9505581]
37. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press; 2008.
38. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
39. Acock A. A Gentle Introduction to Stata, Revised Third Edition. College Station, TX: Stata Press; 2012.
40. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
41. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care 2000 Sep;38(9 Suppl):II84-II90. [Medline: 10982093]
42. Howden L, Meyer J. Age and Sex Composition: 2010. Washington, DC: US Census Bureau; 2011 May. URL: http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf [accessed 2013-11-04] [WebCite Cache ID 6KsrcT4Gu]
43. Humes K, Jones N, Ramirez R. Overview of Race and Hispanic Origin: 2010. Washington, DC: US Census Bureau; 2011 Mar. URL: http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf [accessed 2013-11-04] [WebCite Cache ID 6KsqkEgl9]
44. Schnabl GK, Hassard TH, Kopelow ML. The assessment of interpersonal skills using standardized patients. Acad Med 1991 Sep;66(9 Suppl):S34-S36. [Medline: 1930521]
45. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. Acad Med 1996 Jan;71(1 Suppl):S87-S89. [Medline: 8546794]
46. van Zanten M, Boulet JR, McKinley D. Using standardized patients to assess the interpersonal skills of physicians: six years' experience with a high-stakes certification examination. Health Commun 2007 Jan;22(3):195-205. [doi: 10.1080/10410230701626562] [Medline: 17967142]
47. Winterbottom AE, Bekker HL, Conner M, Mooney AF. Patient stories about their dialysis experience biases others' choices regardless of doctor's advice: an experimental study. Nephrol Dial Transplant 2012 Jan;27(1):325-331. [doi: 10.1093/ndt/gfr266] [Medline: 21642512]
48. DeSalvo KB, Jones TM, Peabody J, McDonald J, Fihn S, Fan V, et al. Health care expenditure prediction with a single item, self-rated health measure. Med Care 2009 Apr;47(4):440-447. [doi: 10.1097/MLR.0b013e318190b716] [Medline: 19238099]
49. Littman AJ, White E, Satia JA, Bowen DJ, Kristal AR. Reliability and validity of 2 single-item measures of psychosocial stress. Epidemiology 2006 Jul;17(4):398-403. [doi: 10.1097/01.ede.0000219721.89552.51] [Medline: 16641618]
50. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jan;6:CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
51. Griffin SJ, Kinmonth AL, Veltman MW, Gillard S, Grant J, Stewart M. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. Ann Fam Med 2004;2(6):595-608 [FREE Full text] [doi: 10.1370/afm.142] [Medline: 15576546]

Abbreviations

CPHS: Committee for the Protection of Human Subjects

ICC: intraclass correlation
IP: Internet Protocol
IRB: Institutional Review Board
PICS-DFS: 5-item Doctor Facilitation subscale of the PICS
PICS: Perceived Involvement in Care Scale
RPAD: Rochester Participatory Decision-Making Scale
SDM-Q-9: 9-item Shared Decision Making Questionnaire
SDM: shared decision making

Edited by G Eysenbach; submitted 05.11.13; peer-reviewed by A Giguere, E Micaiah, L Kriston; comments to author 26.11.13; accepted 14.12.13; published 03.01.14.

Please cite as:
Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G
The Psychometric Properties of CollaboRATE: A Fast and Frugal Patient-Reported Measure of the Shared Decision-Making Process
J Med Internet Res 2014;16(1):e2
URL: http://www.jmir.org/2014/1/e2/
doi:10.2196/jmir.3085
PMID:24389354

Original Paper

Impact of Patient Access to Internet Health Records on Glaucoma Medication: Randomized Controlled Trial

Kenji Kashiwagi[1], MD, PhD; Shigeo Tsukahara[1], MD, PhD

Faculty of Medicine, University of Yamanashi, Chuo, Japan

Corresponding Author:
Kenji Kashiwagi, MD, PhD
Faculty of Medicine
University of Yamanashi
1110 Shimokato
Chuo, 409-3898
Japan
Phone: 81 552731111 ext 2372
Fax: 81 552736757
Email: kenjik@yamanashi.ac.jp

Abstract

Background: Glaucoma is one of the leading causes of blindness. Reduction of intraocular pressure is the only proven way to prevent progression of glaucomatous optic neuropathy. The majority of glaucoma patients need to use antiglaucoma ophthalmic solutions over the course of their life. Thus, good adherence and persistency of glaucoma treatment are important factors for better glaucoma care.

Objective: The purpose of this study was to investigate the impact of an Internet-based glaucoma care support system on glaucoma medication use.

Methods: Patients were randomly divided into two groups. The non–Internet access (NIA) group consisted of patients who had access to the Internet-based glaucoma care support system during the 4-year period only when they were examined by ophthalmologists. The Internet access (IA) group consisted of patients who had the same Internet-based glaucoma care support system access as the NIA group for the first 2 years following enrollment but who were also given free access to the glaucoma care support system for the remaining 2 years. Changes in glaucoma medication use were investigated.

Results: In total, 81 patients in the IA group and 90 patients in the NIA group satisfied the study protocol. The number of antiglaucoma ophthalmic solutions used during the study period significantly increased in the NIA group (*P*<.03) but not in the IA group. The percentages of patients with unchanged, increased, and decreased antiglaucoma ophthalmic solution use during the study period were 61.1% (55/90), 17.8% (16/90), and 3.3% (3/90), respectively, in the NIA group, and 56.8% (46/81), 8.6% (7/81), and 13.6% (11/81), respectively, in the IA group (*P*<.001). Internet access significantly shifted from an increasing intraocular pressure trend to a decreasing trend in the IA group (*P*=.002) among the patients who did not have any medication changes.

Conclusions: Allowing patients to browse their medical data may reduce the use and improve the effectiveness of glaucoma medication.

Trial Registration: UMIN-CTR Clinical Trial Number: UMIN000006982; https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?function=brows&action=brows&type=summary&recptno=R000008238&language=E (Archived by WebCite at http://www.webcitation.org/6MRPQeEAv).

(J Med Internet Res 2014;16(1):e15) doi:10.2196/jmir.2795

KEYWORDS

Internet; glaucoma; intraocular pressure; personal health record; medication

Introduction

Glaucoma is one of the leading causes of acquired blindness and reduction of intraocular pressure (IOP) is the only proven therapy. Glaucoma treatment generally consists of antiglaucoma ophthalmic solutions—medications that are required for long periods of time [1]. Recent studies have revealed that a number of patients fail to comply with proper glaucoma medication regimens [2] because many glaucoma patients lack any glaucoma-related symptoms.

Treatment adherence is a major concern in many chronic diseases [3], and improving patients' understanding of their diseases has been found to be vital for appropriate treatment [4-6]. A variety of interventions have been used to increase understanding, including patient education during medical examinations, public lectures, distributing medical information pamphlets, and mass media advertising [7-11]. Okeke et al have reported that a multifaceted intervention significantly increased adherence to glaucoma medication during a 3-month trial [11]. Significant efforts have been devoted to promoting glaucoma education, which is considered critical to improving treatment effectiveness [8,12]. However, these efforts are expensive and require human resources. In addition, the effects of these efforts are transient [9,13,14], and more efficient and effective long-term patient education systems are needed.

In recent years, a personalized health record (PHR) system has been proposed. This system uses information and communication technology to allow patients to manage their own health conditions [15,16]. The main purpose of many proposed PHR systems is to share medical records among clinicians [16,17].

Unfortunately, the effects of PHR systems on clinical outcomes in previous studies have not been consistent. Miller et al reported that PHR-enabled self-management did not improve care in multiple sclerosis patients [18]. In their review, Tenforde et al concluded that the evidence supporting the clinical value of PHR remains limited, despite its potential to improve chronic disease management and patient outcomes [19].

Some previous studies have focused on the effects of PHR in glaucoma therapy [6,20]. Gray et al reported that individualized patient care improved glaucoma knowledge, pre-existing beliefs, and management of a daily eye drop regimen [20]. However, there is insufficient evidence on the benefits of medical record self-management in glaucoma treatment.

In 2005, we introduced an Internet-based glaucoma care support system (GSS) in Japan. This system allowed glaucoma patients to view their own medical records via the Internet at any time and from any location. The goal of this system was to deepen patient understanding of glaucoma and encourage active involvement in treatment.

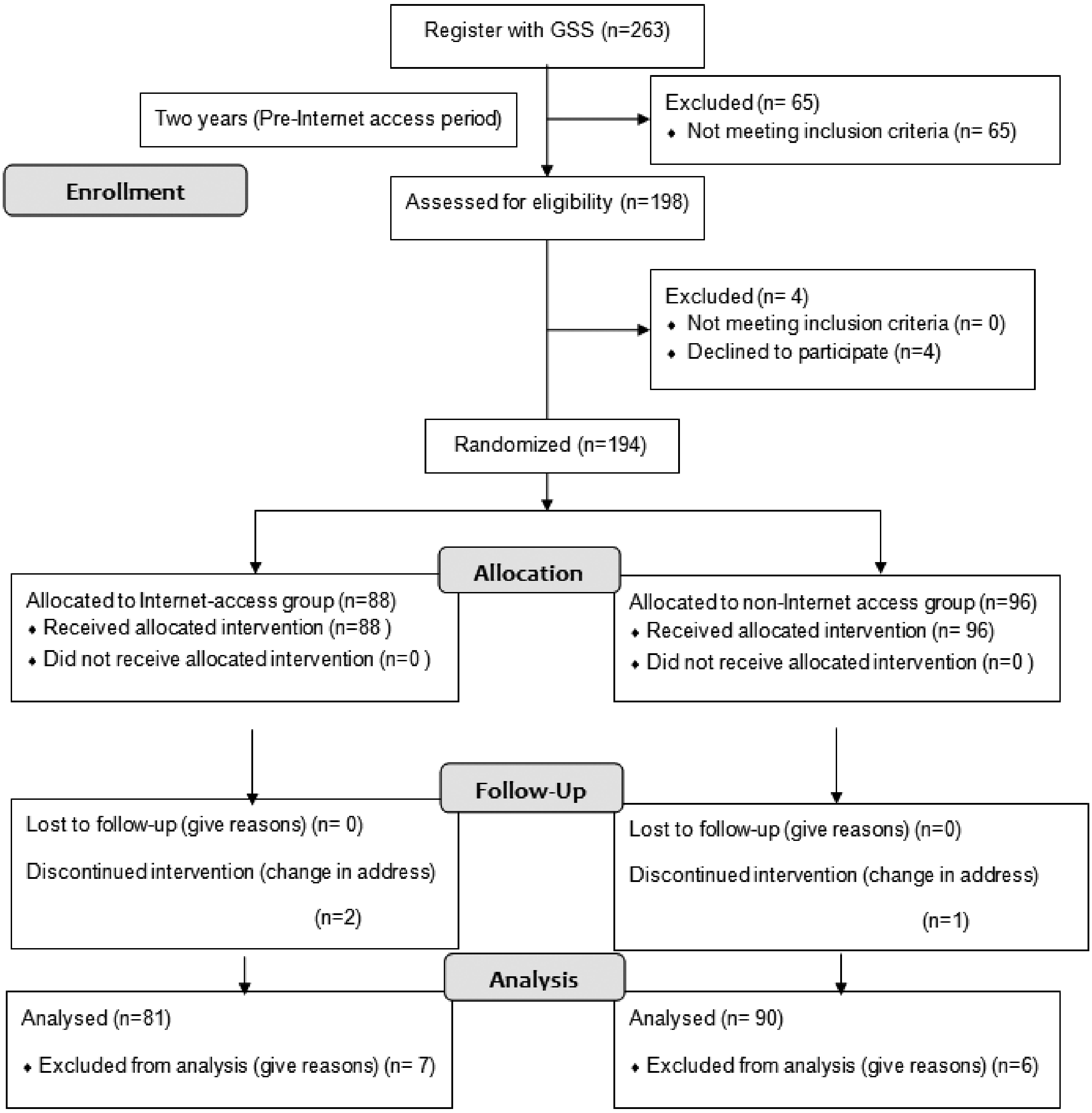
In this study, we targeted patients who had used the GSS for more than 4 years and examined how access to the contents of their own glaucoma medical records ("Internet access") affected their glaucoma treatment.

Methods

Ethics Statement

This randomized, observer-blinded, prospective trial study was performed in accordance with the Helsinki Treaty and was approved by the University of Yamanashi Ethical Review Board. Written informed consent was obtained from all of the patients (see Figure 1 for the design of this study and Multimedia Appendix 1 for the CONSORT checklist; trial registration number UMIN000006982).

Figure 1. Flow diagram of study (GSS: glaucoma care support system).

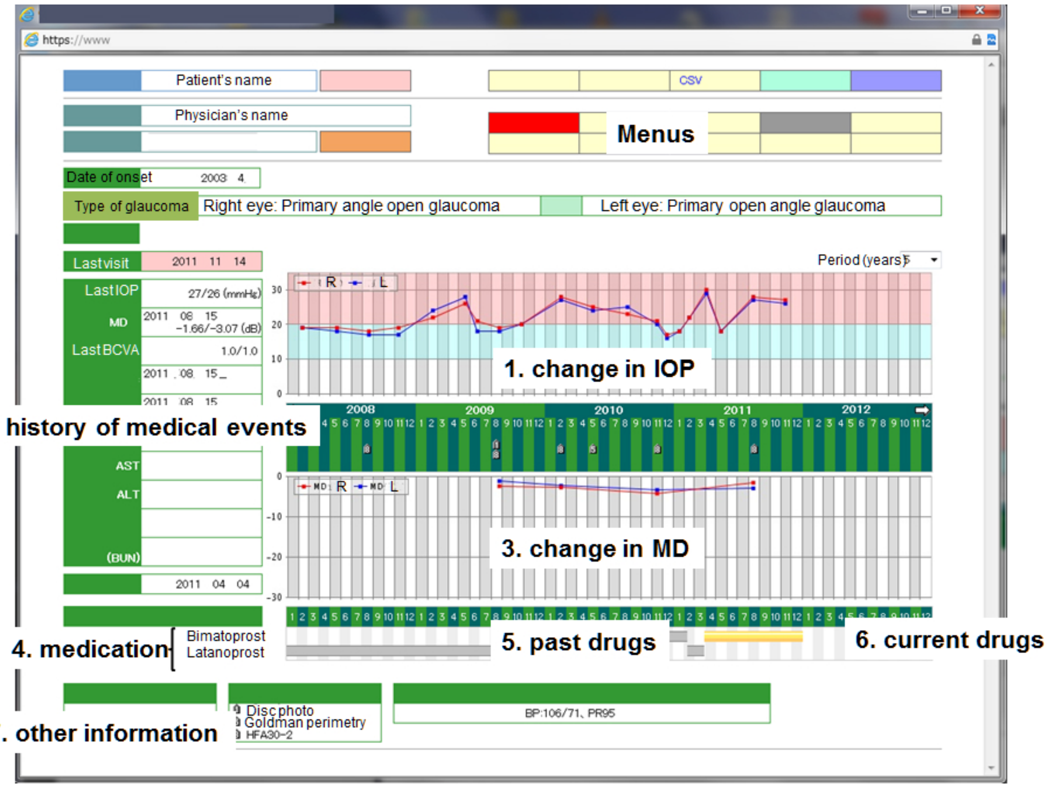


Glaucoma Care Supporting System

The GSS is a secure online PHR system. Registered patients can use the GSS to view their IOP values, visual field test results, current and prior medications, and other findings from their examination histories related to glaucoma. Figure 2 shows the first page of the GSS. Graphs are used to facilitate patient understanding of their medical records. Patients can view the detailed results of visual field tests on different pages (see Figures 3 and 4).

The GSS is based on the concept of an information security management system. The data are stored in a database server that is located in a facility at the University of Yamanashi. The registered GSS data are periodically updated by automated medical chart extraction or manual data entry by physicians and medical staff. All the data transferred from medical charts to the GSS are managed offline using locally developed data management programs. Patient registration for the GSS began in 2005, and approximately 1600 glaucoma patients were registered as of November 2011. Our plan was to allow registered patients to access their medical records by themselves. The GSS was the first PHR system routinely used in Japanese clinical ophthalmology care.

Figure 2. First page of GSS. History of medical events are marked on the calendar with numbers explained in a box (arrow). Changes in MD values indicate possibility of deterioration of glaucoma stage. Prescription details are shown by gray or colored bar, respectively. IOP: intraocular pressure, MD: mean deviation, R: right eye, L: left eye.



https://www

Patient's name

Physician's name

Date of onset 2003 4

Type of glaucoma Right eye: Primary angle open glaucoma Left eye: Primary open angle glaucoma

Lastvisit 2011 11 14

Last IOP 27/26 (mmHg)

MD 2011 08 15 −1.66/−3.07 (dB)

Last BCVA 1.0/1.0

2011 08 15

2011 08 15

2011 04 04

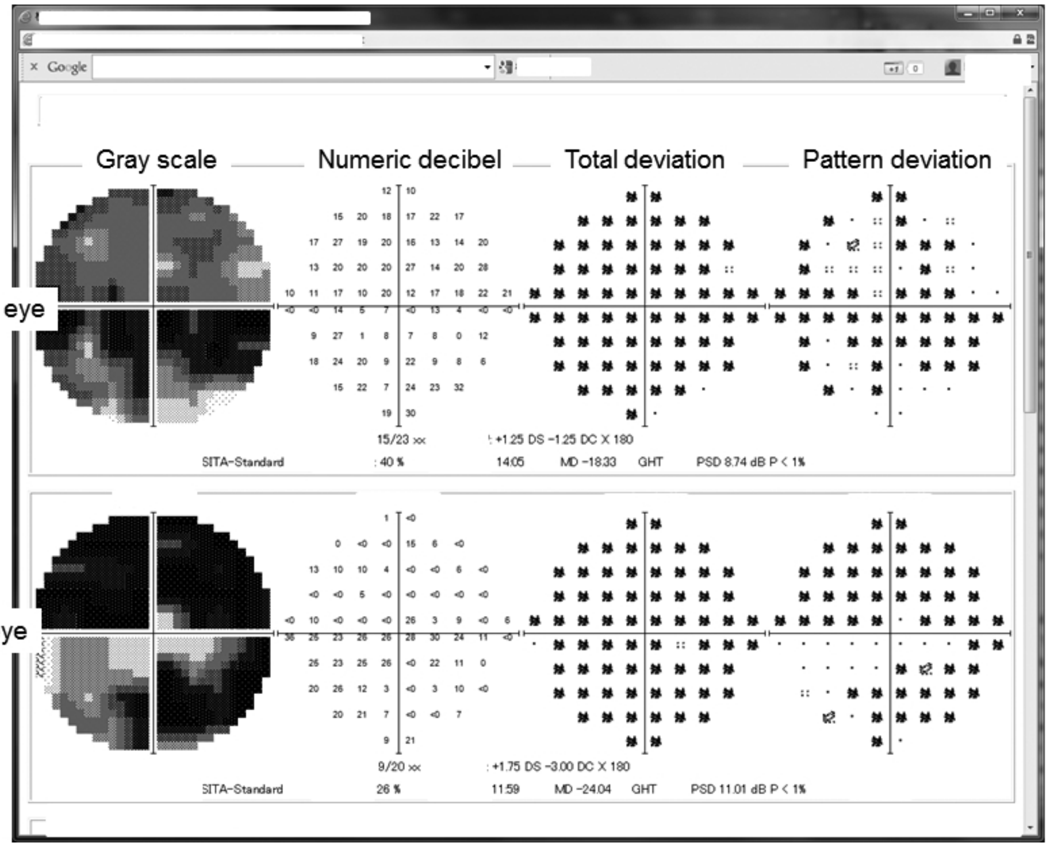
Period (years) 5

Bimatoprost
Latanoprost

Disc photo
Goldman perimetry
HFA30-2

BP:106/71, PR95

Figure 3. Results of the static visual field test using the Humphrey visual field test, including gray scale maps, numeric decibels, and the total and pattern deviation. MD: mean deviation, PSD: pattern standard deviation.



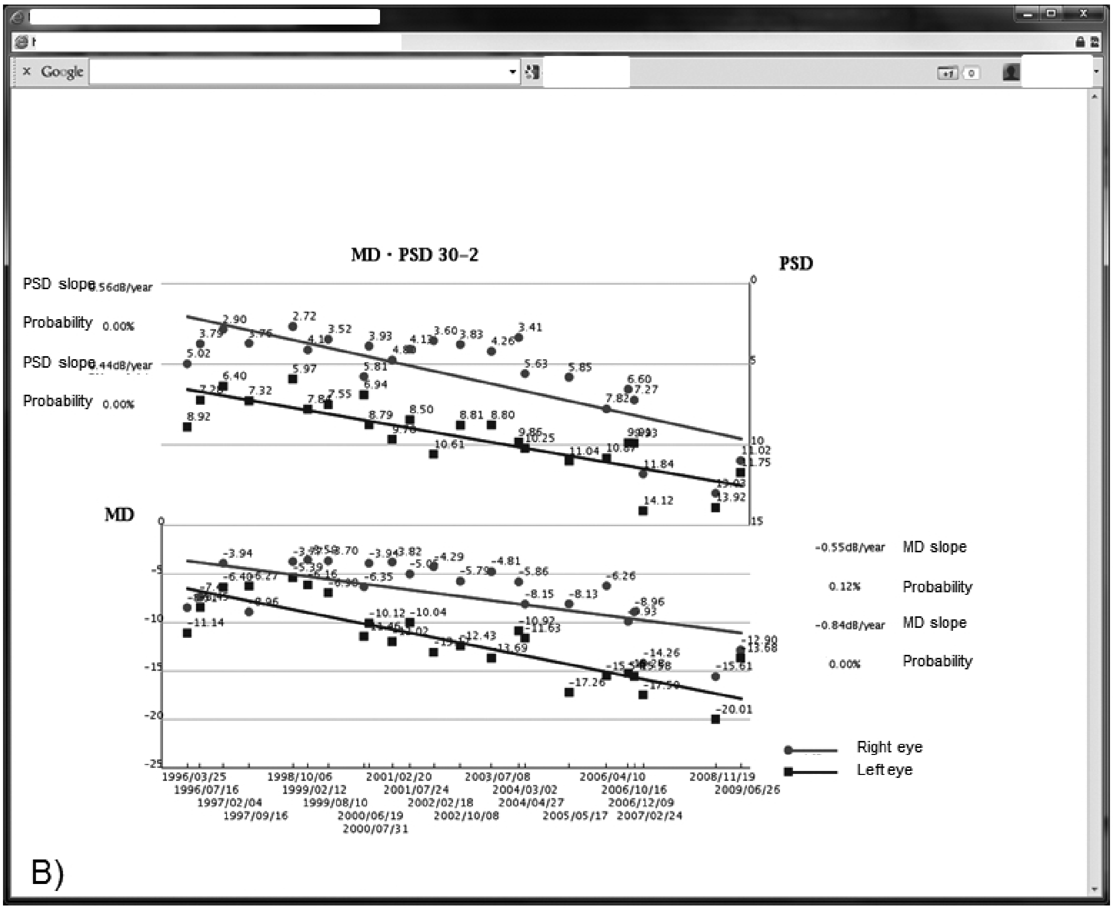
15/23 xx +1.25 DS −1.25 DC X 180

SITA-Standard 40 % 14:05 MD −18.33 GHT PSD 8.74 dB P < 1%

9/20 xx +1.75 DS −3.00 DC X 180

SITA-Standard 26 % 11:59 MD −24.04 GHT PSD 11.01 dB P < 1%

Figure 4. The PSD (upper graph) and MD slopes (lower graph) show the possibility of glaucoma progression with a probability analysis. MD: mean deviation, PSD: pattern standard deviation.



Patients

Glaucoma patients who had been periodically treated at the University of Yamanashi Hospital Glaucoma Outpatient Clinic and expressed interest in registering with the GSS were subsequently registered for the GSS. Patients were informed of the benefits and risks of the GSS by the ophthalmologist and the medical staff. The patients were also informed that the registered medical records would not be available until security was well established. Registration for the GSS began in 2005. By 2008, only ophthalmologists were allowed to browse the registered medical records, and patients could view their registered medical records in the presence of the attending ophthalmologist at the clinic.

Study Design

Internet access by patients began in 2008. We randomly selected patients for Internet access between January 2008 and December 2008. The inclusion criteria included the following: patients who had been registered in the GSS database for 2 years and who were diagnosed with either primary open angle glaucoma (POAG), normal tension glaucoma (NTG), or ocular hypertension (OH). The exclusion criteria included the following: patients who were under 20 years old at the time of registration; patients who had a history of intraocular surgery in both eyes; patients for whom accurate IOP measurement was deemed difficult; patients who had received oral glaucoma therapy, such as carbonic anhydrase inhibitors; patients who had a disease other than glaucoma that caused visual field defects; patients with a visual acuity of less than 20/60 or a mean deviation (MD) value lower than –20 dB in the worse eye, as determined by the Humphrey Field Analyzer (HFA) central 30-2 program (Zeiss Inc); and patients diagnosed with dementia whose use of GSS was judged by an ophthalmologist to be difficult.

After re-confirming the patients’ preferences for browsing their registered medical records, identification numbers and temporary passwords were sent to the patients to allow them Internet access. The patients began to browse the GSS after attending lectures by the system staff.

The users were asked to replace their temporary passwords and register their user names with the system. To maintain patient anonymity, no information that could identify a patient—such as name, age, gender, telephone number, home address, or business address—was accessible through the Internet.

The outcomes of the patients who were assigned to the Internet access (IA) group were compared with those of patients who viewed their medical records only with an ophthalmologist during their medical examinations (the non–Internet access group, or NIA group).

Patients who met any of the following criteria were excluded during the 2-year period post Internet access: those who died; had glaucoma surgery, including laser treatment; received oral glaucoma therapy, such as carbonic anhydrase inhibitors;

developed a disease other than glaucoma that caused visual field disturbances or IOP changes; and developed severe visual impairment or dementia that caused GSS use to become difficult as judged by an ophthalmologist.

Ophthalmological Examination

All of the patients visited the Glaucoma Outpatient Clinic at the University of Yamanashi Hospital approximately every 3 months. A best-corrected visual acuity (BCVA) measurement, an IOP measurement using a Goldmann applanation tonometer, a slit-lamp examination, and a fundus examination were performed as part of the routine examinations. The HFA visual field test was usually performed every 6-10 months and disc photography was performed every year. No differences were found in the ophthalmological examination protocol or follow-up schedule between the NIA and IA groups. The patient and the attending ophthalmologist discussed the results of the glaucoma examination using data displayed by the GSS. The ophthalmologists did not know whether the patient browsed the GSS data at home or elsewhere.

Role of Changing Glaucoma Medication

Any glaucoma medication changes were determined by glaucoma specialists with the patient's consent. The medical glaucoma treatment was determined by glaucoma specialists using a targeted IOP strategy. The specialists increased the number of glaucoma medication treatments in response to the following test results: two or more consecutive IOP values exceeding the target IOP, glaucomatous neuropathy deterioration suggested by a visual field test, and imaging tests focused on the optic nerve head and nerve fiber layer thickness. If two or more consecutive IOP values were sufficiently below the target IOP and the glaucoma specialists judged that reducing the glaucoma medication would not elevate the IOP over the target value, the glaucoma medication was reduced.

Investigated Parameters

This study compared the amount of antiglaucoma ophthalmic solution used, MD values, and BCVA values between the two groups. In addition, we performed a subanalysis of IOP changes over the study period in those patients who did not report any changes in the amount of antiglaucoma ophthalmic solution used during the study. The IOP profiles were compared in the NIA and IA groups. IOP changes from the pre-IA to the post-IA period were also compared in the IA groups. The right eye was chosen for the analysis. If the right eye met exclusion criteria, then the left eye was subject to the analysis.

Definition of Change in the Number of Antiglaucoma Ophthalmic Solutions Used

We analyzed changes in the amount of antiglaucoma ophthalmic solution used for individual patients and individual eyes. If the change in the amount of antiglaucoma ophthalmic solution used during the study period differed between the right and left eyes, the eye that showed greater change was used in the analysis. Any change in the concentration of the antiglaucoma ophthalmic solution used was considered to be a change in the antiglaucoma medication. However, a change to a different antiglaucoma ophthalmic solution having the same pharmacological action and similar IOP-reducing potential was not considered to be a change. For example, changing from a latanoprost ophthalmic solution to another prostaglandin-related antiophthalmic solution (other than an isopropyl unoprostone solution) and changing from a timolol maleate ophthalmic solution to another beta-blocker ophthalmic solution were not considered to be changes in antiglaucoma medication.

Change in Medication Possession Ratio

We investigated the change in medication possession ratio (MPR) among patients whose medication was not changed through the study period. A useful definition of MPR as a parameter to measure adherence is the ratio of the days of prescription supply dispensed over the number of days between the first and last prescription refill [2,21,22]. The number of days of supplied medication was calculated from the actual number of drops per bottle, based on Fiscella's report [23] or from the manufacturer's estimates for the products, summarized by Friedman et al [2]. See Multimedia Appendix 2 for information on the days of supplied medication for the antiglaucoma ophthalmic solutions used.

Statistical Analysis

The data were analyzed using the JMP 8.0 software package (SAS Institute Inc), and the values are presented as the mean and standard deviation (SD). Changes in the amount of antiglaucoma ophthalmic solutions used were analyzed using the Wilcoxon signed rank test or a contingency table analysis. The IOP, BCVA (expressed as logMAR), MD values of the HFA central 30-2 program, and MPR were compared between the IA and NIA groups using the Mann-Whitney U test. The IOP changes within a group were analyzed using the Student *t* test, and the type of glaucoma and patient gender were compared using Fisher's exact probability test. The effect of Internet access on the IOP changing trend was analyzed using the analysis of covariance (ANCOVA) and Pearson correlation coefficient; *P*<.05 was considered to be statistically significant.

Results

Characteristics of Enrolled Patients

In total, 194 patients were randomly assigned to the two groups, and 81 IA and 90 NIA patients completed the study. Tables 1 and 2 show detailed information on patient dropouts and the characteristics of the patients who completed the study. The mean age, type of glaucoma, and gender did not differ significantly between the two groups. Both the IA and NIA patients had no significant changes in their BCVA and MD values obtained with the HFA central 30-2 program between the initial and final examinations.

Table 1. Characteristics of all enrolled patients.

Group	Start of IA			Details of dropout during the IA period			
	Patients, n	Age, years mean (SD)	Male, %	PE[a]	TLE[b]	Oral CAI[c]	Change in address
IA (Internet access)	88	62.6 (15.1)	62.6	2	2	1	2
NIA (non−Internet access)	96	64.6 (12.5)	87.1	1	2	2	1

[a] PE: phacoemulcification.

[b] TLE: trabeculectomy.

[c] CAI: carbonic anhydrase inhibitor.

Table 2. Characteristics of completed patients (numbers in parentheses are standard deviation).

Group	Patients, n	Age, years mean (SD)	Type of glaucoma[a] (POAG:NTG:OH)	Male (%)	Initial BCVA[b]	Final BCVA	Initial MD[c] (dB)	Final MD (dB)
IA (Internet Access)	81	61.8 (15.3)	36:42:3	63.0	0.09 (0.39)	0.1 (0.39)	–6.47 (7.36)	–7.16 (7.66)
NIA (Non−Internet Access)	90	63.4 (11.8)	38:47:5	54.4	0.03 (0.22)	0.03 (0.22)	–6.27 (7.36)	–6.83 (7.43)

[a] POAG: primary open angle glaucoma, NTG: normal tension glaucoma, OH: ocular hypertension.

[b] BCVA: best-corrected visual acuity.

[c] MD: mean deviation.

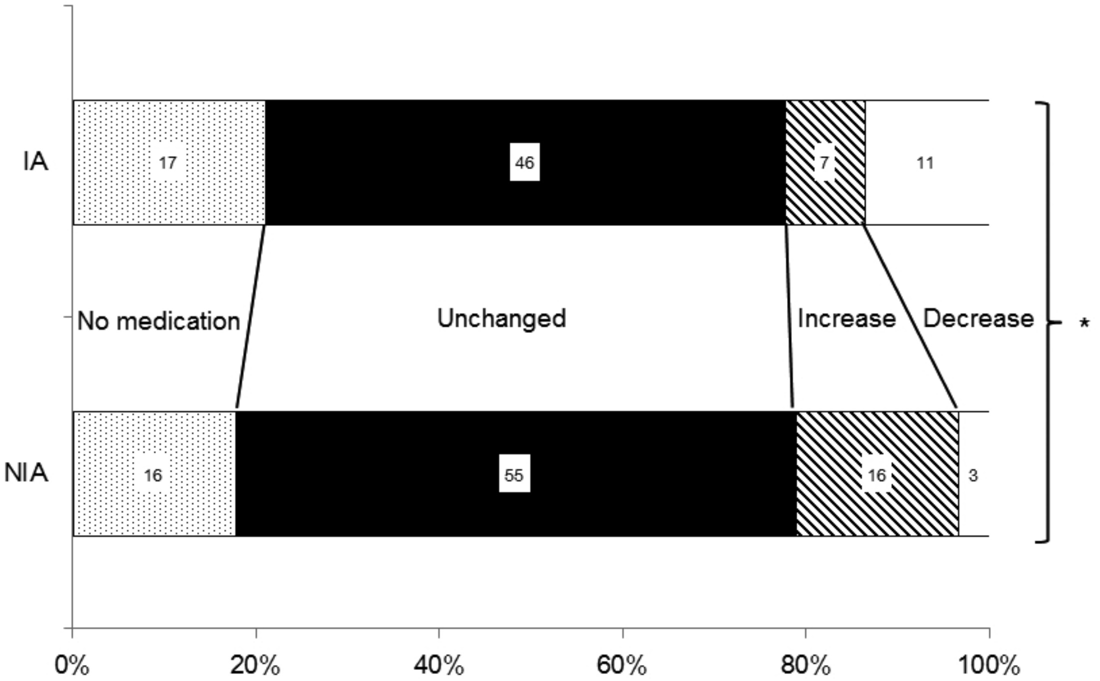
Changes in the Amount of Anti-Glaucoma Ophthalmic Solution Used

The number of antiglaucoma ophthalmic solutions used at the initial examination in the IA group was mean 1.2 (SD 1.2) in the right eye and mean 1.3 (SD 1.1) in the left eye, while the same number in the NIA group was mean 1.0 (SD 1.0) in the right eye and mean 1.1 (SD 1.0) in the left eye. The number of antiglaucoma ophthalmic solutions used did not differ significantly between the IA and NIA groups. In the NIA group at the final examination, it was mean 1.2 (SD 1.1) in the right eye and mean 1.2 (SD 1.0) in the left eye, which were significantly greater than the values at the initial examination (*P*<.001 for the right eye, *P*=.03 for the left eye, Wilcoxon signed-rank test). In contrast, the number of antiglaucoma ophthalmic solutions used in the IA group decreased by mean 0.1 (SD 0.4) in the right eye and by mean 0.1 (SD 0.6) in the left eye. The changes in the number of antiglaucoma ophthalmic solutions used in the NIA and IA groups were mean 0.3 (SD 0.4) and mean –0.1 (SD 0.4), respectively, which is significantly different (*P*<.001, Wilcoxon signed-rank test). None of the patients reported using less antiglaucoma ophthalmic solution in one eye and more in the other eye during the investigation period.

Distribution of Change in the Amount of Antiglaucoma Ophthalmic Solution Used

Figure 5 shows the distribution of the medication used by the glaucoma patients during the course of the study. In the NIA group (n=90), the amount of antiglaucoma ophthalmic solution used did not change in 55 patients (61.1%), increased in 16 patients (17.8%), and decreased in 3 patients (3.3%). No medications were administered to 16 patients (17.8%). In the IA group (n=81), the amount of antiglaucoma ophthalmic solution used did not change in 46 patients (56.8%), increased in 7 patients (8.6%), and decreased in 11 patients (13.6%). No medications were administered to 17 patients (21.0%). Compared to the NIA group, 10.3% more IA patients reported using decreased number of antiglaucoma ophthalmic solutions and 9.2% fewer patients reported using increased number of antiglaucoma ophthalmic solutions. A significant difference was found in the distribution of medications received between the two groups of glaucoma patients (*P*=.008, contingency table analysis).

Figure 5. Changes in glaucoma medication during the study period (*P=.008, 2 x 4 contingency table analysis).



Changes in Intraocular Pressure Values

IOP values at the first examination of the IA and NIA groups were mean 15.3 (SD 4.2) mmHg and mean 15.1 (SD 3.1) mmHg, respectively (*P*=.69). IOP values at the last examination of the IA and NIA groups were mean 14.6 (SD 3.6) mmHg and mean 15.1 (SD 3.3) mmHg, respectively (*P*=.26). Changes in IOP over the study period were not significant in either the IA (*P*=.23) or the NIA group (*P*=.83).

Analysis of the Patients Whose Antiglaucoma Ophthalmic Solution Use Did Not Change During the Study Period

We examined the patients whose antiglaucoma ophthalmic solution use did not change over the course of the study. The subjects included 46 patients in the IA group and 55 patients in the NIA group. The distribution of the cases is shown in Table 3. The two groups did not differ significantly by age, gender, type of glaucoma, or the amount of antiglaucoma ophthalmic solutions used. The two groups also did not differ significantly by their BCVA and MD values measured at the beginning of the study or by changes in their BCVA and MD values during the study period. See Multimedia Appendices 3 and 4 for information on the antiglaucoma ophthalmic solutions used.

Table 3. Characteristics of patients whose antiglaucoma ophthalmic solution use did not change during the study period (numbers in parentheses are standard deviation).

Group	Patients, n	Age, years mean (SD)	Type of glaucoma[a] (POAG:NTG:OH)	Male, %	# of antiglaucoma ophthalmic solution	Initial BC-VA[b]	Final BC-VA	Initial MD[c] (dB)	Final MD (dB)
IA (Internet Access)	46	63.7 (12.8)	21:24:1	66.7	1.5 (0.9)	0.08 (0.38)	0.10 (0.40)	–6.01 (7.32)	–7.04 (7.61)
NIA (Non−Internet Access)	55	66.4 (10.7)	21:32:2	61.1	1.2 (0.8)	0.09 (0.31)	0.10 (0.25)	–6.67 (7.24)	–8.07 (7.94)

[a] POAG: primary open angle glaucoma, NTG: normal tension glaucoma, OH: ocular hypertension.

[b] BCVA: best-corrected visual acuity.

[c] MD: mean deviation.

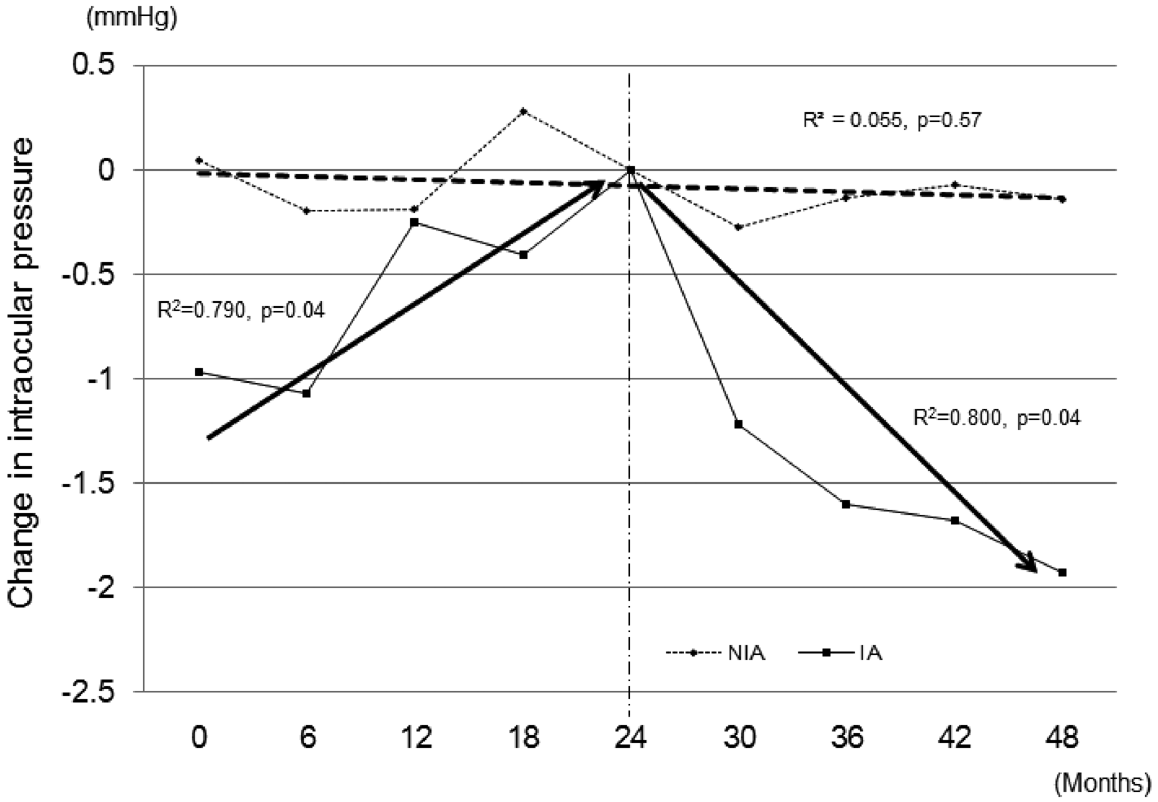
Internet Access and Intraocular Pressure Changes

The IOP values in the IA and NIA groups were mean 15.2 (SD 4.2) mmHg and mean 14.5 (SD 2.1) mmHg at the first examination, respectively. No significant IOP differences were observed between the IA and NIA groups (*P*=.31).

Differences in the IOP values between the first examination and the final examination in the IA group and NIA group were mean 0.2 (SD 1.4) mmHg (*P*=.32) and –0.8 (SD 3.0) mmHg (*P*=.10), respectively. The IA group showed a tendency toward IOP reduction during the study period.

In the IA group, the IOP of both eyes changed significantly between the pre-IA period and the post-IA period. Figure 6 shows the changing IOP profiles during the study period. In the IA group, the IOP was increasing before the start of Internet access (R^2=.790, *P*=.04) and decreased after the start of Internet access (R^2=.800, *P*=.04). By contrast, no significant IOP changes were observed during the study period in the NIA group (R^2=.055, *P*=.57). The ANCOVA showed that Internet access significantly shifted from an increasing IOP trend to a decreasing trend in the IA group (*P*=.002). The IOP increased by mean 0.1 (SD 0.1) mmHg/month during the pre-IA period, while the IOP decreased by mean 0.1 (SD 0.1) mmHg/month during the post-IA period.

Figure 6. A change in the intraocular pressure at 24 months after the start of the study was defined as zero mmHg. IA: Internet access group, NIA: non-Internet access group.



Change in Medication Possession Ratio

In the IA group, the MPRs before and after the start of Internet access were mean 82.3% (SD 30.7) and 91.1% (SD 40.3), respectively, which indicates a significant improvement (*P*=.03). In the NIA group, the MPRs before and after the start of Internet access were mean 84.0% (SD 28.5) and 82.9% (SD 31.1), respectively, which does not indicate significant improvement (*P*=.56).

Discussion

Principal Findings

The utility of patient education and medical information services has been previously discussed [7,8]. Despite multiple efforts, however, there is insufficient evidence supporting the utility of patient education and medical information services in glaucoma treatment [8,24,25]. The current study showed that allowing glaucoma patients to self-browse their clinical data contributed to two favorable effects: using less antiglaucoma ophthalmic solution and improving the effectiveness of the medication in reducing IOP.

It was common for the patients to increase their medication use during the treatment period. Indeed, 17.8% (16/90) of the NIA patients increased their glaucoma medications during the study period, while only 8.6% (7/81) of the IA patients did. In addition, 13.6% (11/81) of the IA patients reduced their medications during the study period. Notably, among the patients whose antiglaucoma ophthalmic solution use did not change during the study period, the Internet access group shifted from an increasing IOP trend to a decreasing trend. It is possible that the amount of antiglaucoma ophthalmic solutions used was increased in the patients who did not start Internet access. It is unclear whether the current outcomes were the direct results of Internet access or were due to other mechanisms, such as improved patient-physician communication during office visits. Improving adherence may contribute to these outcomes, although we were unable to monitor changes in adherence using subjective and quantitative methods, such as electronic monitoring systems, which are not available in Japan.

The current study showed that Internet access significantly improved the MPR among patients whose medication was not changed during the study and that Internet access may contribute to a reduction in IOP over time. The MPR in the current study is higher than that in Friedman's report [2]. A possible explanation for this difference could be that the understanding of glaucoma and the motivation for glaucoma treatment may have been stronger for the patients in this study than those in the previous study. In the present study, patients had been treated for glaucoma before their enrollment and were willing to use the GSS.

Health literacy is an important consideration in using electronic PHRs appropriately. We examined glaucoma literacy by administering a questionnaire to another set of patients who had been registered with the GSS for more than 2 years [26]. The patients who had been self-checking their registered data for more than 1 year exhibited a much better understanding of aspects of their glaucoma, such as their glaucoma severity, IOP values, and medications, than did the patients who had been provided their registered data by the ophthalmologists (see Figures 7 and 8). The IA patients in the current study may have improved their eye health literacy by checking their data through the Internet.

Although the current study demonstrated the effects of Internet access on glaucoma treatment, these effects are not consistent with those observed in previous studies. In their review, Tenforde et al [19] concluded that despite PHR's potential to improve chronic disease management and patient outcomes, the evidence supporting the clinical value of PHR remains limited. Many of the previous studies have enrolled patients with diabetes mellitus or other systemic diseases in which interventions may influence clinical outcomes through complicated processes. By contrast, glaucoma is a disease in which proper medication is the only intervention proven to influence IOP control. This difference could be a possible explanation for inconsistencies with previous studies.

We used a prospective, randomized study design, but the NIA patients were slightly older than the IA patients. The effect of age on adherence is controversial. Dietlein et al have reported that adherence to therapy with antiglaucoma ophthalmic solutions deteriorated with age [27,28], while other previous studies have reported that younger patients have poor adherence [29,30]. In the present study, no significant differences were found in the MD value, the number of antiglaucoma ophthalmic solutions used, or the BCVA value, suggesting that there were no marked differences in glaucoma severity between the two groups. Although we eliminated the patients who met the exclusion criteria, no significant differences were found in the reasons for loss to follow up between the two groups. Altogether, the influence of any factors, except for Internet access, between the two groups may have been limited.

Figure 7. Comparison of the rates of glaucoma patients' understanding of their target intraocular pressure (*$P<.001$, Fisher exact test).

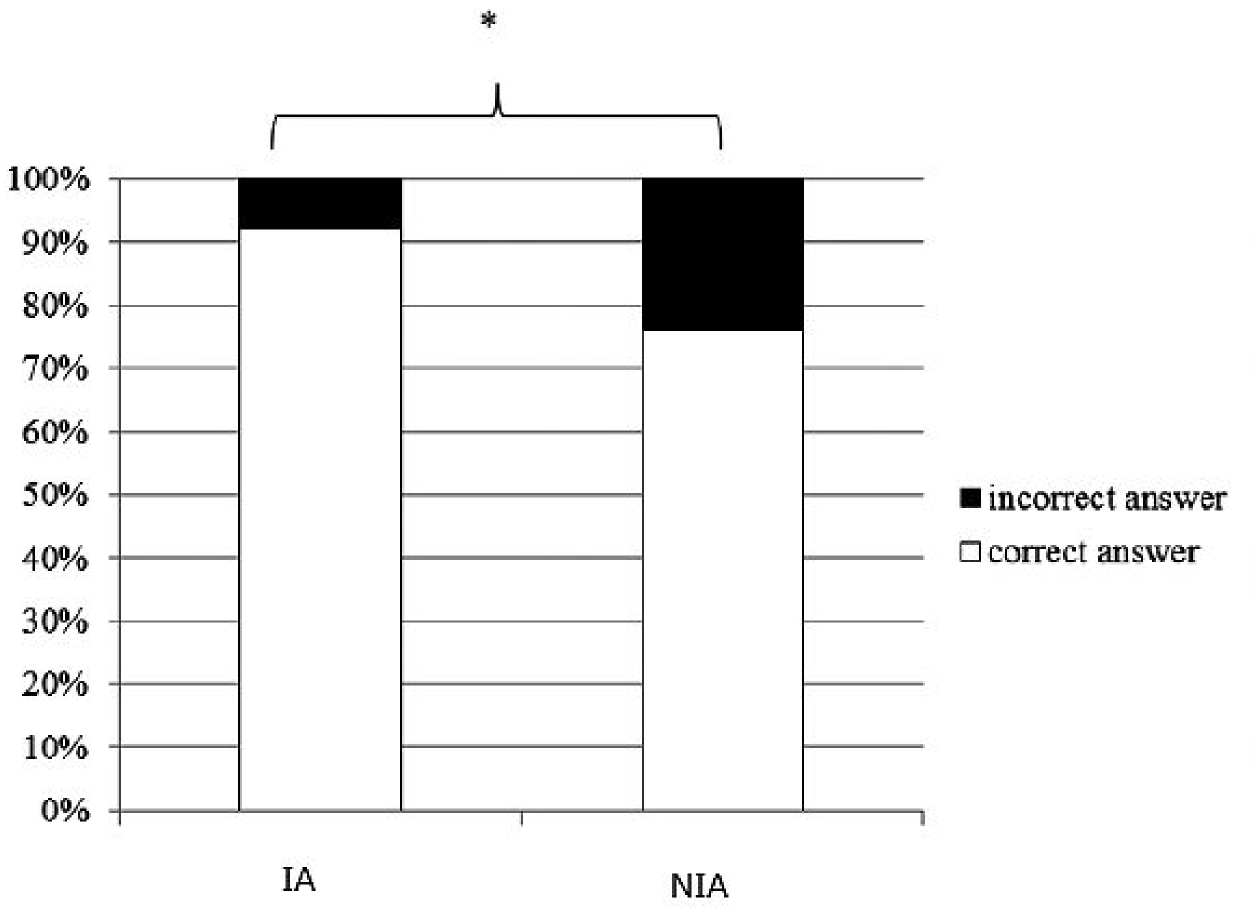
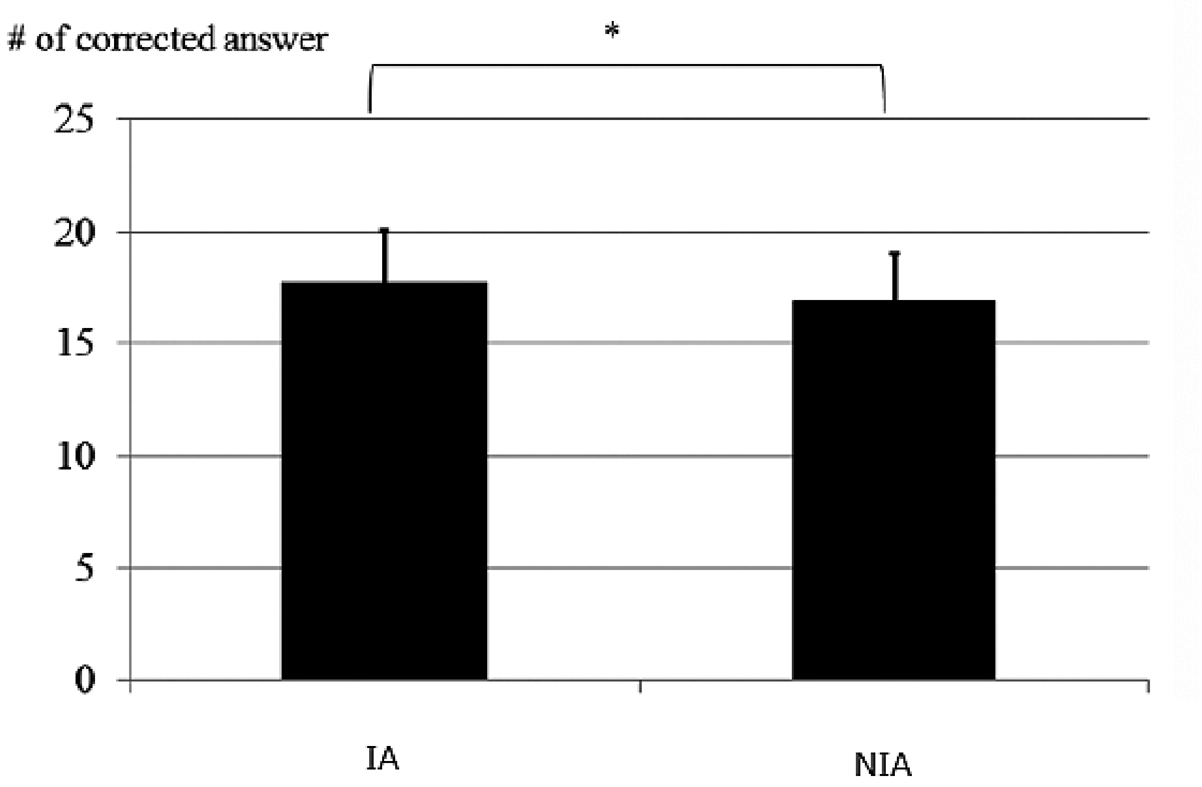


Figure 8. Comparison of number of correct answers to questions regarding glaucoma (*P=.007, Wilcoxon signed-rank test, RE).



Limitations

The patients in this study were GSS users. Given that many patients cannot use a personal computer or the Internet due to their age, poverty, or other reasons, it is possible that the patients in this study are not representative of general glaucoma patients. The present study did not examine how frequently the patients accessed the GSS. Therefore, further investigations should be performed to verify the relationship between the frequency of use and the improvements in glaucoma treatment associated with using the system.

Conclusion

The current study confirmed that patients' understanding of their glaucoma care status may play an important role in better management.

Acknowledgments

The authors thank Ms Kazumi Okudaira and Ms Kunie Ishihara who contributed greatly to the operation and management of the GSS. The authors are grateful to Dr Harmish Fraser for manuscript editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH Checklist V1.6.2 [31].

[PDF File (Adobe PDF File), 2MB - jmir_v16i1e15_app1.pdf]

Multimedia Appendix 2

Days of supplied medication for antiglaucoma ophthalmic solutions used.

[TIF File, 97KB - jmir_v16i1e15_app2.tif]

Multimedia Appendix 3

Antiglaucoma ophthalmic solutions used (1).

[TIF File, 97KB - jmir_v16i1e15_app3.tif]

Multimedia Appendix 4

Antiglaucoma ophthalmic solutions used (2).

[TIF File, 111KB - jmir_v16i1e15_app4.tif]

References

1. Quigley HA. Glaucoma. Lancet 2011 Apr 16;377(9774):1367-1377. [doi: 10.1016/S0140-6736(10)61423-7] [Medline: 21453963]
2. Friedman DS, Quigley HA, Gelb L, Tan J, Margolis J, Shah SN, et al. Using pharmacy claims data to study adherence to glaucoma medications: methodology and findings of the Glaucoma Adherence and Persistency Study (GAPS). Invest Ophthalmol Vis Sci 2007 Nov;48(11):5052-5057 [FREE Full text] [doi: 10.1167/iovs.07-0290] [Medline: 17962457]
3. Vanelli M, Pedan A, Liu N, Hoar J, Messier D, Kiarsis K. The role of patient inexperience in medication discontinuation: a retrospective analysis of medication nonpersistence in seven chronic illnesses. Clin Ther 2009 Nov;31(11):2628-2652. [doi: 10.1016/j.clinthera.2009.11.028] [Medline: 20110007]
4. Friedman DS, Hahn SR, Gelb L, Tan J, Shah SN, Kim EE, et al. Doctor-patient communication, health-related beliefs, and adherence in glaucoma results from the Glaucoma Adherence and Persistency Study. Ophthalmology 2008 Aug;115(8):1320-7, 1327.e1. [doi: 10.1016/j.ophtha.2007.11.023] [Medline: 18321582]
5. Juzych MS, Randhawa S, Shukairy A, Kaushal P, Gupta A, Shalauta N. Functional health literacy in patients with glaucoma in urban settings. Arch Ophthalmol 2008 May;126(5):718-724. [doi: 10.1001/archopht.126.5.718] [Medline: 18474786]
6. Muir KW, Ventura A, Stinnett SS, Enfiedjian A, Allingham RR, Lee PP. The influence of health literacy level on an educational intervention to improve glaucoma medication adherence. Patient Educ Couns 2012 May;87(2):160-164 [FREE Full text] [doi: 10.1016/j.pec.2011.09.009] [Medline: 22000272]
7. Gray TA, Fenerty C, Harper R, Lee A, Spencer AF, Campbell M, et al. Preliminary survey of educational support for patients prescribed ocular hypotensive therapy. Eye (Lond) 2010 Dec;24(12):1777-1786. [doi: 10.1038/eye.2010.121] [Medline: 20829888]
8. Hahn SR, Friedman DS, Quigley HA, Kotak S, Kim E, Onofrey M, et al. Effect of patient-centered communication training on discussion and detection of nonadherence in glaucoma. Ophthalmology 2010 Jul;117(7):1339-47.e6. [doi: 10.1016/j.ophtha.2009.11.026] [Medline: 20207417]
9. Kharod BV, Johnson PB, Nesti HA, Rhee DJ. Effect of written instructions on accuracy of self-reporting medication regimen in glaucoma patients. J Glaucoma 2006 Jun;15(3):244-247. [doi: 10.1097/01.ijg.0000212213.18018.8f] [Medline: 16778648]
10. Okeke CO, Quigley HA, Jampel HD, Ying GS, Plyler RJ, Jiang Y, et al. Interventions improve poor adherence with once daily glaucoma medications in electronically monitored patients. Ophthalmology 2009 Dec;116(12):2286-2293 [FREE Full text] [doi: 10.1016/j.ophtha.2009.05.026] [Medline: 19815286]
11. Rees G, Leong O, Crowston JG, Lamoureux EL. Intentional and unintentional nonadherence to ocular hypotensive treatment in patients with glaucoma. Ophthalmology 2010 May;117(5):903-908. [doi: 10.1016/j.ophtha.2009.10.038] [Medline: 20153902]
12. Lunnela J, Kääriäinen M, Kyngäs H. The views of compliant glaucoma patients on counselling and social support. Scand J Caring Sci 2010 Sep;24(3):490-498. [doi: 10.1111/j.1471-6712.2009.00739.x] [Medline: 20210901]
13. Goudswaard AN, Stolk RP, Zuithoff NPA, de Valk HW, Rutten GE. Long-term effects of self-management education for patients with Type 2 diabetes taking maximal oral hypoglycaemic therapy: a randomized trial in primary care. Diabet Med 2004 May;21(5):491-496. [doi: 10.1111/j.1464-5491.2004.01153.x] [Medline: 15089797]
14. Pladevall M, Brotons C, Gabriel R, Arnau A, Suarez C, de la Figuera M, Writing Committee on behalf of the COM99 Study Group. Multicenter cluster-randomized trial of a multifactorial intervention to improve antihypertensive medication adherence and blood pressure control among patients at high cardiovascular risk (the COM99 study). Circulation 2010 Sep 21;122(12):1183-1191 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.892778] [Medline: 20823391]
15. Fernandez-Luque L, Karlsen R, Bonander J. Review of extracting information from the Social Web for health personalization. J Med Internet Res 2011;13(1):e15 [FREE Full text] [doi: 10.2196/jmir.1432] [Medline: 21278049]
16. Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med 2008 Apr 17;358(16):1732-1737. [doi: 10.1056/NEJMsb0800220] [Medline: 18420506]
17. Schargus M, Grehn F, Glaucocard Workgroup. The European Glaucoma Society Glaucocard project: improved digital documentation of medical data for glaucoma patients based on standardized structured international datasets. Graefes Arch Clin Exp Ophthalmol 2008 Dec;246(12):1757-1768. [doi: 10.1007/s00417-008-0929-x] [Medline: 18766366]
18. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee JC, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
19. Tenforde M, Jain A, Hickner J. The value of personal health records for chronic disease management: what do we know? Fam Med 2011 May;43(5):351-354 [FREE Full text] [Medline: 21557106]

20. Gray TA, Fenerty C, Harper R, Spencer AF, Campbell M, Henson DB, et al. Individualised patient care as an adjunct to standard care for promoting adherence to ocular hypotensive therapy: an exploratory randomised controlled trial. Eye (Lond) 2012 Mar;26(3):407-417 [FREE Full text] [doi: 10.1038/eye.2011.269] [Medline: 22094303]
21. Sikka R, Xia F, Aubert RE. Estimating medication persistency using administrative claims data. Am J Manag Care 2005 Jul;11(7):449-457 [FREE Full text] [Medline: 16044982]
22. Andrade SE, Kahler KH, Frech F, Chan KA. Methods for evaluation of medication adherence and persistence using automated databases. Pharmacoepidemiol Drug Saf 2006 Aug;15(8):565-74; discussion 575. [doi: 10.1002/pds.1230] [Medline: 16514590]
23. Fiscella RG, Green A, Patuszynski DH, Wilensky J. Medical therapy cost considerations for glaucoma. Am J Ophthalmol 2003 Jul;136(1):18-25. [Medline: 12834665]
24. Amro R, Cox CL, Waddington K, Siriwardena D. Glaucoma expert patient programme and ocular hypotensive treatment. Br J Nurs 2010;19(20):1287-1292. [Medline: 21072015]
25. Watkinson S. Improving care of chronic open angle glaucoma. Nurs Older People 2010 Oct;22(8):18-23. [doi: 10.7748/nop2010.10.22.8.18.c7993] [Medline: 21049865]
26. Kashiwagi K. Effects of Web-based medical data providing on disease literacy of glaucoma patients. Effects of Web-based medical data providing on disease literacy of glaucoma patients J Jpn Telemed Telecare 2011;7(1):30-34.
27. Dietlein TS, Jordan JF, Lüke C, Schild A, Dinslage S, Krieglstein GK. Self-application of single-use eyedrop containers in an elderly population: comparisons with standard eyedrop bottle and with younger patients. Acta Ophthalmol 2008 Dec;86(8):856-859. [doi: 10.1111/j.1755-3768.2007.01155.x] [Medline: 18494743]
28. Lacey J, Cate H, Broadway DC. Barriers to adherence with glaucoma medications: a qualitative research study. Eye (Lond) 2009 Apr;23(4):924-932. [doi: 10.1038/eye.2008.103] [Medline: 18437182]
29. Ashaye AO, Adeoye AO. Characteristics of patients who drop out from a glaucoma clinic. J Glaucoma 2008;17(3):227-232. [doi: 10.1097/IJG.0b013e31815768b3] [Medline: 18414110]
30. Ngan R, Lam DL, Mudumbai RC, Chen PP. Risk factors for noncompliance with follow-up among normal-tension glaucoma suspects. Am J Ophthalmol 2007 Aug;144(2):310-311. [doi: 10.1016/j.ajo.2007.04.005] [Medline: 17659968]
31. 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
BCVA: best-corrected visual acuity
CAI: carbonic anhydrase inhibitor
GSS: glaucoma care support system
HFA: Humphrey Field Analyzer
IA: Internet access
IOP: intraocular pressure
MD: mean deviation
MPR: medication possession ratio
NIA: non–Internet access
NTG: normal tension glaucoma
OH: ocular hypertension
PE: phacoemulcification
PHR: personalized health record
POAG: primary open angle glaucoma
TLE: trabeculectomy

Edited by G Eysenbach; submitted 27.06.13; peer-reviewed by S Berger, M Meglic, G Paré; comments to author 17.09.13; revised version received 07.10.13; accepted 05.12.13; published 15.01.14.

Please cite as:
Kashiwagi K, Tsukahara S
Impact of Patient Access to Internet Health Records on Glaucoma Medication: Randomized Controlled Trial
J Med Internet Res 2014;16(1):e15
URL: http://www.jmir.org/2014/1/e15/
doi:10.2196/jmir.2795
PMID:24429379

Original Paper

Sources of Information and Behavioral Patterns in Online Health Forums: Observational Study

Fabian Sudau[1], MSc; Tim Friede[2], PhD; Jens Grabowski[1], PhD; Janka Koschack[3], PhD; Philip Makedonski[1], MSc; Wolfgang Himmel[3], PhD

[1]Institute of Computer Science, Georg-August-University Göttingen, Göttingen, Germany
[2]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[3]Department of General Practice, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Fabian Sudau, MSc
Institute of Computer Science
Georg-August-University Göttingen
Goldschmidtstraße 7
Göttingen, 37077
Germany
Phone: 49 551 39 172022
Fax: 49 551 39 14415
Email: fabian.sudau@cs.uni-goettingen.de

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2014/1/e39

Abstract

Background: Increasing numbers of patients are raising their voice in online forums. This shift is welcome as an act of patient autonomy, reflected in the term “expert patient”. At the same time, there is considerable concern that patients can be easily misguided by pseudoscientific research and debate. Little is known about the sources of information used in health-related online forums, how users apply this information, and how they behave in such forums.

Objective: The intent of the study was to identify (1) the sources of information used in online health-related forums, and (2) the roles and behavior of active forum visitors in introducing and disseminating this information.

Methods: This observational study used the largest German multiple sclerosis (MS) online forum as a database, analyzing the user debate about the recently proposed and controversial Chronic Cerebrospinal Venous Insufficiency (CCSVI) hypothesis. After extracting all posts and then filtering relevant CCSVI posts between 01 January 2008 and 17 August 2012, we first identified hyperlinks to scientific publications and other information sources used or referenced in the posts. Employing *k*-means clustering, we then analyzed the users’ preference for sources of information and their general posting habits.

Results: Of 139,912 posts from 11,997 threads, 8628 posts discussed or at least mentioned CCSVI. We detected hyperlinks pointing to CCSVI-related scientific publications in 31 posts. In contrast, 2829 different URLs were posted to the forum, most frequently referring to social media, such as YouTube or Facebook. We identified a total of 6 different roles of hyperlink posters including Social Media Fans, Organization Followers, and Balanced Source Users. Apart from the large and nonspecific residual category of the “average user”, several specific behavior patterns were identified, such as the small but relevant groups of CCSVI-Focused Responders or CCSVI Activators.

Conclusions: The bulk of the observed contributions were not based on scientific results, but on various social media sources. These sources seem to contain mostly opinions and personal experience. A small group of people with distinct behavioral patterns played a core role in fuelling the discussion about CCSVI.

(J Med Internet Res 2014;16(1):e10) doi:10.2196/jmir.2875

KEYWORDS

Internet utilization; information dissemination; data mining; social media; social networks; multiple sclerosis; CCSVI

Introduction

In the past few decades, we have witnessed a powerful movement toward an active, self-managing, and responsible patient, coined the “expert patient” [1,2]. A key element in this process has been unlimited access to and intelligent use of health-related information, particularly that which is widely available on websites and online forums on the Internet and in online social media [3-6]. This movement has consequences for the traditional way of information dissemination. Today, laypeople, self-support groups, patient advocates, and other stakeholders can raise their voice and can even influence both public and scientific debates. This shift is welcome as an act of patient autonomy and freedom to seek alternatives to the standard therapeutic regimens and the paternalistic doctor-patient relationship. At the same time, there is considerable concern that patients can be easily misguided by pseudoscientific research, because typically they do not have the expertise to assess the reliability of scientific information and because of their circumstances may often accept any suggested solution no matter how unlikely and unrealistic it may seem. Nettleton et al [7] call for a strictly empirical analysis to examine people’s accounts of their use of online health resources.

Most studies in this area have investigated how often people use the Internet for retrieving health information [8], how they access health information on the Internet [9], which factors are important for laypeople when using Internet resources for health issues [10], and how to assess the quality of health information for laypeople on the Internet [11]. There is another area of research that seems promising—social network analysis [12]. Health-related online communities, as one form of social network, are thought to develop their own quasi-professional knowledge of their health conditions [12] and to personalize support [13]. Following applications in marketing, research has investigated diffusion processes of successful new products with the aim of targeting “influential” members of a network [14]. In the medical area, for example, a recent study showed how a social network of parents influenced decision making on vaccination in an unfavorable manner [15]. Similar concerns about misinformation via Twitter arose about flu treatment requiring antibiotics [16].

However, we still know very little about what mechanisms of information dissemination are effective as well as what sources of information people in online forums rely on, how they form their opinions, and how they act. A better understanding of these mechanisms may help to assess their influence on laypeople and to forecast the benefits and dangers of these new forms of information dissemination and exchange.

One promising area for such research is the recently proposed Chronic Cerebrospinal Venous Insufficiency (CCSVI) hypothesis in multiple sclerosis (MS) and its repercussions in patient communities. In short, this hypothesis was first proposed by Paolo Zamboni [17], who suggested that obstruction to venous drainage in the neck and spinal cord [17], termed chronic cerebrospinal venous insufficiency, was linked to MS [18,19]. Although the association between MS and sonographic features of CCSVI is variable [18], some institutions have even begun to offer angioplasty and endovenous stenting of CCSVI, often referred to commercially as “The Liberation Procedure” [20]. The intensity of the CCSVI debate reached such a point that the Society for Interventional Radiology released a position statement regarding endovascular management of CCSVI [21]; the MS Society of Canada funded a study of the prevalence of extracranial venous narrowing, which found evidence neither for a high prevalence of CCSVI nor for its causal relationship to MS [22]. Several studies report a wave of complications following venous stenting and angioplasty [23,24]. The CCSVI hypothesis is also fiercely debated in online patient communities, such as the online forum of the German MS Society (DMSG, Deutsche Multiple Sklerose Gesellschaft) [25] and the United Kingdom’s MS Society online forum [26], as well as numerous other dedicated websites and blogs on the Internet [27,28]. It has even found its way to the popular video-sharing website YouTube [29], with more than 23,800 videos posted up to July 2013, one of them with more than 200,000 views. The CCSVI waves seem to have calmed down and some consider the hypothesis—in a retrospective view of the CCSVI hype—as a waste of valuable time, money, and intellectual energy [30]; others emphasize that the debate has stimulated the need for studies that should contribute to a better understanding of the function and role of the extracranial venous system [31].

Before we can make a statement on whether and how this multitude of information sources and opinions may contribute to the enlightenment of some participants in the debate or the confusion of others, we need to know more about the sources of information used in online health forums and how users and participants use this information, including their different roles and contribution behavior in such forums. To examine these questions, we can build on a UK study on online self-harm discussion forums [32]. Using “social networking metrics”, the authors found different types of online discussion participants and roles: the Caretaker (being always watchful, participating to some degree but not initiating many new threads in discussions), the Butterfly (logging on very frequently with quick looks around and then logging off again), the Discussant (initiating many discussion threads), and the Here for You (initiating few discussion threads but posting the most comments).

Our observational study takes advantage of free access to a large German online forum related to multiple sclerosis, with the aim of identifying (1) sources of information used in online health forums, and (2) roles and patterns of behavior of people actively engaging in the forum in introducing and disseminating this information.

Methods

Design

In this observational study, we extracted the content from an online health forum, using a custom implementation of a Web crawler, with the aim of collecting a large database of discussions from an online health forum. Furthermore, we used an Information Retrieval algorithm (specifically designed and implemented for this particular task) to identify a comprehensive sample of posts dealing with CCSVI.

Database and Retrieval of Relevant Posts

The database for the study comprised contributions posted to the online forum of the Deutsche Multiple Sklerose Gesellschaft (DMSG, German Multiple Sclerosis Society) [25]. On its website, the DMSG presents itself as a non-profit stakeholder of MS patients and their families, founded by clinical and scientific experts in MS in 1952. It is a registered charity with 16 regional branches and over 900 community contact groups. Among other things, the DMSG provides on its website two different kinds of freely accessible forums: one expert forum with time-limited chats between experts and users about different issues (eg, cognitive deficits in MS or pregnancy in MS). The other forum is unstructured, not moderated, and open for anonymous registration. It is targeted at laypeople, mostly people with MS. The forum consists of threads, which in turn contain sequences of posts. These posts can contain hyperlinks and can cite any number of previous posts. A screenshot of such a post is shown in Figure 1.

Between 01 January 2008 and 17 August 2012, all 139,912 posts from 11,997 threads were extracted. Because the forum is about MS in general, only a fraction of the extracted posts were expected to be about CCSVI. Preliminary analysis showed that the assumption of "one thread discusses one topic" does not hold in the observed forum. Instead, users tended to deviate from the original topic as time progressed. Therefore, a custom Information Retrieval algorithm was developed to classify individual posts as either relevant ("discussing CCSVI at least partially") or irrelevant. For details on the algorithm design, training, and evaluation, see Multimedia Appendix 1. The algorithm identified 8628 posts as relevant, which yields a distinction important for further analysis steps.

Figure 1. A screenshot of a forum post.

Das DMSG-Forum · **CCSVI: Venenverengung im Kopfbereich keineswegs MS-spezifisch!**

MazSha

30.10.2013, 15:08

Eine aktuelle Studie aus dem renommierten „Lancet" überführt bisherige Messmethoden bei venöser Insuffizienz als völlig ungeeignet und findet überraschende Ergebnisse zum Vorkommen dieses vermeintlichen Risikofaktors.

Zusammenfassung der Studie:
http://www.navigator-medizin.de/...hema-ccsvi.html

Kommentar des Ärzte-Teams von Navigator Medizin:
http://www.navigator-medizin.de/...lerose_348.html

Gruß MazSha

Search for Scientific Publications

Because the term "expert patient" implies intelligent use of scientific information, we aimed to assess to what degree the use of scientific sources was present in the forum. Users occasionally included hyperlinks in their posts and these links referred to content the users based their opinions on. We analyzed which of these links were defined references to scientific papers in order to get an overview of the kind of papers cited and the temporal citation patterns. Two steps were necessary for this identification process.

First, we generated a presumably exhaustive list of publications dealing with CCSVI. A citation network starting from Zamboni's original publication and using the CiteXplore Web service was constructed [33]. CiteXplore, which is an interface to the PubMed search engine, was used due to easy accessibility of citations. These publications were then merged with a second list that was obtained by a search for "CCSVI" in the PubMed database via the Entrez interface [34]. The merging algorithm removed duplicate publications as identified by their PMID identification number. Our final publication list does not include publications that deal with CCSVI but do not include the CCSVI acronym, or had not shown up in the citation network. We assume this number of publications to be low and prefer our method over manual approaches.

Second, a program fetched every hyperlink (also those in “irrelevant” posts) from the corpus, extracted the textual content from the referenced webpage or PDF document and searched it for titles or publication IDs from the publication list. In the case of a hit, one of the authors verified whether one of the publications was indeed referenced and, if so, which one. Every match was also classified as either a direct reference or an indirect reference. An indirect reference in this context was regarded as a resource that solely discussed or explained a certain publication, not including other work based on the publication. A direct reference linked to the publication itself.

Search for Other Web Resources Used and Their Classification

Apart from searching for scientific information sources in the posts, we also strived to identify other information sources used or referenced in the posts. In order to obtain an overview of the wide spectrum of referenced websites, we defined a classification scheme. First, we reduced every URL found in the reduced corpus to the basic domain part of the URL (ie, only “domainname.com” was used—if the URL included additional content after the domain name, such as directories, folders, webpages, file extensions, that content was removed from the URL). Second, we classified the remaining domains into the 8 classes shown in Table 1. These classes were defined based on content type and authorship provided under the respective domain. A plot was then generated showing the number of URLs from each class posted per month.

Table 1. Primary domain classes.

Organization	Meant in a broader sense, including foundations, associations, and unions. These are sometimes professional and often promote some kind of agenda.
Commerce	Private business selling products or services that do not include treatment.
News	Commercial news providers.
Other	Various content not fitting into the other classes.
Personal	Static content from a single person.
Scientific	Sources of scientific work and knowledge including Wikipedia. We included the latter in this class, because its reliability was established in [35]. We believe that Wikipedia, in contrast to sources from other classes, is perceived as a factual source by most of the users.
Social	Social media websites revolving around communication and user-generated content.
Health care providers	Doctors’ offices, clinics, Q&A by professionals. Not limited to Multiple Sclerosis.

User Behavior

To characterize user behavior, we tried to identify distinct behavior patterns. Since nothing was known in advance about the behavior patterns of forum users, we employed a method of exploratory data analysis to reveal possible patterns. A clustering algorithm groups users based on their similarity according to a set of predefined features. We thus wanted to define two separate feature sets with the aim of describing two different aspects of user behavior and revealing patterns in these features through clustering. We employed the popular *k*-means clustering algorithm (originally proposed in [36]) to group the data vectors representing users together based on how close they were to each other in the Euclidean hyperspace. The algorithm was chosen due to its simplicity and widespread use. Users within one cluster were thus assumed to display similar behavioral patterns, different from those patterns prevalent in other clusters. We defined the user cluster names based on manual inspection of descriptive cluster statistics.

Two behavioral aspects in particular were analyzed in detail by separate clusterings: (1) the preference for discussed sources of information, and (2) the general contribution behavior or posting habits. In the first clustering, we focused on the hyperlinks from each of the 8 domain classes. A user was represented by a vector in 8-dimensional space: for example, a value of 3 for the 2nd dimension meant the user had posted 3 hyperlinks from the domain class “Organization”. The second clustering focused on 9 quantitative features describing what and how a user had posted. The features (measures) were either taken from similar approaches discussed in the literature [37] or defined according to metadata that has not been used previously. Specifically, the literature does not use measures based on the distinction between on-topic and off-topic talk and does not make use of possibly insightful metadata such as hyperlinks or citations. The features and the reasoning behind them are described in Table 2. All of them are defined over the entire contribution period.

In both cases, the *k*-means clustering algorithm was used in the form of a custom implementation. We employed a heuristic initialization step to compensate for adverse effects of bad initial centroid placement. The clustering terminated when no more cluster memberships changed. For details on the employed algorithm, see Multimedia Appendix 2.

In the first clustering, we had to compensate for different general activity levels of users because we wanted to group the users according to their information source preferences only. We divided every vector by its Euclidean norm in order to obtain unit vectors showing only “taste” (preference), but not “activity”. In the second clustering, the different features had different scales. For instance, users often showed several hundred days of activity, but the fraction of their initiated threads can by definition not exceed 1. We thus performed a *z*-score normalization of the data before the second clustering. This means we modified every feature value of every user vector as follows. First, we subtracted the feature mean (over all users) and then we divided by feature standard deviation. The *k*-means algorithm requires that the number of clusters (*k*) is specified in advance. Preliminary experiments with different values of *k* showed that *k*=6 was a good choice in both cases, judged by manual inspection of internal cluster evaluation metrics and resulting cluster sizes.

We visualized the resulting clusters in radar charts [38], also known as spider charts or kiviat diagrams. A radar chart has a "spoke" for each feature; the data length of a spoke is proportional to the magnitude of the variable for the data point relative to the maximum magnitude of the variable across all data points so that multivariate observations with an arbitrary number of variables can be displayed and compared. Star-like figures indicate normalized feature means across the members of a cluster. Although it is difficult to compare lengths of different spokes visually, striking differences as well as commonalities between clusters can be captured easily and therefore the characteristics of the different clusters are thus easily comparable. We gave names to the clusters based on manual assessment of the radar charts and the tables of Multimedia Appendix 3. The definition of these names is based on which feature values "stand out" for a given cluster.

Table 2. Definition of behavior features.

Measure	Definition	Rationale
Average message length (from [39])	Average post content length in characters without counting references.	The message length is an indicator of the amount of effort that is put into a post by a user and it also tells us something about the discussion style of a user. Some users prefer elaborate, essay-like contributions while others use the forum in a more conversational way.
Average number of posts per day (from [32])	Average number of posts per day that a user made.	This is the most important activity feature of a user and it also provides an insight into the selectiveness of the user. A user with a high number of posts per day over a long time period can be expected to be a frequent visitor, who makes posts regardless of outside events.
Average number of references per post	Average number of unique references that are included in a post.	The feature describes the tendency of a user to bring new sources of information to the forum and may also describe the ability to support the stance of the user with evidence.
Average number of threads per day (from [32])	Average number of different threads a user posts to per day.	While this is also an activity feature, it provides an insight into the focus of interest a user has. A low value may indicate a preference to discuss only specific topics while a high value may indicate a preference to join any sort of discussion.
Days active (from [39])	Number of days between the first post and the last one.	The feature indicates the consistency of the contribution behavior and posting habits of a user and is an important piece of context information when interpreting the other features.
Fraction of posts that were cited	Fraction of the posts that have been cited at least once.	While it can only be assumed what users try to express when they use the citation function, the feature is expected to show the tendency to provoke direct responses from other forum participants.
Fraction of relevant posts	Fraction of the posts that were classified as relevant by the Information Retrieval algorithm.	This feature is a solid indicator of the user's interest in CCSVI[a]. While it cannot be inferred from this feature alone whether the user has a pro-CCSVI or anti-CCSVI stance, it seems plausible that users with a high interest in CCSVI believe in the hypothesis.
Fraction of initiated threads (from [37])	Fraction of the threads the user initiated based on the total number of threads the user contributed to.	This feature measures the tendency of a user to start discussions, which is often related to the introduction of new information to the forum.
Coverage of users in relevant parts per post	Number of users the user discussed CCSVI with divided by the total number of posts the user made. An uninterrupted sequence of relevant posts is regarded a single discussion. The users that co-occurred in these discussions are counted as discussion partners.	This feature can be described as the efficiency in opinion exchange about CCSVI.

[a] CCSVI: Chronic Cerebrospinal Venous Insufficiency

Results

Search for Scientific Publications

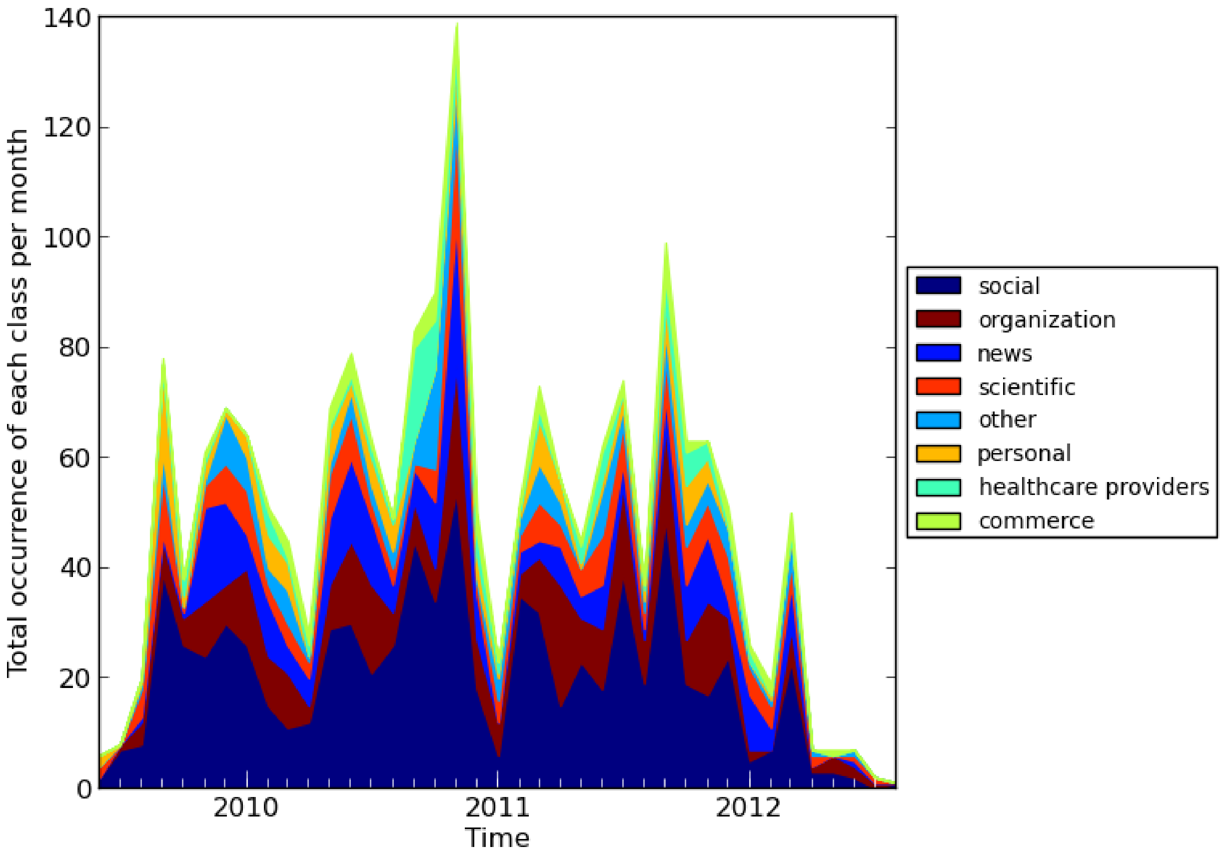
We detected hyperlinks pointing to CCSVI-related scientific publications in 31 posts. Multimedia Appendix 4 gives the 13 different publications referenced by the forum users. Each publication is shown in a separate area where the red star indicates the publication date and the green diamonds show dates where links to the publication were posted. Light green diamonds indicate indirect references. Interestingly, Zamboni’s original publication [17] was brought to the forum no later than two months after publication and referenced repeatedly, often indirectly. Another 4 publications in favor of the CCSVI hypothesis [40-43] were cited by September 2010. The position of the publications was identified manually. Judging from the referenced scientific publications alone and ignoring post content as well as other references, the period from July 2009 to September 2010 can thus be described as a “boom phase” of the CCSVI hypothesis in the forum. However, after September 2010, critical publications appeared and were brought to the forum. In fact, all except one of the referenced publications after September 2010 [44-50] strongly oppose CCSVI. At this time, the series of repeated references to Zamboni’s original publication stopped.

Search for Other Web Resources Used and Their Classification

Figure 2 shows how many hyperlinks of each domain class were posted each month, sorted by overall domain class popularity. At any given point in time, social media websites were the most widely used type of Web resource. Similar to the low number of referenced scientific publications, science-based resources were generally not used very often. About half of the posted hyperlinks from the domain class “scientific” refer to Wikipedia articles. Organization-related websites and news sites were the second and third most important ones.

The large differences in the total number of posted references per month correlate roughly with the total number of relevant posts. Interestingly, the highest peak (September 2010 - November 2010) was observed when the aforementioned phases shifted. The external events causing the other significant fluctuations are not known. However, when the total number of posted references rose from a given point in time to another, the change was typically reflected in all of the domain classes, which indicates a certain echo of external events equally affecting the different types of resources. The plot also shows how quickly the topic caught on in the layperson forum and that users seemed to have lost interest in the debate, as suggested by the few references posted in 2012.

Figure 2. Timeline of posted hyperlinks for each domain class.



User Behavior

We included only a fraction of the users in the clustering because we wanted to focus on those who took part in CCSVI discussions. Furthermore, a sufficient amount of information about each user was required. Therefore, we clustered only users who had posted at least 5 relevant hyperlinks, in the case of hyperlink use (first clustering). In the case of posting habits (second clustering), we included only users who had made at least 5 relevant posts. The filtering process is shown as a flow diagram in Figure 3 and resulting sets are shown in Figure 4. The fraction of users who were active enough for meaningful analysis is rather low, which is typical for online communities. Nearly two-thirds of the DMSG forum users posted only once.

The first clustering of the users into 6 groups revealed clusters shown in Figure 5. Roughly half of the users (29/64) can be described as Social Media Fans. Figure 6 shows the information sources preferred by members of each cluster. Social Media Fans, for example, prefer video-sharing websites (such as YouTube.com), Facebook pages, and blogs over more traditional sources. Balanced Source Users cite sources from different classes equally often, including scientific ones. Organization Followers mainly refer to content published by organizations; we also identified a group that uses sources that do not fit well into the classification scheme. Homepage Promoters post links to websites featuring static content authored by a single person. These traditional sites already existed in the early era of the Internet. Seekers of Healthcare discuss doctors and clinics. Users of Uncommon Sources focus on religion, esoterism, complementary or alternative medicine, or unrelated resources.

Clustering users, who had made at least 5 relevant posts, revealed the 6 groups shown in Figure 7. The cluster names were derived manually without a prespecified algorithm from the corresponding table in Multimedia Appendix 3 and from Figure 8, which shows the feature means normalized to a [0;1] range. About two-thirds of the users could only be described as “average”. This means that they do not stand out, but the characteristics of these users provide a baseline for comparison with the other user roles. Twenty-eight users were CCSVI-Focused Responders, who were active for less than a year on average. What defines them is the low level of posts per day, the low fraction of initiated threads, and the high fraction of CCSVI-related posts. Ten users were Highly Active Relational Posters, who show the highest level of posting activity (about 4 posts per day). They posted in lots of different threads, but rarely initiated them. Another 17 users are CCSVI Activators, who stand out due to their high fraction of initiated threads, their high percentage of CCSVI-related posts, and the fact that they included 3 times as many references as the average user. The 4 Sophisticated Contributors are known for making posts that are 3 times as long as those of average users and include 5 times more references. The remaining 4 Short-Lived CCSVI Spammers were active for a few days only and, during their short contribution period, created many posts about CCSVI. The posts were short and included few references.

Figure 3. Flowchart of the sampling procedure for clusterings.

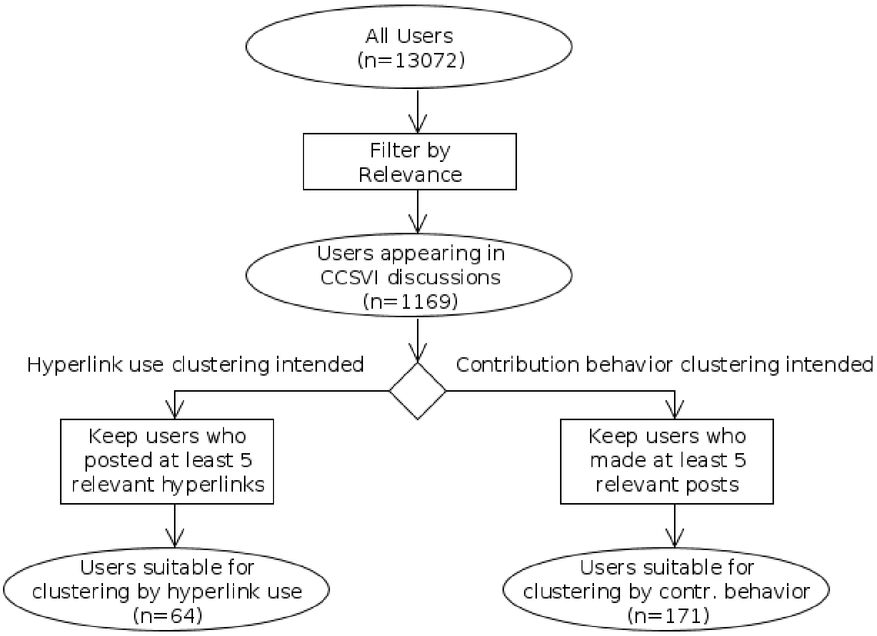


Figure 4. Venn diagram showing the user sets used in the clusterings.

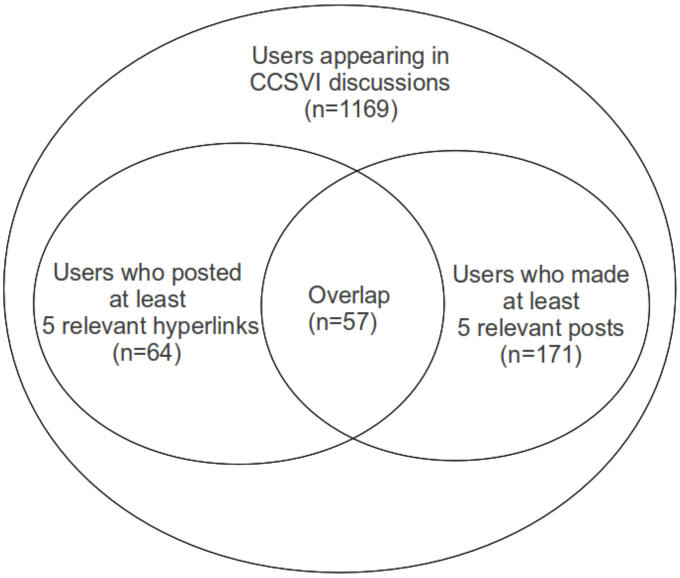


Figure 5. Reference use clusters with number of users in each cluster (n=64 included cases).

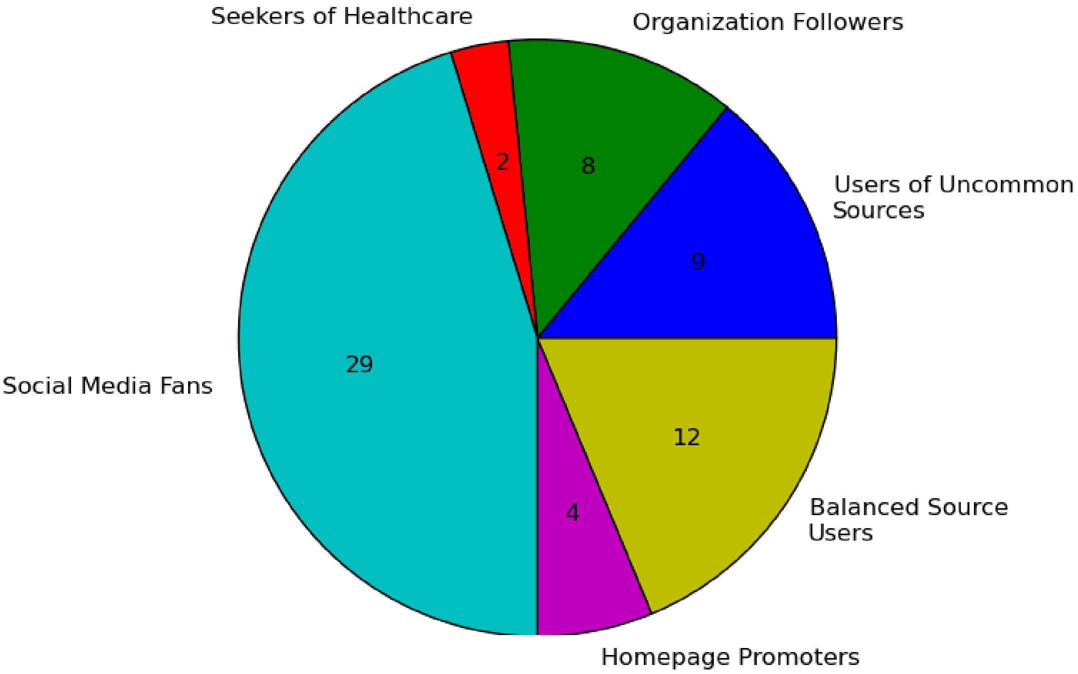


Figure 6. Radar chart showing aggregated domain class use of each cluster (the user vectors belonging to the cluster are summed up). Each cluster vector is a normalized to be a unit vector. The length of a spoke is proportional to the value it represents.

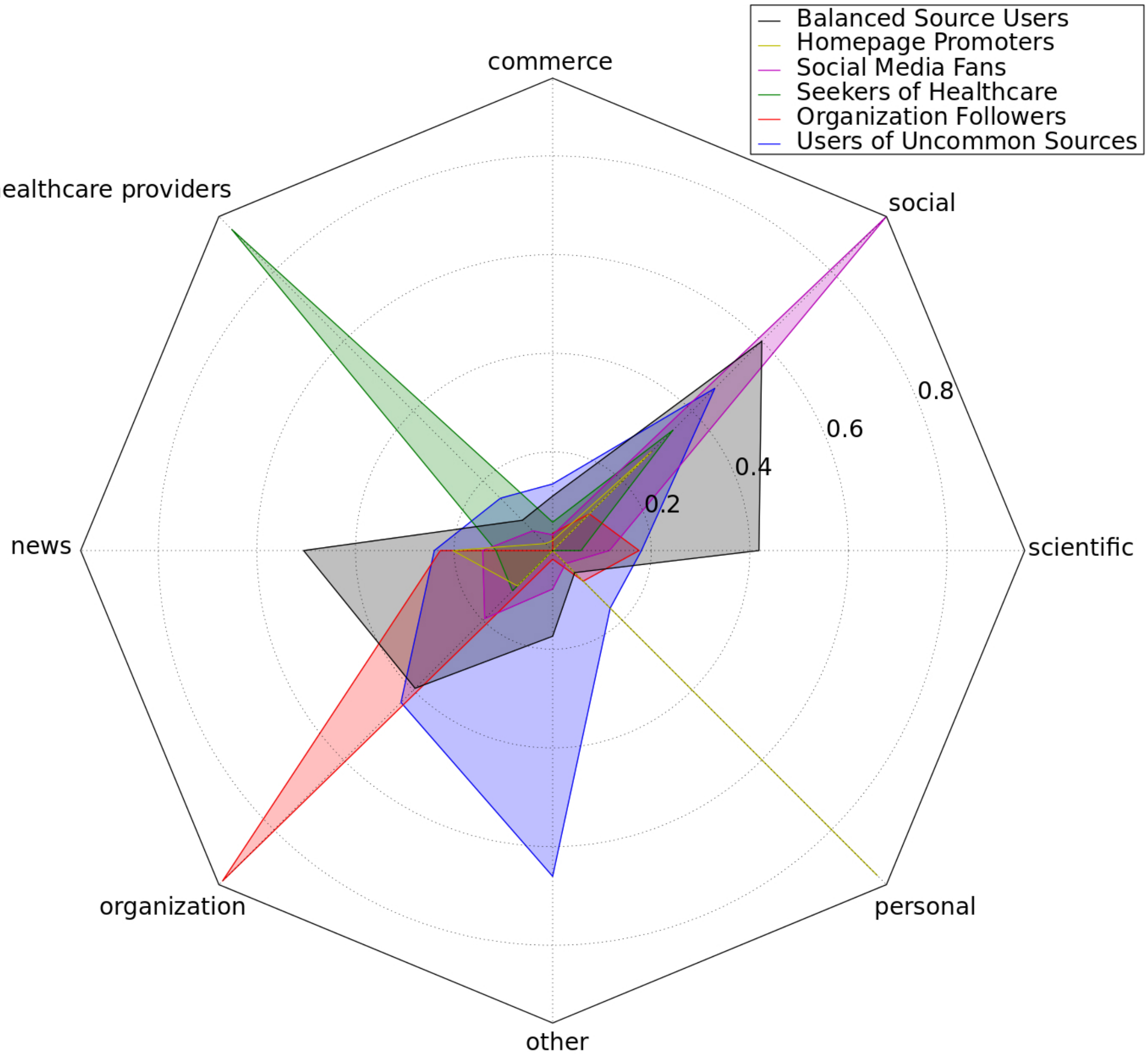


Figure 7. Posting behavior, according to the second clustering, with number of users in each cluster (n=171 included cases).

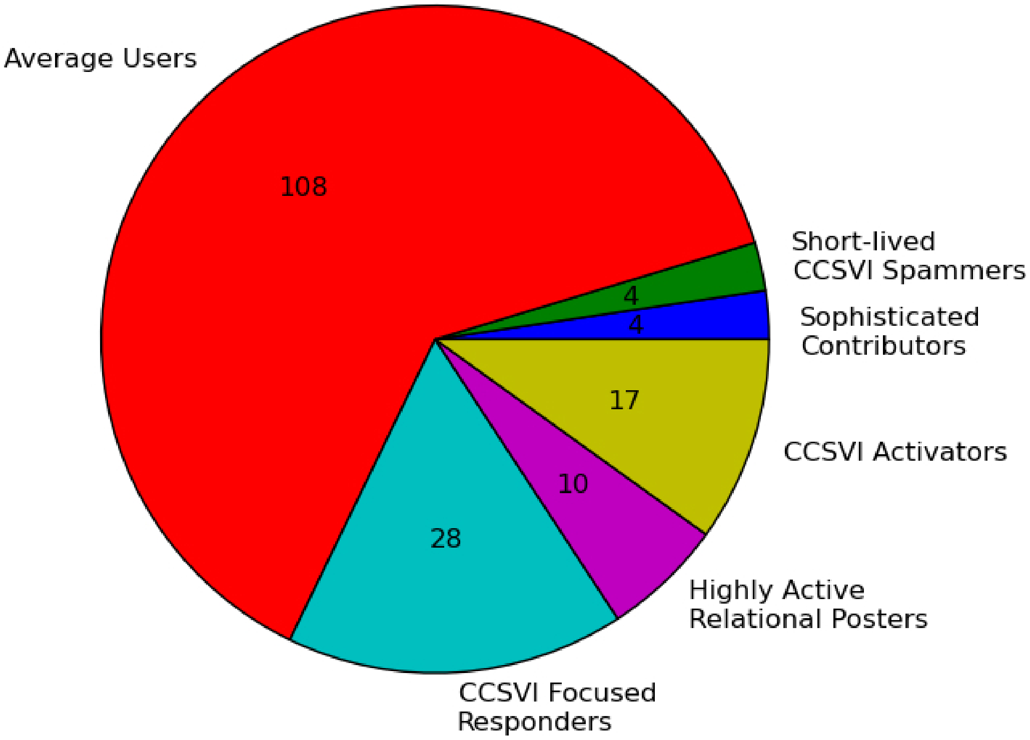
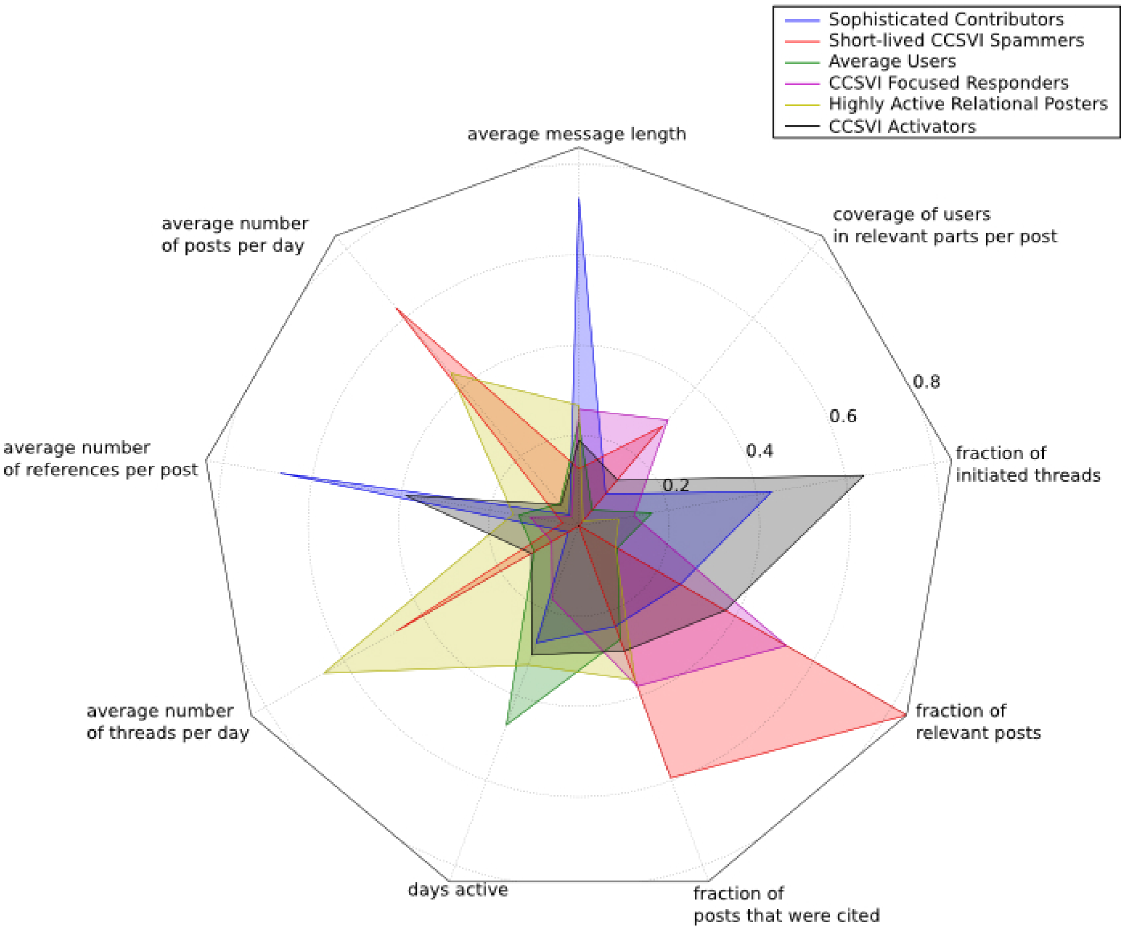


Figure 8. Radar chart showing feature means (overall users within a cluster) of the contribution behavior clusters. The means are min-max-normalized to a [0;1] range. The length of a spoke is proportional to the value it represents.



Discussion

Summary of Main Findings

The bulk of the observed contributions were not based on scientific results, but on various social media sources. These sources seem to contain mostly opinions and personal experience. A small group of people with distinct behavioral patterns played a core role in fuelling the discussion about CCSVI, as identified by their behavior. The identification of this group of people was an unintended consequence of our exploratory analysis technique. Our identification method is behavior driven and thus provides a viable alternative to the influence-based identification of so called “opinion leaders” in forums, as discussed in [51-54].

Meaning of the Results and Comparison With Literature

Scientific publications were brought to the forum at a “boom phase” of CCSVI discussion, followed by a phase of critical views, beginning September 2010 with the opponents of the CCSVI hypothesis getting the upper hand in the forum. Although scientific and lay discourse seem to go hand in hand, it is obvious that scientific publications and scientific sources such as Wikipedia played, in the end, only a minor role in the layperson forum. Instead, social media were the most important source of information. The nature of social media content varies, but we believe that social media are often about personal experiences and exchange of opinions. This is further illustrated by the reference use patterns we identified, such as Social Media Fans or Homepage Promoters. We would suggest characterizing the nature of this lay discourse more as an elementary discourse or an interdiscourse [55] than a special or scientific discourse.

Our 6 groups of posting behavior are based on a careful inspection of different characteristics and are similar to the participants in 5 online forums on self-harm [32]. The CCSVI-Focused Responders, characterized by a low level of posts per day, the low fraction of initiated threads, and the high fraction of CCSVI-related posts may compare to the Here For You user [32], who was very supportive in the self-harm forum. Discussants [32] may compare to our CCSVI Activators, who stand out due to their high fraction of initiated threads, their high percentage of CCSVI-related posts, and their introduction and placing of references. The Highly Active Relational Posters, who posted very actively, but rarely initiated threads, compare to Jones et al [32] Caretakers. The Short-Lived CCSVI Spammers remind us of Jones et al [32] Butterflies. The 4 Sophisticated Contributors remind us of researchers and the emergence of online expert patient groups [12]. The largest fraction of contributors could only be classified as “average”.

Only a small set of the involved users showed enough activity to be suitable for meaningful descriptions of their behavior. This is consistent with the common observation of significant participation inequality in social media. Typically, activity levels are characterized by the power law with about 1% of the users exercising the core influence on a community [56]. For example, studying the community structure of online diabetes forums, Chomutare et al [57] found very low user participation rates and suggested high levels of the few users who participated actively. About 37% (63/171) of the users participating in CCSVI discussions showed distinct patterns in their posting behavior: 28 CCSVI-Focused Responders seemed to wait for CCSVI discussions to come up and then contribute. Doing so, their post-wise efficiency of CCSVI-related information spread is the highest of all roles. This seems to be an interesting new aspect to the usually performed network analysis in online forums. In these analyses, knowledge of, and personal experience with, the disease play an important role in gaining central positions and becoming authorities [57,58]. Obviously, a good command of scientific sources of information may also be one characteristic of a group of influential figures.

The Highly Active Relational Posters are expected to be important community builders, as a substantial amount of personal “small talk” is attributed to them. Interestingly, a group of 17 people, the CCSVI Activators, played a core role in fuelling the discussion about CCSVI, because they often initiated threads about CCSVI and included many hyperlinks. While there is considerable concern that social media and Internet applications permit a minority of individuals to spread misinformation and damage useful interactions as recently discussed in the case of anti-vaccinationism [59], our results show that the CCSVI discussion in the MS forum follows the ups and downs of the scientific debate and does not promote dangerous practices or prevent novel technologies to a dangerous degree.

Some Sophisticated Contributors were identified, but these people did not participate in CCSVI discussions very often. Additionally, 4 very short-lived and CCSVI-focused accounts were identified. One possible explanation is that they were the temporarily-used alternative accounts of some users.

Strengths and Limitations of the Study

One major advantage of this study is its observational nature. Real-world data was observed in an unobtrusive way. We analyzed a public Internet forum, which was unstructured and unmoderated, over a 3-year period of CCSVI discussion. We thus avoided self-reporting biases and artificial setups. Furthermore, we applied a Machine Learning approach in order to shed some light on the complex nature of user interaction.

However, there are several limitations. There is no demographic data available for the forum users and it is even possible that some persons used different accounts. Furthermore, before 27 August 2010, users were able to choose their aliases freely for every individual contribution. Due to the lack of a log-in mechanism, it is possible that different individuals posted under the same name.

The identification of relevant content was non-trivial and did not have 100% accuracy, which resulted in a possibly biased database. The reduction of URLs to the basic domain was a simplification. When assessing user patterns, we had to deal with small sample sizes (N=64 and N=171). The clustering approach itself relied on several assumptions. We assumed that constant behavioral patterns exist, that we defined appropriated features to describe them, and that they are linearly separable in the feature space. The interpretation of the assigned roles is subjective, but based solely on the quantitative data documented in this study.

We had to decide how to identify scientific sources of information in the posts. To be on the safe side, we accepted only the posting of URLs with a link to scientific publications as a use of scientific publications. Of course, other users may have discussed scientific publications in a rather elaborate way without posting URLs. Moreover, publications are often hidden behind a paywall, which may make posting URLs unpopular. They are also written in English, which may pose a language barrier. Our approach underestimates the discussion of scientific publications in online health forums but is highly specific in identifying the introduction of scientific publications.

Our description of participants in this online health forum was based solely on "metrics", similar to the Jones et al study [32]. A full and reliable description of the participants' views would require an elaborated semantic analysis of their contributions.

Implications and Future Research

Scientific sources were by far less important than social media in the posts and forum discussions. While some of the uncovered evidence may indicate the successful propagation of scientific results into discussions among laypeople within an online health forum, scientific results represented by far a rather small fraction of the information sources that were discussed in the particular forum under study. Whether this is any indication of the rise of the "expert patient" remains the subject of further studies. Some of the participants in the forum, especially the Sophisticated Contributors, could be considered experts based on the nature of their contribution behavior and their overall behavior, with rather extensive posts often including scientific and other references. They, however, also represent only a tiny fraction and before we can draw reliable conclusions we need to conduct semantic analyses of their statements. In contrast, the majority of overall users tend to rely on social media-based sources of information, which often feature personal experiences and opinions.

The health care system can be described as a two-sided network: a network with large components linked to each other through multiple platforms so that clinicians, health care institutions, and companies can interact with patients and communities [12]. While we studied one component of this network and how it is affected by the other side of the network, further research should also focus on the opposite direction and mutual influences between the components of this two-sided network.

Our study has used some sophisticated methods for extracting information on the posting behavior in online forums to address important questions in this field. To eliminate some of the limitations of the study and to determine more precisely the role and behavior of forum contributors with regard to scientific information, a qualitative approach is needed, preferably a discourse analysis of the social exchange processes and argumentative strategies in online health forums, similar to a Canadian study of online social support forums for gamblers, in which the interaction of the participants, their common discussions, and how they constructed identities and negotiated legitimacy were analyzed [60]. We are in the midst of a change due to technology where health provision and education can increasingly benefit from using the Web, in an environment in which individuals and communities become more able and responsible for their own health and treatment [61].

Acknowledgments

We are indebted to Lara Weibezahl for critical input and review of the final draft of the paper and to Dr Richard Nicholas for a neurological perspective on CCSVI.

The Ethics Committee of the University Medical Center Göttingen confirmed (ref 11/5/13) that ethical approval was not necessary due to the nature of the data (secondary data analysis of anonymized data).

Some of the methods used and results are part of a Master's thesis available online [62].

Conflicts of Interest

None declared.

Multimedia Appendix 1

Forum information retrieval.

[PDF File (Adobe PDF File), 158KB - jmir_v16i1e10_app1.pdf]

Multimedia Appendix 2

K-means clustering.

[PDF File (Adobe PDF File), 128KB - jmir_v16i1e10_app2.pdf]

Multimedia Appendix 3

Clustering results data tables.

[PDF File (Adobe PDF File), 223KB - jmir_v16i1e10_app3.pdf]

Multimedia Appendix 4

Referenced scientific publications.

[PDF File (Adobe PDF File), 26KB - jmir_v16i1e10_app4.pdf]

References

1. Donaldson L. Expert patients usher in a new era of opportunity for the NHS. BMJ 2003 Jun 14;326(7402):1279-1280 [FREE Full text] [doi: 10.1136/bmj.326.7402.1279] [Medline: 12805129]
2. Greenhalgh T. Patient and public involvement in chronic illness: beyond the expert patient. BMJ 2009;338:629-631. [Medline: 19223339]
3. Hoffman-Goetz L, Donelle L, Thomson MD. Clinical guidelines about diabetes and the accuracy of peer information in an unmoderated online health forum for retired persons. Inform Health Soc Care 2009 Mar;34(2):91-99. [doi: 10.1080/17538150902865136] [Medline: 19412842]
4. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
5. 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]
6. Vance K, Howe W, Dellavalle RP. Social Internet sites as a source of public health information. Dermatol Clin 2009 Apr;27(2):133-136. [doi: 10.1016/j.det.2008.11.010] [Medline: 19254656]
7. Nettleton S, Burrows R, O'Malley L. The mundane realities of the everyday lay use of the Internet for health, and their consequences for media convergence. Sociol Health Illn 2005 Nov;27(7):972-992. [doi: 10.1111/j.1467-9566.2005.00466.x] [Medline: 16313525]
8. 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]
9. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
10. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of Internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
11. Wilson P. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the Internet. BMJ 2002 Mar 9;324(7337):598-602 [FREE Full text] [Medline: 11884329]
12. Griffiths F, Cave J, Boardman F, Ren J, Pawlikowska T, Ball R, et al. Social networks--the future for health care delivery. Soc Sci Med 2012 Dec;75(12):2233-2241. [doi: 10.1016/j.socscimed.2012.08.023] [Medline: 22985490]
13. Weiss JB, Berner ES, Johnson KB, Giuse DA, Murphy BA, Lorenzi NM. Recommendations for the design, implementation and evaluation of social support in online communities, networks, and groups. J Biomed Inform 2013 Dec;46(6):970-976. [doi: 10.1016/j.jbi.2013.04.004] [Medline: 23583424]
14. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. 2003 Presented at: Proc Ninth ACM SIGKDD Int Conf Knowl Discov Data Min; 24.08.2003; Washington, D. C p. 137.
15. Brunson EK. The impact of social networks on parents' vaccination decisions. Pediatrics 2013 May;131(5):1397-1404. [doi: 10.1542/peds.2012-2452] [Medline: 23589813]
16. 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]
17. Zamboni P, Galeotti R, Menegatti E, Malagoni AM, Tacconi G, Dall'Ara S, et al. Chronic cerebrospinal venous insufficiency in patients with multiple sclerosis. J Neurol Neurosurg Psychiatry 2009 Apr;80(4):392-399 [FREE Full text] [doi: 10.1136/jnnp.2008.157164] [Medline: 19060024]
18. Thapar A, Lane T, Nicholas R, Friede T, Ellis M, Assenheim J, et al. Systematic review of sonographic chronic cerebrospinal venous insufficiency findings in multiple sclerosis. Phlebology 2011 Dec;26(8):319-325. [doi: 10.1258/phleb.2011.011098] [Medline: 22021635]
19. Thapar A, Lane TRA, Nicholas R. Chronic Cerebrospinal Venous Insufficiency: It is an entity and a subset of multiple sclerosis - Against The Motion. In: Vascular and Endovascular Controversies Update. London: Greenhalgh RM; 2012:696-705.
20. Turner L. Beyond "medical tourism": Canadian companies marketing medical travel. Global Health 2012;8(16):11 [FREE Full text] [doi: 10.1186/1744-8603-8-16] [Medline: 22703873]
21. Vedantham S, Benenati JF, Kundu S, Black CM, Murphy KJ, Cardella JF, Society of Interventional Radiology, Canadian Interventional Radiology Association. Interventional endovascular management of chronic cerebrospinal venous insufficiency in patients with multiple sclerosis: a position statement by the Society of Interventional Radiology, endorsed by the Canadian

Interventional Radiology Association. J Vasc Interv Radiol 2010 Sep;21(9):1335-1337. [doi: 10.1016/j.jvir.2010.07.004] [Medline: 20800776]

22. Traboulsee AL, Knox KB, Machan L, Zhao Y, Yee I, Rauscher A, et al. Prevalence of extracranial venous narrowing on catheter venography in people with multiple sclerosis, their siblings, and unrelated healthy controls: a blinded, case-control study. Lancet 2013 Oct 8:e8. [doi: 10.1016/S0140-6736(13)61747-X] [Medline: 24119384]
23. Imperial College CCSVI Investigation Group, Thapar A, Lane TR, Pandey V, Shalhoub J, Malik O, et al. Internal jugular thrombosis post venoplasty for chronic cerebrospinal venous insufficiency. Phlebology 2011 Sep;26(6):254-256. [doi: 10.1258/phleb.2011.011052] [Medline: 21803799]
24. Burton JM, Alikhani K, Goyal M, Costello F, White C, Patry D, et al. Complications in MS patients after CCSVI procedures abroad (Calgary, AB). Can J Neurol Sci 2011 Sep;38(5):741-746. [Medline: 21856578]
25. Deutsche Multiple Sklerose Gesellschaft/German Multiple Sclerosis Society. DMSG-Forum URL: http://www.dmsg.de/multiple-sklerose-forum/index.php?kategorie=msforen [accessed 2013-08-06] [WebCite Cache ID 6IfNcEH1V]
26. Forum home | Multiple Sclerosis Society UK. URL: http://www.mssociety.org.uk/forum [accessed 2013-08-06] [WebCite Cache ID 6IfNlvI4I]
27. The CCSVI Foundation of Canada. URL: http://www.ccsvifoundationcanada.org/ [accessed 2013-08-06] [WebCite Cache ID 6IfNrwcGu]
28. CCSVI Locator - Locate CCSVI Social Network Sites. URL: http://ccsvi-ms.ning.com/ [accessed 2013-08-06] [WebCite Cache ID 6IfNxH8GX]
29. CCVSI - YouTube. URL: http://www.youtube.com/results?search_query=ccsvi [accessed 2013-08-06] [WebCite Cache ID 6IfOF8KzQ]
30. Ghezzi A. Funding CCSVI research is/was a waste of valuable time, money and intellectual energy: yes. Mult Scler 2013 Jun;19(7):855-857. [doi: 10.1177/1352458513479825] [Medline: 23712524]
31. Zivadinov R, Weinstock-Guttman B. Funding CCSVI research is/was a waste of valuable time, money and intellectual energy: no. Mult Scler 2013 Jun;19(7):858-860. [doi: 10.1177/1352458513480252] [Medline: 23712525]
32. Jones R, Sharkey S, Smithson J, Ford T, Emmens T, Hewis E, et al. Using metrics to describe the participative stances of members within discussion forums. J Med Internet Res 2011;13(1):e3 [FREE Full text] [doi: 10.2196/jmir.1591] [Medline: 21239373]
33. CiteXplore. 2013 May 31. URL: http://www.ebi.ac.uk/citexplore/webservice.jsp [accessed 2013-12-20] [WebCite Cache ID 6M0vmBswA]
34. Sayers E. NCBI (National Center for Biotechnology Information). 2008. Entrez programming utilities help URL: http://www.ncbi.nlm.nih.gov/books/NBK25501/ [accessed 2013-12-20] [WebCite Cache ID 6M0w3Mlp4]
35. Fallis D. Toward an epistemology of Wikipedia. J Am Soc Inf Sci Technol 2008:1662-1674.
36. MacQueen J. Some methods for classification and analysis of multivariate observations. 1967 Presented at: Proc 5th Berkeley Symp Math Stat Probab; 1967; Berkeley p. 281-297.
37. Chan J, Hayes C, Daly EM. Decomposing discussion forums and boards using user roles. 2010 Presented at: AAAI Conf Weblogs Soc Media; 2010; Washington DC p. 215-218.
38. Chambers JM, Cleveland WS, Tukey PA. Graphical methods for data analysis. Duxbury: Duxbury Press; 1983:158-162.
39. Brush AJ, Wang X, Turner TC, Smith MA. Assessing differential usage of Usenet social accounting meta-data. 2005 Presented at: Proc SIGCHI Conf Hum Factors Comput Syst; 2005; Portland, Oregon p. 889-898.
40. Singh AV, Zamboni P. Anomalous venous blood flow and iron deposition in multiple sclerosis. J Cereb Blood Flow Metab 2009 Dec;29(12):1867-1878. [doi: 10.1038/jcbfm.2009.180] [Medline: 19724286]
41. Zamboni P, Galeotti R, Menegatti E, Malagoni AM, Gianesini S, Bartolomei I, et al. A prospective open-label study of endovascular treatment of chronic cerebrospinal venous insufficiency. J Vasc Surg 2009 Dec;50(6):1348-1358. [doi: 10.1016/j.jvs.2009.07.096] [Medline: 19958985]
42. Al-Omari MH, Rousan LA. Internal jugular vein morphology and hemodynamics in patients with multiple sclerosis. Int Angiol 2010 Apr;29(2):115-120. [Medline: 20351667]
43. Malagoni AM, Galeotti R, Menegatti E, Manfredini F, Basaglia N, Salvi F, et al. Is chronic fatigue the symptom of venous insufficiency associated with multiple sclerosis? A longitudinal pilot study. Int Angiol 2010 Apr;29(2):176-182. [Medline: 20351673]
44. Doepp F, Paul F, Valdueza JM, Schmierer K, Schreiber SJ. No cerebrocervical venous congestion in patients with multiple sclerosis. Ann Neurol 2010 Aug;68(2):173-183. [doi: 10.1002/ana.22085] [Medline: 20695010]
45. Wattjes MP, van Oosten BW, de Graaf WL, Seewann A, Bot JC, van den Berg R, et al. No association of abnormal cranial venous drainage with multiple sclerosis: a magnetic resonance venography and flow-quantification study. J Neurol Neurosurg Psychiatry 2011 Apr;82(4):429-435. [doi: 10.1136/jnnp.2010.223479] [Medline: 20980483]
46. Yamout B, Herlopian A, Issa Z, Habib RH, Fawaz A, Salame J, et al. Extracranial venous stenosis is an unlikely cause of multiple sclerosis. Mult Scler 2010 Nov;16(11):1341-1348. [doi: 10.1177/1352458510385268] [Medline: 21041329]
47. Bastianello S, Romani A, Viselner G, Tibaldi EC, Giugni E, Altieri M, et al. Chronic cerebrospinal venous insufficiency in multiple sclerosis: clinical correlates from a multicentre study. BMC Neurol 2011;11(132):e7 [FREE Full text] [doi: 10.1186/1471-2377-11-132] [Medline: 22029656]

48. Baracchini C, Perini P, Calabrese M, Causin F, Rinaldi F, Gallo P. No evidence of chronic cerebrospinal venous insufficiency at multiple sclerosis onset. Ann Neurol 2011 Jan;69(1):90-99. [doi: 10.1002/ana.22228] [Medline: 21280079]
49. Mayer CA, Pfeilschifter W, Lorenz MW, Nedelmann M, Bechmann I, Steinmetz H, et al. The perfect crime? CCSVI not leaving a trace in MS. J Neurol Neurosurg Psychiatry 2011 Apr;82(4):436-440 [FREE Full text] [doi: 10.1136/jnnp.2010.231613] [Medline: 21296899]
50. Centonze D, Floris R, Stefanini M, Rossi S, Fabiano S, Castelli M, et al. Proposed chronic cerebrospinal venous insufficiency criteria do not predict multiple sclerosis risk or severity. Ann Neurol 2011 Jul;70(1):51-58. [doi: 10.1002/ana.22436] [Medline: 21786298]
51. Zhai Z, Xu H, Jia P. Identifying opinion leaders in BBS. 2008 Presented at: Int Conf Web Intell Intell Agent Technol WI-IAT08 IEEEWICACM; 2008; Sydney p. 398-401.
52. Zhang X, Dong D. Ways of identifying the opinion leaders in virtual communities. Int J Bus Manag 2008;3(7):21-27.
53. Xiao Y, Xia L. Understanding opinion leaders in bulletin board systems: Structures and algorithms. : Structures and algorithms. 2010 IEEE 35th Conf Local Comput Networks LCN 2010.–; 2010 Presented at: 2010 IEEE 35th Conference on Local Computer Networks (LCN); 2010; Denver p. 1062-1067.
54. Ziyi L, Jing CFS, Donghong S, Yongfeng H. Research on methods to identify the opinion leaders in Internet community. 2013 Presented at: 4th IEEE Int Conf Softw Eng Serv Sci ICSESS; 2013; Beijing p. 934-937.
55. Waldschmidt A, Klein A, Tamayo Korte M, Dalman-Eken S. Discourse in everyday life - the everyday life of discourse: Towards an empirical grounded methodology of discourse research in social sciences [in German]. Forum Qual Sozialforschung FQS 2007;8(2):e29.
56. Horowitz B. Elatable. Creators, synthesizers, and consumers URL: http://blog.elatable.com/2006/02/creators-synthesizers-and-consumers.html [accessed 2013-08-06] [WebCite Cache ID 6IfOi4gNA]
57. Chomutare T, Arsand E, Fernandez-Luque L, Lauritzen J, Hartvigsen G. Inferring community structure in healthcare forums. An empirical study. Methods Inf Med 2013;52(2):160-167. [doi: 10.3414/ME12-02-0003] [Medline: 23392282]
58. Dias A, Chomutare T, Botsis T. Exploring the community structure of a diabetes forum. Stud Health Technol Inform 2012;180:833-837. [Medline: 22874309]
59. 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]
60. Mudry TE, Strong T. Doing recovery online. Qual Health Res 2013 Mar;23(3):313-325. [doi: 10.1177/1049732312468296] [Medline: 23208202]
61. Denecke K, Brooks E. Web science in medicine and healthcare. Methods Inf Med 2013;52(2):148-151. [Medline: 23508344]
62. Sudau F. Master’s Thesis in Applied Computer Science at the Institute of Computer Science, ZAI-MSC--04, ISSN 1612-6793.: Georg-August-University of Göttingen; 2013. Analysis of controversial debates in online fora - a showcase analysis of the CCSVI discussion in the DMSG layperson forum URL: http://www.swe.informatik.uni-goettingen.de/publications/analysis-controversial-debates-online-fora-showcase-analysis-ccsvi-discussion-dmsg [accessed 2013-12-21] [WebCite Cache ID 6M2Em93Xk]

Abbreviations

CCSVI: Chronic Cerebrospinal Venous Insufficiency
DMSG: Deutsche Multiple Sklerose Gesellschaft/ German Multiple Sclerosis Society
MS: multiple sclerosis

Edited by G Eysenbach; submitted 12.08.13; peer-reviewed by K Denecke, C Smith; comments to author 10.09.13; revised version received 24.10.13; accepted 11.11.13; published 14.01.14.

Please cite as:
Sudau F, Friede T, Grabowski J, Koschack J, Makedonski P, Himmel W
Sources of Information and Behavioral Patterns in Online Health Forums: Observational Study
J Med Internet Res 2014;16(1):e10
URL: http://www.jmir.org/2014/1/e10/
doi:10.2196/jmir.2875
PMID:24425598

Original Paper

The Influence of Two Different Invitation Letters on Chlamydia Testing Participation: Randomized Controlled Trial

Gill ten Hoor[1], MSc; Christian JPA Hoebe[2,3], PhD; Jan EAM van Bergen[4,5], PhD; Elfi EHG Brouwers[2], MSc; Robert AC Ruiter[1], PhD; Gerjo Kok[1], PhD

[1]Maastricht University, Department of Work & Social Psychology, Maastricht, Netherlands
[2]Public Health Service South Limburg, Department of Sexual Health, Infectious Disease and Environmental Health, Geleen, Netherlands
[3]Maastricht University, Department of Medical Microbiology, Maastricht, Netherlands
[4]STI AIDS Netherlands, Amsterdam, Netherlands
[5]AMC - University of Amsterdam, Department of General Practice, Amsterdam, Netherlands

Corresponding Author:
Gill ten Hoor, MSc
Maastricht University
Department of Work & Social Psychology
PO Box 616
Maastricht, 6200MD
Netherlands
Phone: 31 433881617
Fax: 31 433884199
Email: gill.tenhoor@maastrichtuniversity.nl

Abstract

Background: In the Netherlands, screening for chlamydia (the most prevalent sexually transmitted infection worldwide) is a relatively simple and free procedure. Via an invitation letter sent by the public health services (PHS), people are asked to visit a website to request a test kit. They can then do a chlamydia test at home, send it anonymously to a laboratory, and, within two weeks, they can review their test results online and be treated by their general practitioner or the PHS. Unfortunately, the participation rates are low and the process is believed to be not (cost-) effective.

Objective: The objective of this study was to assess whether the low participation rate of screening for chlamydia at home, via an invitation letter asking to visit a website and request a test kit, could be improved by optimizing the invitation letter through systematically applied behavior change theories and evidence.

Methods: The original letter and a revised letter were randomly sent out to 13,551 citizens, 16 to 29 years old, in a Dutch municipality. Using behavior change theories, the revised letter sought to increase motivation to conduct chlamydia screening tests. The revised letter was tailored to beliefs that were found in earlier studies: risk perception, advantages and disadvantages (attitude), moral norm, social influence, and response- and self-efficacy. Revisions to the new letter also sought to avoid possible unwanted resistance caused when people feel pressured, and included prompts to trigger the desired behavior.

Results: No significant differences in test package requests were found between the two letters. There were also no differences between the original and revised letters in the rates of returned tests (11.80%, 581/4922 vs 11.07%, 549/4961) or positive test results (4.8%, 23/484 vs 4.1%, 19/460). It is evident that the new letter did not improve participation compared to the original letter.

Conclusions: It is clear that the approach of inviting the target population through a letter does not lead to higher participation rates for chlamydia screening. Other approaches have to be developed and pilot tested.

(J Med Internet Res 2014;16(1):e24) doi:10.2196/jmir.2907

KEYWORDS

invitation letter; chlamydia; screening; testing; behavior change theories

Introduction

In a 3-year systematic register-based yearly chlamydia screening project in three regions in the Netherlands, all 16 to 29 year old citizens were given the opportunity at no charge to test for Chlamydia trachomatis. Via an invitation letter sent by the public health services (PHS), they were asked to visit a website [1] where they could request a test kit. Subsequently, they could do a chlamydia test at home, send it anonymously to a laboratory, and, within two weeks, they would be able to review their test results online and could be treated by their general practitioner or the PHS. Despite this free and relatively simple procedure, receiving the letter triggered only a small number of young people to participate. In the first round in 2008, the participation rate was 16.1%. The rate decreased over subsequent rounds (10.8% in 2009 and 9.5% in 2010) [2]. With these participation rates, screening was believed to be not (cost-) effective and therefore further nationwide implementation of the program was discontinued [3]. In the present study, we assessed whether the participation rate could be improved by optimizing the invitation letter through systematically applied behavior change theories and evidence [4].

The first step in planned behavior change is to identify the reasons or determinants of the behavior. In an earlier study [5], we assessed the reasons for non-participation by asking 713 people within the age range of 16 and 29 years about their intention to participate in chlamydia screening and included measures of attitude, subjective norm, self-efficacy, the moral norm, susceptibility, the descriptive norm, outcome expectations, and unrealistic optimism toward chlamydia testing. Questions asked were based on literature reports suggesting that the targeted young people felt invulnerable to chlamydia infection, did not compare themselves with people who get infected, and had no time or interest in participating [6,7], experienced barriers such as lack of knowledge, worries, and lack of guidance [8], were afraid of doing it wrong, found the procedure unpleasant, were afraid of the consequences of a possible positive outcome, and expressed fear of negative reactions from a partner and others [9-11]. Furthermore, three single category implicit association tasks (scIAT) [12] were included to identify impulsive reactions towards chlamydia in terms of annoyance, threat, and reassurance. All data were gathered without giving any information in advance about chlamydia or chlamydia testing. The results showed that people have a very low intention to participate in chlamydia screening (mean 1.42, SD 0.76 on a scale of 1-5), low risk perception, in particular low susceptibility, and high unrealistic optimism (most young people do not think they have ever run the risk of being infected with chlamydia and they do not identify themselves with people who test positive for chlamydia). The intention was correlated with the subjective norm, the moral norm, susceptibility, the descriptive norm, one's attitude, outcome expectations, and unrealistic optimism. Furthermore, chlamydia screening was implicitly associated with reassurance, as well as with threat and annoyance.

Also in the same study, a first attempt was made to optimize the invitation letter by assessing the influence of the original PHS invitation letter versus a letter that was adapted to improve readability and increase a positive response. The results showed no differences between the effects of the two letters; however, receiving a letter had, compared to not receiving a letter, a positive effect on people's evaluations and intentions to request a test package [5]. There was no measure of testing behavior in that study.

Interventions targeting behavior change have a higher chance of success when theories are systematically applied [13-17]. In the current study, the research question is whether another newly developed invitation letter, systematically written based on theory and adapted to the new evidence from our earlier study has a positive influence on people's chlamydia screening behavior.

Methods

Study Population and Study Procedure

The PHS sent 13,551 letters to all 16 to 29 year old citizens of the Dutch municipality Sittard-Geleen. They received randomly either a newly developed letter or the letter that was already used in the Dutch national Chlamydia Screening Implementation program (CSI). Both invitation letters offered the recipient the opportunity to anonymously request a free chlamydia test kit via a website. At the website, visitors logged in using an anonymous personalized ID and first filled in an 8-item risk questionnaire [18]. Only participants with at least a minimum level of chlamydia risk could proceed to request a test kit. When requested online, they received a chlamydia test, could perform the test at home (urine sample or vaginal swab), send it anonymously to a laboratory, and, within two weeks, they could review their test results online. The study design was approved by the Research Ethics Board of the Faculty of Psychology & Neurosciences of Maastricht University. Registration of this trial was not required.

Two Invitation Letters

Overview

Influencing behavior is more successful when theory is applied and when the content of the message is tailored to the target group [4]. Elaboration Likelihood Theory [19] suggests that people only process information seriously when they are motivated and able to do so. The Reasoned Action Approach [20] suggests that people will change if the right beliefs are changed: beliefs identified through elicitation research. The new letter therefore provides personally relevant information (increase motivation) in such a way that it is easy to process (increase ability) and is tailored to the beliefs that were found in earlier studies (elicitation): risk perception, advantages and disadvantages (attitude), moral norm, social influence, and response- and self-efficacy. There is also anticipation of possible unwanted reactance when people feel pressured and there are prompts to trigger the desired behavior.

The two invitation letters were similar in layout and information content. For the new letter, the order in which information was presented was changed and the content was simplified to increase comprehension and adapted based on the findings of our earlier study [5] and appropriate theories [4]. To keep the new letter short, readers were referred to the website for detailed instructions. Logos on the new letter were identical but fonts were slightly larger to increase readability. Table 1 shows the differences in letters. In this table, the new letter content is chronologically displayed. The PHS letter is not chronologically displayed, but shows how comparable information was given. The (Dutch) invitation letters can be found in Multimedia Appendices 1-4 and [21].

First Impression Bias, Primacy Effect, and Self-Affirmation

People's attitudes or opinions towards specific information are colored by their first impression [22]. Furthermore, the primacy effect suggests that information that is presented first is often remembered best [23]. Therefore, possible negative triggers, as shown in the PHS letter (AIDS, STI), were removed from the top of the letter and added to the footnote in the new letter (see Table 1, #1). Receiving an invitation letter to participate in chlamydia screening can be seen as threatening health information. It is known that people rarely change their behavior after receiving threatening health messages and sometimes the information leads to defensive responses [24]. Self-affirmation is used to decrease the chance of defensive reactions to the threat, or reactance [25]. Applying self-affirmation theory, people were made aware of the value of their personal relationships, thereby increasing their self-identity and promoting a constructive response (see Table 1, #2).

Advantages and Convenience, Reactance (I), Efficacy, and the Prompt to Action

Both the advantages (attitude) and ease of testing (self-efficacy) are correlated with the intention to request a chlamydia test package [5]. Therefore, both were highlighted in the newly developed letter (see Table 1, #3). Further, the use of the wording "sexually active" (see Table 1, #2) might be interpreted defensively by receivers as having sex with multiple partners. Moreover, reactance theory suggests that people respond negatively to others' attempts to limit their freedom [26]. In the new letter, that phrase was deleted and, to prevent possible reactance, the invitation was presented as a general request to all people in that age group, along with a rationale (see Table 1, #4). Also, adaptive behavior is promoted by stressing the belief that the behavior is effective in reducing threat (response efficacy) and the confidence that one can accomplish this behavior (self-efficacy) [24]. Furthermore, a trigger to action was given [27] (see Table 1, #5).

Negative Consequences, Severity, and Moral Norm

Threat is the combination of severity (how bad are the consequences?) and susceptibility (do I personally run a risk?) [24]. The severity of the negative consequences of chlamydia is not always recognized [5]; in the letter, the need for early treatment was stressed (see Table 1, #6). Also, in the earlier study, the personal moral norm (a person's judgment as to whether they themselves think they should or should not perform a certain behavior [28]) was highly correlated with one's intention to screen. To activate a moral norm, the possibility of unintentionally infecting someone else was mentioned in the letter (see Table 1, #7).

Perceived Risk, Unrealistic Optimism, and Reactance (II)

In the earlier study, people scored very low on the perceived risk of getting chlamydia. Furthermore, people thought that other people's risks were higher than their own. Therefore, it was important to emphasize that all sexually active people, not only people with multiple partners and people who have unsafe sex, can get chlamydia; risk is a matter of risk behavior rather than of risk groups [27] (see Table 1, #8). To minimize a possible reactance (see Reactance (I) above), it was highlighted again that the invitation for a chlamydia test was not a targeted invitation, but part of a regional screening (see Table 1, #9).

Self-Efficacy (II) and Procedure, Descriptive Norm, and Implicit Attitudes

In our earlier study, a large majority of people stated that they were unable to test because they did not have time. Therefore, it was important to explain that the procedure would be very simple and would take less than five minutes (see Table 1, #10). Also, people's behavior, and especially young people's behavior, is influenced by the behavior of peers [29,30]. Therefore, the letter mentioned that many comparable young people had already tested for chlamydia. Because the earlier study showed an implicit association with annoying, threatening, as well as reassuring, those comparable others were reported to evaluate the test as reassuring and not as threatening or annoying (see Table 1, #11).

Moral Norm (II), Anticipated Regret, Privacy, and Response Efficacy

Moral norm, as well as anticipated regret (having people imagine how they would feel after they behaved in a risky way contrary to their own intentions [31]) may both lead to attitude and behavior change (see Table 1, #12). On the topic of privacy, doing a chlamydia test is for many people a private procedure that should not be observable by others [32]. Therefore, the privacy of the testing procedure was stressed in the letter (see Table 1, #13). Also, as mentioned before, a threat may lead to an appropriate behavioral response when people believe that such a response is available and easy to do (see Table 1, #14).

Log-In Code, Prompt to Action, Sender Information, and Footnote Information

The information about the log-in procedure was simplified in the new letter (see Table 1, #15). Also included were prompts to form a plan for action, which may increase the number of people performing the testing behavior [33] (see Table 1, #16). Sender information was identical in both letters (see Table 1, #17), but footnote information was simplified and, to avoid primacy effects, the AIDS/STD info was given here (see First Impression Bias and Primacy Effect above and Table 1, #18).

Table 1. Adaptation of the new letter compared to the original PHS letter.

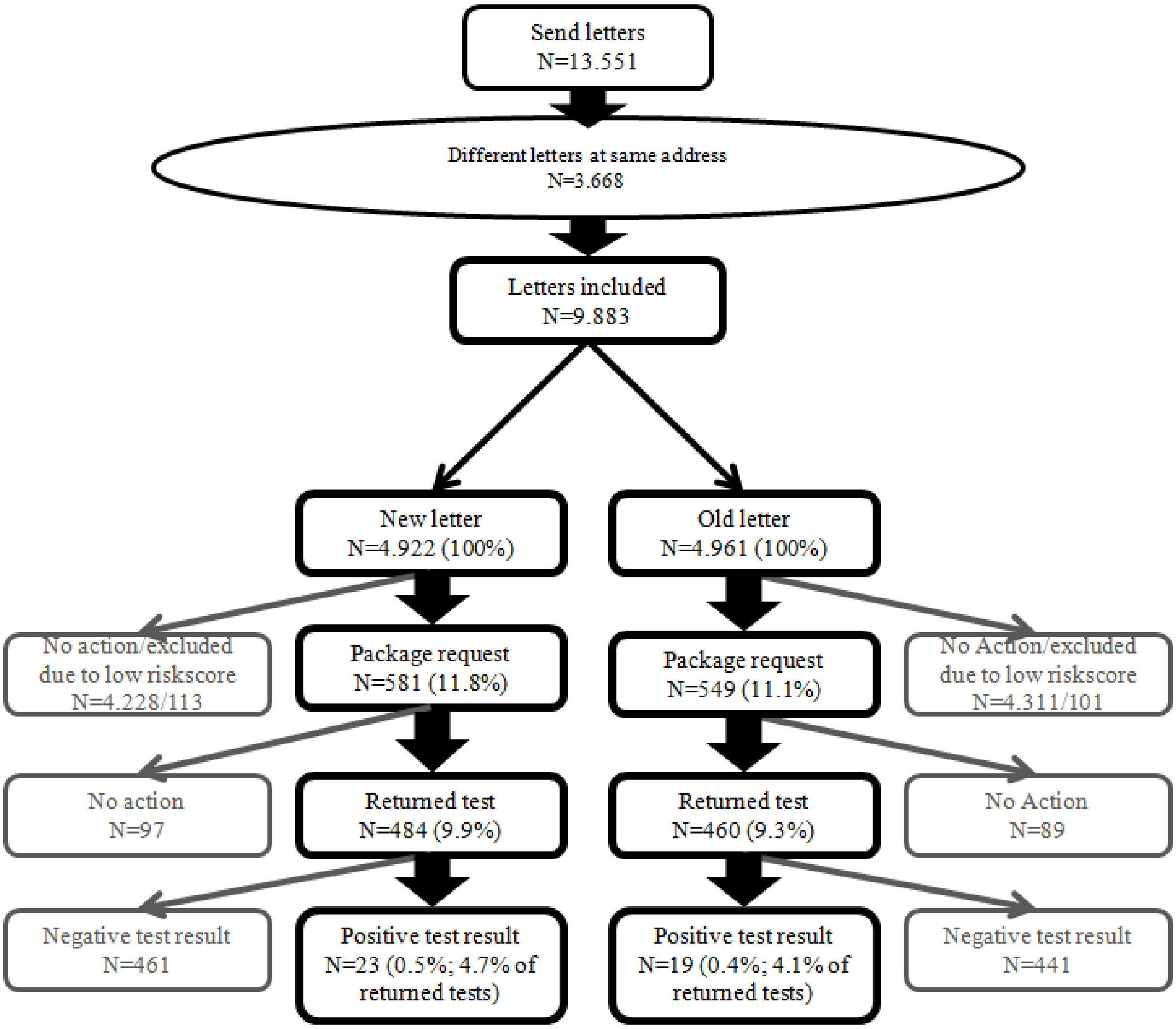
New letter	Comparable information from PHS letter
1. First Impression Bias and Primacy Effects	
Visiting address: Het Overloon 2 6411 TE Heerlen 045-8506613 (9.00h-12.00h)	Visiting Address: Het Overloon 2 6411 TE Heerlen 045-850 66 13 (9.00h-12.00h) Aids STD Info line (for questions about sex, STD and The Pill): 0900-2042040
2. Self-Affirmation	
“Your health is very important. Not only for yourself, but also for a possible partner and family.”	“When you are (or have been) sexually active, it is important to do this test, even if you’ve done this test before or if you’re in a solid relationship.”
3. Advantages and Convenience	
“Therefore, you should take the opportunity to take a free and easy chlamydia test at home.”	“Via this letter, we invite you again to participate in the Chlamydia Screening South-Limburg.”
4. Reactance (I)	
“When you have sex, it is possible to contract chlamydia without realizing it.”	“That [ie, treat quickly] is difficult, because most people don’t know whether they are infected with chlamydia or not.”
5. Efficacy and Prompt to Action	
“Chlamydia is simple and easy to trace, and very trouble-free to treat. However, it is important to do the test as early as possible!”	“It is easy to treat, but important to be quick. Severe and frustrating health issues can be prevented by testing for chlamydia.”
6. Negative Consequences, Severity	
“If you wait too long for treatment, you can get severe and permanent health issues like infertility.”	“If chlamydia is not treated in time, men can get epididymitis and women can become infertile.”
7. Moral Norm	
“Above all, you can infect others without knowing it.”	“That [ie, treat quickly] is difficult, because most people don’t know whether they are infected with chlamydia or not.”
8. Perceived Risk, Unrealistic Optimism	
“Chlamydia does not only occur in people who have unsafe sex with many partners, but also in people with a few partners or just one partner.”	“Chlamydia is an STD (sexually transmitted disease) with a high prevalence in the Netherlands, especially in young people aged 16 to 29.”
9. Reactance (II)	
“To decrease the number of chlamydia infections, all 16-29 year olds from your region are invited to request a free chlamydia test via www.chlamydiatest.nl.”	“On www.chlamydiatest.nl The PHS South Limburg invites all men and women, aged 16-29 years, to participate in the annual free chlamydia test. On www.chlamydiatest.nl, ”
10. Self-efficacy (II) and Procedure	
“When you have requested and received the test package, you can do the test in less than five minutes at home, after which you can send it to the laboratory.”	“How does it work: You log in on with your personal log-in code from this letter. There you can create immediately your own username and password, making sure that no one else can log in. Subsequently, you can request a test package. You will receive this in a blank package at the address of your choice. You will find instructions in the test package about how to collect your test sample. Subsequently, you can send the package back to the laboratory at no cost. The test result is available within two weeks and will be online for three months via www.chlamydiatest.nl. With your username and password, you can request your test result. If you’ve forgotten your personal details, you need your personal log-in code from this letter. Therefore, you should keep this letter!”
11. Descriptive Norm and Implicit Attitudes	
“A lot of your peers have already tested for chlamydia, not finding it annoying or threatening, but more reassuring.”	N/A
12. Moral Norm (II) and Anticipated Regret	

New letter	Comparable information from PHS letter
“Also, some of them said that they would feel guilty if they didn’t do the test.”	N/A
13. Privacy	
“The entire test can be performed anonymously and at no charge. Your personal details are kept confidential. Only you, with your username and password, are able to request your test result (within two weeks!). No one, not your parents nor your general practitioner, is notified of the results.”	“All your personal details are kept confidential. Only you, with your username and password, are able to request your test result. And only you, not your parents nor your general practitioner, are notified of the results. In the folder, attached to this letter, and on www.chlamydiatest.nl, you can find more information about chlamydia and the test.”
14. Response Efficacy	
“If you are infected with chlamydia, it is very easily treated with antibiotics.”	N/A
15. Log-in code	
“Your strictly personal log-in code for www.chlamydiatest.nl is: ”	“To log in, you need a personal log-in code. Your strictly personal log-in code for www.chlamydiatest.nl is: ”
16. Prompt to Action (II)	
“Don’t delay − request your test package today!”	“Think about your health and participate in the chlamydia screening!”
17. Sender Information	
“With kind regards, Dr. Christian J.P.A. Hoebe Doctor/Epidemiologist, Infectious Diseases Project Leader, Chlamydia Screening South Limburg”	“With kind regards, Dr. Christian J.P.A. Hoebe Doctor/Epidemiologist, Infectious Diseases Project Leader, Chlamydia Screening South Limburg”
18. Footnote Information	
“For questions and information: www.chlamydiatest.nl. For questions about participation: ms H.L.G. ter Waarbeek (independent GP: 045-8506264). For questions about sex, STDs and The Pill: Aids STD info line: 0900-2042040.”	“There (ie, on www.chlamydiatest.nl can read about all the things you have to do for this test. P.S. This invitation is sent district by district across 10 municipalities to all individuals aged 16 to 29 years, so not everybody receives the invitation at the same time. If you want to know which municipalities, see www.chlamydiatest.nl. If you want to discuss questions about participation in this study with an independent general practitioner: contact ms H.L.G. ter Waarbeek (045-8506264).”

Results

In total, 13,551 letters were sent to all 16 to 29 year old citizens of the municipality Sittard-Geleen (population 94,024 [34]) in the south of the Netherlands, randomly divided over the new and the original letter. When two different letters were delivered at one unique address (which could only be checked afterward), or when letters were returned as undeliverable, these data were excluded from further analyses (n=3668). Of the 9883 included respondents, 11.43% (1130/9883) requested a test package. No significant differences in test package requests were found between the two letters (χ^2_1=1.33, P=.25, phi= −.012). There were also no differences between the two letters for the rates of returned tests (χ^2_1=0.05, P=.82, phi= .007), and the number of positive test results (χ^2_1=0.21, P=.64, phi= −.015) (see Figure 1). It is evident that the new letter did not improve participation compared to the original letter. In acknowledgement of recent concerns regarding lack of disclosure in scientific research [35], and to aid future meta-analyses, all data, syntax files, and output files are available in Multimedia Appendices 1- 6 and [21].

Figure 1. Flowchart of responses to the new letter and the original PHS letter.



Discussion

Principal Findings

In this study, the effect of a new theory- and evidence-based adaptation of an invitation letter for chlamydia screening was compared to the effect of the original letter. Both letters resulted in small percentages of participation, comparable to other screening projects in the Netherlands [3] and outside the Netherlands [36]. However, contrary to expectations, there was no significant difference between the two letters. The new letter did not stimulate more young people to go for the test.

Strengths and Limitations

The strength of this study is that actual behavior was measured, while the weakness is that there was only observational data and no data on how people processed the information or on possible changes in the determinants of behavior. It is, however, difficult to imagine how a mass media letter could be improved differently to reach a substantial larger number of young people participating in the screening. It is obvious that the approach to invite the target population via a general letter does not lead to sufficient participation rates for chlamydia screening.

Conclusions

Why is it so hard to convince young people to participate in chlamydia screening? The strongest determinants of chlamydia screening participation in earlier studies seem to be (low) risk perception, in particular low susceptibility, and high unrealistic optimism [5]. That means that the basic proposition for action is fully lacking [24] and, as it turns out, it seems to be very difficult to convince people that they indeed are at risk. Risk perception and unrealistic optimism can be changed but not easily. Bartholomew et al [4] suggest a number of methods including scenario-based risk information, consciousness raising, or self-affirmation, but those methods require more individual tailoring, more attention, and more time than is feasible in one general letter (page 333 [4]). Other behavior change approaches may be needed. There are some suggestions in the literature: the use of the Internet independent of geographic area [37], financial incentives [38], a focus on self-identity [39], and tailoring on risk perception [40]. Schmid et al [3] suggest retesting people who were found positive and intensifying partner notification. In that approach, the focus is on people who already know they are at risk. Based on the results in our studies, possible alternative strategies for people who do not see themselves at risk might involve the use of social media in targeting high-risk groups. Social circles around people who test positive for chlamydia are shown to be at higher risk [41,42]. Young people who tested positive in chlamydia screening could serve as role models for other young people in their social circles. If this approach is used, the target group should see those models as someone from their own circle that they can identify with, who had to overcome some personal resistance to participate, who is reinforced for participating in the screening by reporting reassurance, and who explains the ease of participation [43]. This alternative approach should be tried out in a randomized study comparable to this study.

Acknowledgments

We would like to acknowledge the Chlamydia Screening Implementation (CSI) project group: JEAM van Bergen, IVF van den Broek, EEHG Brouwers, JSA Fennema, HM Götz, CJPA Hoebe, RH Koekenbier, ELM Op de Coul, LL Pars, and SM van Ravesteijn. Financial support was provided by ZonMw.

Conflicts of Interest

None declared.

Multimedia Appendix 1

New letter.

[JPG File, 3MB - jmir_v16i1e24_app1.jpg]

Multimedia Appendix 2

New letter translated.

[PDF File (Adobe PDF File), 23KB - jmir_v16i1e24_app2.pdf]

Multimedia Appendix 3

PHS letter.

[PDF File (Adobe PDF File), 154KB - jmir_v16i1e24_app3.pdf]

Multimedia Appendix 4

PHS letter translated.

[PDF File (Adobe PDF File), 34KB - jmir_v16i1e24_app4.pdf]

Multimedia Appendix 5

Datafile.

[SAV File, 725KB - jmir_v16i1e24_app5.sav]

Multimedia Appendix 6

SPSS syntax.

[SPS File, 1KB - jmir_v16i1e24_app6.sps]

References

1. Chlamydia Screening Website. URL: http://www.chlamydiatest.nl [accessed 2014-01-07] [WebCite Cache ID 6MSNcdoAQ]
2. van den Broek IV, van Bergen JE, Brouwers EE, Fennema JS, Götz HM, Hoebe CJ, et al. Effectiveness of yearly, register based screening for chlamydia in the Netherlands: controlled trial with randomised stepped wedge implementation. BMJ 2012;345:e4316 [FREE Full text] [Medline: 22767614]
3. Schmid BV, Over EAB, van den Broek IVF, Op de Coul ELM, van Bergen JEAM, Fennema JSA, et al. Effects of population based screening for Chlamydia infections in the Netherlands limited by declining participation rates. PLoS One 2013;8(3):e58674 [FREE Full text] [doi: 10.1371/journal.pone.0058674] [Medline: 23527005]
4. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, Fernández ME. Planning Health Promotion Programs: An Intervention Mapping Approach, 3rd Edition. San Francisco, CA: Jossey-Bass; 2011.

5. Ten Hoor GA, Ruiter RA, van Bergen JE, Hoebe CJ, Houben K, Kok G. Non-participation in chlamydia screening in the Netherlands: determinants associated with young people's intention to participate in chlamydia screening. BMC Public Health 2013;13:1091 [FREE Full text] [doi: 10.1186/1471-2458-13-1091] [Medline: 24266906]
6. Duncan B, Hart G, Scoular A, Bigrigg A. Qualitative analysis of psychosocial impact of diagnosis of Chlamydia trachomatis: implications for screening. BMJ 2001 Jan 27;322(7280):195-199 [FREE Full text] [Medline: 11159612]
7. Greenland KE, Op de Coul EL, van Bergen JE, Brouwers EE, Fennema HJ, Götz HM, et al. Acceptability of the internet-based Chlamydia screening implementation in the Netherlands and insights into nonresponse. Sex Transm Dis 2011 Jun;38(6):467-474. [doi: 10.1097/OLQ.0b013e318204546e] [Medline: 21217416]
8. McNulty CA, Freeman E, Bowen J, Shefras J, Fenton KA. Barriers to opportunistic chlamydia testing in primary care. Br J Gen Pract 2004 Jul;54(504):508-514 [FREE Full text] [Medline: 15239912]
9. PHS, Public Health Service Amsterdam (2010). Kwalitatief onderzoek naar redenen voor het niet terugsturen van een aangevraagd Chlamydia testpakket; Chlamydia Screening Implementation (CSI) A qualitative study on reasons for not returning a requested chlamydia screening test. Public Health Service Amsterdam (2010) (study report) 2012:1.
10. Blake DR, Kearney MH, Oakes JM, Druker SK, Bibace R. Improving participation in Chlamydia screening programs: perspectives of high-risk youth. Arch Pediatr Adolesc Med 2003 Jun;157(6):523-529. [doi: 10.1001/archpedi.157.6.523] [Medline: 12796231]
11. Rose SB, Smith MC, Lawton BA. "If everyone does it, it's not a big deal." Young people talk about chlamydia testing. N Z Med J 2008 Apr 4;121(1271):33-42. [Medline: 18392060]
12. Karpinski A, Steinman RB. The single category implicit association test as a measure of implicit social cognition. J Pers Soc Psychol 2006 Jul;91(1):16-32. [doi: 10.1037/0022-3514.91.1.16] [Medline: 16834477]
13. Albarracín D, Gillette JC, Earl AN, Glasman LR, Durantini MR, Ho MH. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull 2005 Nov;131(6):856-897 [FREE Full text] [doi: 10.1037/0033-2909.131.6.856] [Medline: 16351327]
14. de Bruin M, Viechtbauer W, Schaalma HP, Kok G, Abraham C, Hospers HJ. Standard care impact on effects of highly active antiretroviral therapy adherence interventions: A meta-analysis of randomized controlled trials. Arch Intern Med 2010 Feb 8;170(3):240-250. [doi: 10.1001/archinternmed.2009.536] [Medline: 20142568]
15. Mullen PD, Green LW, Persinger GS. Clinical trials of patient education for chronic conditions: a comparative meta-analysis of intervention types. Prev Med 1985 Nov;14(6):753-781. [Medline: 2418436]
16. Peters LW, Kok G, Ten Dam GT, Buijs GJ, Paulussen TG. Effective elements of school health promotion across behavioral domains: a systematic review of reviews. BMC Public Health 2009;9:182 [FREE Full text] [doi: 10.1186/1471-2458-9-182] [Medline: 19523195]
17. van Achterberg T, Huisman-de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162 [FREE Full text] [doi: 10.1093/heapro/daq050] [Medline: 20739325]
18. van den Broek IV, Brouwers EE, Götz HM, van Bergen JE, Op de Coul EL, Fennema JS, et al. Systematic selection of screening participants by risk score in a Chlamydia screening programme is feasible and effective. Sex Transm Infect 2012 Apr;88(3):205-211. [doi: 10.1136/sextrans-2011-050219] [Medline: 22215696]
19. Petty RE, Barden J, Wheeler SC. The Elaboration Likelihood Model of Persuasion: Developing health promotions for sustained behavioral change. In: DiClemente RJ, Crosby RA, Kegler M, editors. Emerging Theories in Health Promotion Practice and Research. San Francisco, CA: Jossey-Bass; 2009:185-214.
20. Fishbein M, Ajzen I. Predicting and changing behavior: the reasoned action approach. New York: Psychology Press; 2010.
21. ScienceRep Repository (Multimedia Appendices). URL: http://sciencerep.org/12 [accessed 2014-01-07] [WebCite Cache ID 6MSNqOPIf]
22. Lim KH, Benbasat I, Ward LM. The role of multimedia in changing first impression bias. Information Systems Research 2000;11(2):115-136.
23. Petty RE, Wegener DT. Attitude change: multiple roles for persuasion variables. In: Gilbert DR, Fiske ST, Lindzey G, editors. The handbook of social psychology. Boston: McGraw-Hill; 1998:323-390.
24. Peters GJ, Ruiter RA, Kok G. Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. Health Psychol Rev 2013 May;7(Suppl 1):S8-S31 [FREE Full text] [doi: 10.1080/17437199.2012.703527] [Medline: 23772231]
25. Harris PR, Mayle K, Mabbott L, Napper L. Self-affirmation reduces smokers' defensiveness to graphic on-pack cigarette warning labels. Health Psychol 2007 Jul;26(4):437-446. [doi: 10.1037/0278-6133.26.4.437] [Medline: 17605563]
26. Laurin K, Kay AC, Fitzsimons GJ. Reactance versus rationalization: divergent responses to policies that constrain freedom. Psychol Sci 2012 Feb;23(2):205-209. [doi: 10.1177/0956797611429468] [Medline: 22241813]
27. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. San Francisco, CA: Jossey-Bass; 2008:45-65.
28. Godin G, Fortin C, Michaud F, Bradet R, Kok G. Use of condoms: Intention and behaviour of adolescents living in juvenile rehabilitation centres. Health Education Research 1997;12:289-300. [doi: 10.1093/her/12.3.289]
29. Cialdini RB. Influence: science and practice. Boston: Pearson Education; 2009.

30. Mollen S, Rimal RN, Ruiter RA, Jang SA, Kok G. Intervening or interfering? The influence of injunctive and descriptive norms on intervention behaviours in alcohol consumption contexts. Psychol Health 2013;28(5):561-578. [doi: 10.1080/08870446.2012.752827] [Medline: 23289490]
31. Richard R, van der Pligt J, de Vries N. Anticipated affective reactions and prevention of AIDS. Br J Soc Psychol 1995 Mar;34 (Pt 1):9-21. [Medline: 7735735]
32. Westmaas AH, Kok G, Vriens P, Götz H, Richardus JH, Voeten H. Determinants of intention to get tested for STI/HIV among the Surinamese and Antilleans in the Netherlands: results of an online survey. BMC Public Health 2012;12:961 [FREE Full text] [doi: 10.1186/1471-2458-12-961] [Medline: 23136830]
33. Milkman KL, Beshears J, Choi JJ, Laibson D, Madrian BC. Following through on good intentions: The power of planning prompts. National Bureau of Economic Research, MA: Cambridge 2012:NBER Working Paper No 17995 [FREE Full text]
34. Deerenberg I, Elzinga H, van Houwelingen C, Keuning - van Oirschot H, Geurden - Slis M, Beeckman D, et al. Centraal Bureau voor de statistiek. 2011. Gemeente op Maat, Sittard-Geleen URL: http://www.cbs.nl/NR/rdonlyres/BC2ADC41-2F75-45CA-A795-8435A729DD38/0/SittardGeleen.pdf [accessed 2014-01-10] [WebCite Cache ID 6MWFjsSXe]
35. Peters GJY, Abraham C, Crutzen R. Full disclosure: doing behavioural science necessitates sharing. European Health Psychologist 2012;14(4):77-84.
36. Jamil MS, Hocking JS, Bauer HM, Ali H, Wand H, Smith K, et al. Home-based chlamydia and gonorrhoea screening: a systematic review of strategies and outcomes. BMC Public Health 2013;13:189 [FREE Full text] [doi: 10.1186/1471-2458-13-189] [Medline: 23496833]
37. Woodhall SC, Sile B, Talebi A, Nardone A, Baraitser P. Internet testing for Chlamydia trachomatis in England, 2006 to 2010. BMC Public Health 2012;12:1095 [FREE Full text] [doi: 10.1186/1471-2458-12-1095] [Medline: 23253518]
38. Currie MJ, Schmidt M, Davis BK, Baynes AM, O'Keefe EJ, Bavinton TP, et al. 'Show me the money': financial incentives increase chlamydia screening rates among tertiary students: a pilot study. Sex Health 2010 Mar;7(1):60-65. [doi: 10.1071/SH08091] [Medline: 20152098]
39. Booth AR, Norman P, Harris PR, Goyder E. Using the theory of planned behaviour and self-identity to explain chlamydia testing intentions in young people living in deprived areas. Br J Health Psychol 2013 Feb 13:epub. [doi: 10.1111/bjhp.12036] [Medline: 23406510]
40. Mevissen FE, Ruiter RA, Meertens RM, Zimbile F, Schaalma HP. Justify your love: testing an online STI-risk communication intervention designed to promote condom use and STI-testing. Psychol Health 2011 Feb;26(2):205-221. [doi: 10.1080/08870446.2011.531575] [Medline: 21318930]
41. Rosenberg NE, Pettifor A, Kamanga G, Bonongwe N, Mapanje C, Hoffman I, et al. Social networks of STI patients have higher STI prevalence than social networks of community controls. Sexually Transmitted Infections 2013 Jul 13;89(Suppl 1):A65-A65. [doi: 10.1136/sextrans-2013-051184.0198]
42. Theunissen K, Hoebe C, Crutzen R, Niekamp A, Kara-Zaitri C, Vries ND, et al. A targeted Web-based Chlamydia Trachomatis screening strategy for testing in young people at risk using social and sexual networks. Sexually Transmitted Infections 2013 Jul 13;89(Suppl 1):A157-A157. [doi: 10.1136/sextrans-2013-051184.0490]
43. McAlister AL, Perry CL, Parcel GS. How individuals, environments,health behaviors interact: Social Cognitive Theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice (4th edition). San Francisco, CA: Jossey-Bass; 2008:169-188.

Abbreviations

CSI: Chlamydia Screening Implementation program
PHS: public health services
STD: sexually transmitted disease

Edited by G Eysenbach; submitted 26.08.13; peer-reviewed by P Norman, S Kalwij, M Currie; comments to author 22.10.13; revised version received 15.11.13; accepted 21.11.13; published 30.01.14.

Please cite as:
ten Hoor G, Hoebe CJPA, van Bergen JEAM, Brouwers EEHG, Ruiter RAC, Kok G
The Influence of Two Different Invitation Letters on Chlamydia Testing Participation: Randomized Controlled Trial
J Med Internet Res 2014;16(1):e24
URL: http://www.jmir.org/2014/1/e24/
doi:10.2196/jmir.2907
PMID:24480721

Original Paper

Engaging With a Wiki Related to Knowledge Translation: A Survey of WhatisKT Wiki Users

Deepa Mathew[1], MSc; K Ann McKibbon[1], MLIS, PhD; Cynthia Lokker[1], MSc, PhD; Heather Colquhoun[2], PhD

[1]Health Information Research Unit, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada
[2]Centre for Practice-Changing Research (CPRC), Ottawa Hospital Research Institute, Ottawa Hospital – General Campus, Ottawa, ON, Canada

Corresponding Author:
Cynthia Lokker, MSc, PhD
Health Information Research Unit
Department of Clinical Epidemiology and Biostatistics
McMaster University
CRL 125
1280 Main St W
Hamilton, ON, L8S 4K1
Canada
Phone: 1 905 525 9140 ext 22208
Fax: 1 905 546 0401
Email: lokkerc@mcmaster.ca

Abstract

Background: In 2008, WhatisKT wiki was launched as a collaborative platform for knowledge translation (KT) researchers and stakeholders to debate the use and definitions of KT-related terms. The wiki has definitions for over 110 terms from disciplines including health care, information technology, education, accounting, and business. WhatisKT wiki has over 115 registered users. Approximately 73,000 unique visitors have visited the wiki since 2008. Despite annual increases in visitors and regular maintenance of the wiki, no visitors have contributed content or started a discussion.

Objective: We surveyed wiki users to gain an understanding of the perceived value of the website, reasons for not engaging in the wiki, and suggestions to facilitate collaboration and improve the usability of the wiki.

Methods: We surveyed three cohorts: KT Canada members who were previously invited to join the wiki, registered wiki members, and unregistered visitors. The first two cohorts completed a Web-based survey that included the System Usability Scale (SUS) questionnaire to assess usability; additionally 3 participants were interviewed. Unregistered wiki visitors were surveyed with polls posted on the wiki. The study received ethics approval from the McMaster University Faculty of Health Sciences Research Ethics Board.

Results: Twenty-three participants completed the Web-based and SUS surveys; 15 participants indicated that they would collaborate on the wiki. The mean SUS score of 67 (95% CI 56-77) indicated that the wiki could be considered for design improvements. Study participants indicated that the wiki could be improved by email notification regarding new terms, better grouping of terms, user friendly interface, and training for users interested in editing content.

Conclusions: The findings from this survey will be used to enhance the design and content of WhatisKT wiki. Further feedback from participants will be used to make the wiki an ideal collaboration platform for KT researchers interested in terminology.

(J Med Internet Res 2014;16(1):e21) doi:10.2196/jmir.3001

KEYWORDS

knowledge translation; wiki; usability

Introduction

Knowledge Translation

Interest in knowledge translation (KT) has increased considerably in the past several years. Research communities use over 100 KT terms worldwide. Terms such as knowledge to action, knowledge transfer, knowledge exchange, research utilization, implementation, quality improvement, dissemination, and diffusion are often used by stakeholders in the field. A relatively widely used definition of KT was developed by the

Canadian Institutes of Health Research (CIHR), which defines KT as “the exchange, synthesis and ethically-sound application of knowledge—within a complex system of interactions among researchers and users—to accelerate the capture of the benefits of research for Canadians through improved health, more effective services and products, and a strengthened health care system” [1]. However, even health care funding agencies vary in the operational definition of KT [2].

The existence of multiple terms and definitions related to KT is a challenge for researchers trying to identify previous research and for communicating with others. Despite the increase in popularity of KT research, finding meaningful and consistent definitions for KT terms is a challenge [3]. McKibbon and colleagues reviewed over 2600 articles published in health care journals and identified 100 individual terms to describe KT research and a lack of consistency in the use of terms [4].

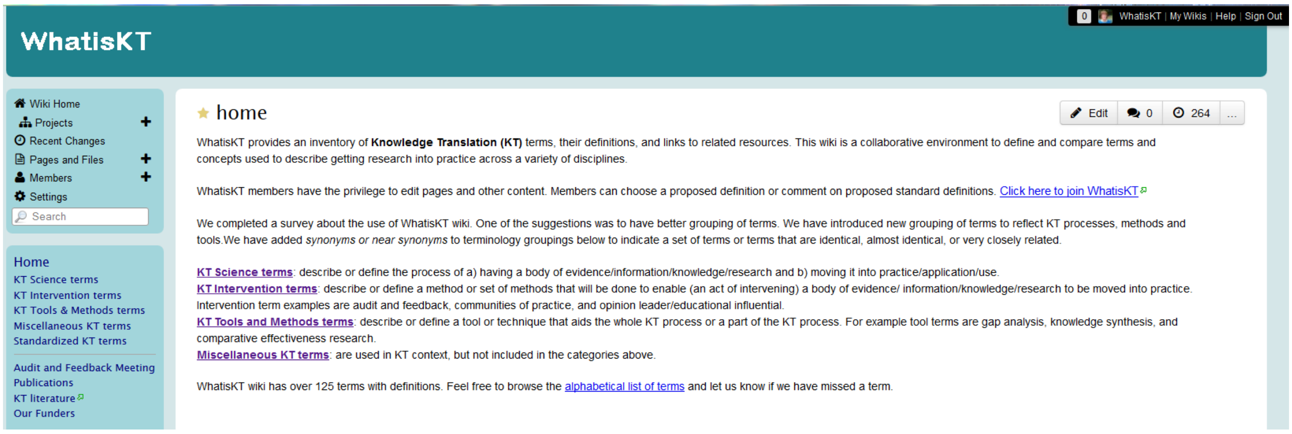
Wikis

Wikis are an online, Web-based, collaborative platform where anyone with proper access rights can modify content and contribute to online discussions. The most well-known wiki is Wikipedia. Wikis facilitate ease of documentation, learning, and ad-hoc collaboration, thereby developing online communities of practice and user groups, regardless of organizational affiliation [5]. Wikis have been used for collaborative content development for clinical decision support [6], and by patients contributing to the development of clinical practice guidelines [7] and providing input in developing care plans [8]. Some studies have researched how wikis are or could be used in health care, such as the intention of physicians to use social media including wikis to share medical content [9] and emergency care professionals’ beliefs around the use of a wiki to share reminders that promote best practices in trauma [10].

In 2008, WhatisKT wiki was launched as a collaborative platform for KT researchers and stakeholders to debate the use and definitions of KT-related terms (Figure 1). The wiki has over 110 terms with definitions from a variety of disciplines including health care, social sciences, information technology, education, accounting, and business. Each term has a dedicated page with a short description of definitions and the discipline where the term originated. In 2011, in an effort to generate discussion around terms and definitions, wiki organizers analyzed definitions for the 13 most frequently visited terms and, by a process of discussion and consensus, 2 team members selected one definition as being representative of that term. The definitions were selected based on having the most of 12 identified concepts within the CIHR definition of KT, clarity, comprehensiveness, reputation of source, and breadth of coverage [11]. The preferred published definition was presented at the top of the list of definitions on the wiki page for that term and highlighted. In September 2012, to provide better clarity and structure to the wiki, all terms were grouped into “Core KT” terms (eg, implementation research, quality improvement, research utilization) and “Additional KT” terms (eg, adoption, best practice, change). Core terms directly associate with the broad field of KT and moving knowledge into practice. Additional KT terms are those more specific terms used in the KT context but not representative of the field. The 13 terms with standardized definitions were grouped under “Standardized KT” terms (eg, research utilization, knowledge mobilization, diffusion of innovation). The wiki’s navigation bar facilitates access to core, standardized, and additional terms. The wiki also has links to sources of KT publications and literature.

Almost 73,000 unique visitors have visited the wiki since 2008. There are over 115 registered members. The average number of visitors per month to the wiki has increased steadily from about 196 in 2008, to 1900 in 2012 [11], and more than 2400 per month in 2013. The percentage of returning visitors also increased from less than 3% in 2008 to over 10% in 2012 [11]. The organizers regularly update the wiki, and the most recent changes are readily available for viewing by all wiki visitors. Despite the increase in visitors and regular maintenance of the wiki, no member has contributed source material or started a discussion on the wiki.

Figure 1. Screenshot of WhatisKT.



WhatisKT provides an inventory of **Knowledge Translation (KT)** terms, their definitions, and links to related resources. This wiki is a collaborative environment to define and compare terms and concepts used to describe getting research into practice across a variety of disciplines.

WhatisKT members have the privilege to edit pages and other content. Members can choose a proposed definition or comment on proposed standard definitions. Click here to join WhatisKT

We completed a survey about the use of WhatisKT wiki. One of the suggestions was to have better grouping of terms. We have introduced new grouping of terms to reflect KT processes, methods and tools.We have added *synonyms or near synonyms* to terminology groupings below to indicate a set of terms or terms that are identical, almost identical, or very closely related.

KT Science terms: describe or define the process of a) having a body of evidence/information/knowledge/research and b) moving it into practice/application/use.
KT Intervention terms: describe or define a method or set of methods that will be done to enable (an act of intervening) a body of evidence/ information/knowledge/research to be moved into practice. Intervention term examples are audit and feedback, communities of practice, and opinion leader/educational influential.
KT Tools and Methods terms: describe or define a tool or technique that aids the whole KT process or a part of the KT process. For example tool terms are gap analysis, knowledge synthesis, and comparative effectiveness research.
Miscellaneous KT terms: are used in KT context, but not included in the categories above.

WhatisKT wiki has over 125 terms with definitions. Feel free to browse the alphabetical list of terms and let us know if we have missed a term.

Usability

Usability can apply to any system or product with which a user interacts. Usability is defined as the “extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [12] and depends on factors such as the product design, ease with which users can learn the product, efficiency of the product in helping users achieve their objective, ease with which users can memorize product features for future use, and user satisfaction. Usability tests allow quantification of how well a product satisfies the user’s needs. In health care systems, usability studies can reveal use-related hazards [13]. Usability evaluations are usually conducted by experts and provide subjective data regarding user satisfaction. Surveys and questionnaires, interviews, focus groups, and user observation can be used to gather a range of usability information. The System Usability Scale (SUS) is a 10-item, 5-point Likert scale used for subjective assessment of usability [14]. Each item has a scale position of 1 for strongly disagree, 2 for disagree, 3 for neutral, 4 for agree, and 5 for strongly agree. SUS scoring uses transformed data so that the final output can range from 0 to 100 representing the overall usability of the system under evaluation.

In this study, we surveyed KT Canada members, registered wiki members, and unregistered wiki visitors to gain a better understanding of our users, especially their experience with the wiki. The objective of this study is to gather the following information about the wiki: user's perceptions, value, reasons for not engaging, suggestions to facilitate collaboration, and assessment of usability. The study was funded by KT Canada, which is an organization made up of Canadian researchers and educators working in the field of KT and is supported by funding from CIHR.

Methods

Web-Based Survey

To get a better idea of how and why people use the WhatisKT wiki, we surveyed three cohorts: KT Canada members, registered wiki members, and unregistered wiki visitors. A subset of survey participants agreed to a follow-up interview via Skype. The study received ethics approval from the McMaster Faculty of Health Sciences Research Ethics Board.

KT Canada members were invited to join the wiki soon after launch in 2008. Occasional email reminders of the wiki's existence have been sent to this group since that time. This cohort was invited to take part in an online survey via the KT Canada weekly newsletter December 12, 2012. Survey participants were asked to fill in the same Web-based questionnaire. To obtain more in-depth information on user perceptions, the participants were also asked if they would be willing to volunteer for a virtual interview using Skype or another Web-based communication tool.

The wiki currently has over 115 registered members. This cohort was sent a link to the Web-based survey via the wiki interface on November 26, 2012, with a reminder sent on December 13, 2012.

The survey (Multimedia Appendix 1) included questions regarding their membership status, frequency of visits, experience with other wikis, and likelihood of collaboration on WhatisKT wiki. To understand user perception of the wiki, participants were asked to choose from a list of descriptors that describes the wiki: “a reference tool for KT terms”, “a collaboration platform”, and “a source of KT literature and links”. Participants also answered questions about barriers to engaging in the wiki and features that would make wiki collaboration easier. Respondents were not required to access the wiki during the survey. Only those participants who indicated they had visited the wiki completed the SUS questionnaire to assess the usability of the wiki and assessed content-related statements (Textbox 1) scored on a 5-point Likert scale. The score contributions for items 1, 3, 5, and 7 are calculated by deducting 1 from the scale position. The score contributions for items 2, 4, 6, and 8 are calculated by deducting the scale position from 5. The SUS score is 2.5 times the sum of scores of all 10 items [10]. Sauro reported an average SUS score across 500 usability evaluations with data from over 5000 users was 68 [15]. A SUS score <68 indicates below average usability, and a value equal or higher indicates above average usability. Descriptive statistics were performed on survey data.

Textbox 1. System Usability Scale (SUS) and content-related survey questions for KT Canada and wiki members.

SUS items:

1. I think I would like to use this application frequently
2. I found the application unnecessarily complex
3. I thought the application was easy to use
4. I think I would need Tech Support help to use this application
5. I found the various functions in the application to be well integrated
6. I thought there was too much inconsistency in this application
7. I believe that most people would learn to use this application very quickly
8. I found the application very cumbersome to use
9. I felt very confident in using the application
10. I need to learn a lot about this application before I could effectively use it

Content-related items (not included in the SUS scoring):

1. It is easy to find the information that I need
2. The information provided by the wiki is easy to understand
3. The organization of information on the wiki is clear

Interview

KT Canada or registered wiki members who volunteered to be interviewed were contacted and scheduled for a short, online interview. A standardized list of questions including membership status, frequency of visits, and involvement with other wikis was asked during the interview. Interviewees provided their overall impression about the wiki and a letter grade for the wiki. In addition, the participants were asked to provide three words to describe the wiki and comment on the best and least preferred characteristic of the wiki. Interviewees also commented on open-ended questions such as "what are your suggestions to improve the wiki?" and "If you were the wiki administrator, what would be the first thing you would do to improve the wiki?" The audio component of each interview session was recorded and transcribed. The planned analysis was a directed content analysis [16] performed by DM. Interview questions were targeted based on usability and wiki research with the intent to describe perceptions of the wiki, barriers and facilitators to contributing to the wiki, and usability of the interface. Interviews were transcribed and the data were grouped according to the key concepts of perceptions, barriers, facilitators, and possible improvements. Themes were developed within these concepts and supported by quotes from the interviewees. CL reviewed the content analysis with any disagreements resolved through consensus.

Poll of Unregistered Visitors

Much of the wiki traffic is from unregistered visitors who come to the wiki via searches engines such as Google and from direct visits [11]. Short polls were displayed on a dedicated wiki page from January to March 2013 to capture responses from unregistered wiki visitors. The home page and highly accessed pages had prominent links to the poll page, which included two questions: (1) Which group describes you the best (researcher, clinician, decision/policy maker, educator, student, or other)? (2) What is the objective of your visit to the wiki (find definitions(s), find publication(s), learn more about KT, learn more about WhatisKT, none of the above)? Since our visitors tend to stay for a less than one minute, we opted to pose 2 short questions to get a sense of the types of unregistered visitors that view the wiki rather than limit responses with the burden of the longer survey.

Results

Web-Based Survey

Twenty-five KT Canada and registered wiki members began the Web-based survey over a period of approximately 50 days. Two sets of responses were incomplete and were excluded from analyses. Among the 23 respondents who completed the survey, most were researchers (Table 1) and 17 indicated that they had no prior experience using a wiki. The majority of the study participants became aware of WhatisKT wiki through KT Canada (n=12). Survey participants described the wiki as a reference tool (n=21), collaboration platform (n=10), and source of KT literature (n=9).

Sixteen participants indicated that they had used the wiki at least once and provided subjective usability assessment using the SUS questionnaire. The mean SUS score was 67 (95% CI 56-77), which was not statistically different from the average score of 68 reported from 500 usability evaluations [15] indicating that the wiki could be considered for design improvements. For content-related items, average score per question for ease of finding information was 3.69 (95% CI 3.67-3.70), ease of understanding information was 4 (CI 3.99-4.01), and clear organization of information was 3.63 (CI 3.61-3.64) out of a potential score of 5. Among the six participants who had previously collaborated on a wiki, only two visited WhatisKT wiki. They gave SUS scores of 97.5 and 60 and were not different in their ratings from non-wiki collaborators. Given the low number of responses, we were unable to look at subgroups of participants with and without wiki experience.

Fifteen participants indicated that they would likely collaborate on the wiki. However, two participants commented that time might be a significant barrier to collaboration. Other barriers included "complicated login" (n=1), "wiki edits are time consuming" (n=13), "lack of incentive" (n=3), "anxiety in editing the wiki" (n=4), and "need additional training" (n=4).

Participants indicated that "removal of user login" (n=5), "email notification regarding new terms or polls" (n=13), and "detailed tutorials" (n=8) could enhance collaboration. Six participants provided free-text comments on enhancing wiki collaboration such as "incentive to use wiki", "requirement to use the wiki and protected time to do so", and "email notifications on any new content such as new publications".

Participants also indicated that they would like to see latest KT publications (n=17), discussion forums (n=10), user ratings for standardized KT terms (n=8), and more terms related to KT (n=5). Study participants indicated that wiki could be improved by providing email notification regarding new terms, better grouping of terms, user friendly interface, and training for users interested in editing the wiki.

Table 1. Demographic group of online survey and poll participants.

Demographic group	Survey participants (n=23)	Poll participants (n=10)
Researcher	10	6
Clinician	2	0
Decision or policy maker	0	1
Educator	1	0
Student	1	1
Other	9	2

Interview

Six survey participants indicated a willingness to be interviewed but only 3, including a clinician and 2 researchers, returned follow-up emails and were interviewed via Skype. Two participants were wiki members. Two of the participants learned about the existence of WhatisKT wiki through word of mouth, and one participant found the wiki through online searching. The frequency of visits to the wiki ranged from couple of times since the wiki's launch to a couple of times a month. Two participants visited or contributed to other wikis. Two participants were from Canada, and one was from the United States. Content analysis produced the following themes: reference resource, positive impressions with room for improvement, limitations, engaging contributors, and suggestions for improvements.

Perceptions

All participants mentioned that the best characteristic of the wiki was the collection of definitions. It was seen as a valuable resource when writing manuscripts and for teaching purposes: "I find it very helpful to use as a reference point when I try to define terms in the KT/KM field to other people" and "So what is most useful for us is a collection of different definitions that also comes with a citable source".

Overall, participants described the wiki as "linkable, helpful, and health focused", "innovative, open, and complete", and "informative, not as much used, and efficient". A theme that emerged was that the wiki is viewed positively, but that it still needs work. Their overall impression of the wiki was positive, but when asked to rate the wiki they gave it a B or B– and 6 or 7 (out of 10). Comments included:

> *It is good—a great tool. It has a lot of visits. When you Google "Knowledge Translation" it is the first or second link that comes up.*
>
> *B–because it answers the question, "What is KT?" But, may be if it was broadly used or had more diverse opinions or more discussion, I am not sure what is missing. I guess there is room for improvement. It is pretty good.*
>
> *I think we really like it as a repository especially for different definitions. Where we disagree a bit—or where we think is too early—is that lot of the work that has been done on consensus and finding a definition that is agreed—because we think it is an evolving field.*
>
> *My impression is that it is a wiki designed to collect, synthesize and create consensus about the definition of "What is KT"…and how people interpret the term implementation science, knowledge translation, knowledge-to-action and all sorts of varied terms that are used for the definition of implementation science and knowledge translation.*

Participants mentioned that the Canadian perspective, health focus, and theoretical content were the least preferred characteristics of the wiki. The limitations in what the wiki provides emerged as a separate theme:

> *It's more health focused than the work I am doing, so with sort of that lens in mind it is not always totally appropriate", "…the definitions that are up there have a very Canadian flavor about them.*
>
> *I consulted it in the past, I felt it a very interesting resource with many terms for KT, but beyond that I might not have had the time to push to analyze what the KT was used for. I didn't feel any incentive to add more to what was already there.*

Possible Improvements

Participants suggested expanding the wiki's content to include definitions from various disciplines and research communities. Another suggestion for improvement was to include practical KT tools. All three participants mentioned the need to engage the users and increase their participation in the wiki. However, two participants mentioned that lack of incentives might be a reason why members are not contributing to the wiki. These suggestions could be used in generating user interaction with the wiki in the future.

> *Trying to improve visibility, but also getting more content and more editing going on which involves forming active community. An interesting way to do it would be to have an ongoing cohort of graduate student class taking some sort of knowledge translation class curating the wiki which would keep it cleaned up.*
>
> *What would be useful is to show difference and to show the different definitions that are out there—not a consensus—showing the variety of terms and variety in definitions.*

Barriers and Facilitators

Participants were specifically asked to describe barriers and facilitators for contributing to the wiki. Incentives, providing credit, developing critical mass, and defining a clear goal for the wiki emerged as ideas.

> *The reasons for not contributing: Not paid. We spend so much time worrying about full publication that we do not have the energy to contribute to another outlet.*
>
> *People wondered—is there a critical mass involved that contribute to it? KT is an evolving field—too early to agree on definition.*
>
> *I would try to see how I can engage the community that you are trying to get people to contribute to; to find the incentive that would encourage people to contribute to it, find ways of giving credits; continuing professional development credits for contributing to the wiki or ways of creating some sort of interest in getting people to contribute; recognizing contributions to it; ways of getting people to contribute to it.*
>
> *Trying to engage the target knowledge users that you are aiming to get involved in the wiki so they understand where the [WhatisKT] wiki is going and what is the end goal of the wiki.*

Poll

Over the course of 2 months, we received only 10 responses from unregistered wiki visitors through the polls posted on the wiki. During that time, Google analytics tracked 2929 unique visitors, so the response rate was extremely low at 10/2929 (0.3%). Six respondents were researchers, 1 decision/policy maker, 2 students, and 2 identified as “other”. Participants indicated that the objective of their visit was to learn more about KT (n=6), learn more about WhatisKT (n=2), find definitions (n=1), and none of the above (n=1).

Discussion

Principal Results

In our Web-based survey, 16 participants provided a subjective assessment of the wiki’s usability via the SUS questionnaire, which provides the most reliable results with sample sizes of at least 12 to 14 participants [17]. The usability results from our study can be considered reliable despite the small sample size compared to typical clinical studies, though the confidence interval for the mean SUS score was fairly wide (56% to 77%). The online poll had only 10 responses over a period of 2 months, a response rate of only 0.3% of unique visitors. The low response rate is lower than studies involving recruitment of research participants through the Internet [18]. We had hoped to provide a pop-up poll to all unregistered visitors to maximize awareness of the poll and participation. Unfortunately, the wikispaces platform is not designed to accommodate such polls. Our small sample size for the polls, surveys, and interviews did not allow us to reach saturation, which is a primary limitation.

More researchers participated in both the online survey and the poll than any other demographic group. This could be an indication that researchers are prominent stakeholders interested in the wiki. The majority of the survey participants indicated that WhatisKT wiki could be best described as a reference tool, which was reiterated by interviewees. On the other hand, poll results indicate that the primary objective for some visitors is to learn more about KT and very few (10%) indicated that they were interested in the definitions, though with a sample size of 10, this observation is limited. These visitors could have reached the wiki through a search engine while searching for a KT-related term or concept. Such visitors could expect more KT-related content than just a repository of terms.

Despite the fact that one of the goals of our work was to understand the lack of uptake for the discussion function of the wiki, we were unable to resolve this issue. Although 10 respondents indicated an interest in discussion forums on the wiki, a functionality that is already available, no discussion has yet been started despite the wiki being 5 years old. This is not an unusual situation in collaborative writing projects, and frequently reported barriers include unfamiliarity and lack of skills with the technologies, time and work constraints, and concerns about quality of contributions, and legal ramifications [18]. Since participants indicated interest in these forums, we will continue to encourage and develop this function of the wiki. Barriers to contributing to discussion forums or content identified in the survey include time, issues related around the technology (complicated login and need for training), and lack of incentives. We will be considering how we can overcome some of these barriers.

Previous research has identified a number of potential facilitators such as training, ease of use of the system, having a moderator or champion to monitor the content and ensure quality, having a critical mass of content providers, creating a community of practice or learners, and providing incentives [5]. Currently, the wiki offers no incentive for edits although users are able to tag their contributions to identify themselves. For clinical members, continuing medical education credits could be provided as an incentive. For students, instructors could be encouraged to include wiki contributions as a required task, as was suggested by one of our interviewees. This guided interaction with the wiki could also generate the community of learners as a step towards reaching critical mass. For researchers, however, incentives and credit are more difficult to devise since their focus is on peer-reviewed publication and receiving grants. The specific facilitators or barriers to interactivity is something that requires further research and would benefit from moving beyond asking people if they are interested in interactivity to determining the specific conditions in which interactivity occurs.

Our study found that the wiki is being used as a reference tool. Survey participants were pleased with the wiki content and its quality. This was evident from the above average score given for organization, ease of finding, and ease of understanding the information provided on the wiki. However, the wiki users preferred better grouping of terms and having a friendly user interface. The term groupings, ”Core”, “Additional” and “Standardized” terms, could be re-organized into more intuitive groupings to facilitate ease of reference of terms and definitions in the KT domain. One limitation of wikis is that as the site is presented one page at a time, it can be difficult to get the overall picture of the wiki [19]. The wiki can be modified with easily accessible groupings of terms and an alphabetical list to provide a comprehensive view to all visitors.

Email notification regarding new terms or polls was mentioned as a feature that will enhance collaboration. In the future, it would be ideal to send wiki members email notifications when new terms or polls are added. Addition of latest KT publications can increase potential visits to the wiki.

Limitations

Our primary limitation was the low response rate, which reduces the generalizability of the findings. This limitation demonstrates a common problem inherent in research in online tools [5]. However, this study comprises some of the first user-based knowledge summaries related to collaborative platforms. Given the growth in the use of these technologies, advancing what is known about optimizing these types of platforms is needed. By their nature, wikis present technical limitations for understanding “best wiki practices” through user engagement. We experienced technical limitations related to the wiki platform; our poll was included as a link on select pages in the wiki rather than a pop-up visible to all visitors. A pop-up containing the poll questions, for all visitors regardless of landing page, could have resulted in better response rate from wiki visitors. Additionally, since we anticipated fewer response rates from unregistered wiki visitors and were concerned about burden, we posted only two questions. A more elaborate poll could have probed into the specific needs and usability concerns of non-registered visitors but at the potential risk of response rate. Improved understanding of weighing these issues in online data collection is needed. Due to the nature of using a commercial wiki platform, we are unable to describe our cohort. This further limits our ability to contextually analyze the data that we collect. The present study described the usability assessment from KT Canada and registered wiki members. The general wiki visitor was not provided the opportunity for usability assessment. An elaborate survey targeting all wiki visitors could give an indication of the issues and needs of the general wiki visitor.

Future Work

In response to this survey, enhancements have already been made to the wiki. For example, we have regrouped our KT terms into categories: KT science terms, KT intervention terms, KT tools and methods, and miscellaneous terms based on consensus within our team. KT science terms describe or define the process of (1) having a body of evidence, information, knowledge, research and (2) moving it into practice, application, use. KT intervention terms describe or define a method or set of methods that will be done to enable (an act of intervening) a body of evidence, information, knowledge, research to be moved into practice. Intervention term examples are audit and feedback, communities of practice, and opinion leader/educational influential. KT tools and methods terms describe or define a tool or technique that aids the whole KT process or a part of the KT process. For example, tool terms are gap analysis, knowledge synthesis, and comparative effectiveness research. Miscellaneous KT terms are used in KT context, but not included in the categories above.

However, content changes will not be enough to improve activity on the wiki. We will look at the previously reported facilitators and determine which might be best to generate contributions to the wiki. We have started discussing ways to facilitate improved interaction with the wiki, including conversations with graduate course providers, but these actions have not yet been implemented. Also, by having a dedicated staff person to moderate and verify any new content, we can ensure the quality of the information used to build the wiki.

Conclusions

This study has shown that WhatisKT wiki is being used mainly as a reference tool by users. The wiki scored reasonable for usability, but study participants indicated a number of barriers and facilitators to adding content and contributing to discussion forums. We will focus our efforts on improving the usability and testing strategies to remove barriers.

Acknowledgments

Funding for this project was provided by the Canada Institutes of Health Research (FRN-86258).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Web survey completed by KT Canada members and registered wiki members.

[PDF File (Adobe PDF File), 58KB - jmir_v16i1e21_app1.pdf]

References

1. Canadian Institutes of Health Research. Knowledge Translation Strategy 2004-2009. URL: http://www.cihr-irsc.gc.ca/e/26574.html [accessed 2013-09-30] [WebCite Cache ID 6K1e0Bblz]
2. Tetroe JM, Graham ID, Foy R, Robinson N, Eccles MP, Wensing M, et al. Health research funding agencies' support and promotion of knowledge translation: an international study. Milbank Q 2008 Mar;86(1):125-155 [FREE Full text] [doi: 10.1111/j.1468-0009.2007.00515.x] [Medline: 18307479]
3. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
4. McKibbon KA, Lokker C, Wilczynski NL, Ciliska D, Dobbins M, Davis DA, et al. A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: a Tower of Babel? Implement Sci 2010;5:16 [FREE Full text] [doi: 10.1186/1748-5908-5-16] [Medline: 21080976]
5. Archambault PM, van de Belt TH, Grajales FJ, Faber MJ, Kuziemsky CE, Gagnon S, et al. Wikis and collaborative writing applications in health care: a scoping review. J Med Internet Res 2013;15(10):e210 [FREE Full text] [doi: 10.2196/jmir.2787] [Medline: 24103318]
6. Wright A, Bates DW, Middleton B, Hongsermeier T, Kashyap V, Thomas SM, et al. Creating and sharing clinical decision support content with Web 2.0: Issues and examples. J Biomed Inform 2009 Apr;42(2):334-346. [doi: 10.1016/j.jbi.2008.09.003] [Medline: 18935982]
7. den Breejen EM, Nelen WL, Knijnenburg JM, Burgers JS, Hermens RP, Kremer JA. Feasibility of a wiki as a participatory tool for patients in clinical guideline development. J Med Internet Res 2012;14(5):e138 [FREE Full text] [doi: 10.2196/jmir.2080] [Medline: 23103790]

8. Gupta S, Wan FT, Newton D, Bhattacharyya OK, Chignell MH, Straus SE. WikiBuild: a new online collaboration process for multistakeholder tool development and consensus building. J Med Internet Res 2011;13(4):e108 [FREE Full text] [doi: 10.2196/jmir.1833] [Medline: 22155694]
9. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
10. Archambault PM, Bilodeau A, Gagnon MP, Aubin K, Lavoie A, Lapointe J, et al. Health care professionals' beliefs about using wiki-based reminders to promote best practices in trauma care. J Med Internet Res 2012;14(2):e49 [FREE Full text] [doi: 10.2196/jmir.1983] [Medline: 22515985]
11. McKibbon KA, Lokker C, Keepanasseril A, Colquhoun H, Haynes RB, Wilczynski NL. WhatisKT wiki: a case study of a platform for knowledge translation terms and definitions--descriptive analysis. Implement Sci 2013;8:13 [FREE Full text] [doi: 10.1186/1748-5908-8-13] [Medline: 23347357]
12. International Organization for Standardization. ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability. 1998. URL: http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf [accessed 2013-09-30] [WebCite Cache ID 6K1eQRU8P]
13. Fairbanks RJ, Caplan SH, Bishop PA, Marks AM, Shah MN. Usability study of two common defibrillators reveals hazards. Ann Emerg Med 2007 Oct;50(4):424-432. [doi: 10.1016/j.annemergmed.2007.03.029] [Medline: 17498847]
14. Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, editor. Usability evaluation in industry. London: Taylor & Francis; 1996.
15. Sauro J. Measuring Usability. 2011 Feb 2. Measuring Usability With The System Usability Scale (SUS) URL: http://www.measuringusability.com/sus.php [accessed 2013-09-30] [WebCite Cache ID 6K1eeanhu]
16. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
17. Tullis T, Stetson J. A comparison of questionnaires for assessing website usability. 2004 Jun Presented at: Proceedings of the Usability Professional Association Conference; June 7-11, 2004; Minneapolis, MN URL: http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf
18. Im EO, Chee W. Recruitment of research participants through the Internet. Comput Inform Nurs 2004;22(5):289-297. [Medline: 15520599]
19. Désilets A, Paquet S, Vinson NG. Are wikis usable? In: International Symposium on Wikis. 2005 Oct 16 Presented at: Wiki Sym '05; October 16-18, 2005; San Diego, CA p. 3-15 URL: http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=8913757&lang=en [doi: 10.1145/1104973.1104974]

Abbreviations

CIHR: Canadian Institutes of Health Research
KT: knowledge translation

Edited by G Eysenbach; submitted 03.10.13; peer-reviewed by M Hamm, S Scott, P Archambault, E Santoro; comments to author 30.10.13; revised version received 02.12.13; accepted 21.12.13; published 21.01.14.

Please cite as:
Mathew D, McKibbon KA, Lokker C, Colquhoun H
Engaging With a Wiki Related to Knowledge Translation: A Survey of WhatisKT Wiki Users
J Med Internet Res 2014;16(1):e21
URL: http://www.jmir.org/2014/1/e21/
doi:10.2196/jmir.3001
PMID:24449712

Review

The Effects on Health Behavior and Health Outcomes of Internet-Based Asynchronous Communication Between Health Providers and Patients With a Chronic Condition: A Systematic Review

Catharina Carolina de Jong[1,2], MSc Nursing; Wynand JG Ros[1], PhD; Guus Schrijvers[1], PhD

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands
[2]Stichting Transmurale Zorg Den Haag EO, The Hague, Netherlands

Corresponding Author:
Catharina Carolina de Jong, MSc Nursing
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
University of Utrecht
PO Box 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 623908387
Fax: 31 887568099
Email: c.c.dejong-8@umcutrecht.nl

Abstract

Background: In support of professional practice, asynchronous communication between the patient and the provider is implemented separately or in combination with Internet-based self-management interventions. This interaction occurs primarily through electronic messaging or discussion boards. There is little evidence as to whether it is a useful tool for chronically ill patients to support their self-management and increase the effectiveness of interventions.

Objective: The aim of our study was to review the use and usability of patient-provider asynchronous communication for chronically ill patients and the effects of such communication on health behavior, health outcomes, and patient satisfaction.

Methods: A literature search was performed using PubMed and Embase. The quality of the articles was appraised according to the National Institute for Health and Clinical Excellence (NICE) criteria. The use and usability of the asynchronous communication was analyzed by examining the frequency of use and the number of users of the interventions with asynchronous communication, as well as of separate electronic messaging. The effectiveness of asynchronous communication was analyzed by examining effects on health behavior, health outcomes, and patient satisfaction.

Results: Patients' knowledge concerning their chronic condition increased and they seemed to appreciate being able to communicate asynchronously with their providers. They not only had specific questions but also wanted to communicate about feeling ill. A decrease in visits to the physician was shown in two studies (P=.07, P=.07). Increases in self-management/self-efficacy for patients with back pain, dyspnea, and heart failure were found. Positive health outcomes were shown in 12 studies, where the clinical outcomes for diabetic patients (HbA1c level) and for asthmatic patients (forced expiratory volume [FEV]) improved. Physical symptoms improved in five studies. Five studies generated a variety of positive psychosocial outcomes.

Conclusions: The effect of asynchronous communication is not shown unequivocally in these studies. Patients seem to be interested in using email. Patients are willing to participate and are taking the initiative to discuss health issues with their providers. Additional testing of the effects of asynchronous communication on self-management in chronically ill patients is needed.

(J Med Internet Res 2014;16(1):e19) doi:10.2196/jmir.3000

KEYWORDS

chronic disease; telecommunications; Internet; telemedicine; health services; delivery of health care; medical informatics; electronic mail; self-care; self-efficacy

Introduction

Background

Self-management is a central concept of health care that is increasing in popularity. This is because people strive to be autonomous and also because there is an increase in the global population, in the number of chronic diseases people have, and in the length of time people endure a chronic disease [1]. To make health care attainable for all, it is necessary to implement self-management on as large a scale as is justifiable. The level of patient participation is a key aspect in the realization of self-management. Research shows that patient participation results in improved adherence and health outcomes [2-4]. It is therefore important to understand what makes patients participate in their own health care.

The transition towards self-management is also visible in the new definition of health that experts have recommended to the World Health Organization (WHO): “the ability to adapt and self-manage in the face of social, physical, and emotional challenges” [5]. The earlier definition, which dates to 1948 [6], describes health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”, which today seems unattainable. Ursum gives a clear operationalization of self-management as “the individual ability of a person to prevent health problems from arising or, if they do arise, to adapt to the symptoms, the treatment, the physical, psychological and social consequences of the health problem and adjust their lifestyle. In this way persons are able to monitor their own health and to react in a way that results in a satisfactory quality of life” [7]. Thus, self-management is an ongoing process in the lives of people, even when they are not receiving care from health care providers. It requires that all people, but certainly patients, need to have open communication with health care providers to be able to take responsibility. This definition suggests that an important focus of health care research should be evidence-based ways of improving self-management by patient participation [8].

Chronic Patients and Modern Communication Technology

Modern information and communication technologies (ICT) provide new ways for patients to participate in their own health care. Internet interventions have been developed to record, measure, monitor, and manage the delivery of health care [9]. These interventions enable patients to remotely supply providers with personal health information and for providers to remotely deliver instructions. For instance, patients with cardiac problems can play an active role in diagnosis by monitoring and communicating their health data so that the professional can make a diagnosis [10,11]. Another option is to give patients access to specific interactive self-care techniques such as back pain management training. A third option is inviting the patient to view their electronic health record (EHR). The patient can log in to the EHR and comment on the content. Some evidence shows that interventions via the Internet also improve self-care behavior and health outcomes in patients with chronic diseases [12-16]. Especially in the case of chronically ill patients living at home, these Internet-based interventions are seen as a promising development to improve the quality and safety of health care [17]. However, robust research on the effectiveness and consequences of these interventions is needed to guide large-scale implementation [18,19].

Specifically Asynchronous Communication

The above-mentioned Internet interventions are created according to a standard pattern based on one-way communication with not much scope for interaction. However, Internet interventions can also be combined with interactive communication tailor-made for individual patients [12,20,21]. Interactive communication can be synchronous (concurrent by telephone consultation or videoconferencing) or asynchronous (non-concurrent by, for example, email or discussion board). This latter form of digital communication has the advantage that the patient and the provider do not need to use it concurrently. For example, patients can pose a question about the organization of their care or a health concern at the moment that it worries them and do not have to wait until the next planned consultation. For health care professionals, it means they can react to patient questions at a time that is suitable for them. Asynchronous communication is not suitable in urgent situations because there is a time gap in the communication. The asynchronous options make it possible to deliver tailor-made self-management support to large numbers of patients with a chronic disease [20].

Current Investigation

This review examines publications that describe the effect of digital asynchronous communication between chronically ill patients and health care providers. The first research question is whether this type of communication works: do patients and providers actually use this form of interactive communication and how do they evaluate the usability? The second question reviewed is whether this form of interactive communication helps: does it have an effect on health behavior, health outcomes, and patient satisfaction?

Methods

Definitions

Digital asynchronous communication is defined as electronically mediated communication in which the participants do not communicate concurrently. Examples of asynchronous communication in health care are electronic messaging (email) and bulletin boards. Patients with chronic conditions have one or more chronic diseases, which are defined as diseases with a long duration and generally slow progression [1].

Literature Searching Methods

The systematic review was conducted using the PICO method [22]. The keywords (MeSH terms [Medical Subject Headings]) used were chronic disease, telecommunications OR Internet OR telemedicine OR health services OR delivery of health care OR medical informatics OR electronic mail, self-care, self-efficacy. The search was filtered for Randomized Controlled Trials (RCTs), adults, English language, and publication period of 2001-2013. The search was limited to studies conducted from 2001 onward because Internet access for individuals from their homes has increased since the turn of the century [23] and interactive asynchronous communication thus became an option for more people. The search procedure consisted of the following steps:

1. PubMed and Embase databases were searched.
2. Duplicates were removed.
3. Titles and abstracts were scanned for Internet-based interventions.
4. Full text analysis was undertaken to select studies
    - describing asynchronous communication between patient and provider, alone or as part of an intervention;
    - where patients were able to initiate communication at any time of the day (24/7);
    - directed at self-management;
    - where control groups were free of any digital intervention and received usual care.
5. The bibliographies of the articles included were manually searched to identify additional relevant articles.

Quality Appraisal

The methodological quality of the studies was evaluated by applying the National Institute for Health and Clinical Excellence (NICE) criteria for RCTs [1,24,25]. Selection bias, performance bias, attrition bias, and detection bias were assessed.

Research Questions

Do Patients and Providers Use Asynchronous Communication Within Internet-Based Interventions and Do They Find It Usable?

To answer this question, the use of asynchronous communication in digital interventions by patients was assessed by determining frequency of access; number of patients who use the digital intervention; number of messages; and usability of interventions, including asynchronous communication (this refers to experiences regarding the actual (technical) use of the ICT with a focus on ease of use, clarity, and attractiveness).

Does the Use of Asynchronous Communication Within Internet-Based Interventions Affect Health Behavior and Health Outcomes?

To answer this question, the present study assessed the effects of using digital interventions, including asynchronous communication, on health behavior (knowledge, health care utilization, and self-efficacy/self-management); health outcomes (clinical parameters, physical symptoms, quality of life); and patient satisfaction.

The results were reported as significant if $P<.05$.

Results

Characteristics of the Studies

As shown in Figure 1, the search identified 311 studies in PubMed and 231 in Embase. Four additional studies were retrieved via the bibliographies of the retrieved studies. The elimination of duplicates resulted in 385 studies. All abstracts were screened and 27 studies were identified that specifically focused on Internet-based intervention. After full text analysis, 20 studies were identified that described asynchronous digital communication between patients and providers as part of the intervention. Furthermore, five studies in which the control group received usual care via the Internet were eliminated. Ultimately, 15 studies were found to meet the inclusion criteria and were thus included in the review (Figure 1).

Table 1 shows the characteristics of the samples of the studies. The studies varied in geographic location, sample size, mean age of the sample, and nature of the chronic disease (Table 1). Eleven of the 15 studies were performed in the United States, two in Europe (Portugal, The Netherlands), one in Australia, and one in Asia (Korea). The sample sizes varied from 0-50 [26], 50-100 [9,27-29], 100-150 [30-34], and more than 150 [35-39].

In 9 studies, the mean age of participants was approximately 50 years (range 45-57) [28-31,34-36,38,39]. In four studies, the mean age was over 60 years [9,27,32,33], and in two studies, the mean age was lower (range 29-36) [26,37]. The nature of the chronic disease differed in the samples, including unspecified chronic illnesses [30,35,38,39], chronic pain [27,36], diabetes [9,29,31,32], asthma [26,37], chronic obstructive pulmonary disease [33], chronic neurological conditions [28], and congestive heart failure [34].

Table 2 shows the characteristics of the interventions of the studies. The interventions were diversely directed. Eight studies focussed on self-care techniques [27,28,30,33,35,36,38,39], six on monitoring disease and symptoms [9,26,29,31,32,37], and three on sharing an EHR [29,31,34].

An assessment was carried out to determine whether the interventions were complementary or a substitute for usual care because of the consequences for the interpretation of the results. One might evaluate interventions that are complementary to usual care as effective if the study shows improved outcomes, whereas interventions that are a substitute for usual care can be evaluated as effective if the study shows no differences compared with usual care. Most interventions were complementary (n=11), but some were a substitute for usual care (n=4).

Table 1. Characteristics of the sample.

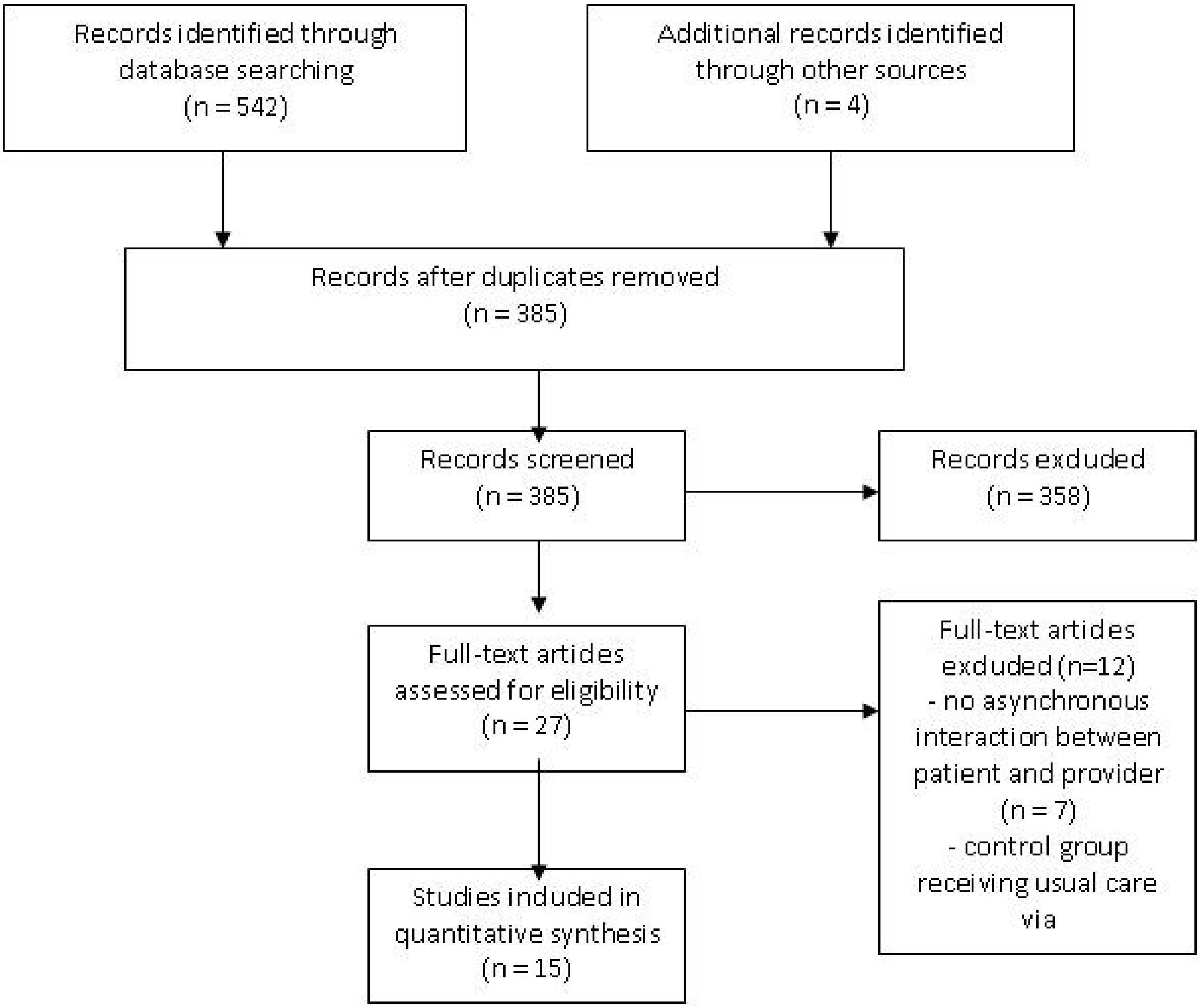
Primary author	Country	Sample size[a], n	Chronic disease	Mean age, yrs (range)	Female, %	Recruitment
Berman [27]	United States	I: 41 C: 37	Chronic pain	65.8 (55-91)	87.2	Mass media
Bond [9]	United States	I: 31 C: 31	Diabetes, type 1 and 2	67.2	45	Provider
Cruz [26]	Portugal	Cross-over I: 21	Asthma	29 (18-62)	71	Provider
Ghahari [28]	Australia	IA: 34 IB: 28 C: 33	Chronic neurological conditions: MS, Parkinson's, post-polio	50.25 (23-90)	81.1	Mass media
Hill [30]	United States	I: 61 C: 59	Chronically ill rural women	52.2 (35-65)	100	Mass media
Kwon [31]	Korea	I: 51 C: 50	Diabetes type 2	54.1	30.9	Provider
Lin [35]	United States	I: 305 C: 301	Acad. internal medicine practice	51	51	Provider
Lorig [36]	United States	I: 296 C: 284	Chronic back pain	45.5	38.5	Mass media
McMahon [32]	United States	I: 52 C: 52	Diabetes, HbA1c ≥9	63.5	0	Provider
Meer [37]	Netherlands	I: 101 C: 99	Asthma	36.5 (19-50)	69.5	Provider
Nguyen [33]	United States	IA: 43 IB: 41 C: 41	COPD	68.7	46	Mass media
Ralston [29]	United States	I: 42 C: 41	Diabetes, type 2, HbA1c≥7	57.3	49.4	Provider
Ross [34]	United States	I: 54 C: 53	Congestive heart failure	56	77	Provider
Weinert [39]	United States	IA: 54 IB: 58 C: 64	Chronically ill rural women	51.8 (30-69)	100	Mass media
Weinert [38]	United States	I: 155 C: 154	Chronically ill rural women	55.5	100	Mass media

[a] I=intervention, C=control.

Table 2. Characteristics of the intervention.

Primary author	Intervention				Complement or substitute for usual care	Length, wks
	Self-care technique	Monitor disease/ symptom	Share EHR	Description		
Berman [27]	x			Online mind-body: Facilitator sends prompts and answers questions using asynchronous communication	complement	6
Ghahari [28]	x			A. Online fatigue self-management program: Facilitators logged on daily and responded to participant entries, posed questions and provided encouragement to the intervention group. B. Online information-only self-management program: In the information-only group, facilitators checked for technical problems and sent standardized weekly reminders to read the information.	complement	13
Hill [30]	x			Computer intervention on psychological status: Online peer-led support group with health teaching; included an asynchronous chat room and an email function giving access to each other and to research team.	complement	22
Lin [35]	x			Patient portal: To send secure messages directly to their physicians and to request appointments, prescription refills and referrals; also included clinical messaging.	complement	26
Nguyen [33]	x			A. Internet-based dyspnea self-management program (eDSMP): Web diary, personalized reinforcement, feedback emails, and a discussion board. B. Face-to-face fDSMP: Paper diary, personalized reinforcement, and feedback telephone calls	complement	52
Weinert [38]	x			Peer-led support group and a self-study health coaching unit	complement	24
Weinert [39]	x			A. Computer-based intervention providing online support and health information B. Health information only	complement	53
Lorig [36]	x			Closed, moderated email discussion group; book, videotape.	complement	52
Bond [9]		x		Web-based intervention with disease monitoring, coaching, motivational, and social support	complement	26
Ralston [29]		x	x	Shared electronic record and email with providers	complement	52
Ross [34]			x	Web-based online review of EHR and email messaging directed at clarifying doctor's assessment and instructions	complement	52
Cruz [26]		x		E-diaries and asthma self-management with PIKO-1 e-tool: Immediate feedback using secure messaging	substitute	8
McMahon [32]		x		Care website with educational modules, monitoring system, and internal messaging system	substitute	52
Meer [37]		x		Internet-based self-management plus education and communication with an asthma nurse	substitute	52
Kwon [31]		x	x	Website for monitoring and communication	substitute	13

Figure 1. Search results.



Methodological Quality of Studies Included

In Multimedia Appendix 1, the methodological quality of the RCTs in this study is presented.

Risk for selection bias, or systematic differences between the comparison groups, existed in three studies [30,36,38]. Two studies did not describe how the participants were randomized and up to which point the investigators were blinded to allocation [30,36]. Two studies did not describe how the groups compared at baseline [30,38]. In all three studies, the participants were invited through mass media, which means that the assignment of patients to a group was partially concealed because the investigators did not know the participants [30,36,38]. The partial concealment was not considered a high risk for bias.

Risk for performance bias, or systematic differences between groups in the care provided (apart from the intervention under investigation), was low. In all studies, the comparison groups received the same care except for the intervention studied. In all 15 studies, it was clear that participants were not blinded to the treatment allocation due to the physical character of the intervention. In the studies by Ghahari [28], Nguyen [33], and Weinert [38], where two interventions (intense and less intense) were compared with usual care, patients in the intervention groups knew that they were taking part in the intervention but did not know whether they were participating in the intense or less intense intervention group.

The investigators were kept blinded in three studies. In the study by Bond [9], it was possible to keep the investigators blinded because the outcomes were measured at the beginning and the end of the intervention during a home visit. Cruz [26] used a crossover design in which all participants took part in the treatment and control group sequentially so that blinding of investigators was not an issue. In the study by Ghahari [28], the investigators were partially blind to how the outcomes were ascertained because they were involved only in one arm of the three-armed study and were blind to the survey results.

Risk of attrition bias, or systematic differences between the comparison groups with respect to dropouts, hardly existed in the studies. All groups within the studies were followed for an equal period of time. They were comparable with respect to the availability of outcome data and for treatment completion. Only Weinert [38] did not describe data on treatment completion.

Risk of detection bias, or bias in how outcomes are ascertained, diagnosed, and verified, showed more diversity. The length of follow-up was found to be on the short side in Berman (6 weeks) [27] and Cruz (8 weeks) [26]. The validity and reliability of the outcome measures was unclear in Cruz [26] because the psychometric qualities were not discussed. In Weinert [38], the quality of the outcome measures was not described. In three studies [9,28,30], the investigators were kept blind to the participants' exposure to the intervention and to other important confounding prognostic factors.

In conclusion, a total score for the studies was derived that summed up the risk for the four categories of bias, showing that Berman [27], Hill [30], and Weinert [38] have some risk to methodological quality. The other 12 studies show low risk.

Does It Work? Use and Usability

In Multimedia Appendix 2, the results describing the use of the intervention, health behaviour, and health outcomes are presented. Although all 15 studies studied an intervention including asynchronous communication, outcomes concerning use of the intervention, such as accessing the intervention and the use of electronic messaging, were not reported in all studies. Twelve studies described the use of the intervention by patients. In six studies, the specific use of the asynchronous communication was also subject of the study.

To use the asynchronous communication, the patients accessed a website and then logged in to a patient portal. It was then possible for them to communicate asynchronously with their providers by using email, viewing their EHR, or using a discussion board. Three studies that examined sharing EHR [29,31,34] described data about how the intervention was accessed. These patients accessed the intervention between 1.5 times a month [29,34] and 16 times a month [31]. Ross [34] reports that the use declined and leveled off from 1.5 times per month in the beginning to 0.4 times per month after 12 months. In two studies, the percentage of the sample that used the intervention was reported as 76% (32/43) [29] and 80% (43/54) [34]. The pages in the EHR that were the most reviewed by patients were the clinical notes and the lab results [29,34].

The accessing of the intervention was also reported by Berman [26], who found that patients accessed their intervention 16 times a month. Lin reported that the percentage of the sample accessing the intervention was 31% (95/305) [35], and Nguyen reported that 75% of patients (28/37) accessed the intervention [33].

The use of electronic messaging was reported on in six studies [26,29,31,34-36]. Three were publications on sharing EHR [29,31,34], two were on self-care techniques [35,36], and one was on monitoring disease/symptoms [37].

A large percentage of subjects used electronic messaging. Ralston [29] found that 100% of patients (39/39) used email, Ross [34] found 76% (41/54) did, and Cruz [26] found 90% (19/21). The intensity of messaging use varied per patient per year from 1.2 [34], to 5.9 [37], to 8.7 [35], and to 55.2 [31] times. In the latter study, electronic reminders were sent. In one study, it was reported that 73% of messages (302/414) were sent outside of office hours [35].

In three studies, the content of the asynchronous communication [31,34,35] was reported. Kwon [31] reported that the main topics were nutrition, diabetic complications, exercise, and other aspects of diabetes management. Lin described that the main topic of 42% of messages (32/76) was biomedical concerns, and 14 messages were "for your information" (18%), and the latter type of message was significantly more common in electronic messaging than in telephone messaging [34]. Urgent messages were conveyed by telephone [35]. Ross [34] reported that electronic messages appear to supplement telephone messages. The main topics in electronic messaging were scheduling appointments, getting medication refills, asking questions about medication, getting test results, reporting "feeling ill", and getting assistance to interpret test results.

The usability of electronic messaging was shown for a variety of experiences, and overall, patients were positive about using electronic messaging. In one study, patients found the intervention helpful, easy to navigate, and would recommend it to others [27]. In another study, patients felt that the doctor understood their problems better and explained the information better when using email [34]. Patients prefer email as a way to send information and psychosocial messages to their doctor. In one sample, 75% of patients (132/175) thought they would use this in the future, and 85% (149/175) preferred email to telephone messaging [35]. Additionally, 162 out of 341 (48%) were willing to pay for online correspondence with their physician. Of those willing to pay, the median amount cited was US $2 per message [35]. Cruz found that patients preferred using the Internet rather than paper when monitoring their health [26]. Health care utilization through a patient portal led to higher patient satisfaction [35]. Patients seemed to appreciate the fact that they could communicate with the clinic and conduct administrative actions asynchronously instead of using the telephone. In one study, the content of the communication was studied, revealing that patients had specific questions about medication and tests but also wanted to communicate about "feeling ill" [34]. Patients seemed able to estimate correctly when to use the portal or when to use the telephone for messages, as the telephone was used for urgent messages [35].

Problems concerning usability can be expected when using technology, and these problems might influence usage. Minor usability problems were described in three studies in this review [26,27,39], but none had a large effect on the use.

Does It Help? Effects on Health Behavior

Multimedia Appendix 2 shows that of the 15 studies aimed at improving health behavior with Internet-based interventions including asynchronous communication, seven reported results on health behavior. The outcomes were in three different areas of health behavior: knowledge [38], health care utilization [35-37,40], and self-efficacy/self-management [27,33,34,36].

Increased knowledge was realized in a study involving rural women with a chronic disease who lived a long distance from the clinic [38].

Health care utilization in the form of visits to physicians did not decrease significantly, but a decrease was shown in the case of back pain patients (*P*=.07) [36] and asthma patients (*P*=.07) [37], although not statistically significant.

In the area of Internet-based support of self-efficacy/self-management, several results were reported. In the case of pain management, Internet-based interventions seemed to increase patients' self-efficacy in using non-medical techniques [27], self-care orientation in back pain [36], and managing dyspnea [33]. In patients with congestive heart failure, the general adherence to therapy increased when patients shared their EHR with their providers and communicated asynchronously about the content and implications of the EHR [34].

Does It Help? Effects on Health Outcomes

Health outcomes are important indicators for providers to guide the therapy of chronically ill patients. Health outcomes as a result of using an intervention with asynchronous communication were described as clinical and physical symptoms, psychosocial outcomes, and satisfaction. They were reported in 12 studies [9,27-33,35-37,39,41].

Outcomes for clinical symptoms were shown in four studies of diabetic patients. Improvements were shown in HbA1c level [9,29,31,32], weight, cholesterol, high-density lipoproteins [9], and blood pressure [32]. In a study of asthma patients, the forced expiratory volume (FEV) and the control of asthma were shown to increase with Internet-based support [37]. These were positive results for Internet-based interventions with asynchronous communication, but only McMahon looked at specific aspects of the intervention. He found that improved health outcomes were related to the frequency of use of the intervention [32].

Improved physical symptoms were also observed when using Internet-based self-care techniques. Berman and Lorig demonstrated a decrease in back pain in patients [27,36], and Ghahari demonstrated a decrease in fatigue-impact for patients with multiple sclerosis [28]. Nguyen reported increased arm endurance with exercise [33]. Berman discussed a relationship with a specific aspect of the intervention: logging on to the intervention seemed to decrease the patients' pain immediately [27].

Varied psychosocial outcomes were shown in the studies of interventions with asynchronous communication for chronically ill patients. In one study, personal well-being increased [28] for both intervention groups (interactive and information only) in comparison to the control group with usual care. Meer showed improvement in quality of life for asthma patients [37]. An increased acceptance of the illness was also shown [39], as well as increased self-esteem [30,39], empowerment, and social support [30]. Weinert described a decrease in stress, depression, and loneliness [39]. Lorig found that patients felt less disabled, whereas role functioning improved and health-related distress decreased [36]. Patients seemed to feel better when they had an Internet-based connection with their providers.

Does It Help? Effects on Patient Satisfaction

Satisfaction with the overall care from the clinic increased in one study when patients used the Internet-based connection with their provider via the patient portal [35].

Discussion

Results of Search

The literature search revealed that there are few studies of the effects of asynchronous communication on self-management of chronically ill patients. Only Lin had an Internet-based patient-provider communication system as the focus of an RCT [35]. The literature search yielded another 14 studies in which asynchronous communication was described as part of the intervention. The RCTs were mainly performed in the United States with patients with specific chronic conditions. To glean information about the introduction of asynchronous communication for tailor-made health care, more evidence from other countries and patients with diverse chronic conditions is required.

Quality Appraisal

When reviewing the methodological quality of the studies, a certain amount of lack of blinding was noted. This lack of blinding is inevitable for technical reasons when the use and effect of digital communication on health behavior and health outcomes is being examined. In assessing the performance bias in RCTs using these techniques, it can be argued that not too much weight should be given to this aspect of quality appraisal.

Results on Use and Usability

Although 12 studies report on the frequency of use of the intervention, none specifically examined why and when patients log on to the intervention. The results of this review suggest that it might be interesting to find out more about the meaning of the frequency of use by patients. Comparisons with other publications about the frequency of access of Internet and communication technology by patients show varied results from increasing use [42] to declining use [34,43], but these results did not differ from those for usual care [44]. Kwon shows that the frequency of use increases when reminders are sent by the provider [31], resulting in increased health outcomes. Ross [34] showed a decline in use but an increase in adherence. Possibly the effect is not in the actual use but in having the connection to the provider, who can be contacted if necessary. It is also possible that the patients' questions have been answered and they know what to do.

In the studies where email use was measured, a high percentage of patients (>75%) used it [29,34]. Patients seem to be interested, but it is not yet clear when they feel the need to use it or whether being connected is enough to feel satisfied and more in control of their health. Perhaps patients experience the connection as a supporting factor in their self-management. The virtual presence of the professional through the digital connection might have effects that could be interesting. Perhaps patients do not need to "check" the digital connection by using it after a time, as they feel confident knowing the digital connection with the provider can be made whenever they need it. The provider is always present and can be approached if the need arises. In research on social support, it is shown that merely the availability of support is helpful and related to higher levels of well-being [45,46]. More understanding of the effects on patients of asynchronous communication could lead to increasingly tailor-made health care.

From the viewpoint of transition to integration of modern communication technologies, several threats to successful integration can be identified. It is certainly a threat if patients are not able to distinguish issues that are acute or (life)threatening versus non-acute, but we found no evidence to support this. We found that patients know when to use the telephone, because urgent messages were conveyed by the telephone [35]. They also know when to use asynchronous communication; patients preferred sending non-urgent messages such as "for your information/feeling ill" by email [34,35]. In addition, informational and psychosocial messages are sent via email as opposed to the telephone [35].

Also with regard to integration of ICT, it is interesting that patients are prepared to pay for email service with their provider [35]. The fact that 73% of messages (302/414) are sent outside of office hours suggests that time and place might be a factor [35].

The content of the electronic messages suggests that patients are willing to participate actively because they share more information than strictly necessary. They have a variety of issues they want to communicate about when using asynchronous communication. Understanding what these issues are is important for further implementation. In addition to "for your information/feeling ill" messages, they have health issues that they want to clarify, such as biomedical and medication concerns and receiving test results and assistance to interpret them. What seems to be happening is that asynchronous communication is used to communicate information that may or may not be relevant, but it satisfies patients to send it. They have taken action by sending the information, and it is now up to the health care provider to say whether it is relevant and if action is needed. This is the start of "shared-decision making". With asynchronous communication, patients seem to make use of the option to share their worries and their psychosocial condition with their provider. This is an indication for willingness for further patient participation.

Age does not seem to be a factor in the use of asynchronous communication, given the advanced age of the participants. This is relevant because most patients with a chronic disease belong to the older segment of the population [1]. In the literature, older age has been identified as a barrier to the use of Internet communication technology [47-49]. It can of course be argued that patients who are included in the sample must have Internet access and minimum competencies to use it. It does show, however, that age is not an unsurpassable barrier. It may even be so that the benefits of having a virtual connection with the provider stimulates patients to use the Internet connection or at least understand how to use it, no matter what their age.

Results Relating to Health Behavior

In the seven studies where health behavior is described, improvements are shown when using the interventions [27,33-38]. The Internet-based intervention is therefore an option in regard to providing support self-management at a distance.

The meaning to the patient of the digital connection to the provider is interesting. Does having the connection at your fingertips give a reassuring feeling? Does sitting down and logging on to the connection feel like the first step in self-management and being assertive about your needs? Some results on health behavior may point in this direction. The two studies that used an intense (with an online coach) and a less intense intervention (without an online coach) show that a less intense intervention is just as effective [28,38]. This again raises the question of whether the connection alone is enough to improve health behavior, or whether a more complex intervention is necessary to gain an effect. In another study, it is suggested that logging on has an immediate impact on pain reduction [27]. This stresses the need to clarify which aspects of eHealth interventions are effective for patients. Is it the direct connection via Internet with the provider, is it the online coach, or is it the tailored information? The latter may be a very interesting point, as this interactive communication makes it possible to obtain detailed information about health management from the provider and the patient. In all other settings, such as in the consulting room, using the telephone or writing letters, it requires more effort to obtain the information necessary for tailor-made intervention.

Two studies found a trend to a significant decrease of health care utilization in the form of visits to physicians when using an Internet-based intervention [36,37]. It may be that asynchronous communication plays a role in this change in health behavior because patients can discuss their health concerns interactively with their provider. The triggers for this health behavior might be less time spent travelling to the physician and in the waiting room; the convenience and fact that no travel is required make the interaction more economical. However, this aspect has not been studied.

The content of the communication in the study by Ross shows that patients have precise questions about medication and tests, but that they also want to communicate about "feeling ill". The results show that general adherence increases. However, the question of whether asynchronous communication about these issues affects the general adherence is not raised [34]. There may be some support for this conclusion in the Lin study, where patient satisfaction increased when they were able to communicate "for your information" messages through electronic messaging [35].

Results on Health Outcomes

Improved health outcomes were shown in 11 studies using Internet-based interventions, including electronic messaging. In studies where the intervention was complementary to usual care, an improvement could be expected. In studies with an intervention as a substitute for usual care, similar outcomes from intervention and usual care can be seen as a positive result. However, improvements were also reported in studies in which the intervention was a substitute, thus showing that the Internet intervention has better results than face-to-face care. This calls for more testing of these interventions as a substitute for usual care in larger samples of diverse patients.

Limitations

The Internet-based interventions in the studies consisted of different components, such as peer-support groups, sharing medical records, self-management programs, and patient portals. The component they all had in common was electronic messaging. The multicomponent aspect of the studies made it difficult to trace the exclusive effect of the intervention to the asynchronous communication.

Electronic messaging was not evaluated separately in these studies. It can be concluded that the effect of asynchronous communication is not adequately shown in these studies and that many questions can be raised about the precise effect of Internet-based asynchronous communication between patients with a chronic condition and their providers. This is a limitation. However, there were many positive findings about electronic messaging in relation to telephone messages and other forms of messaging.

Future Research

Research is needed to determine the technical characteristics of effective asynchronous communication with patients for specific disease categories where specific health behavior is needed by specific patients. The meaning of the virtual connection with the provider should be explored. This may be a very basic intervention with a large effect.

Additional testing is needed to clarify what patients want to discuss with their providers and how shared decision making about these issues can be effective.

The effects of asynchronous communication on self-management for larger samples of diverse patients with a chronic condition require examination. A clear definition of desirable outcomes is needed. The desired results for health behavior should be operationalized with regard to “the ability to adapt and self-manage in the face of social, physical, and emotional challenges” [5]. In this review, three categories of health behavior could be distinguished as outcomes: knowledge, health care utilization, and self-management/self-efficacy.

We also advise further testing of Internet interventions as a substitute for usual care because significant health outcomes were found in this review.

Conclusions

It can be concluded that using asynchronous communication in health care may be an important instrument to increase patient participation leading to self-management. After reviewing the literature, the answer to both research questions seems to be positive: asynchronous communication is used by patients and it helps to increase the effects on health behavior and health outcomes, at least for some. Patients seem to be interested in using email and understand how to use it. They use email for questions about biomedical concerns, medication, and test results, as well as to inform the providers about non-urgent health issues. They tend to prefer email to telephone for this communication. They also understand when they can use email or when contact by telephone is needed.

From the viewpoint of the new definition of health with an emphasis on self-management and patient participation, it seems possible to take steps towards sustainable health care by implementing asynchronous communication, as it enables patients to communicate effectively about their perceived health problems and their adaptation to health problems.

Acknowledgments

This work was supported by Fonds NutsOhra, Menzis, and CZ funds in the Netherlands.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Methodological quality randomized clinical trials.

[DOCM File, 23KB - jmir_v16i1e19_app1.docm]

Multimedia Appendix 2

Results reported on use of the intervention, health behavior, and health outcomes.

[PDF File (Adobe PDF File), 36KB - jmir_v16i1e19_app2.pdf]

References

1. World Health Organization. Health topics - chronic diseases URL: http://www.who.int/topics/chronic_diseases/en/ [accessed 2014-01-05] [WebCite Cache ID 6MOrWXsKm]
2. Bastiaens H, Van Royen P, Pavlic DR, Raposo V, Baker R. Older people's preferences for involvement in their own care: a qualitative study in primary health care in 11 European countries. Patient Educ Couns 2007 Sep;68(1):33-42. [doi: 10.1016/j.pec.2007.03.025] [Medline: 17544239]

3. Griffin SJ, Kinmonth AL, Veltman MW, Gillard S, Grant J, Stewart M. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. Ann Fam Med 2004;2(6):595-608 [FREE Full text] [doi: 10.1370/afm.142] [Medline: 15576546]
4. Edwards A, Elwyn G, Hood K, Atwell C, Robling M, Houston H, Study Steering Group. Patient-based outcome results from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. Fam Pract 2004 Aug;21(4):347-354 [FREE Full text] [doi: 10.1093/fampra/cmh402] [Medline: 15249521]
5. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ 2011;343:d4163. [Medline: 21791490]
6. World Health Organization. Official Records of the World Health Organization. 2004. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States URL: http://www.who.int/governance/eb/who_constitution_en.pdf [accessed 2014-01-05] [WebCite Cache ID 6MOsT7pjn]
7. Ursum J, Rijken M, Heijmans M. NIVEL Overzichtstudies: zorg voor chronisch zieken: organisatie van zorg, zelfmanagement, zelfredzaamheid en participatie. Utrecht: NIVEL; 2011. URL: http://www.nivel.nl/node/2430?database=ChoicePublicat&priref=1002102
8. Lambregts J, Grotendorst A. Leren van de toekomst Verpleegkundigen & Verzorgenden 2020. Houten: Bohn Stafleu van Loghum; 2012.
9. Bond GE, Burr R, Wolf FM, Price M, McCurry SM, Teri L. The effects of a Web-based intervention on the physical outcomes associated with diabetes among adults age 60 and older: a randomized trial. Diabetes Technol Ther 2007 Feb;9(1):52-59 [FREE Full text] [doi: 10.1089/dia.2006.0057] [Medline: 17316098]
10. Oudshoorn N. Diagnosis at a distance: the invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol Health Illn 2008 Mar;30(2):272-288. [doi: 10.1111/j.1467-9566.2007.01032.x] [Medline: 18290936]
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. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
13. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
14. 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]
15. Pare´ G. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension : systematic review. Montre´al: HEC Montre´al, Chaire de recherche du Canada en technologie de l'information dans le secteur de la sante´; 2009.
16. Radhakrishnan K, Jacelon C. Impact of telehealth on patient self-management of heart failure: a review of literature. J Cardiovasc Nurs 2012;27(1):33-43. [doi: 10.1097/JCN.0b013e318216a6e9] [Medline: 21558862]
17. Weiner M, Callahan CM, Tierney WM, Overhage JM, Mamlin B, Dexter PR, et al. Using information technology to improve the health care of older adults. Ann Intern Med 2003 Sep 2;139(5 Pt 2):430-436. [Medline: 12965971]
18. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
19. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
20. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
21. Allen M, Iezzoni LI, Huang A, Huang L, Leveille SG. Improving patient-clinician communication about chronic conditions: description of an Internet-based nurse E-coach intervention. Nurs Res 2008;57(2):107-112. [doi: 10.1097/01.NNR.0000313478.47379.98] [Medline: 18347482]
22. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
23. Eurostat. URL: http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/isoc_bde15c_esms.htm [accessed 2013-11-17] [WebCite Cache ID 6LCWIAuHG]

24. 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]
25. National Institute for Clinical Excellence. The Guideline manual appendices (09/18). 2013. URL: http://publications.nice.org.uk/the-guidelines-manual-appendices-bi-pmg6b/ap [accessed 2014-01-06] [WebCite Cache ID 6MQsOaKwY]
26. Cruz-Correia R, Fonseca J, Lima L, Araújo L, Delgado L, Castel-Branco MG, et al. Web-based or paper-based self-management tools for asthma--patients' opinions and quality of data in a randomized crossover study. Stud Health Technol Inform 2007;127:178-189. [Medline: 17901611]
27. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
28. Ghahari S, Leigh Packer T, Passmore AE. Effectiveness of an online fatigue self-management programme for people with chronic neurological conditions: a randomized controlled trial. Clin Rehabil 2010 Aug;24(8):727-744. [doi: 10.1177/0269215509360648] [Medline: 20543022]
29. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
30. Hill W, Weinert C, Cudney S. Influence of a computer intervention on the psychological status of chronically ill rural women: preliminary results. Nurs Res 2006;55(1):34-42 [FREE Full text] [Medline: 16439927]
31. Kwon HS, Cho JH, Kim HS, Song BR, Ko SH, Lee JM, et al. Establishment of blood glucose monitoring system using the Internet. Diabetes Care 2004 Feb;27(2):478-483. [Medline: 14747232]
32. McMahon GT, Gomes HE, Hickson Hohne S, Hu TM, Levine BA, Conlin PR. Web-based care management in patients with poorly controlled diabetes. Diabetes Care 2005 Jul;28(7):1624-1629 [FREE Full text] [Medline: 15983311]
33. Nguyen HQ, Donesky D, Reinke LF, Wolpin S, Chyall L, Benditt JO, et al. Internet-based dyspnea self-management support for patients with chronic obstructive pulmonary disease. J Pain Symptom Manage 2013 Jul;46(1):43-55. [doi: 10.1016/j.jpainsymman.2012.06.015] [Medline: 23073395]
34. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a Web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
35. Lin CT, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
36. Lorig KR, Laurent DD, Deyo RA, Marnell ME, Minor MA, Ritter PL. Can a Back Pain E-mail Discussion Group improve health status and lower health care costs?: A randomized study. Arch Intern Med 2002 Apr 8;162(7):792-796. [Medline: 11926853]
37. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, SMASHING (Self-Management in Asthma Supported by Hospitals, ICT, NursesGeneral Practitioners) Study Group. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [Medline: 19620163]
38. Weinert C, Cudney S, Hill W. Health knowledge acquisition by rural women with chronic health conditions: a tale of two Web approaches. Aust J Rural Health 2008 Oct;16(5):302-307 [FREE Full text] [doi: 10.1111/j.1440-1584.2008.01004.x] [Medline: 18808489]
39. Weinert C, Cudney S, Comstock B, Bansal A. Computer intervention impact on psychosocial adaptation of rural women with chronic conditions. Nurs Res 2011;60(2):82-91. [doi: 10.1097/NNR.0b013e3181ffbcf2] [Medline: 21358372]
40. Lorig KR, Ritter PL, González VM. Hispanic chronic disease self-management: a randomized community-based outcome trial. Nurs Res 2003;52(6):361-369. [Medline: 14639082]
41. Reed RL, Battersby M, Osborne RH, Bond MJ, Howard SL, Roeger L. Protocol for a randomised controlled trial of chronic disease self-management support for older Australians with multiple chronic diseases. Contemp Clin Trials 2011 Nov;32(6):946-952. [doi: 10.1016/j.cct.2011.08.001] [Medline: 21864719]
42. Välimäki M, Nenonen H, Koivunen M, Suhonen R. Patients' perceptions of Internet usage and their opportunity to obtain health information. Med Inform Internet Med 2007 Dec;32(4):305-314. [doi: 10.1080/14639230701819792] [Medline: 18072007]
43. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
44. Ros WJ. MyUMC Patient Portal: patient and professional perspectives. Int J Integr Care 2012;12(1):e90.
45. Komproe IH, Rijken M, Ros WJ, Winnubst JA, `tHart H. Available Support and Received Support: Different Effects Under Stressful Circumstances. Journal of Social and Personal Relationships 1997 Feb 01;14(1):59-77. [doi: 10.1177/0265407597141003]
46. Cutrona CE, Russell D. The provisions of social relationships and adaptation to stress. In: Jones WH, Perlman D, editors. Advances in personal relationships. Greenwich, CT: JAI Press; 1987:37-67.
47. Kruse RL, Koopman RJ, Wakefield BJ, Wakefield DS, Keplinger LE, Canfield SM, et al. Internet use by primary care patients: where is the digital divide? Fam Med 2012 May;44(5):342-347 [FREE Full text] [Medline: 23027117]

48. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
49. Eley R, Fallon T, Soar J, Buikstra E, Hegney D. Barriers to use of information and computer technology by Australia's nurses: a national survey. J Clin Nurs 2009 Apr;18(8):1151-1158. [doi: 10.1111/j.1365-2702.2008.02336.x] [Medline: 19320784]

Abbreviations

EHR: electronic health record
FEV: forced expiratory volume
ICT: information and communications technology
NICE: National Institute for Health and Care Excellence
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 01.10.13; peer-reviewed by J Clemensen, R Shaw, E Rubenstein; comments to author 31.10.13; revised version received 11.12.13; accepted 21.12.13; published 16.01.14.

Please cite as:
de Jong CC, Ros WJG, Schrijvers G
The Effects on Health Behavior and Health Outcomes of Internet-Based Asynchronous Communication Between Health Providers and Patients With a Chronic Condition: A Systematic Review
J Med Internet Res 2014;16(1):e19
URL: http://www.jmir.org/2014/1/e19/
doi:10.2196/jmir.3000
PMID:24434570

Original Paper

Virtual Patients in Primary Care: Developing a Reusable Model That Fosters Reflective Practice and Clinical Reasoning

Helena Salminen[1], MD, PhD; Nabil Zary[2], MD, PhD; Karin Björklund[1], OT; Eva Toth-Pal[1], MD, PhD; Charlotte Leanderson[1], MD, PhD

[1]Centre for Family Medicine, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden
[2]Digital Patient Lab, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Helena Salminen, MD, PhD
Centre for Family Medicine
Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels Allé 12
Huddinge, 14183
Sweden
Phone: 46 8 52488701
Fax: 46 8 52488706
Email: helena.salminen@ki.se

Abstract

Background: Primary care is an integral part of the medical curriculum at Karolinska Institutet, Sweden. It is present at every stage of the students' education. Virtual patients (VPs) may support learning processes and be a valuable complement in teaching communication skills, patient-centeredness, clinical reasoning, and reflective thinking. Current literature on virtual patients lacks reports on how to design and use virtual patients with a primary care perspective.

Objective: The objective of this study was to create a model for a virtual patient in primary care that facilitates medical students' reflective practice and clinical reasoning. The main research question was how to design a virtual patient model with embedded process skills suitable for primary care education.

Methods: The VP model was developed using the Open Tufts University Sciences Knowledgebase (OpenTUSK) virtual patient system as a prototyping tool. Both the VP model and 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 the medical program. The students' opinions on the VP were investigated through focus group interviews with 14 students and the results analyzed using content analysis.

Results: The VP primary care model was based on a patient-centered model of consultation modified according to the Calgary-Cambridge Guides, and the learning outcomes of the study program in medicine were taken into account. The VP primary care model is based on Kolb's learning theories and consists of several learning cycles. Each learning cycle includes a didactic inventory and then provides the student with a concrete experience (video, pictures, and other material) and preformulated feedback. The students' learning process was visualized by requiring the students to expose their clinical reasoning and reflections in-action in every learning cycle. Content analysis of the focus group interviews showed good acceptance of the model by students. The VP was regarded as an intermediate learning activity and a complement to both the theoretical and the clinical part of the education, filling out gaps in clinical knowledge. The content of the VP case was regarded as authentic and the students appreciated the immediate feedback. The students found the structure of the model interactive and easy to follow. The students also reported that the VP case supported their self-directed learning and reflective ability.

Conclusions: We have built a new VP model for primary care with embedded communication training and iterated learning cycles that in pilot testing showed good acceptance by students, supporting their self-directed learning and reflective thinking.

(J Med Internet Res 2014;16(1):e3) doi:10.2196/jmir.2616

KEYWORDS

virtual patients; clinical reasoning; reflection; primary care; medical education

Introduction

Virtual Patients

The development of reliable and acceptable information technology offers opportunities to create new types of learning activities that enhance professional interaction across time and space [1]. In a review, Cook and Triola defined virtual patients as a “specific type of computer programme that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions” [2]. Virtual patients (VPs) are currently introduced into many health care programs worldwide but are still sparsely used in medical school curricula [3]. Designing a VP is a delicate process, and the production of learning objects for multimedia consumes both time and resources [4]. VPs are found to be useful in teaching students clinical reasoning skills [2,5,6]. In clinical reasoning, both analytic and nonanalytic approaches are used [7]. Doctors in primary care often use strategies in clinical reasoning that are not always explicitly taught, and few studies have explored how these skills are acquired [8]. Examples are an important part of clinical reasoning, and VPs have something to contribute here as a complement to the real-life clinical practice. Using video in education may stimulate clinical diagnostic reasoning [9]. Primary care, with its complexity and its structure as an educational arena for students, may gain from a blended learning structure where virtual patients may help the student be introduced to learning situations similar to those they encounter during their clinical placements in primary care.

Learning in Primary Care

The same learning theories that are applicable for learning in primary care are also applicable for virtual patients. Education in primary care has a holistic perspective on patient care, and the learning is based on theories of adult learning and self-directed learning. During a clinical placement in primary care, the students meet patients with a wide variety of medical problems and have a strong internal motivation to learn from these patients. Adults learn most effectively when they do things in a real-life setting and have active roles in a meaningful context [10,11]. Students’ learning is an active process where the starting point is their pre-understanding [12]. The virtual learning experience is very similar to what may happen in real clinical practice. A virtual patient may be described as an approximation to reality and is often experienced as very authentic by learners. The knowledge the students gain from the virtual patient can be directly applied to real patients that they meet at the primary health care centre. The “sit-in” method, where students act as doctors during part of or an entire patient’s encounter supervised by a tutor-observer, is a frequently used method in primary care to teach both clinical reasoning and communication skills. “Sit-in” gives the opportunity for students to learn their professional role and reflect upon it, but time constraints make its use limited. This kind of reflection may be useful as a tool for meaningful learning, and reflections are best done in an authentic context [13]. Reflections around patients in primary care are an important part of the education and may also be used in a virtual patient setting.

The aim of this study was to create a model for a virtual patient in primary care that enhances medical students’ reflective practice and clinical reasoning and training in communication skills. The main research questions were:

1. How can we embed clinical reasoning, communication skills, and reflection in a VP primary care model?
2. How can the VP primary care model be represented as a formative learning activity?
3. How do students perceive the VP, based on the model developed?

Methods

Context of the Study

Primary care is present at different stages of the study program at Karolinska Institutet (KI), Sweden. Medical students have primary care in 9 semesters of the 11-semester study program. The structure of primary care—with about 200 units that participate in clinical supervision of approximately 1500 medical students every semester and student placements in primary care at many different stages of the program—is a huge challenge when learning activities are designed. The primary care placements of medical students are only about 1 week per semester, and the students change primary health care centres three to four times during the study program.

Modeling of the Virtual Patient Primary Care Model

A reference group of 10 experienced primary care teachers from the Centre for Family Medicine were involved in developing the new model of VP for primary care with the aim of embedding communication skills, clinical reasoning, and reflection in the model. The group of teachers met on several occasions and explored different current systems for VPs. The VP system (Open Tufts University Sciences Knowledgebase, OpenTUSK) was chosen after consensus in the reference group. The decision was based on the possibilities this system offered from a pedagogical family medicine perspective. All teachers were involved in creating the VP case material about the patient with multiple diagnoses. A working group of 4 teachers (authors of this paper), together with a VP expert (NZ), created a VP prototype model based on these discussions.

The prototype was then tested with final-semester medical students to eliminate design and information ambiguities and to ensure that the case was meaningful, self-directed, and followed the learning outcomes of the study program in medicine.

The OpenTusk virtual patient system was used as a prototyping tool. This tool allows open-ended questions and free-text answers that may promote students’ own reflections and make it possible to adapt the learning with the VP case to how diagnostic decision making takes place in a primary care context. The tool also allows the use of video clips. Videos are already widely used in primary care education, and the students are trained in analyzing videotaped consultations as a part of their education in primary care. The preformulated comments from teachers after the free text provided possibilities to give students insight into experienced family doctors’ clinical reasoning. The preformulated comments (100–250 words) from teachers preceded the students’ reflective writing in each part. The structure of the case was visualized in the program VUE, and the case was written in a text document before it was entered in OpenTUSK.

Representation of the Virtual Patient Primary Care Model as a Formative Learning Activity

Experience-based education in communication skills is one of the responsibilities of the primary care part of the medical study program at KI and one of the most important learning outcomes of the primary care part of the program. The communication skills training program is based on the Calgary-Cambridge Guides [14]. A Danish patient-doctor communication model (PRACTICAL) is applied when giving feedback on patient encounters [15]. PRACTICAL has further developed the core aspects of patient-centeredness and does not only stress the importance of taking into account the patient’s ideas, concerns, and expectations, but has six more relevant aspects related to different parts of the patient encounter. The training in communication skills has a clear progression through the study program in medicine. During the first semesters, the students have training in early clinical contact and have focus on the patient’s agenda. From the third semester, they also train the doctor’s part of the patient encounter with specific questions a doctor needs to ask the patients in order to get the correct diagnosis. At later stages, the students practice how to involve the patient in decision making and how to build a common ground for shared understanding and follow-up. How to motivate the patient and how to mobilize the patient’s internal capacities is an important part of that training. The same strategy in communication training was adapted to the virtual patient environment (see Figure 1). The communication, physical examination, and clinical reasoning were highlighted in the VP model through teachers’ comments and embedded hyperlinks. To visualize and enhance the learning process, a cycle for self-directed learning was constructed (see Figure 2). The VP case learning outcomes are in alignment with the learning objectives of the study program in medicine.

Screenshots of the first learning cycle of gathering information are presented in Figures 3 to 6 with answers written by one student who performed the VP case. All text was translated from Swedish to English.

Figure 1. Example of how the different dimensions of the VP case were structured to be congruent with the Calgary Cambridge Guides and the PRACTICAL training model in communication skills.

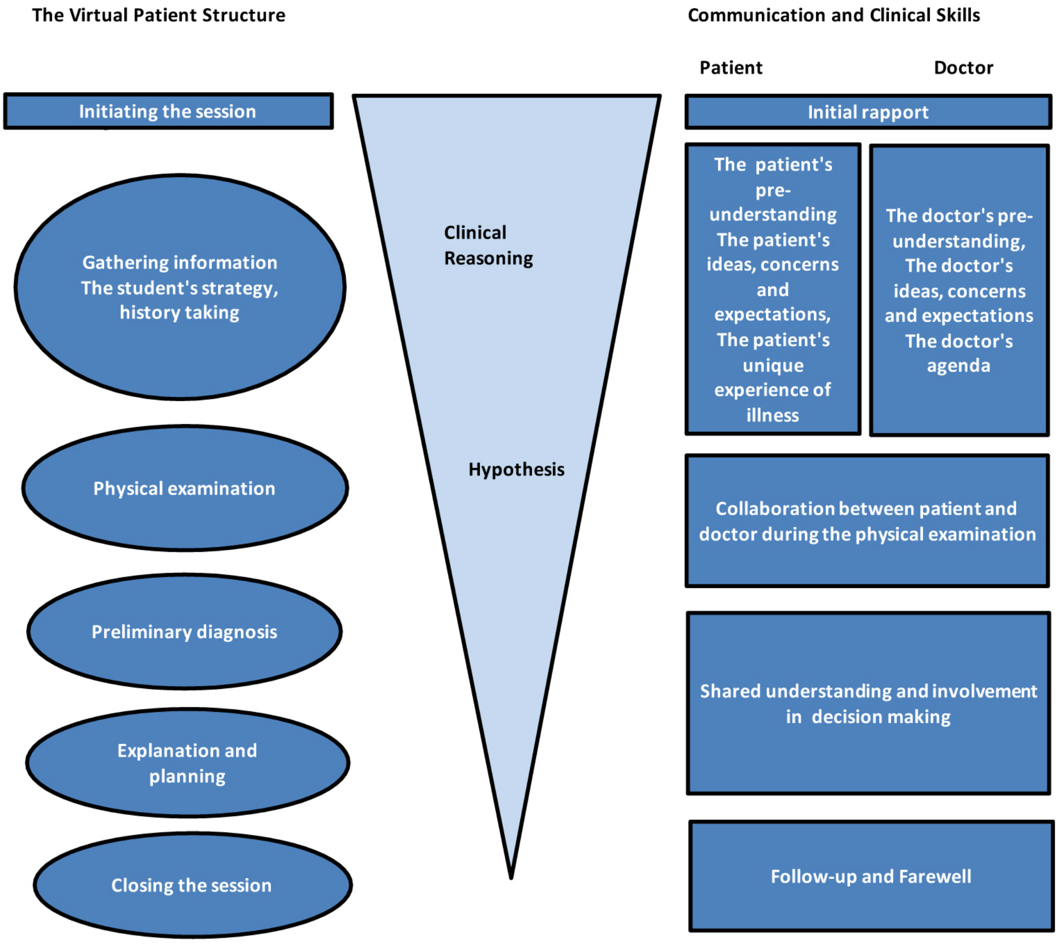


Figure 2. Iterated learning cycle (before the first cycle of the VP case started, the student formulated their objectives for performing the case; after finishing the last cycle, the student wrote a final reflection about their educational outcomes during the VP encounter).

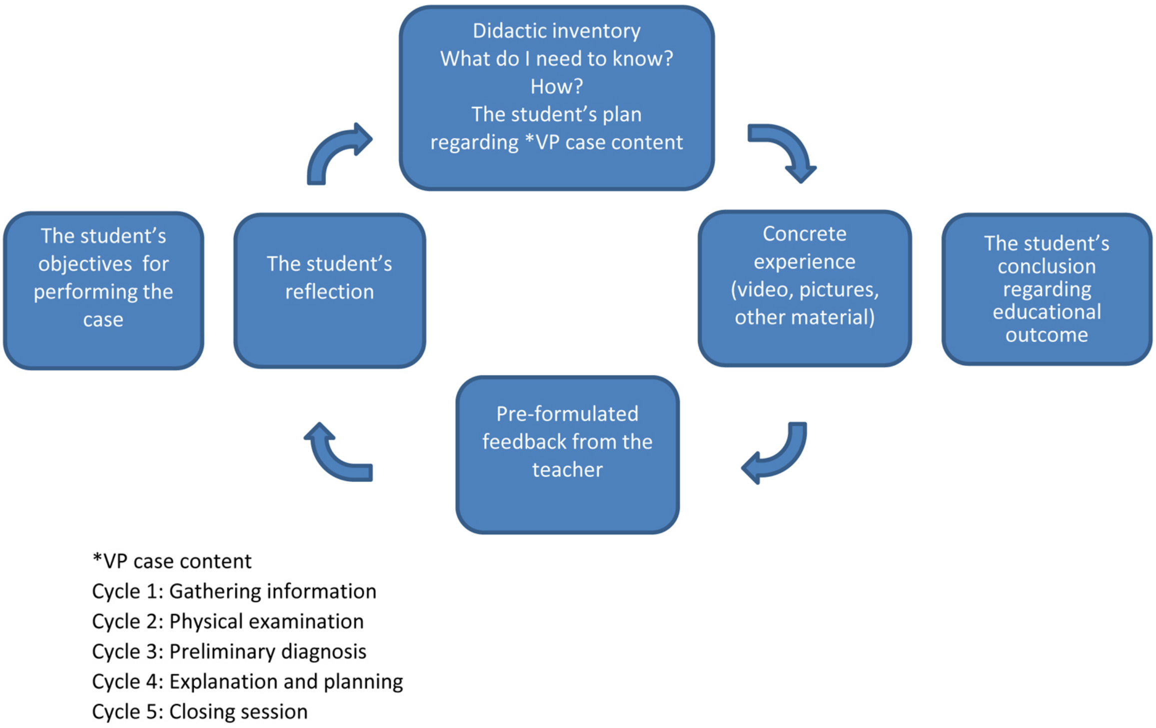
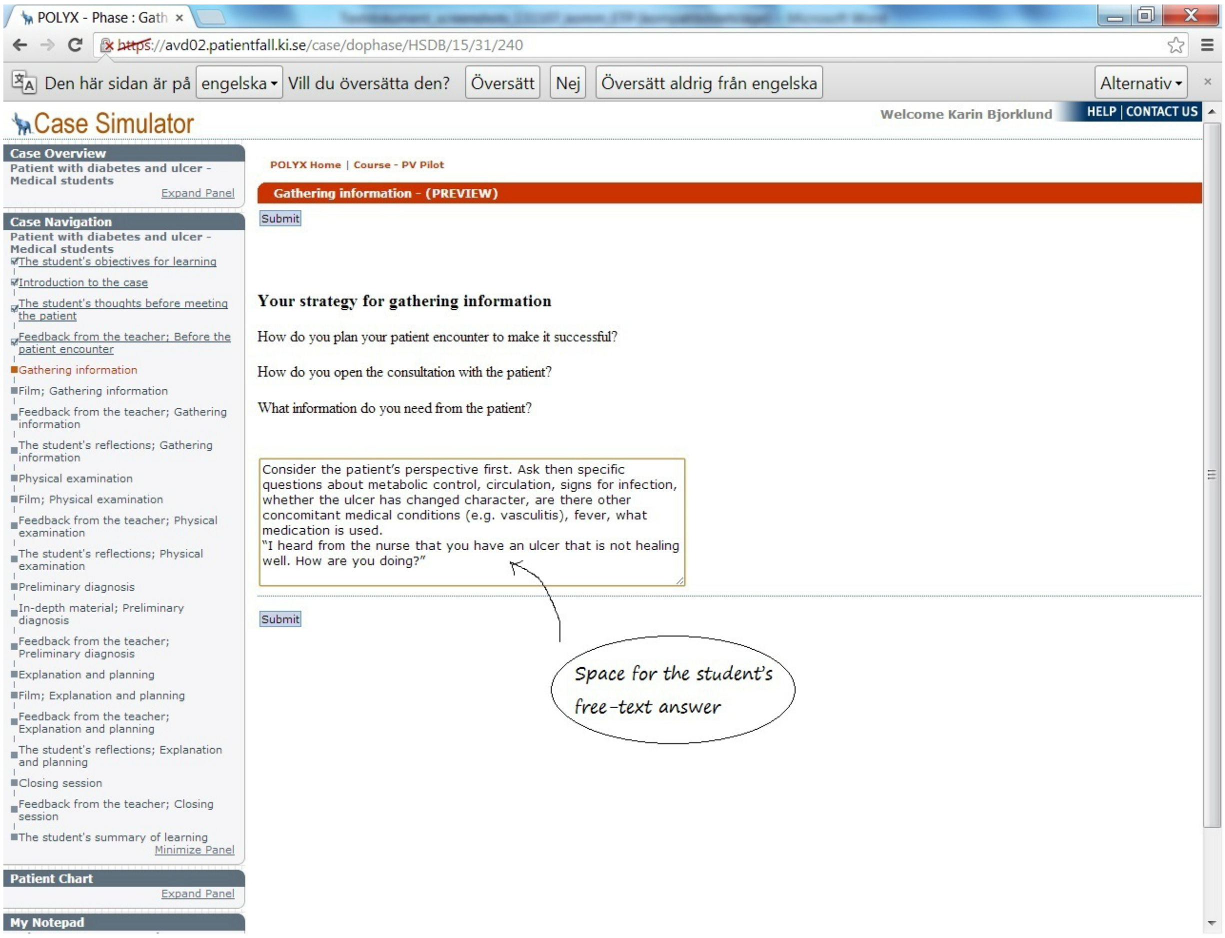


Figure 3. Didactic inventory.



Your strategy for gathering information

How do you plan your patient encounter to make it successful?

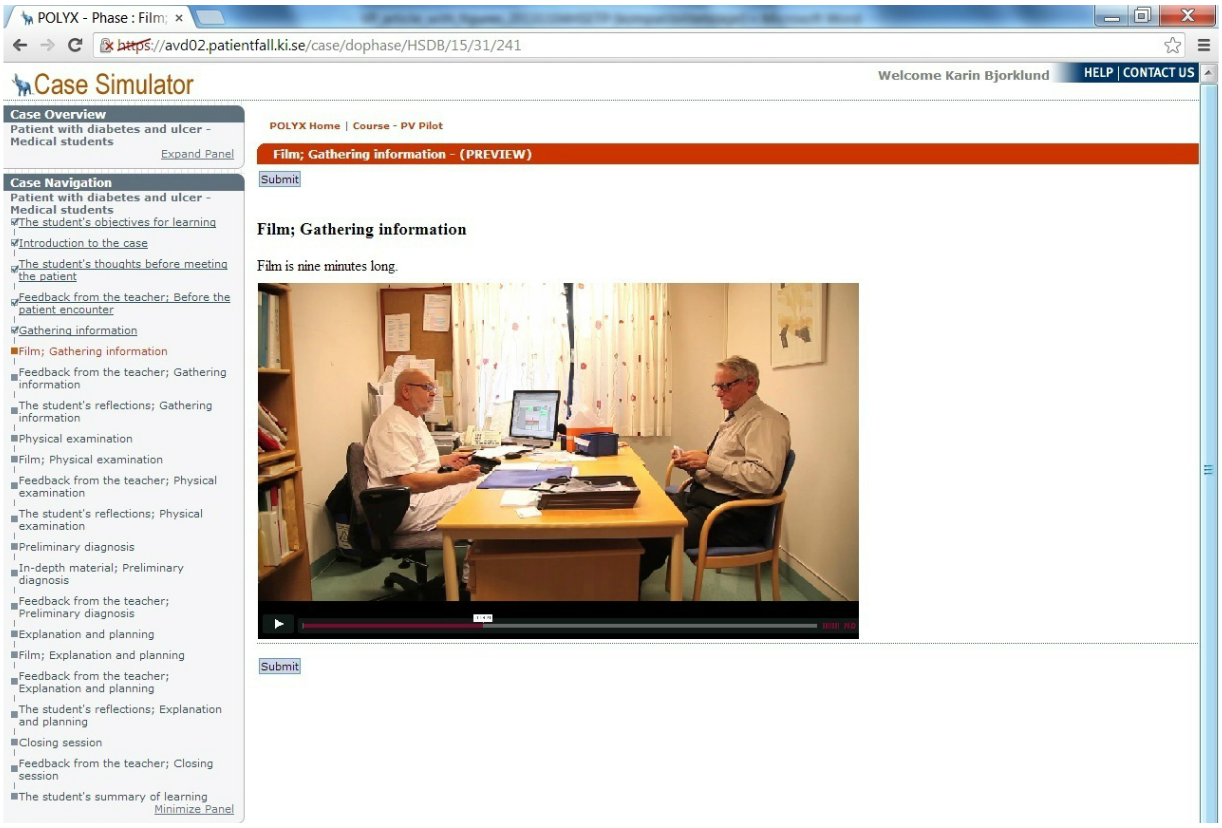
How do you open the consultation with the patient?

What information do you need from the patient?

Consider the patient's perspective first. Ask then specific questions about metabolic control, circulation, signs for infection, whether the ulcer has changed character, are there other concomitant medical conditions (e.g. vasculitis), fever, what medication is used.
"I heard from the nurse that you have an ulcer that is not healing well. How are you doing?"

Space for the student's free-text answer

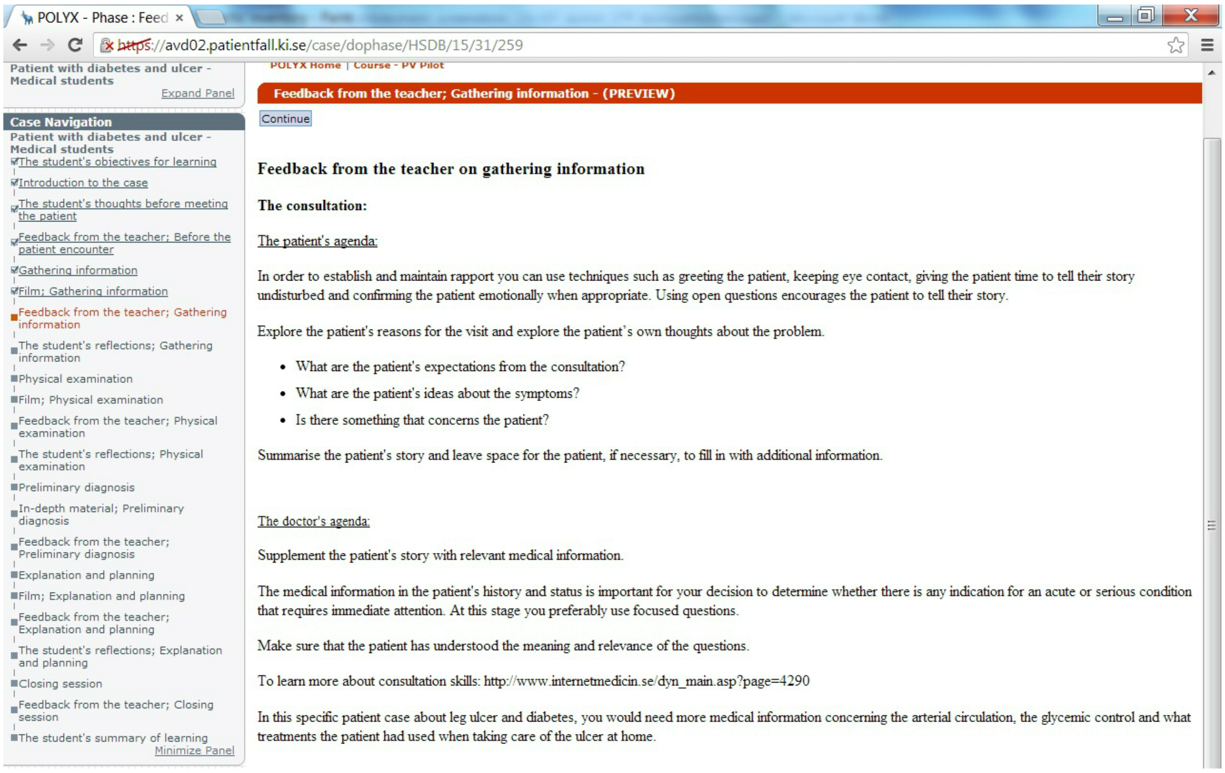
Figure 4. Concrete experience.



Film; Gathering information

Film is nine minutes long.

Figure 5. Pre-formulated feedback from the teacher.



Feedback from the teacher on gathering information

The consultation:

The patient's agenda:

In order to establish and maintain rapport you can use techniques such as greeting the patient, keeping eye contact, giving the patient time to tell their story undisturbed and confirming the patient emotionally when appropriate. Using open questions encourages the patient to tell their story.

Explore the patient's reasons for the visit and explore the patient's own thoughts about the problem.

- What are the patient's expectations from the consultation?
- What are the patient's ideas about the symptoms?
- Is there something that concerns the patient?

Summarise the patient's story and leave space for the patient, if necessary, to fill in with additional information.

The doctor's agenda:

Supplement the patient's story with relevant medical information.

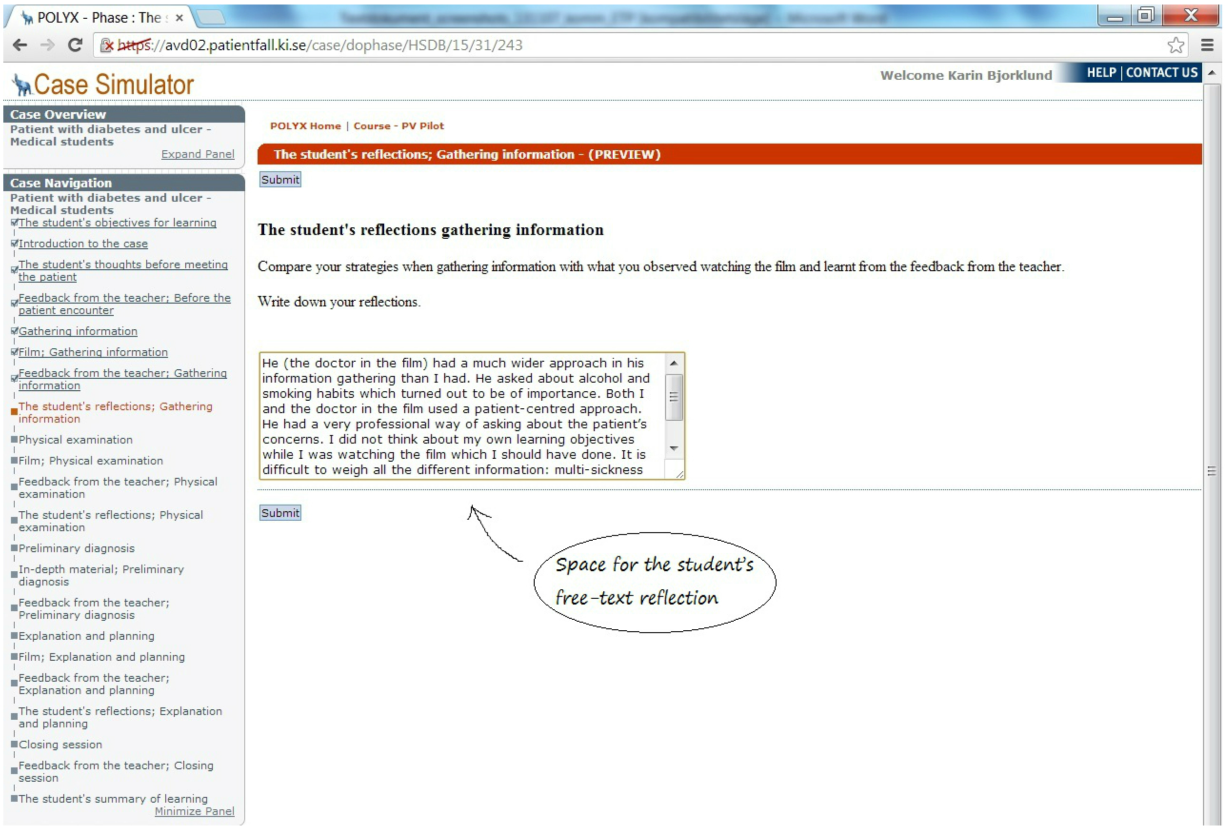
The medical information in the patient's history and status is important for your decision to determine whether there is any indication for an acute or serious condition that requires immediate attention. At this stage you preferably use focused questions.

Make sure that the patient has understood the meaning and relevance of the questions.

To learn more about consultation skills: http://www.internetmedicin.se/dyn_main.asp?page=4290

In this specific patient case about leg ulcer and diabetes, you would need more medical information concerning the arterial circulation, the glycemic control and what treatments the patient had used when taking care of the ulcer at home.

Figure 6. The student's reflection.



The student's reflections gathering information

Compare your strategies when gathering information with what you observed watching the film and learnt from the feedback from the teacher.

Write down your reflections.

He (the doctor in the film) had a much wider approach in his information gathering than I had. He asked about alcohol and smoking habits which turned out to be of importance. Both I and the doctor in the film used a patient-centred approach. He had a very professional way of asking about the patient's concerns. I did not think about my own learning objectives while I was watching the film which I should have done. It is difficult to weigh all the different information: multi-sickness

Evaluation of the Virtual Patient Primary Care Model

All final-semester students at the KI medical program were invited to participate in a pilot session of the VP model. The VP model was evaluated by three groups (n=14) working individually on a computer. After finishing the VP session, a semistructured interview with each group was performed. The interview followed an interview guide on the following areas: technical design/structure, authenticity, learning process, feedback, and feedback-content. The interviews were recorded, transcribed, and analyzed using content analysis [16].

Results

The Virtual Patient Primary Care Model

The VP model was designed and divided into five cycles in alignment with an authentic patient consultation: Gathering information, Physical examination, Preliminary diagnosis, Explanation and planning, and Closing session (see Figure 1).

The VP started with the introduction of the learning activity, including user instructions, and the medical case was presented briefly. The first student task was to formulate their learning goals. In general, the questions were open-ended and students were asked to formulate their answers in free text. At the beginning of the first learning cycle, the student was requested to plan the patient encounter. The patient encounter was then introduced through a film showing the information gathering, followed by preformulated teachers' comments that were given as feedback to the student. The learning cycle was concluded by the student's reflection. The VP continued with iterated cycles (physical examination, preliminary diagnosis, explanation and planning, and closing session). Students finished the case by a final reflection about their educational outcome during the VP encounter. The iterative reflective writing parts could be reread and revised throughout the VP and resubmitted several times. After the VP was finalized, no further alterations could be made, and all written content was displayed on the screen.

Characteristics of the Primary Care Virtual Patient Model as a Learning Activity

The VP case was designed based on an authentic patient story that was modified in order to hide the identity of the patient and to highlight the intended learning outcomes of the activity. All activities were plausible in a real clinical environment. Our overall intention was to create a structure in the VP case as close as possible to what might happen in a real clinical context in primary care (see Figure 1).

The student's iterative reflections drove the VP case forward. Reflections were designed to be part of every learning cycle (see Figure 2).

Films and other material from the patient encounter were embedded in the VP case. Video films were produced in an authentic clinical setting with a family doctor and a standardized patient. The films were based on manuscripts in which the intention was to show good communication and clinical skills. The camera viewed both the family doctor and the patient simultaneously in order to visualize the interaction in all parts of the patient encounter. The content of the VP case was created in congruence with the Calgary-Cambridge Guides [14] and the PRACTICAL [15] training model in communication skills. Open-ended questions were used as an essential part of these models.

The teachers' preformulated comments reflected evidence-based medicine relevant for the case and the tacit knowledge of an experienced clinician. The comments were written as a short text (100-150 words) following each part of the encounter. Further in-depth knowledge was accessible through embedded hyperlinks pointing to continuously updated evidence-based pages.

The average time for students to complete the VP was about 1.5 hours. The content analysis of the three group interviews resulted in the following themes.

Evaluation of the Virtual Patient Primary Care Model

The Learning Process

Students reported that they had had a generally positive learning experience working with the VP. They felt that the VP-model supported self-directed learning as it could activate students and stimulate reflection. The opportunity not only to read text, but also to actually see the doctor's actions on films, enriched the learning experience. The students quite liked the fact that they received immediate feedback to their answers and reflections. However, they also saw some potential to develop the model by more comprehensive information on the goals and intended learning outcomes. Some student wanted more possibilities for individualized structure; there they could exclude some points that they felt they had already learned.

The Contents

The students described the VP as an intermediate learning activity between theory and practice with real patients, also suitable to work with in an interprofessional student group. They appreciated that the VP dealt with common conditions that they rarely met on hospital wards. They found the texts and comments concise and comprehensive. The tutors' comments were very relevant and gave the students deepened medical knowledge. The comments also worked well as feedback, though they were general and not individualized to the answer from each student. Students wanted more emphasis on the most important and potentially dangerous conditions when reflecting on differential diagnoses. They also wanted information on how comprehensive the answers they were expected to write should be. They also pointed out some minor incongruencies in the content.

The Status of the Virtual Patient Model in Clinical Education in Primary Care

Students described it as being a very good complement to the theoretical part of the course and the clinical placement, as it gave an opportunity to fill out gaps in their clinical experience. Working with the VP gave also a welcome opportunity to meaningful learning when no other organized activities were available. Students saw a great flexibility in the VP model that, with some adaptation, could be used in teaching on all levels of the medical program in primary health care, for example, to teach about typical medical cases. Such typical cases could also be used as material for peer/group discussions and seminars. Comments included "It would be beneficial to create such VP scenarios on all levels of the medical education." and

> *During lectures...you don't listen all the time...they show PowerPoint presentations and half of the class is sleeping...but here you have to listen and you have to be sharp in your listening...you get the opportunity to relate theory to practice. If I hadn't met a diabetic patient with a leg ulcer before, then I have the opportunity here, how it looks, how I should handle it...I can follow the whole process of care...*

The Technical Design

The students described the model structure as being easy to follow. They appreciated the possibility to move both forward and backward in the model. They also provided some suggestions for improvement of the structure. A box with a summary of all patient data shown on each page might facilitate learning. The video film about the clinical examination could be shortened, as it contained some redundant information.

Discussion

Principal Findings

We present the preliminary results from a new model for virtual patients based on the learning theories applicable to primary care as a learning environment for medical students. The model has embedded communication training and a funnel of clinical reasoning through the case. Reflection-in-action embedded through the case is emphasized and ends with a reflection-on-action. The features that make this VP model suitable for primary care education include the way the model works with open-ended questions, allows free-text answers, and enables the visualization of communication and clinical reasoning processes in a family medicine context.

Our evaluation with student interviews shows this to be a promising model. The students found working with the primary care VP to be active and meaningful with a sense of authenticity. Realistic video clips were used, which was seen as an important part of the VP case design. Other studies have also shown that patient videos can add value to virtual educational tools, such as Adams and colleagues who advocate for the creation of a whole library of videos used as VPs in a primary care context [17]. An adaptation of the VP model as described by Hansen and colleagues [18], for example, as an app for mobile device, could facilitate the implementation of the present virtual patient concept.

For the students, it is important that they find the activity understandable, valuable, and meaningful [11,19]. The VP activity is designed to be self-instructed and easy to use so that students can take responsibility for their own education and learning process in line with the theories of self-directed learning [19]. One important learning activity by the students throughout the VP case is how they formulate the questions and problems related to each loop in their own words. Upon this they write their reflection that they then compare with the teacher's comments.

The context is also important for the students to become motivated to achieve meaningful learning; here the students can work with the VP case in a clinical setting at the primary health care centre. For a student's ability to apply concepts to solve new problems, active learning with multiple examples can have major effects [20]. VPs may offer students multiple examples and visualize how family doctors reason—something that is often considered as tacit knowledge. During the clinical placements, the students may then practice their knowledge from the VP case in real patient encounters so there is no long transfer process of knowledge. The integration of theory and practice should promote the students' individual construction of meaningful learning that is readily available without a long process of transfer from a theoretical context to a clinical environment. In alignment with Mitchell and colleagues [21], our findings show that the virtual patient model presents an effective learning environment, where education in patient-centered communication and counseling skills are facilitated.

The student is guided to work with critical reflection during the case. The reflection enables the student to identify their learning needs [13]. This is a way for the student to act and think professionally as an integrated part of the formative learning throughout the case. The iterated self-regulated learning cycles, one of the essential parts of the VP model were inspired by Kolb's learning cycle. For learning to occur, the student has to proceed through all stages of the cycle, which is based on a continuous flow of actions: Doing, Observing, Thinking, and Doing again are repeated. In our learning cycles, the student started out by planning their actions in the current VP section (doing), followed by a concrete experience via multimedia where they were prompted to reflective observation (observing), and after having read the preformulated feedback, wrote down their reflections (thinking). The student could also return to their text and revise it (doing again).

For learning to occur, the student has to proceed through all stages of the cycle. External feedback (ie, the teachers' preformulated comments) helped the cycle go forward, especially in the "thinking" and "doing" stages. The feedback focused on the main ways of reasoning in primary health care to guide students to the next step.

Strengths

The starting point and emphasis of our model has been the primary care medical education. In our model, we have embedded ways to promote meaningful, deep learning for the students, such as reflection, clinical reasoning, and deepened subject knowledge. The chosen tool for creating the VP is especially suitable for primary care because it allows both open-ended questions and free-text answers, and the use of video clips. This also makes it possible to visualize the tacit clinical reasoning processes of family doctors. The VP case used is an authentic patient case that was slightly modified for the purpose. It is important that the case and context are as authentic as possible, contain adequate subject knowledge, and are related to a real-life situation; otherwise, the students will not understand the meaning of the task and will not be motivated to perform the case.

The students work with the case independently, without help from their teachers. They are given relevant questions related to a case that resembles the patients in their practice at the primary health care centre. Their learning with the VP is self-directed; they are able to work with the case at their own pace and go backwards and forwards in the case pathway.

Limitations

We have also experienced some aspects that can be regarded as limitations of the model. We have inserted a step in each cycle where students have to reflect and write down their thoughts, which could take time. This limits the number of cycles that can be inserted. We used five cycles, which took 1.5 hours to perform for the students. This is still doable but can be regarded as a time limit for an effective learning activity.

In the chosen prototype VP system, there is no synchronous interaction between the student and a tutor. All feedback has to therefore be preformulated. This was inherent in the otherwise best technical solution we found, but also one of our concerns when we started the project. Surprisingly, all the students participating in the pilot test reported that they perceived the feedback as natural, as from a real tutor. They described their experience with the VP as a dynamic process that could deepen their knowledge.

Future Directions

A further development of the model would encompass the progression dimension across a whole medical curriculum where the VPs are presented in all semesters of the student's education and help to integrate theory and practice from different parts of the program. The VPs may also have a "virtual life" through the program, and progression can be designed in order to be used as an educational tool to acquire in-depth knowledge, reflective skills, and to assess the students' developmental path towards improved competencies in medical education. The integrated learning cycle, within the VP model, can easily be adapted, not only to medical education, but also to various other health care professional educations and for interprofessional use. The VP model design enables the insertion of VP cases in various professional settings and for different learning outcomes. The content of the VP cases can be reused, repurposed, and shared in different educational contexts. A wider implication and dissemination of the present VP model could be facilitated by use of Web-based platforms such as suggested in the mEducator project [22,23]. The VP model might also serve as a tool for international collaboration and shared understanding in health care education in culturally disparate structures.

Conclusions

We have built a new VP model for primary care education with a patient-centered approach that was congruent with the Calgary Cambridge Guides and contained embedded communication training by allowing the student to act and think professionally and then reflect upon a visual presentation of the different parts of the patient encounter handled by an experienced family doctor. The model contained iterated learning cycles with a didactic inventory, a concrete virtual experience assisted by multimedia, preformulated feedback from experienced teachers, and the student's reflections in free text. Pilot testing showed good acceptance by students who regarded the VP case as authentic, and the model supporting their self-directed learning and reflective thinking. Further evaluation of the model is needed.

Acknowledgments

The study was supported in part by grants from Stockholm County Council.

Conflicts of Interest

None declared.

References

1. van Soeren M, Devlin-Cop S, Macmillan K, Baker L, Egan-Lee E, Reeves S. Simulated interprofessional education: an analysis of teaching and learning processes. J Interprof Care 2011 Nov;25(6):434-440. [doi: 10.3109/13561820.2011.592229] [Medline: 21899398]
2. 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]
3. Cendan J, Lok B. The use of virtual patients in medical school curricula. Adv Physiol Educ 2012 Mar;36(1):48-53 [FREE Full text] [doi: 10.1152/advan.00054.2011] [Medline: 22383412]
4. Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med 2007 May;82(5):446-451. [doi: 10.1097/ACM.0b013e31803e8a0a] [Medline: 17457063]
5. Gesundheit N, Brutlag P, Youngblood P, Gunning WT, Zary N, Fors U. The use of virtual patients to assess the clinical skills and reasoning of medical students: initial insights on student acceptance. Med Teach 2009 Jan;31(8):739-742. [doi: 10.1080/01421590903126489]
6. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. The New England Journal of medicine 2006;355(21):0028-4793. [doi: 10.1056/NEJMra054782.ISSN:1533-4406] [Medline: 17124019]
7. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005 Jan;39(1):98-106. [doi: 10.1111/j.1365-2929.2004.01972.x] [Medline: 15612906]
8. Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, et al. Diagnostic strategies used in primary care. BMJ 2009;338:b946 [FREE Full text] [Medline: 19380414]
9. Hansen M, Oosthuizen G, Windsor J, Doherty I, Greig S, McHardy K, et al. Enhancement of medical interns' levels of clinical skills competence and self-confidence levels via video iPods: pilot randomized controlled trial. J Med Internet Res 2011;13(1):e29 [FREE Full text] [doi: 10.2196/jmir.1596] [Medline: 21447471]
10. Kolb D. Experiential learning: experience as the source of learning and development. Englewood Cliffs, N.J: Prentice-Hall; 1984:231-241.
11. Knowles M. The adult learner. In: The adult learner: a neglected species. Houston, TX: Houston [Tex.] Gulf Pub. Co; 1984.
12. Fyrenius A, Bergdahl B, Silén C. Lectures in problem-based learning--why, when and how? An example of interactive lecturing that stimulates meaningful learning. Med Teach 2005 Jan;27(1):61-65 (forthcoming). [doi: 10.1080/01421590400016365] [Medline: 16147772]
13. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2009 Oct;14(4):595-621. [doi: 10.1007/s10459-007-9090-2] [Medline: 18034364]
14. Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. Acad Med 2003 Aug;78(8):802-809. [Medline: 12915371]
15. Larsen JH, Risør O, Putnam S. P-R-A-C-T-I-C-A-L: a step-by-step model for conducting the consultation in general practice. Fam Pract 1997 Aug;14(4):295-301 [FREE Full text] [Medline: 9283850]
16. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
17. Adams EC, Rodgers CJ, Harrington R, Young MD, Sieber VK. How we created virtual patient cases for primary care-based learning. Med Teach 2011;33(4):273-278. [doi: 10.3109/0142159X.2011.544796] [Medline: 21456983]
18. Hansen M, Oosthuizen G, Windsor J, Doherty I, Greig S, McHardy K, et al. Enhancement of medical interns' levels of clinical skills competence and self-confidence levels via video iPods: pilot randomized controlled trial. J Med Internet Res 2011;13(1):e29 [FREE Full text] [doi: 10.2196/jmir.1596] [Medline: 21447471]
19. Silén C, Uhlin L. Self-directed learning – a learning issue for students and faculty!. Teaching in Higher Education 2008 Aug;13(4):461-475. [doi: 10.1080/13562510802169756]
20. Norman G. Research in clinical reasoning: past history and current trends. Med Educ 2005 Apr;39(4):418-427. [doi: 10.1111/j.1365-2929.2005.02127.x] [Medline: 15813765]
21. Mitchell S, Heyden R, Heyden N, Schroy P, Andrew S, Sadikova E, et al. A pilot study of motivational interviewing training in a virtual world. J Med Internet Res 2011;13(3):e77 [FREE Full text] [doi: 10.2196/jmir.1825] [Medline: 21946183]
22. Bamidis PD, Kaldoudi E, Pattichis C. mEducator: A best practice network for repurposing and sharing medical educational multi-type content. IFIP Advances in Information and Communication Technology 2009;307:769-776. [doi: 10.1007/978-3-642-04568-4_78]

23. Kaldoudi E, Dovrolis N, Konstantinidis ST, Bamidis PD. Depicting educational content repurposing context and inheritance. IEEE Trans Inf Technol Biomed 2011 Jan;15(1):164-170. [doi: 10.1109/TITB.2010.2092442] [Medline: 21134819]

Abbreviations

KI: Karolinska Institutet
OpenTUSL: Open Tufts University Sciences Knowledgebase
VP: virtual patient

Edited by P Bamidis, D Giordano, C Pattichis, N Zary; submitted 14.03.13; peer-reviewed by G Kopp, L Woodham; comments to author 05.04.13; revised version received 16.05.13; accepted 15.11.13; published 06.01.14.

Please cite as:
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
URL: http://www.jmir.org/2014/1/e3/
doi:10.2196/jmir.2616
PMID:24394603

Original Paper

A Framework for Different Levels of Integration of Computational Models Into Web-Based Virtual Patients

Andrzej A Kononowicz[1,2], PhD; Andrew J Narracott[3,4], PhD; Simone Manini[5,6], BEng; Martin J Bayley[3,4], PhD; Patricia V Lawford[3,4], PhD; Keith McCormack[3,4], PhD; Nabil Zary[1], MD, PhD

[1]Digital Patient Lab, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Department of Bioinformatics and Telemedicine, Faculty of Medicine, Jagiellonian University, Kraków, Poland

[3]Medical Physics Group, Department of Cardiovascular Science, Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield, United Kingdom

[4]INSIGNEO Institute for in silico Medicine, University of Sheffield, Sheffield, United Kingdom

[5]Department of Biomedical Engineering, IRCSS Mario Negri Institute, Bergamo, Italy

[6]Orobix Srl, Bergamo, Italy

Corresponding Author:
Andrzej A Kononowicz, PhD
Digital Patient Lab
Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Tomtebodavägen 18A
Stockholm, 171 77
Sweden
Phone: 46 8 524 83626
Fax: 46 8 524 800 00
Email: andrzej.kononowicz@ki.se

Abstract

Background: Virtual patients are increasingly common tools used in health care education to foster learning of clinical reasoning skills. One potential way to expand their functionality is to augment virtual patients' interactivity by enriching them with computational models of physiological and pathological processes.

Objective: The primary goal of this paper was to propose a conceptual framework for the integration of computational models within virtual patients, with particular focus on (1) characteristics to be addressed while preparing the integration, (2) the extent of the integration, (3) strategies to achieve integration, and (4) methods for evaluating the feasibility of integration. An additional goal was to pilot the first investigation of changing framework variables on altering perceptions of integration.

Methods: The framework was constructed using an iterative process informed by Soft System Methodology. The Virtual Physiological Human (VPH) initiative has been used as a source of new computational models. The technical challenges associated with development of virtual patients enhanced by computational models are discussed from the perspectives of a number of different stakeholders. Concrete design and evaluation steps are discussed in the context of an exemplar virtual patient employing the results of the VPH ARCH project, as well as improvements for future iterations.

Results: The proposed framework consists of four main elements. The first element is a list of feasibility features characterizing the integration process from three perspectives: the computational modelling researcher, the health care educationalist, and the virtual patient system developer. The second element included three integration levels: basic, where a single set of simulation outcomes is generated for specific nodes in the activity graph; intermediate, involving pre-generation of simulation datasets over a range of input parameters; advanced, including dynamic solution of the model. The third element is the description of four integration strategies, and the last element consisted of evaluation profiles specifying the relevant feasibility features and acceptance thresholds for specific purposes. The group of experts who evaluated the virtual patient exemplar found higher integration more interesting, but at the same time they were more concerned with the validity of the result. The observed differences were not statistically significant.

Conclusions: This paper outlines a framework for the integration of computational models into virtual patients. The opportunities and challenges of model exploitation are discussed from a number of user perspectives, considering different levels of model integration. The long-term aim for future research is to isolate the most crucial factors in the framework and to determine their influence on the integration outcome.

(J Med Internet Res 2014;16(1):e23) doi:10.2196/jmir.2593

KEYWORDS

computer simulation; computer-assisted instruction; education, medical; medical informatics applications

Introduction

Background

Computers and Internet technologies have already entered the mainstream of health care education [1]. Although a gap still exists between what is technically possible in pilot studies and the realities of educational practice, we have reached a stage where Web-based training is regarded as routine [2,3]. The use of such technology is driven not only by the urge for innovation, a willingness to improve teaching quality, and limited personnel, but also by cost-effectiveness analysis [4].

The use of virtual patients is undeniably one of the techniques most often associated with application of computer-aided training in health care [2,5] but, as is often the case with new concepts, the understanding of the term “virtual patient” varies depending on the research community. We define a virtual patient as “interactive computer simulation of real-life clinical scenarios for the purpose of health care and medical training, education or assessment” [6]. This excludes other methods used in medical education such as human role-playing, computerized mannequins, part-task trainers, and systems requiring specialized equipment [7], as well as all non-educational virtual patients.

While complex, immersive virtual reality scenarios are technologically possible (eg, [8]), the routine use of virtual patients often focuses on technically simple solutions. Huwendiek et al attempted to classify this type of virtual patient solution with four example systems: CAMPUS, CASUS, Open Labyrinth, and Web-SP [9]. While these systems differ in many respects, they have in common the presentation of a clinical, case-based scenario divided into discrete steps displayed either (1) linearly, with a single final outcome, or (2) branched, enabling different narration paths depending on user choice. All these systems are Web-based and simply require an Internet connection through a standard browser. By common agreement, the ANSI-accredited MedBiquitous Virtual Patient (MVP) standard [10] has been adopted by the medical education community for the representation of virtual patient data. Implementation of support for the standard in four systems was achieved as part of the eViP project [11].

There are at least two significant advantages of this type of virtual patient. The first is the possibility of significant teacher involvement in the development process. This has been enabled by investment in the user-friendliness of authoring tools and the simplification of technical workflows to enable medical experts to focus on the content of virtual patients. As a result, virtual patients of this class are generally tailored to the needs of a particular teacher and institution, to the type of educational activity, and to the specific learning objectives. This might not be the case with technologically complex virtual patients. The second advantage is the high level of accessibility of this type of virtual patient to learners over the Internet with personal computers, or even “just-in-time” access with mobile devices. Accessibility is also enhanced by the low cost of licensing of these virtual patient systems.

There are several possible “next steps” in the development of virtual patients from the class described above. Our chosen step is to increase the interactivity of virtual patients by enriching them with computational models of physiological and pathological processes. Here we use the term “computational model” following the definition given by Garrido [12] as a “mathematical model implemented in a computer system that requires high performance computational resources to execute”. This usually requires the numerical solution of non-linear models. Being cognizant of the long research tradition of modeling and simulation in biomedical engineering (eg, [13]), it is not our intention to create new models but rather to seek opportunities to integrate existing models into present virtual patient systems.

Here we propose a framework to support integration of existing computational models of physiological function within virtual patients. This framework is expected to define concepts that help in formalizing the integration and evaluating the feasibility of the integration. Extending virtual patients with computational models is recognized as a promising trend in the current virtual patient literature (eg, [3,14]).

While integration frameworks for simulation in health care education have been published previously (eg, [15-17]), the focus of these earlier publications is different from that envisaged here. The “Practica continua” framework, developed by Ellaway et al [15], discusses integration of a broad range of simulation modalities, covering methods such as human role-playing (eg, standardized patients), physical equipment (eg, computerized mannequins, part-body task trainers, and haptic devices), teleconferencing, and 3D visualization tools. The framework we present is focused solely on integration of physiological computational models derived from existing research projects, such as the Virtual Physiological Human (VPH) initiative [18], within screen-based, narrative virtual patient systems (as used in the eViP project [9]).

The RICORDO framework has a more technical scope, with the aim of improving the accessibility of simulation data and model resources for physiology and pharmacology research, as well as health care education [16]. This involves selection of technical standards for the use of controlled vocabularies, ontologies, and metadata encoding representation. In addition, an open toolkit for creation of ontology composites and annotation of resources was developed. A critical appraisal of technical standards for sharing educational resources in medical education was also undertaken by the mEducator Best Practice Network [19]. These contributions can complement the methods described in the current paper by facilitating the discovery of new models to be integrated in existing virtual patients and by promoting dissemination of these based on clinical classification of the content.

Finally, frameworks are described within the literature, which facilitate the integration of various software libraries for high-fidelity medical simulation. For instance, Halic et al proposed a framework called SoFMIS for rendering 3D scenes in surgical simulation involving the use of haptic devices [17]. A related framework adopts the SoFMIS architecture for simulation in Web environments [20]. In contrast to this approach, our framework deals more with virtual patient authoring strategies than with concrete software solutions. The SoFMIS framework demonstrates how different software libraries may be grouped in modules supporting reusability. We intend to extend the existing, well-established systems with new elements. Furthermore, our virtual patients are directed towards improving clinical reasoning skills (as diagnostic and treatment selection processes) and do not focus on the psychomotor and procedural skills that would be required in high-fidelity surgical simulations.

Objective

The aim of this paper is to propose a conceptual framework to organize research and development towards the integration of computational models within virtual patients. In particular, we focus on (1) characteristics to be addressed while preparing the integration, (2) the extent of the integration, (3) strategies to achieve integration, and (4) methods for evaluating the feasibility of integration.

The framework will be considered successful if its application leads to practical recommendations and predictions enabling the medical education and simulation communities to collaborate in the integration of computational models within virtual patients. Our long-term aim is to isolate the crucial factors and determine their influences on the success of the integration process in supporting the practical use of the virtual patient for teaching. One way of achieving this is to evaluate the reaction of a suitable target audience to virtual patients implemented with different levels of model integration proposed by the framework. For this reason, an additional goal is to pilot an investigation of altered perceptions of virtual patients at two levels of integration.

Methods

Process for Developing the Framework

The framework will be initiated using an iterative process inspired by the Soft System Methodology (SSM) [21]. The SSM method was designed to aid analysis of situations involving groups with interacting perceptions where human-related aspects play an important role. The methodology looks for solutions that are both desirable and feasible considering the components of the system.

While developing virtual patients, we should address human-related aspects. For instance, their construction should be influenced by actual learning needs and preferences of learners and not by the accuracy of simulation results alone. This makes SSM methodology particularly suitable for the purpose of developing the framework. This methodology has been previously applied to design conceptual frameworks to introduce information technologies into health care, for instance by Ruotsalainen et al to define a framework for Trusted Pervasive Health [22].

The SSM method consists of four steps, constituting a learning cycle, which are repeated to improve the system design [21]. This involves (SSM Step 1) characterization of the system in question including description of its features, problems of interest, risks, and challenges, (SSM Step 2) expression of the ideas and planned actions in the form of models encapsulating different perceptions, (SSM Step 3) testing of the model on a real-world example, and (SSM Step 4) synthesis of the system description with verification outcomes to propose improvements for the next development cycle.

Source of Computational Models

The initial framework design is informed by experience acquired by the authors in previous projects dealing with virtual patients, health care education and biomedical modeling simulation, including eViP [6] and the Virtual Physiological Human (VPH) initiative [18]. The recently established collaboration between the VPH Network of Excellence (NoE) and representatives of the Association for Medical Education in Europe (AMEE) health care educational community provided an appropriate environment to develop the framework.

The overarching aim of the VPH is to establish methods and tools for computational analysis of the human body, integrating the diverse nature of physiology and pathophysiology of the different organ systems [18]. The VPH combines expertise in computer modeling and clinical research to deliver a spectrum of advanced simulations of physiological function, disease development and progression, and response to intervention. The ambitious goal is to develop a technological infrastructure based on patient-specific data to deliver prediction of clinical outcomes by means of quantitative models that integrate biophysical processes across diverse scales from the molecular level, to organ systems, and even populations [23]. The clinical targets of VPH projects are diverse and address the challenges of computational modeling in cardiology (eg, euHeart), vascular pathology (eg, ARCH), or various types of cancer (eg, PASSPORT) [24]. Effort has also been allocated to address the technical aspects of the computational infrastructure required by VPH projects (eg, VPH Share).

The VPH NoE [25] has provided a central focus for VPH project output, developing best practice and support for the exposure and sustainability of VPH-related tools, training, and standards. An important task of the VPH NoE has been to ensure that the academic, medical, and industrial domains have access to a workforce that is well prepared to meet the possibilities offered by the VPH. This has been addressed through delivery of training activities (workshops, summer schools) and the development of educational materials to raise awareness of these new technologies [24]. The educational goals of VPH NoE align well with the aims of the virtual patient community for more interactive virtual patients.

Demonstration of the Framework

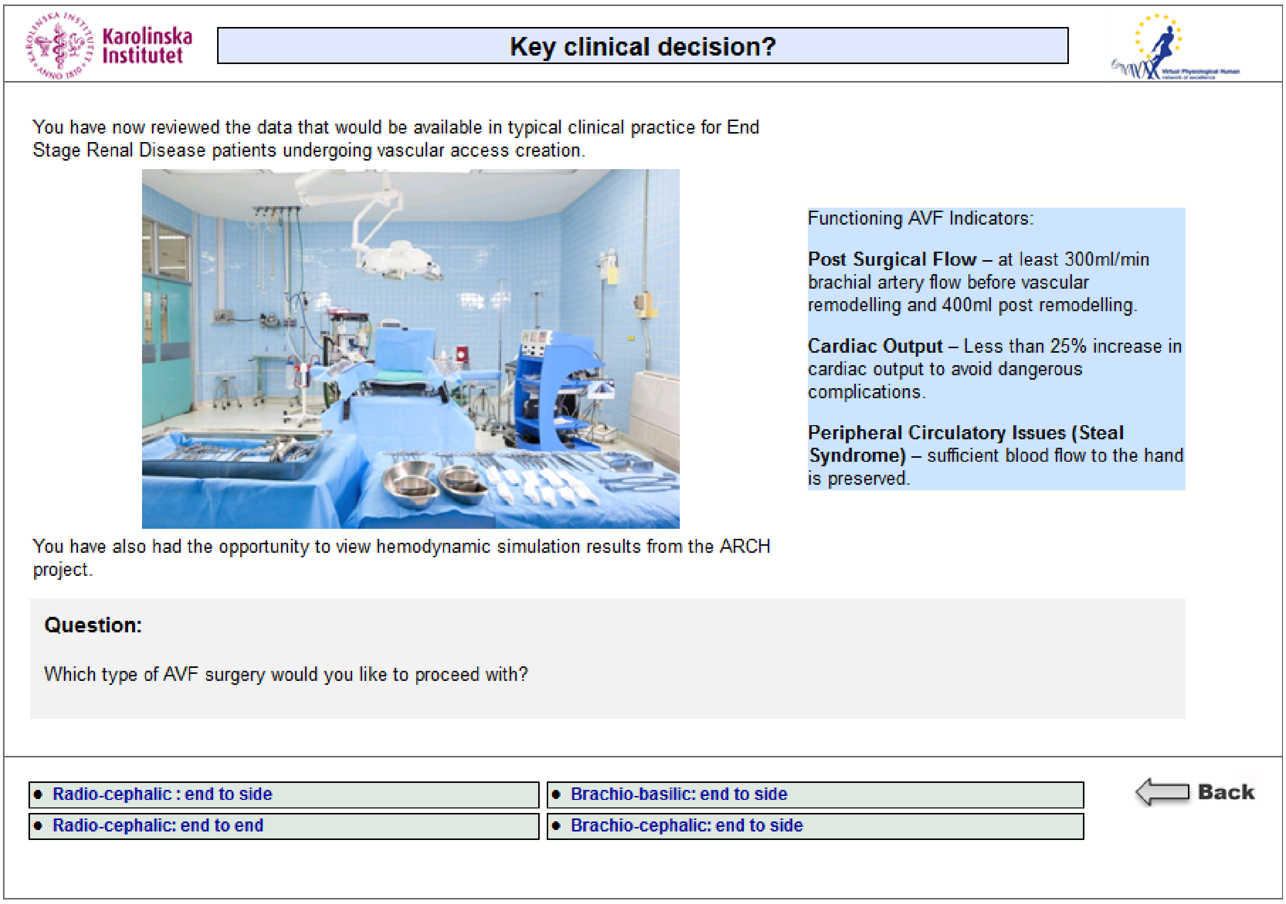
Selected Model for Integration

The proposed framework is showcased using a real-world example: the VPH ARCH project [26]. This project developed vascular access modeling for surgical planning for hemodialysis in end-stage renal disease (ESRD) patients. While there are several methods to create vascular access for ESRD patients, the arteriovenous fistula (AVF) is preferred due to reduced potential complications. An AVF may be created either in the lower arm (radiocephalic fistula) or upper arm (brachiocephalic or brachiobasilic fistula) [27]. Deciding where to create the fistula is informed by preoperative physical examination and duplex ultrasound evaluation of the vasculature of the arm. Within the VPH ARCH project a computational model was developed to support clinical decision making. The simulation is based on a distributed lumped-parameter implementation of a wave propagation model [28], which has been evaluated with clinical data. The computational model is available under an open-source license as part of a software toolkit called archTk [29] and as part of a clinical Web application [30].

The Integration Process

Using the archTk model and a set of patient input parameters, we simulated blood flow and pressure in the upper extremity after implementing different types of AVF, incorporating simulation outcomes into a virtual patient. The case presents the story of John Jones: a 68-year-old male who was diagnosed with chronic kidney disease following an infection. The virtual patient was developed under the supervision of clinical experts by authors of this paper. The basic version contains 28 screen cards connected in a graph structure to provide branching possibilities (Figure 1). The screen cards comprise narratives, medical examination results (including simulation outcomes), and images. The integration was performed using Bit Pathways [31,32]. This authoring tool was developed by one of the paper's authors and allows export of the graph structure, including the case data, as an MVP package [10]. Such reusable learning objects are readable by all MVP-supported virtual patient systems. Alternatively, the content may be wrapped by the Bit Pathways authoring tool into an HTML/Java Script player to form a stand-alone or Web-based application (Figure 1).

Figure 1. The virtual patient case containing archTk simulation outcome wrapped in a stand-alone, Web-enabled virtual patient player.



You have now reviewed the data that would be available in typical clinical practice for End Stage Renal Disease patients undergoing vascular access creation.

Functioning AVF Indicators:

Post Surgical Flow – at least 300ml/min brachial artery flow before vascular remodelling and 400ml post remodelling.

Cardiac Output – Less than 25% increase in cardiac output to avoid dangerous complications.

Peripheral Circulatory Issues (Steal Syndrome) – sufficient blood flow to the hand is preserved.

You have also had the opportunity to view hemodynamic simulation results from the ARCH project.

Question:

Which type of AVF surgery would you like to proceed with?

Evaluation Event

Context Description

To test the potential of the proposed framework as a tool for discriminating between different levels of simulation integration, a virtual patient exposing the archTk simulation results was prepared in two variants, reflecting the different integration levels defined by the framework. Both variants of the case were demonstrated during a dissemination and evaluation workshop organized by the VPH NoE at the European Vascular Course (EVC) conference in Maastricht in March 2013. Access to the virtual patients was offered to the participants by 10

Internet-connected Apple iPad 2 tablets. The virtual patient variants were stored on a remote Web server. The user interface of the Web application was tailored to the size and navigation possibilities of the tablet's touch screen. Participants were allocated to one of the two virtual patient variants at random (Study Groups 1 and 2) based on the tablet the user selected to sit at in the workshop. Participants were blinded to this selection to prevent bias (Hawthorne effect).

Participation in the workshop was voluntary. Participants were recruited by email announcements broadcast by the conference organizers, posters at the conference venue, and direct invitations with pamphlets handed out during coffee breaks. Participation required registration for a specific workshop session due to a limit of 10 individuals per session. Use of the virtual patient was preceded by a 15-minute introduction to the VPH initiative and virtual patient tools. A Web-based evaluation questionnaire, authored with Google Forms, was completed immediately after the session. User identification was based only on the tablet ID and time of questionnaire completion. The questions asked pertained to the level of interest in the virtual patient and agreement with the content. In addition, the study group dealing with the higher integration level were given the possibility of commenting on the added interactivity (as explained in the results section of this paper).

Study Participants

Thirty-eight participants filled in the questionnaire. The distribution of the participants in study groups was well balanced with 20 (53%) in Study Group 1 and 18 (47%) in Study Group 2. Basic demographic and background data are presented in Table 1. Most of the participants were clinicians specializing in vascular surgery or nephrology (28/38, 74%). The remaining participants were nurses, biomedical engineers, and producers of medical equipment (9/38, 24%; 1 missing response).

Table 1. Summary of the study participants' demographic data.

	Group 1	Group 2
Number of participants	20	18
Gender: female/male/no answer, %	75/25/–	61/22/17
Age, yrs, mean (SD)	41.6 (12.6)	36.0 (11.8)
Physician, %	85	61
Work experience, yrs, mean (SD)	11.8 (12.5)	9.5 (9.0)
Involved in teaching, %	80	72

Statistical Analysis

Differences between the two groups were assessed using a non-parametric double-sided Mann-Whitney U-test. The significance level (alpha) was set to .05. The statistical analysis was performed using the R statistical package version 3.0.1 (the R Foundation for Statistical Computing, 2013).

Results

The Integration Framework

System Definition

The central element of SSM methodology is a real-world "problematic situation". In our case, we define it as low level of interactivity of virtual patients to be increased through the integration of existing simulation technologies. We envisage a community where computational models produced by research projects are publicly available on the Internet and selected by the health care education community to extend the interactivity of existing or newly developed virtual patients.

The challenges of this approach are defined by different "worldviews" of the stakeholders (SSM Step 1) that include the biomedical modeling community, health care educators and learners, and virtual patient system developers.

The current perspective of the biomedical modeling community is well represented by the diversity of the VPH initiative. It focuses on the development and validation of patient-specific simulations to address clinical problems relevant to a particular patient population. Model construction is intellectually demanding with, as yet, little formalization. This focus is driven by researcher motivation, clinical drivers, and the availability of research funding, involving state-of-the-art technologies (eg, grid systems, cloud computing, supercomputers). One priority of the VPH initiative is to develop models and simulations that are able to provide a holistic view of the human body and able to deliver a technical infrastructure capable of providing access to different models and tools within a single computational workflow. Authoring of models and simulations is typically carried out in large international teams and requires formal consideration of ethical and intellectual property issues, particularly when industrial partners are involved. VPH researchers are largely unaware of the learning objectives of medical curricula, and the research focus of projects means that resources to adapt or maintain the tools for educational purposes are limited.

The worldviews of health care educators and learners, though clearly distinct, are closely interconnected and can be treated as one subsystem at the current stage of development of this framework. Driven by sound theoretical foundations, such as Kolb's experiential learning theory [33], the community is building up an educational system where the learners experience the subject of interest by active experimentation, observations, and reflection. This group is willing to use simulations that clearly address learning objectives within the contemporary medical curriculum. It is important to consider factors such as the appropriate level of difficulty of the task and to promote high levels of interaction with the content or content-specific feedback [34]. However, caution is advised because too much interactivity may be detrimental to the learning process [35]. Most educators and students are not interested in, or do not have time, to study the technology that underpins the simulation. Their major expectation is to obtain a tool that is intuitive, works without delays, is well aligned with previous learning experience, and returns valid results. This group usually has access to standard IT equipment and Internet access, and neither the faculty nor the learners are willing to cover substantial additional expenses for access to simulation.

The worldview of the virtual patient system developer is, perhaps, not as clear to an outside observer. It is important to consider this perspective as the technical infrastructure in many medical universities has been in development for several years and is well established for educational workflows, particularly in the case of virtual patient systems [36]. Replacing well-functioning systems or adding new systems to an already complex e-learning infrastructure is likely to encounter resistance. These objections will be supported by health care educators who have already invested substantial resources in creating content for existing virtual patient systems. The virtual patient format, while existing in several variants, follows similar design rules [34,37]. In practice, this involves the presentation of a consistent story for a single patient, with new information unraveling over time. Virtual patient systems, which have been implemented in different technologies, have already achieved a common technological denominator in the form of the MVP standard [10].

Feasibility Features

The different perspectives characterized within the system definition were transformed into system features (Table 2). This list contains factors that should be considered when assessing the feasibility of simulation integration. To ensure that the most significant and timely factors have been included, its completeness has been validated against features identified in previous studies presenting the views of the three stakeholder groups (sources [9,11,16,34,37-41] are shown in the third column in Table 2). The list is not assumed to be exhaustive and has the potential to be extended in future.

Table 2. Features relevant for integrating computational models with virtual patients.

Stakeholder group	Identified features relevant for integrating computational models with virtual patients	Previous study
Computational modeling researchers	Availability of high-quality documentation (including a clear description of modeled parameters: their permitted input ranges, simulation steps, and post-processing steps)	[16,38,39]
	Validity of simulation results generated (compatibility with experimental data or expected observations)	
	Availability of model in machine-readable (preferably popular) format	
	Availability (and preferably mobility) of the simulation software for the model	
	Information on the magnitude of computational time required for simulation	
	Information on mobility and required storage space demands for input and output data, model, solver	
	Clearance of copyright issues (information about the authors of the model and terms of use and distribution)	
	Description of confidentiality constraints	
Health care education (educators and students)	Suitable learning objectives	[34,37,40]
	Relevance for study	
	Suitable target group	
	Appropriate level of difficulty	
	High interactivity	
	Availability of specific feedback	
	Optimal use of media	
	Focus of attention on relevant learning points	
	Recapitulation of key learning points	
	Authentic Web-based interface	
	Content tailored to the clinical reasoning process	
	Realistic narration to include the simulation in the case	
	Support for individualized approach to learning	
	Support for collaborative learning	
Virtual patient system developers	Simulation elements supported by the virtual patient system	[9,11,41]
	Simulation elements supported by the MVP standard	

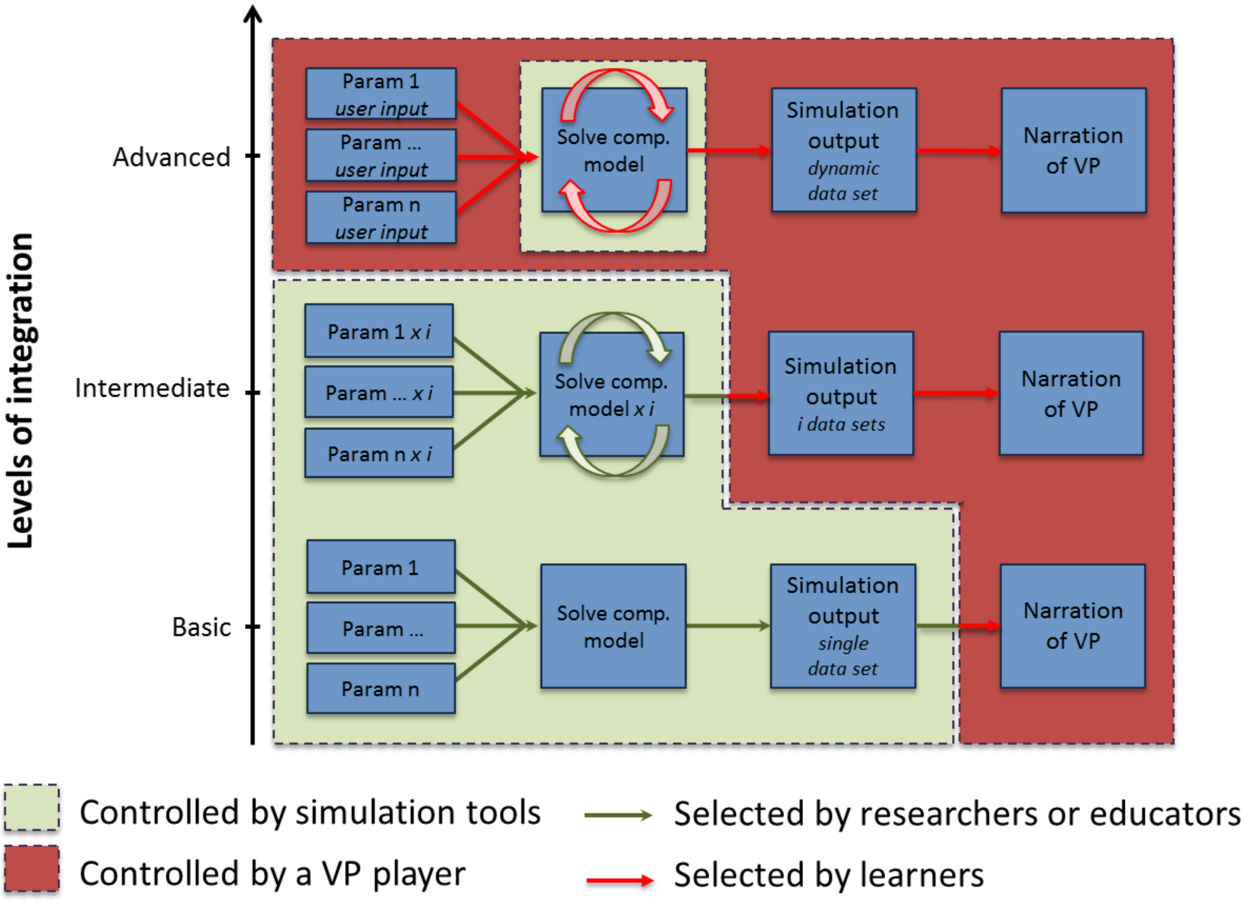
Integration Levels

Analysis of the feasibility features presented above depends largely on the extent to which the VPH-related data, models, or simulations are integrated within the virtual patient. This may be addressed by introducing different levels of integration. We propose three fundamental integration levels: (1) basic, (2) intermediate, and (3) advanced (Figure 2). In the discussion that follows, it is assumed that the virtual patient has a branching navigation model where the learner can select from a number of alternative options, solving the virtual patient with different possible end points [9]. The alternative linear model, where there is just one narration thread and one final end point, can be regarded as a special case of branching.

At the basic integration level, a single set of n input parameters is taken to generate a single outcome (or a single time-dependent outcome series) of biomedical variables for a single (virtual) patient. The results are reported within the narration of the virtual patient. For the intermediate level, pre-generation of simulation data over a range of input parameters is performed and included within the virtual patient package, for instance in the form of a look-up table. The execution environment enables the student to explore i different variants of data combination within the predefined constraints. The advanced level proposes a dynamic solution of the model while allowing the learner to work directly with the solver and freely specify the input within the domain of variables.

Figure 2 shows that, as the level of model integration is increased, the run-time control of the virtual patient simulation shifts from simulation infrastructure (green background in Figure 2) to the virtual patient player (red background in Figure 2). In the case of advanced integration, the model solver becomes an integral part of the virtual patient environment, enabling more self-reliance for the learner and, consequentially, supporting an explorative learning approach to a larger extent.

Figure 2. Three levels of integration of computational models into virtual patients.



Integration Strategies

Overview

Integration strategies provide guidance on how to implement the integration in practice. The basic level integration strategy is to manually copy the output of the solver to the relevant places in the virtual patient narrative (s_0). In this paper, we propose four higher level strategies that apply to the intermediate and advanced levels of integration: (1) narrative integration, (2) integration with branching nodes, (3) characteristics of data, and (4) model location. The strategies are not mutually exclusive and can be applied in parallel. The list of integration strategies has the potential to be extended to encompass additional considerations in the future as both simulation and virtual patient technologies continue to develop.

Narrative Integration (s1)

This strategy determines how simulation outcomes fit within the narrative of the virtual patient. This is highlighted by the situation where a single patient cannot have different values of various anatomical or physiological parameters. Two modes of integration are proposed: "what-if-nodes" and "multiple-case-packages". The what-if-node interrupts the narration of a single activity node in the navigation graph with a discovery learning task. Students are presented with a simulation output following the main narrative of the virtual patient but are encouraged to reflect how the results would differ if the input parameters change. By manipulating input parameter values (eg, by sliders or combo boxes), different simulation outcomes are loaded from a set of pre-generated values (intermediate integration level) or generated by dynamic solution of the model (advanced integration level). Learners can alter the input parameters an unlimited number of times. Unless combined with an integration mechanism that influences the branching nodes, after leaving the what-if-node the case resumes the original narrative thread of the virtual patient. One single case could encompass several what-if-nodes. The multiple-case-package entails dynamic creation of a whole population of virtual patients from one virtual patient package. The narrative of the activity nodes are formulated as templates with empty locations to hold values from either a set of pre-generated values (intermediate integration level) or dynamically generated by an integrated solver (advanced level). The selection of values for input parameters could be influenced by student interaction, follow a predefined range, or be selected at random and would not be changed after virtual patient navigation has started. The generation of these multiple cases would be carried out by the virtual patient player. A similar method has been used by Tworek et al in the Open Labyrinth virtual patient system to produce 97 virtual patients [42]. However, in this latter study the input values were taken from controlled vocabularies, statistical distributions of normal values, and manual correction of pathologies by subject experts and not dynamically created by solving computational models. Kononowicz et al proposed a variant of the multiple-case-package strategy for generating virtual patients from templates in computer-interpretable guidelines [32].

Integration With Branching Nodes (s2)

This strategy specifies whether simulated results have dynamic influence on branching nodes. In the "no influence" mode, the what-if-nodes and multiple-case-packages retain the same static branched navigation structure in all cases. No matter how the input parameter is manipulated in a what-if-node, after the student leaves this node the virtual patient player resumes the original narration thread. Similarly, a multiple-case-package would always have the same solution path. Alternatively, the definition of branching nodes could contain simulation variables encapsulated in a formal logical expression (influence on branching mode). These expressions would be evaluated dynamically by the virtual patient player during the run-time to automatically perform branching or to alter the scoring of decisions made by the student while solving the case. In this mode, the what-if-node would change the succeeding activity sub-trees. A potential scenario might involve a student trying out different levels of drug dosage to discover immediate reactions of the virtual patient before making the final decision. For the multiple-case-package, the same route taken by the student through the activity graph could result in different scoring based on the simulated output of a randomized or pre-selected input set.

Characteristics of Data (s3)

This strategy describes the selection of model input parameters. The "simple data" mode allows selection of any combination of values from permitted input ranges. At the intermediate level of model integration, these ranges are discrete whereas, for the advanced level, they might be close to continuous. The designer of the virtual patient may wish to include input parameters that do not influence simulation outcomes to improve student motivation, include distractors as part of the learning design or increase the realism of the case. Examples of distractor parameters include the name of the virtual patient, description of the hospital setting, or physiological parameters that do not influence the simulation (eg, eye color) [42]. The "simple data with exclusion" level enables specification of a set of excluded input values (to define a subspace of the permitted input space) for which no simulation results will be generated (eg, because they are physiologically impossible). Finally, for "interdependent data" a functional connection (eg, gender specific values, values with a specific non-negligible biological feedback loop, dependencies preserving anatomical continuity, etc) would be defined between some input variables to either guide or restrict the learner's choice of input parameters.

Model Location (s4)

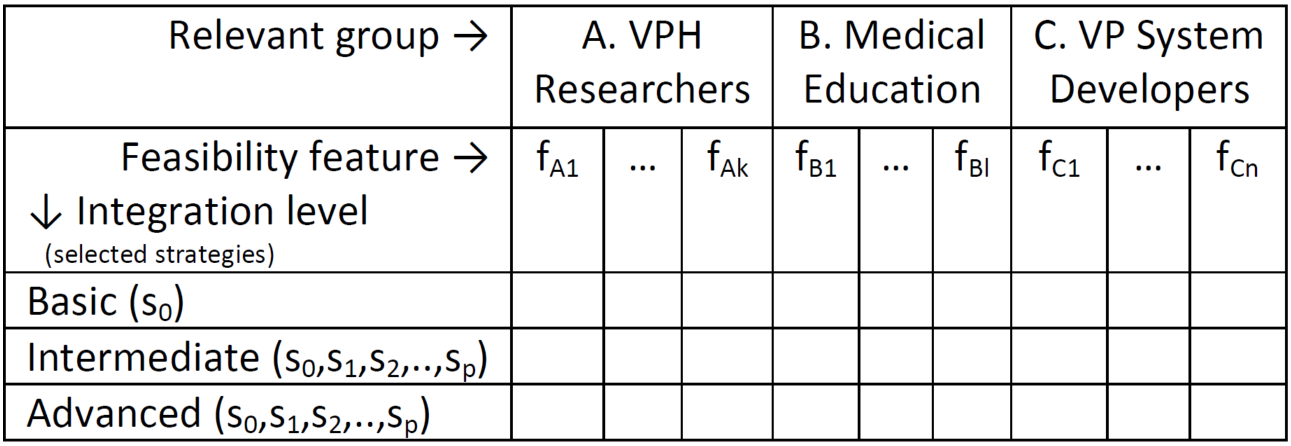
This strategy defines where simulated data or simulation software is located. For the intermediate level of integration operating in a "local" mode, the generated data are located within the virtual patient package. A "distant" mode would involve data dynamically loaded from a central repository of pre-generated data (because of substantial storage space requirements or for confidentiality due to sensitive, patient-specific information, for instance [43]). This could be managed either by the virtual patient system itself or through a service independently accessible on the Internet. For the advanced level of integration, where model solution is integrated with virtual patient navigation, the local mode requires either direct incorporation of the solver within the virtual patient package or inclusion of the solver as part of virtual patient execution environment. The advanced level of integration operating in a distant mode requires communication with an external solver service, for example, by Web Services or some other form of Web interface.

Evaluation Profiles

Decisions about the feasibility and/or desirability of performing a particular level of integration for a given computational model should be made after considering the perspectives of the relevant stakeholders. This is represented visually in Figure 3 and forms a tool for evaluation of the feasibility of integration (SSM Step 2). The columns represent the features identified in Table 2 for particular stakeholder groups, identified by the notation $f_{[\{A,B,C\}]}$, where the subscripts define the stakeholder group and the feature reference number. The integration levels are also parameterized by the integration strategies ($s_0,s_1,s_2,\ldots,s_p$).

The evaluation table can be populated through the use of evaluation profiles to guide consultation with the appropriate stakeholder communities. An evaluation profile is defined by the selection of a subset of feasibility features and specification of acceptance thresholds for critical features to reflect the priorities and requirements of the integration. These evaluation profiles can be used to determine the status of cells using colors (green—threshold passed; red—failed; yellow—borderline or unknown; grey—not relevant for decision). If a threshold is not specified for a given feasibility feature, the feature may be discussed during consultation to provide qualitative feedback, but not quantitatively evaluated. Potential evaluation profiles classes might include "integration for optimal exposure of the VPH tool" (ie, project dissemination purpose), "integration for high interactivity", "integration for high relevance to formative assessment within a particular curriculum", and "integration for optimal use in a particular virtual patient environment" (eg, in "Open Labyrinth" virtual patient player).

Figure 3. Method for systematic evaluation of different levels of simulation outcome integration.



Relevant group →	A. VPH Researchers			B. Medical Education			C. VP System Developers		
Feasibility feature → ↓ Integration level (selected strategies)	f_{A1}	…	f_{Ak}	f_{B1}	…	f_{Bl}	f_{C1}	…	f_{Cn}
Basic (s_0)									
Intermediate ($s_0,s_1,s_2,..,s_p$)									
Advanced ($s_0,s_1,s_2,..,s_p$)									

Demonstration of the Framework on a Real-World Example

Profile

An evaluation profile of the type "integration for optimal exposure of the VPH tool" was defined for the archTk simulation (SSM Step 3). Tables 3 and 4 present details of the profile including feasibility features, their acceptable thresholds (Table 3), and selected integration strategy modes (Table 4). The designated thresholds reflect the requirements for presenting the project outcomes in workshops and postgraduate courses using standard PCs or tablets without significant computational delays (<5 seconds).

Table 3. VPH ARCH evaluation profile—selected feasibility factors and thresholds.

A. VPH researchers	AT[1]	B. Medical education	AT	C. VP system developers	AT
f_{A1}: computational time	<5s[2]	f_{B1}: suitable target group	Any[3]	f_{C1}: VP player support	–
f_{A2}: storage requirements	(<1MB)	f_{B2}: appropriate level of difficulty	–	f_{C2}: MVP standard support	–
f_{A3}: results validity	[4]				
f_{A4}: intellectual property	Yes				

[1] AT: acceptance threshold.

[2] Time measured by the student.

[3] Within the context of health professionals.

[4] Face validity (system performs as a subject matter expert would intuitively expect).

Table 4. VPH ARCH evaluation profile—selected integration strategies.

Strategy	
s_1: narration integration	(one) what-if-node
s_2: integration with branching nodes	No influence
s_3: characteristics of data	Changeable input parameters: gender, age, height, weight
	Output parameters: Predicted brachial artery flow (mL/min) and pressure (mmHg)
	Dependent variables: weight and height values specific to gender, typical distributions of vascular anatomy.
s_4: model location	Local (within virtual patient player)

Computational Modeling Researchers Feasibility Features

The pulse wave propagation model applied in the VPH ARCH project as part of the archTk toolkit was initially developed by Huberts et al [28]. It depends on over 70 parameters, many of which are patient-specific and difficult to measure [44]. To decrease model complexity, the number of parameters may be reduced by generating typical distributions of vascular anatomy and material properties for a male and female of specific age and BMI. In the virtual patient context, the input parameters are expressed as weight, height, gender, and age. In general, the patient-specific nature of such computational models complicates the use of generic parameter sets. However, the use of this approach for the archTk model is supported by published data [45]. Assessment of the influence of the choice of input parameters on the effectiveness of model integration has not been undertaken in this study and is an important topic for further refinement of the framework.

Generation of a single set of output values on a standard PC workstation (Intel Core 2; 1.66GHz; 4GB RAM) for the archTk model takes 5 minutes. At the intermediate level of integration, we have evaluated a test case with a range of input parameters that include two values for the gender parameter and two values of age, weight, and height, which depend on gender, giving a total of 16 result sets. Calculation of this level of simulation data requires 80 minutes of computation prior to virtual patient navigation. This is feasible. The display of pre-generated values takes less than one second, which is sufficiently short to ensure acceptable results at the basic and intermediate level of integration. Dynamic generation of a dataset is not efficient enough on a standard PC with the current solver implementation (pyNS 0.4.2) to support a swift (<5s) response of the graphical user interface. Storage requirements for all levels of integration are negligible as the model output consists of small numbers of ASCII files.

We were able to test the validity of the outcomes for the basic and intermediate level of integration with the experts with whom we are collaborating. The validity of the advanced level of integration cannot be assessed with the current level of development of the simulation software.

Intellectual property issues are not a barrier with this application. The software is available in an open-source form allowing unrestricted use for educational purposes.

Health Care Education Feasibility Features

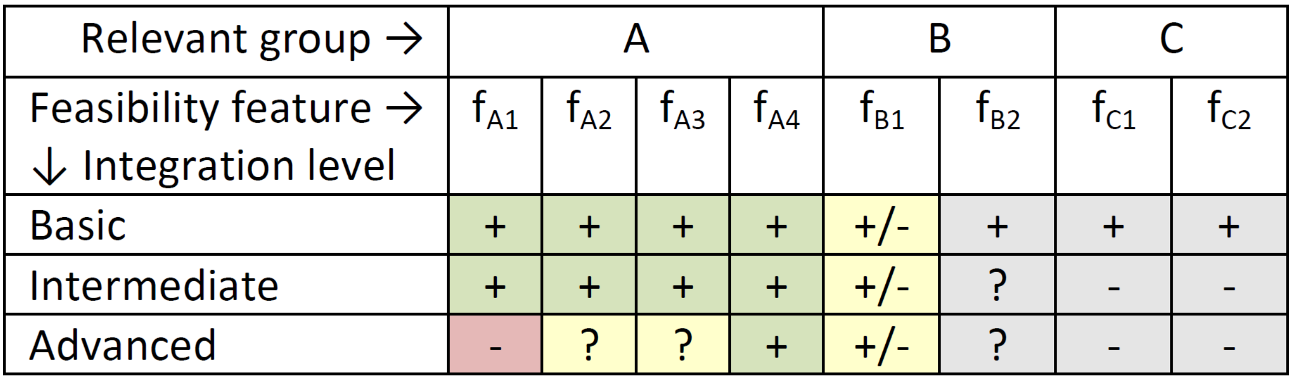
An evaluation was performed of the integration potential of the VPH ARCH project virtual patient before initiating the study through collaboration with AMEE representatives. A challenge was identified during this consultation relating to the exploitation of such simulations for undergraduate level training. The high level of clinical specialization of the content exceeds the scope of the undergraduate curriculum; vascular access surgeon training was recommended as a more suitable target. A second target group identified was medical science or medical informatics students with an interest in the application of information and communications technology methods in medicine.

Virtual Patient System Developers Feasibility Features

Presentation of simulation data at the basic level of integration of the archTk is possible in any virtual patient system compatible with the MVP standard. The what-if-node is not currently supported by many virtual patient systems or by the MVP standard, and while it is possible to generate 18 activity nodes to represent each what-if node, this process is cumbersome for the end user and could result in practical limitations for large input parameter ranges. We do not regard the current status of the f_{C1}feature (virtual patient player support) as critical since we intend to extend existing open-source virtual patient systems to accommodate such features on which the intermediate level of integration relies.

The feasibility analysis of the VPH ARCH simulation integration for "integration for optimal exposure of the VPH tool" profile (Figure 4) concluded that integration is possible for the basic and intermediate levels.

Figure 4. Evaluation profile for the VPH ARCH integration.



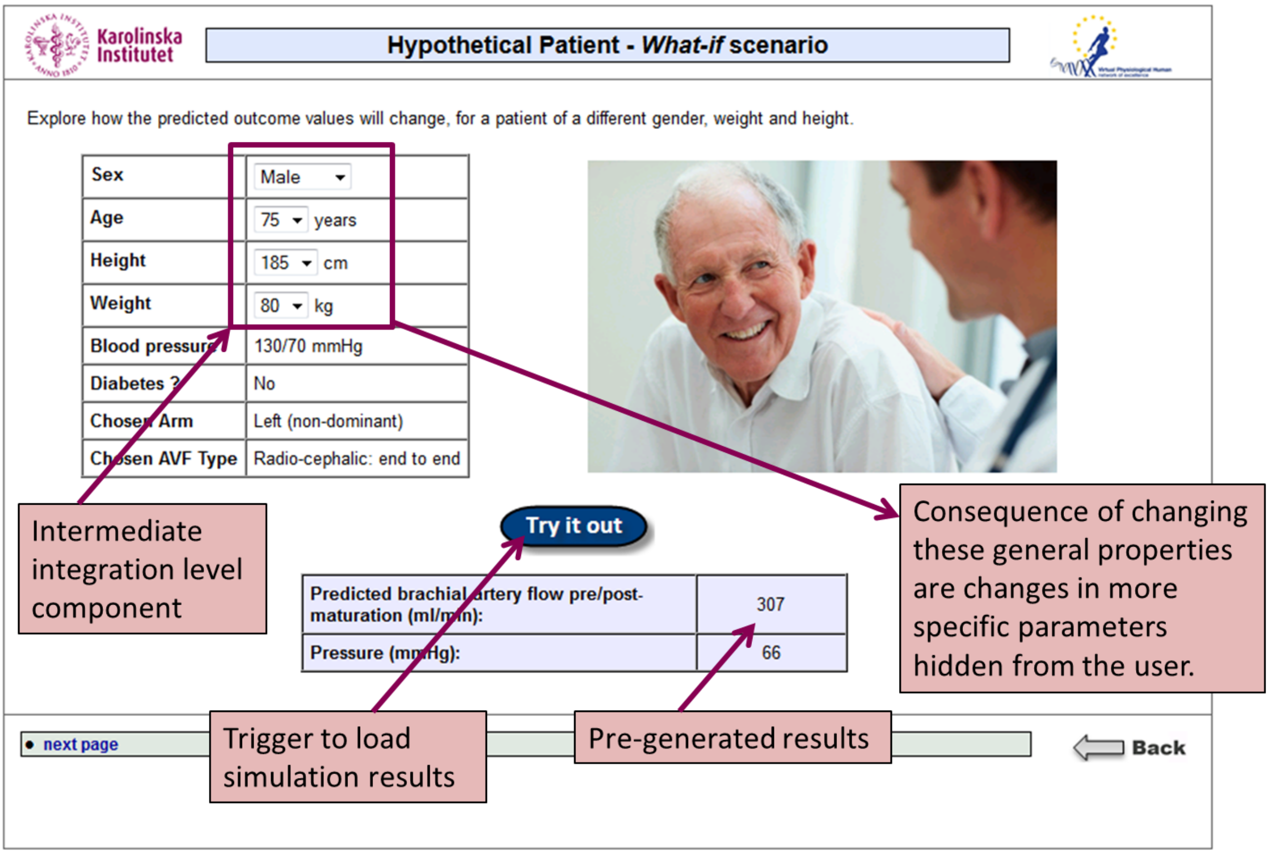
Relevant group →	A				B		C	
Feasibility feature → ↓ Integration level	f_{A1}	f_{A2}	f_{A3}	f_{A4}	f_{B1}	f_{B2}	f_{C1}	f_{C2}
Basic	+	+	+	+	+/-	+	+	+
Intermediate	+	+	+	+	+/-	?	-	-
Advanced	-	?	?	+	+/-	?	-	-

Implementation of Two Levels of Integration

Following the positive evaluation of the integration potential of the VPH ARCH project at the basic and intermediate levels, two variants of the same virtual patient were implemented reflecting these levels. The first case (Study Group 1, Basic Integration Level) presented static simulation results through manual pasting into the case (s_0 integration strategy), whereas the second case (Study Group 2, Intermediate Integration Level) provided the user with the option to manipulate different parameters of the simulation. The narrative integration (s_1) was implemented by adding a what-if-node at the end of the case (Figure 5).

Users of the second group could not circumvent the opportunity to use this feature as it was placed in the central navigation path. There was no influence of the parameter manipulation on the virtual patient branching (s_2). The input parameters of gender, age, height, and weight are interdependent and represent a set of input parameters with typical distributions of vascular anatomy for a male and female of specific age and Body Mass Index (s_3). The simulation data (16 result sets) were generated prior to the experiment on a standard PC workstation (Intel Core 2; 1.66GHz; 4GB RAM), at run-time loaded from a pre-generated look-up table located in the virtual patient player (deployed at a WWW server) and available locally for manipulation in the Web browser through a JavaScript method (s_4).

Figure 5. Second level computational model integration ("what-if-node") for the VPH ARCH project virtual patient.



Explore how the predicted outcome values will change, for a patient of a different gender, weight and height.

Sex	Male
Age	75 years
Height	185 cm
Weight	80 kg
Blood pressure	130/70 mmHg
Diabetes ?	No
Chosen Arm	Left (non-dominant)
Chosen AVF Type	Radio-cephalic: end to end

Predicted brachial artery flow pre/post-maturation (ml/min):	307
Pressure (mmHg):	66

Intermediate integration level component

Consequence of changing these general properties are changes in more specific parameters hidden from the user.

Trigger to load simulation results

Pre-generated results

Perception of the Framework Output by Health Care Workers

Table 5 summarizes the responses of the participants at the EVC conference event. The virtual patient was interesting or very interesting for 84% of participants (32/38), 11% (4/38) were neutral, and 5% (2/38) found the virtual patient “not that interesting” or “not at all interesting”. The average evaluation score (Likert scale, 5=very interesting, 1=not at all interesting) was better for the intermediate integration level (4.22) than the basic level (4.10) in this sample, but the difference was not statistically significant (*P*=.84). None of the evaluators disagreed or strongly disagreed with the presented content, 76% (29/38) agreed or strongly agreed with it, and 24% (9/38) neither agreed nor disagreed. The average evaluation score (Likert scale, 5=strongly agree, 1=strongly disagree) was better in this case for the basic integration level (4.05) than the intermediate level (3.83). Again this difference was not statistically significant (*P*=.32).

In the questions directed to Study Group 2 only, “The possibility of manipulating simulation parameters (gender, age, weight, height) for the hypothetical patient was …”, 28% (5/18) responded with “very interesting”, 39% (7/18) “interesting”, 33% (6/18) “neutral”, and no one answered “not that”, “not at all interesting”. It should be stressed that no participants responded “I have not noticed this possibility” to this question, which was provided as an alternative option. The participants in Study Group 2 were also largely satisfied with the results of manipulating the simulation parameters. To the question “Did the results of manipulating the simulation parameters meet your expectations?”, 11% (2/18) answered “strongly agree”, 72% (13/18) “agree”, and 17% (3/18) “neither agree nor disagree”. No one disagreed with this statement or had not used this function.

Table 5. Summary of answers in the evaluation questionnaire.

Question	Likert scale	Group 1 n=20	Group 2 n=18	*P*
Was the presented virtual patient case interesting for you?	5=very interesting; 1=not at all interesting	4.10	4.22	.84
Do you agree with the presented content of the virtual patient case?	5=strongly agree; 1=strongly disagree	4.05	3.83	.32

Discussion

Summary

Integration of simulation is gaining importance in health care education informatics. After a period where isolated systems were developed, the advantages of combining systems into “continua of simulation” have become apparent [15]. This connection not only augments the value of learning activities by facilitating wider opportunities to synthesize knowledge and skills, but it also fosters reuse of existing models, thereby increasing the return on investment in their development.

We may regard the proposed framework as a specialization of a section of Ellaway’s et al “Practica continua” framework [15]. Some integration strategies we have proposed are direct responses to the desiderata of this framework. For instance, our narrative integration strategy (s_1) has a clear correspondence to the “narrative integration” dimension and requirements (eg, timeline and causality continua) in Ellaway et al. The added value is that we have exemplified their implementation in the context of virtual patients and computational models of physiology by concrete implementation proposals in the form of the “what-if-node” and “multi-case-package” strategies. On the other hand, our framework is not just a specialization of the “Practica continua” framework, since it discusses the integration of computational models that are products of research projects initially outside the educational sector. This consideration is given in our framework by inclusion of tools for systematic evaluation of the feasibility of integration. The framework described here is currently restricted to the integration activity phases, that is, the authoring process. This leaves the run-time and analysis phases for further iterations of framework development. The HSVO project, related to the “Practica continua” framework, demonstrated how the run-time phase may be implemented by a common execution interface and middleware layer [46]. These outcomes will be useful in the further development of our framework.

It is important to stress that current virtual patient systems and the MVP standard are not yet, in general, ready for the higher levels of integration described in this paper. However, the effort involved in extending virtual patient systems to facilitate the intermediate level of model integration does not seem to be obstructive. A graphical user interface component to handle the what-if-node manipulation of model parameters and display the simulation results is of primary importance along with a standard mechanism to manage pre-generated data in virtual patient packages. The advanced level of integration poses a more significant challenge. Many models available today have been implemented using scientific tools such as MatLab (Mathworks) or specific numerical libraries. This limits their portability in virtual patient packages. A viable strategy might be to host the solver on a remote server, but this conflicts with the self-containment rule of content of a virtual patient package, raises security concerns, and requires the maintenance of an additional service. These considerations will fuel further research as suggested by this paper (SSM Step 4).

The proposal of this framework was helpful for the authors in developing, discussing, and reporting the integration of the archTk simulation within the virtual patient. It is challenging to demonstrate, in the short term, the long-term benefits of the framework for the users of virtual patients. The evaluation study carried out at the EVC conference provided an unambiguously positive response to the approach of integrating advanced computational models of physiological processes within virtual patients. Based on the obtained feedback sample, we hypothesize an altered perception of integration depending on changes in the framework’s variables. The queried group of experts found higher integration more interesting while, at the same time, being more concerned with the validity of the results. It has to be stressed, however, that this conclusion cannot be generalized as yet because the difference is not statistically significant and the influences of other factors (eg, age or experience) could not be excluded.

Limitations

The application of the integration framework has been demonstrated by a case study involving a single computational model. Improvements for the next SSM learning cycle should aim to extend the proposed lists of feasibility features, integration strategies and scope of exemplar projects. This cycle may be repeated several times depending on the research outcomes, prioritizing different simulation aspects.

Some proposed integration modes are currently theoretical constructs that require testing on concrete examples. In particular, this applies to the advanced level of integration that was not attempted in this case because the computational load was unsuitable for target application. Future SSM iterations should focus on concrete data representations and software solutions enabling further development of the integration strategies and evaluation process.

The educational impact of the proposed integration levels is still uncertain and needs to be tested in rigorously conducted comparative studies. The integration of the archTk simulation into a virtual patient was tested with clinical experts and health care education specialists, but not on a wider scale with different virtual patient systems. The completeness and priority of the proposed framework elements could be addressed by a Delphi study to formally collect feedback from computational modeling researchers and virtual patient system developers.

Conclusions

The paper has outlined a conceptual framework for the integration of computational models into virtual patients. This includes consideration of feasibility features, levels of integration, integration strategies with various modes, and evaluation profiles. The opportunities and challenges of model exploitation have been discussed in the context of a virtual patient developed from the VPH ARCH project, incorporating archTk simulation results at different integration levels. The empirical evaluation of two variants of the virtual patient provided positive feedback on the value of this type of integration. The responses suggest further investigation of increased user satisfaction, but decreased trust, at higher levels of simulation integration. The long-term research aim is to isolate the most crucial factors in the integration framework and their influence on the integration outcome.

Acknowledgments

The authors would like to thank the ARCH consortium. The authors wish to acknowledge their gratitude to Dr Inga Hege for her comments on the paper and to Anneliese Lilienthal for graphical support. This work is partially funded by the European Community’s 7th Framework Programme (VPH NoE, Grant Agreement No 223920).

Conflicts of Interest

None declared.

References

1. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]
2. Ellaway RH, Poulton T, Smothers V, Greene P. Virtual patients come of age. Med Teach 2009 Aug;31(8):683-684. [Medline: 19811203]
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. Maloney S, Haas R, Keating JL, Molloy E, Jolly B, Sims J, et al. Breakeven, cost benefit, cost effectiveness, and willingness to pay for web-based versus face-to-face education delivery for health professionals. J Med Internet Res 2012;14(2):e47 [FREE Full text] [doi: 10.2196/jmir.2040] [Medline: 22469659]
5. Kononowicz AA, Hege I. Virtual patients as a practical realisation of the e-learning idea in medicine. In: Soomro S, editor. E-learning, experience and future. Vukovar, Croatia: In-Teh; Apr 01, 2010:345-370.
6. Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. Med Teach 2008;30(2):170-174. [doi: 10.1080/01421590701874074] [Medline: 18464142]
7. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
8. Dev P, Heinrichs WL, Youngblood P. CliniSpace: a multiperson 3D online immersive training environment accessible through a browser. Stud Health Technol Inform 2011;163:173-179. [Medline: 21335784]
9. 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]
10. Smothers V, Azan B, Ellaway R. ANSI/MEDBIQ VP.10.1 – 2010. Baltimore, MD: MedBiquitous Consortium; 2010 Apr 06. MedBiquitous Virtual Patient Specifications and Description Documen URL: http://www.medbiq.org/working_groups/virtual_patient/VirtualPatientDataSpecification.pdf [accessed 2013-02-25] [WebCite Cache ID 6EgCCgsho]
11. Zary N, Hege I, Heid J, Woodham L, Donkers J, Kononowicz AA. Enabling interoperability, accessibility and reusability of virtual patients across Europe - design and implementation. Stud Health Technol Inform 2009;150:826-830. [Medline: 19745428]
12. Garrido J. Introduction to Elementary Computational Modeling: Essential Concepts, Principles, and Problem Solving. Boca Raton, FL: CRC Press; 2011.
13. van Meurs W. Modeling and Simulation in Biomedical Engineering: Applications in Cardiorespiratory Physiology. New York: McGraw-Hill Professional; 2011.
14. Schwarz D, Štourač P, Komenda M, Harazim H, Kosinová M, Gregor J, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res 2013;15(7):e135 [FREE Full text] [doi: 10.2196/jmir.2590] [Medline: 23835586]
15. Ellaway RH, Kneebone R, Lachapelle K, Topps D. Practica continua: Connecting and combining simulation modalities for integrated teaching, learning and assessment. Med Teach 2009 Aug;31(8):725-731. [Medline: 19811209]
16. de Bono B, Hoehndorf R, Wimalaratne S, Gkoutos G, Grenon P. The RICORDO approach to semantic interoperability for biomedical data and models: strategy, standards and solutions. BMC Res Notes 2011;4:313 [FREE Full text] [doi: 10.1186/1756-0500-4-313] [Medline: 21878109]

17. Halic T, Venkata SA, Sankaranarayanan G, Lu Z, Ahn W, De S. A software framework for multimodal interactive simulations (SoFMIS). Stud Health Technol Inform 2011;163:213-217. [Medline: 21335791]
18. Coveney PV, Diaz V, Hunter P, Kohl P, Viceconti M. The Virtual Physiological Human. Interface Focus 2011 Mar 31;1(3):281-285. [doi: 10.1098/rsfs.2011.0020]
19. Bamidis PD, Kaldoudi E, Pattichis C. mEducator: a best practice network for repurposing and sharing medical educational multi-type content. In: Leveraging Knowledge for Innovation in Collaborative Networks: 10th IFIP WG 5.5 Working Conference on Virtual Enterprises. Berlin Heidelberg: Springer; 2009 Presented at: PRO-VE 2009; Oct 7-9, 2009; Thessaloniki, Greece p. 769-776.
20. Halic T, Ahn W, De S. A framework for web browser-based medical simulation using WebGL. Stud Health Technol Inform 2012;173:149-155. [Medline: 22356977]
21. Checkland P, Poulter J. Soft Systems Methodology. In: Reynolds M, Holwell S, editors. Systems Approaches to Managing Change: A Practical Guide. London: Springer; 2010:191-242.
22. Ruotsalainen PS, Blobel BG, Seppälä AV, Sorvari HO, Nykänen PA. A conceptual framework and principles for trusted pervasive health. J Med Internet Res 2012;14(2):e52 [FREE Full text] [doi: 10.2196/jmir.1972] [Medline: 22481297]
23. Viceconti M, Clapworthy G. VPH-FET Research Roadmap - Advanced Technologies for the Future of the Virtual Physiological Human.: VPH-FET consortium; 2011 Sep. URL: http://www.vph-institute.org/upload/vph-fet-final-roadmap-1_519244713c477.pdf [accessed 2013-02-24] [WebCite Cache ID 6Eg6MCAl5]
24. Hunter P, Coveney PV, de Bono B, Diaz V, Fenner J, Frangi AF, et al. A vision and strategy for the virtual physiological human in 2010 and beyond. Philos Trans A Math Phys Eng Sci 2010 Jun 13;368(1920):2595-2614 [FREE Full text] [doi: 10.1098/rsta.2010.0048] [Medline: 20439264]
25. Virtual Physiological – Human Network of Excellence. URL: http://www.vph-noe.eu/ [accessed 2013-02-24] [WebCite Cache ID 6Eg6ifKEp]
26. VPH ARCH Project. 2010. URL: http://www.vph-arch.eu/ [accessed 2013-02-24] [WebCite Cache ID 6EgDz6xLN]
27. Bode AS, Huberts W, Bosboom EM, Kroon W, van der Linden WP, Planken RN, et al. Patient-specific computational modeling of upper extremity arteriovenous fistula creation: its feasibility to support clinical decision-making. PLoS One 2012;7(4):e34491 [FREE Full text] [doi: 10.1371/journal.pone.0034491] [Medline: 22496816]
28. Huberts W, Bode AS, Kroon W, Planken RN, Tordoir JH, van de Vosse FN, et al. A pulse wave propagation model to support decision-making in vascular access planning in the clinic. Med Eng Phys 2012 Mar;34(2):233-248. [doi: 10.1016/j.medengphy.2011.07.015] [Medline: 21840239]
29. ArchTk toolkit. 2012. ARCH Consortium URL: http://archtk.github.com/ [accessed 2013-02-24] [WebCite Cache ID 6EgEXzo70]
30. AVF.SIM - Patient-specific computer modelling for computer-based planning of vascular access surgery in haemodialysis patients. 2013. URL: http://avfsim.com/ [accessed 2013-09-26] [WebCite Cache ID 6Jv7GD0Kb]
31. Kononowicz AA, Holler T. Bit pathways–a tool for teaching and learning clinical pathways. Bio-Algorithms and Med-Systems 2008;4(8):33-40.
32. Kononowicz AA, Hege I, Krawczyk P, Zary N. New approaches to linking clinical guidelines to virtual patients. Stud Health Technol Inform 2012;180:958-962. [Medline: 22874335]
33. Kolb DA. Experiential learning: experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall; 1984.
34. Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, 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]
35. Kalet AL, Song HS, Sarpel U, Schwartz R, Brenner J, Ark TK, et al. Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module. Med Teach 2012;34(10):833-839 [FREE Full text] [doi: 10.3109/0142159X.2012.706727] [Medline: 22917265]
36. Hege I, Kopp V, Adler M, Radon K, Mäsch G, Lyon H, et al. Experiences with different integration strategies of case-based e-learning. Med Teach 2007;29(8):791-797. [doi: 10.1080/01421590701589193] [Medline: 18236274]
37. Kim S, Phillips WR, Pinsky L, Brock D, Phillips K, Keary J. A conceptual framework for developing teaching cases: a review and synthesis of the literature across disciplines. Med Educ 2006 Sep;40(9):867-876. [doi: 10.1111/j.1365-2929.2006.02544.x] [Medline: 16925637]
38. Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol 2005 Dec;23(12):1509-1515. [doi: 10.1038/nbt1156] [Medline: 16333295]
39. Waltemath D, Adams R, Beard DA, Bergmann FT, Bhalla US, Britten R, et al. Minimum Information About a Simulation Experiment (MIASE). PLoS Comput Biol 2011 Apr;7(4):e1001122 [FREE Full text] [doi: 10.1371/journal.pcbi.1001122] [Medline: 21552546]
40. Posel N, Fleiszer D, Shore BM. 12 Tips: Guidelines for authoring virtual patient cases. Med Teach 2009 Aug;31(8):701-708. [doi: 10.1080/01421590902793867] [Medline: 19513927]

41. Kononowicz AA, Heid J, Donkers J, Hege I, Woodham L, Zary N. Development and validation of strategies to test for interoperability of virtual patients. Stud Health Technol Inform 2009;150:185-189. [Medline: 19745294]
42. Tworek J, Paget M, McLauglin K, Wright B. How Dungeons & Dragons Made Us Better VPs: Randomizing Physiological Data to Rapidly Produce 97 Clinically Realistic VPs. Bio-Algorithms and Med-Systems 2010;6(11):41-45.
43. Kononowicz AA, Zary N, Davies D, Heid J, Woodham L, Hege I. Push and pull models to manage patient consent and licensing of multimedia resources in digital repositories for case-based reasoning. Stud Health Technol Inform 2011;169:203-207. [Medline: 21893742]
44. Bode AS, Huberts W, Bosboom EM, Kroon W, van der Linden WP, Planken RN, et al. Patient-specific computational modeling of upper extremity arteriovenous fistula creation: its feasibility to support clinical decision-making. PLoS One 2012;7(4):e34491 [FREE Full text] [doi: 10.1371/journal.pone.0034491] [Medline: 22496816]
45. Passera K, Manini S, Antiga L, Remuzzi A. Patient-specific model of arterial circulation for surgical planning of vascular access. J Vasc Access 2013;14(2):180-192. [doi: 10.5301/jva.5000099] [Medline: 23032951]
46. Ellaway R, Topps D, MacDonald J, Copeland B, Olmos A, Spencer B. HSVO: A Functional XML Specification for Integrating Simulation Devices. Bio-Algorithms and Med-Systems 2010;6(11):53-57.

Abbreviations

AVF: arteriovenous fistula
ESRD: end-stage renal disease
EVC: European Vascular Course
MVP: MedBiquitous Virtual Patient Standard
NoE: Network of Excellence
SSM: Soft System Methodology
VP: virtual patient
VPH: Virtual Physiological Human

Edited by P Bamidis, D Giordano; submitted 01.03.13; peer-reviewed by B Lok, L Woodham, T Halic; comments to author 12.05.13; revised version received 28.05.13; accepted 27.12.13; published 23.01.14.

Please cite as:
Kononowicz AA, Narracott AJ, Manini S, Bayley MJ, Lawford PV, McCormack K, Zary N
A Framework for Different Levels of Integration of Computational Models Into Web-Based Virtual Patients
J Med Internet Res 2014;16(1):e23
URL: http://www.jmir.org/2014/1/e23/
doi:10.2196/jmir.2593
PMID:24463466

Original Paper

Online Activity and Participation in Treatment Affects the Perceived Efficacy of Social Health Networks Among Patients With Chronic Illness

Racheli Magnezi[1,2], MBA, PhD; Yoav S Bergman[1,3], PhD; Dafna Grosberg[2], RN, MHA

[1]Department of Public Health and Health Systems Management Program, Bar Ilan University, Ramat Gan, Israel
[2]Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer, Tel Aviv, Israel
[3]Interdisciplinary Department for Social Sciences, Bar Ilan University, Ramat Gan, Israel

Corresponding Author:
Racheli Magnezi, MBA, PhD
Department of Public Health and Health Systems Management Program
Bar Ilan University
Department of Management
Ramat Gan, 52900
Israel
Phone: 972 35317128
Fax: 972 37384041
Email: magnezir@biu.ac.il

Abstract

Background: The use of online health-related social networks for support, peer-to-peer connections, and obtaining health information has increased dramatically. Participation in an online health-related social network can enhance patients' self-efficacy and empowerment, as they are given knowledge and tools to manage their chronic health condition more effectively. Thus, we can deduce that patient activation, the extent to which individuals are able to manage their own health care, also increases. However, little is known about the effects of participation in online health-related social networks and patient activation on the perceived usefulness of a website across disease groups.

Objective: The intent of the study was to evaluate the effects and benefits of participation in an online health-related social network and to determine which variables predict perceived site usefulness, while examining patient activation.

Methods: Data were collected from "Camoni", the first health-related social network in the Hebrew language. It offers medical advice, including blogs, forums, support groups, internal mail, chats, and an opportunity to consult with experts. This study focused on the site's five largest and most active communities: diabetes, heart disease, kidney disease, spinal injury, and depression/anxiety. Recruitment was conducted during a three-month period in which a link to the study questionnaire was displayed on the Camoni home page. Three questionnaires were used: a 13-item measure of perceived usefulness (Cronbach alpha=.93) to estimate the extent to which an individual found the website helpful and informative, a 9-item measure of active involvement in the website (Cronbach alpha=.84), and The Patient Activation Measure (PAM-13, Cronbach alpha=.86), which assesses a patient's level of active participation in his or her health care.

Results: There were 296 participants. Men 30-39 years of age scored higher in active involvement than those 40-49 years (P=.03), 50-64 years (P=.004), or 65+ years (P=.01). Respondents 20-29 years of age scored higher in perceived usefulness than those 50-64 years (P=.04) and those 65+ years (P=.049). Those aged 20-29 years scored significantly lower on the PAM-13 scale than those aged 30-39 years (P=.01) and 50-64 years (P=.049). Men and women had similar PAM-13 scores ($F_{9,283}$=0.17, P=.76). Several variables were significant predictors of perceived usefulness. Age was a negative predictor; younger age was indicative of higher perceived usefulness. Active involvement was a positive predictor. There was a negative relationship found between PAM-13 scores and perceived usefulness, as taking a less active role in one's own medical care predicted higher perceived website usefulness. A trend toward higher frequency of website activity was associated with increased perception of usefulness.

Conclusions: Online health-related social networks can be particularly helpful to individuals with lower patient activation. Our findings add information regarding the social and medical importance of such websites, which are gradually becoming an inseparable part of day-to-day chronic disease management in the community.

(J Med Internet Res 2014;16(1):e12) doi:10.2196/jmir.2630

KEYWORDS

Internet; social health network; Patient Activation Measure (PAM); online health network

Introduction

Overview

The use of search engines and websites to seek medical information is continually growing. As the human life span increases due to novel medical interventions and technologies, more and more individuals are developing chronic illnesses. In this study, we distinguish between general forums where one can ask questions about health conditions, general social networks where the user is identified and "converses" with friends (eg, Facebook), and specific forums related to a unique interest group such as a social health network. The essential difference is that a social network includes many different features (such as blogs and live chats), whereas a forum uses only one of the tools found in an interactive health site. Although information about the usefulness of medical social networks for certain conditions exists, very little attention has been paid to how patient involvement in an online health-related social network is related to how they manage their health-related conditions. Due to the increase in social health networks, this study evaluated patient activation in this context.

Patient activation describes how much a patient is involved in his or her health care. Online health-related social networks provide patients with information about their disease and the ability to learn more about it by interacting with others with a similar problem. We looked at the concepts of patient activation and online health site use in an effort to determine what factors influence how useful patients perceive the site to be. This study aimed to explore the combined effects of patient activation and involvement in an online health-related social network on the perceived usefulness of such websites. By increasing our knowledge regarding the factors that underlie how patients perceive and gain information from the Internet, we can continue to reshape and develop online medical and health information in ways that are compatible with the needs of the target users.

Background

The use of online support groups has increased dramatically in recent years. In 1998, over 50 million American adults turned to the Internet to obtain health information. By 2005, that number had increased to 117 million and in the course of a single year—from 2009 to 2010—this number rose from 154 million to 175 million [1]. Statistics from 2009 indicate that an estimated 58% of all Internet users consulted the Web for health purposes, 61% of American adults looked online for health information, and roughly one-third accessed social media related to health [2].

Individuals draw on many different sources of information for health-related decisions. The Internet adds to available resources by offering specialized blogs and by allowing individuals to seek advice from question-and-answer type health forums hosted by experts and from interactive social health networks. Participation in a social health network can help people feel connected to others by giving them the opportunity to offer and obtain information about their diseases, although individuals are more likely to consume information than they are to contribute to the dialogue [3].

Medical online support groups are designed to improve individuals' understanding of their health conditions, change their health behaviors, and enhance their ability to manage a chronic health condition. They can also enhance self-confidence by providing increased emotional support and by enabling members to better manage their diseases [4-10]. Online health-related social networks have changed the patient-physician relationship and have benefited both parties: patients receive external support and information, whereas clinicians gain increased accessibility to new ideas and alternative therapies and approaches from their patients [11-14].

Patients who are more likely to search online for health care information include women [15,16], those with a higher education, a chronic health condition, more years of Internet experience, and those with broadband access [17,18]. Associations between factors such as income and age with online health-related information seeking are less consistent [19-21].

Patient Activation

The underlying theory behind the Patient Activation Measure is that "*Activation* refers to having the capability and the willingness to take on the role of managing one's own health and health care" [22]. "Patient activation" describes the extent to which individuals are able to manage their own health care. Hibbard and her colleagues [23,24] conceptualized patient activation as encompassing a range of elements important to self-management that extend beyond any single health behavior. As a measure of this concept, they developed the Patient Activation Measure (PAM), a broad construct that encompasses self-efficacy, behavior, and knowledge. It can predict a variety of behaviors including healthy behavior, preventive care measures, disease-specific self-care behavior, and information seeking [22]. The PAM is a 13-item psychometric tool with a 4-level model of health-related behaviors that measures the latent construct of patient activation. It captures the degree to which patients have the beliefs, knowledge, and skills to "manage their condition(s), collaborate with their providers, maintain their health, and access appropriate and high-quality care" [24,25].

Studies have found that supportive social environments lead to increased activation levels [26]. Higher activation scores have also been linked with sociodemographic factors of female gender, younger age, and higher education or income levels.

Several studies have explored the relationship between patient activation and individuals with illnesses such as HIV [27], asthma [28], diabetes [29], inflammatory bowel disease [30], spine surgery [31], multiple sclerosis [32], and mental illnesses [33]. Research has demonstrated that increasing patients' self-management with regard to their illness increased their PAM scores, which in turn was connected with more favorable health outcomes. Additionally, higher PAM scores were associated with greater degrees of optimism, hope, and control [31].

Activated patients strive to understand their health conditions, viewing problems as challenges, and displaying confidence with regard to their positive resolution [25,34]. Acquiring the attributes of self-management is a gradual process that involves the attainment of knowledge and problem-solving skills that enable individuals to confidently engage in decision-making and actions aimed at managing their chronic health condition more effectively [35-38]. Activated patients are more likely to adhere to behaviors that help control their symptoms and the progression of their disease [32]. Higher activation has been associated with less use of health care services [39].

Only two previous studies examined the effect of a Web-based intervention, similar to our site, on patient activation among individuals with a chronic health condition. One, which included patients with diabetes, asthma, or hypertension, found that the Web-based intervention had a positive effect on the patient activation levels of participants in the intervention group [40]. Likewise, a study of patients with type 2 diabetes found that patient activation improved after participation in the online diabetes self-management program [41]. However, we were unable to find a previous study that specifically examined patient activation measures in reference to use of an online health-related social network. Most existing information pertains to simple online medical question-and-answer sites, which provide a more limited scope than interactive websites that offer a range of services including blogs, patient information exchanges, and contact with physicians.

Thus, the purpose of this study was to further evaluate the effects and benefits of participating in an online health-related social network on patient activation and to determine which variables predict perceived usefulness of the site. We hypothesized that patients with higher activation scores would tend to follow the information provided by an interactive website and consequently perceive it as more useful. As patients who use an interactive website frequently become adept at obtaining relevant information quickly and efficiently, we predicted that high levels of website involvement would also predict perceived website usefulness.

Methods

The Platform

Camoni is the first Hebrew-language, nonprofit, online, social network that is targeted to individuals with chronic conditions and assists them in finding others facing similar health issues [42]. The Hebrew word “camoni” means “like me”. The Camoni site is comprised of 16 communities, defined according to the health conditions of diabetes, chronic pain, heart disease, hypertension, obesity, eating disorders, multiple sclerosis, spinal injury, lung disease, kidney disease, stroke, osteoporosis, Crohn’s disease, cancer, obesity, and depression. Figure 1 shows a page from the Camoni website. Each community is headed by a medical expert. Camoni offers advice, the opportunity to consult with experts, and the chance to converse with other patients who face the same health condition. The site includes blogs, forums, support groups, internal mail, and chats. It also explains each health condition, its diagnosis, and offers practical advice on how to maintain one’s health and cope with the disease. Registration is required for active participation on the site, which is open to all. Camoni enables people to share, to learn and gain encouragement from each other, and to provide advice based on their own experiences. Since its launch in August 2009, Camoni has grown exponentially and today has thousands of registered users. This study focused on the five largest communities: diabetes, heart disease, kidney disease, spinal injury, and depression/anxiety.

Figure 1. Screenshot of the Camoni website.

שיקום מהבית בליווי מומחים קליניים

זה לא בריא להתמודד לבד

בואו להיות חלק מקהילה תומכת וגדולה של
חברים המתמודדים עם מצבים רפואיים כרוניים,
בדיוק כמוכם, ולהתייעץ עם עשרות מומחים
הנמצאים כאן בשבילכם

למה כדאי להרשם?

- ✓ מאגר ידע עצום של עשרות אלפי חברים כמוך
- ✓ מומחים מקצועיים מובילים
- ✓ תוכן בריאות עדכני איכותי ומהימן
- ✓ אתר מבית מכון גרטנר, ללא מטרות רווח

הרשם חינם

נדרש רק שם/כינוי ומייל (הרישום אנונימי)

עדכונים אחרונים

עישון בהריון עשוי להעלות סיכון ל-ADHD בילד
מחקר בדק איזה גורמים סביבתיים עשויים להשפיע על הסיכון להפרעות קשב וריכוז ...המשך

חברים מספרים

varduku:
השיתוף הזה הינו על דוגמא לאנשים היקרים והטובים האלו שאני זוכה להכיר כאן באתר, ויש כאן הרבה כאלו אשר התחברו אל חיי בד...

אביגיל:
הדבר הכי נעים שקרה לי מפה זה שיצא שדברים שכתבתי לגולשת אחרת ממש עזרו לה. התחושה

Recruitment

This cross-sectional study is based on a single sampling time. Recruitment was conducted during a 4-month period that began in February 2012 and continued through May 2012. The questionnaire was accessible through Google Docs (see Multimedia Appendix 1 for an English translation). An invitation to participate and a link were available on the Camoni home page of each of the chosen disease communities (diabetes, heart disease, spinal injury, kidney disease, and depression/anxiety). During the three-month recruitment period, reminders were placed in the monthly newsletter sent to all Camoni participants who had not declined the option. An individual could only answer once.

Instruments

Demographic and Health Characteristics

Respondents (n=296) were asked to provide information about basic demographic variables such as age group, gender, income, and diagnosis, as well as information regarding their use of the website. Descriptive information regarding the cohort is shown in Table 1.

Table 1. Demographics and health statistics of the cohort[a] (n=296).

Variable		n	%
Gender			
	Male	135	45.6
	Female	161	54.4
Age range, years			
	20-29	23	7.8
	30-39	20	6.8
	40-49	46	15.5
	50-64	140	47.3
	65+	64	21.6
Income			
	Below average	143	48.3
	Average	59	19.9
	Above average	64	21.6
Illness			
	Diabetes	115	38.9
	Heart	31	10.5
	Kidney	25	8.4
	Spine	59	19.9
	Depression/Anxiety	53	17.9
Number of illnesses			
	1	95	32.1
	2	143	48.3
	3+	58	19.6
Duration of illness			
	Less than 6 months	8	2.7
	6-12 months	13	4.4
	1-2 years	24	8.1
	2-4 years	50	16.9
	Over 4 years	193	65.2
Duration of activity on website			
	Less than 6 months	104	35.1
	6-12 months	73	24.7
	1-2 years	85	28.7
	2-3 years	32	10.8
Frequency of activity on website			
	Every day	30	10.1
	1-2 times a week	89	30.1
	1-2 times every 2 weeks	55	18.6
	1-2 times a month	41	13.9
	Once a month	66	22.3

[a] Does not include missing values.

Perceived Usefulness of Online Groups

Perceived usefulness of the website was measured by 13 items on a scale ranging from 1 (“very little”) to 5 (“very much”). Possible scores ranged from 0 to 65. It is a measure of the extent to which participants found the website helpful and informative (eg, “Conversations/chats with other surfers on Camoni helped me”; “Because of Camoni, I changed the type of treatment”; “On Camoni, I found new information about a treatment type”; “I use the information I obtained when I visit my doctor”). Cronbach alpha was high (.93), and a principal component factor analysis with varimax rotation yielded one main factor (Eigenvalue =7.02), which accounted for 54% of the variance. Accordingly, a mean perceived usefulness score was calculated for each respondent.

Active Involvement in Online Groups

Active involvement was measured by a 9-item scale ranging from 1 (“very little”) to 5 (“very much”), including items measuring the extent to which respondents participated in the website components (eg, reading/writing articles, reading/writing blogs, chats). Possible scores ranged from 0 to 45. Here too, Cronbach alpha was high (.84) and a principal component factor analysis with varimax rotation yielded one main factor (Eigenvalue =4.15), which accounted for 52% of the variance. Accordingly, a mean score of active involvement was calculated for each participant.

Patient Activation

We used the 13-item patient activation measure (PAM-13) to quantify patient activation. The PAM-13 is designed to elicit responses about an individual’s attitudes toward knowledge, skills, and confidence in self-managing health [23,25]. The scale is based on the Guttman technique with items ordered according to level of difficulty (eg, “When all is said and done, I am the person who is responsible for taking care of my health”). Patients are required to indicate their agreement with each item on a scale ranging from 1 (“Disagree strongly”) to 4 (“Agree strongly”), or to indicate that the item is not applicable to them. Possible scores ranged from 0 to 100. Cronbach alpha for this scale was .86. Accordingly, a mean patient activation score was calculated for each participant.

Data Analysis

Data were analyzed with the statistical software package SPSS 20. Age and gender differences were examined by analyses of variance (ANOVA), in which both were independent variables and active involvement, perceived usefulness, and patient activation were the dependent variables. The contribution of independent variables to perceived website usefulness was examined by a hierarchical regression. The first step included demographic variables (gender, age, income, duration and frequency of activity in the website, duration of illness) and the second step included the number and type of illnesses an individual has (diabetes, heart disease, kidney disease, spinal injury, and anxiety/depression). The third step included active involvement in the website and the patient activation measure.

Results

Active Involvement

Main effects for age ($F_{9,260}$=2.16, P=.07, η^2=.03) and gender ($F_{9,260}$=1.51, P=.22, η^2=.01) were not significant. However, a significant interaction of age × gender was found, ($F_{9,260}$=3.36, P=.01, η^2=.05). In order to discover the root of the interaction, we conducted two separate ANOVA for males and females, in which the dependent variable was active involvement and the independent variable was age. The ANOVA revealed a significant effect of age for male participants ($F_{4,120}$=4.02, P=.01, η^2=.12), but not for females ($F_{4,140}$=0.64, P=.637, η^2=.01). Scheffe’s post hoc tests demonstrated that men 30-39 years of age scored significantly higher in active involvement (mean 4.16, SD 0.99) than those 40-49 years (mean 2.62, SD 1.15; P=.03), 50-64 years (mean 2.46, SD 0.99; P=.004), and 65+ years (mean 2.58, SD 0.90; P=.01).

Perceived Usefulness

A significant main effect of age was found, ($F_{9,261}$=3.22, P=.01, η^2=.05). Pairwise comparisons demonstrated that individuals 20-29 years of age scored significantly higher in perceived usefulness (mean 2.26, SD 1.24) than those aged 50-64 years (mean 1.43, SD 1.18; P=.04) and 65+ years (mean 1.38, SD 1.00; P=.049). Gender was not significant ($F_{9,261}$=2.19, P=.13, η^2=.01).

Patient Activation Measure

A significant main effect of age was found ($F_{9,279}$= 4.41, P=.003, η^2=.06). Pairwise comparisons demonstrated that individuals aged 20-29 years scored significantly lower on the PAM-13 scale (mean 48.44, SD 21.25) than those aged 30-39 years (mean 62.28, SD 19.78; P=.01) and 50-64 years (mean 57.50, SD 17.66; P=.049). Gender was not significant ($F_{9,283}$=0.17, P=.76, η^2=.01).

Factors Contributing to Perceived Usefulness

Several variables were significant predictors for perceived usefulness (see Table 2). Age was found to be a negative predictor: younger age was indicative of higher perceived usefulness. Active involvement was a positive predictor, thereby demonstrating that such website involvement promotes perceived usefulness. More interesting, however, was the negative connection between PAM scores and perceived usefulness, as individuals who take a less active role in their care found that the website was more useful and provided them with information regarding their ailment. Additionally, a near-significant association (P=.06) between higher frequency of activity and an increase in perceived usefulness was found.

Table 2. Hierarchal multiple regression analysis predicting perceived website usefulness (n=296).

Predictor		ΔR^2	Step 1		Step 2		Step 3	
			B	*P*	*B*	*P*	*B*	*P*
Step 1		.121[b]						
	Gender		−.33	.04[a]	−.29	.08	−.26	.08
	Age		−.23	.001[b]	−.26	.001[b]	−.26	.001[b]
	Income		−.04	.53	−.05	.44	−.01	.78
	Duration of activity in website		.08	.27	.10	.18	.09	.17
	Frequency of activity in website		−.21	.001[b]	−.21	.001[b]	.10	.06
	Duration of illness		.05	.54	.02	.81	−.05	.48
Step 2		.036						
	Diabetes				−.20	.48	−.12	.65
	Heart				.61	.08	.61	.07
	Kidney				−.05	.85	−.07	.72
	Spine				−.13	.64	−.13	.61
	Depression/anxiety				−.31	.32	−.40	.17
	Number of illnesses				.02	.93	.02	.96
Step 3		.142[b]						
	Active involvement in site						.45	.001[b]
	Patient Activation Measure						−.01	.04[a]
Total R^2		.299						

[a] $P<.05$

[b] $P<.001$

Discussion

Principal Findings

To the best of our knowledge, this is the first study to examine how a patient's degree of active involvement in managing their health and their participation in an online social health forum affected the perceived usefulness of such a website.

Online social groups have the broadest reach and impact when the target population is younger [43,44]. Although chronic diseases are more prevalent among older individuals, fewer use the Internet as a source of information. However, chronic illness is a growing problem among younger individuals as well [45,46]. The increase in the number of younger, computer literate individuals with chronic illness is contributing to the growth of websites like Camoni.

We found that individuals 20 to 29 years of age reported higher perceived usefulness of the website and its contents compared to those over the age of 50. The relatively low perceived website usefulness and lesser involvement among older adults may be a consequence of difficulty in accessing relevant information online. We suggest that younger people experience greater perceived usefulness because they use the Internet more [16] and enjoy sharing personal aspects of their life on the Internet (for example, Facebook). They also belong to a generation that is used to immediate results (instant messaging, instant meals, instant gratification, etc.) and therefore prefer to obtain health information online rather than wait for a physician consultation. In the past, authoritative health information originated within the physician's office [47,48]. This is still true for many older individuals. We see a need for a longitudinal study to determine the effects of information gathering and social support from online communities and from health forums on patient-physician relationships.

Additionally, individuals 20 to 29 years of age scored significantly lower on the PAM-13 scale than those 30 to 39 and 50 to 64 years of age. We believe that these two age groups are more likely to seek health information. Individuals 30 to 39 years of age are often parents of young children and therefore seek health information to preserve their health and well-being as parents and as care providers. They also might seek health information regarding their children. The 50 to 64 age group is likely to demonstrate greater activation because they are beginning to encounter more chronic illness and are interested in preserving their health.

Unlike other studies [3], we found more men using the website than women. It is possible, since we found that men were more interested in using the site to obtain information and improve their health, they use Camoni to learn more about their condition rather than for peer support, as they tend to be less inclined to discuss personal issues and ask for advice than women are.

In this study, both younger age and active involvement in the site were indicative of higher perceived website value. Of interest was the negative relationship between PAM scores and perceived usefulness, as individuals who were less active in their care found that Camoni was more useful compared to those with higher PAM scores. We believe that this finding might have important ramifications regarding patients' ability to actively participate in their own health care management. Perhaps those with lower PAM scores have more to gain from the website, both in terms of medical information and the experiences of others. Therefore, we are planning a study that will measure changes in PAM among new participants in Camoni. Several authors also found that Web-based interventions benefited participants more at the earliest stages of patient activation. They also suggested that these patients have the most to gain from these websites [40,49].

We had anticipated that higher PAM levels would predict greater perceived usefulness, because, by definition, a high PAM level means that the individual is involved in his or her care and will therefore seek a variety of resources, one of which is a health-related social network. Unexpectedly, we found that low PAM scores predicted greater perceived website usefulness. We suspect that those who newly enter the site are at an early stage in their illness and are at a low PAM level. They are more open and willing to participate and receive information regarding their health status through the site. However, at this stage, they still are not ready to actively participate in their self-care and comply with medical advice. Their activation is only at the level of interacting with a peer group with similar problems, who "know what they are going through". We suggest that a future study that follows these individuals would find that PAM scores increase from the time of their initial participation.

Limitations of the Present Study

This study sampled Israelis and was conducted in the Hebrew language. Therefore, it should be expanded to include other nationalities. Moreover, as with many studies that concern Web activity, the sample used was self-selected; consequently, it might be difficult to infer causation from the results. Additionally, this study included five disease groups, but necessarily excluded others. We did not have information about other sources of knowledge or support that the participants might have accessed. For example, we do not know about additional medical consultations, family members who provided support, other forums or Internet sites where they received information, or avenues such as medical books or pamphlets.

Conclusions

The study results suggest that participation in an online health-related social network can be particularly helpful to individuals who have lower levels of patient activation. As these individuals might have less exposure to other sources of online medical information, they find specialized social health networks such as Camoni more useful. We can assume that such individuals are able to benefit from others seeking similar information. It may well be that once they observe how others weigh information from different sources and give support to each other, they are encouraged to do the same and subsequently take a more active role in managing their own health. Thus, we anticipate that this modeling behavior might lead to an increase in PAM measures, which in turn, will be related to better compliance with medical regimens. We are implementing a prospective study to test this theory.

Acknowledgments

We would like to thank the entire team at Camoni who made this work possible, especially Tami Shifter. In particular, we would like to thank Dr Lisa Carroll Bergman for her close reading of the manuscript and Faye Schreiber for editing the text. Additionally, we thank Tali Bendahan for assistance in data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1

English translation of the study questionnaire.

[PDF File (Adobe PDF File), 44KB - jmir_v16i1e12_app1.pdf]

References

1. Tayor, Humphrey. The Harris Poll #95. 2010 Aug 04. &#8220;Cyberchondriacs&#8221; on the Rise? URL: http://www.harrisinteractive.com/vault/HI-Harris-Poll-Cyberchondriacs-2010-08-04.pdf [accessed 2013-12-30] [WebCite Cache ID 6MFUR0zEI]
2. Fox S, Jones S. Washington D. C.: The Pew Internet & American Life Project; 2009 Jun 11. The Social Life of Health Information URL: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx [accessed 2013-12-30] [WebCite Cache ID 6MFV8m49D]
3. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]

4. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
5. Kerr C, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. J Med Internet Res 2006 Jul;8(3):e13 [FREE Full text] [doi: 10.2196/jmir.8.3.e13] [Medline: 16954123]
6. 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 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
7. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
8. Blank TO, Adams-Blodnieks M. The who and the what of usage of two cancer online communities. Computers in Human Behavior 2007 May;23(3):1249-1257. [doi: 10.1016/j.chb.2004.12.003]
9. Owen JE, Klapow JC, Roth DL, Tucker DC. Use of the Internet for information and support: disclosure among persons with breast and prostate cancer. J Behav Med 2004 Oct;27(5):491-505. [Medline: 15675637]
10. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and Internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: 10.1016/j.socscimed.2005.11.016] [Medline: 16361029]
11. 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]
12. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008 Oct;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
13. Weiner JP. Doctor-patient communication in the e-health era. Isr J Health Policy Res 2012;1(1):33 [FREE Full text] [doi: 10.1186/2045-4015-1-33] [Medline: 22929000]
14. 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]
15. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003 Jan;58(1):5-14. [Medline: 12674814]
16. Mackenzie CS, Gekoski WL, Knox VJ. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health 2006 Nov;10(6):574-582. [doi: 10.1080/13607860600641200] [Medline: 17050086]
17. Bansil P, Keenan NL, Zlot AI, Gilliland JC. Prev Chronic Dis, 3:A36. 2006. Health-related information on the Web: Results from the HealthStyles Survey, 2002-2003 URL: http://www.cdc.gov/pcd/issues/2006/apr/05_0155.htm [accessed 2013-03-22] [WebCite Cache ID 6FJE7PKvMhttpwwwcdcgovpcdissues2006apr050155htm]
18. Atkinson NL, Saperstein SL, Pleis J. Using the Internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
19. Dickinson A, Gregor P. Computer Use has no demonstrated impact on the well-being of older adults. International Journal of Human-Computer Studies 2006;64(8):744-753.
20. Huang C. Internet use and psychological well-being: a meta-analysis. Cyberpsychol Behav Soc Netw 2010 Jun;13(3):241-249. [Medline: 20557242]
21. Weaver JB, Mays D, Lindner G, Eroglu D, Fridinger F, Bernhardt JM. Profiling characteristics of Internet medical information users. J Am Med Inform Assoc 2009;16(5):714-722 [FREE Full text] [doi: 10.1197/jamia.M3150] [Medline: 19567794]
22. Hibbard JH, Greene J, Becker ER, Roblin D, Painter MW, Perez DJ, et al. Racial/ethnic disparities and consumer activation in health. Health Aff (Millwood) 2008;27(5):1442-1453 [FREE Full text] [doi: 10.1377/hlthaff.27.5.1442] [Medline: 18780935]
23. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
24. Hibbard JH. Moving toward a more patient-centered health care delivery system. Health Aff (Millwood) 2004;Suppl Variation:VAR133-VAR135 [FREE Full text] [doi: 10.1377/hlthaff.var.133] [Medline: 15471780]
25. Hibbard JH, Tusler M. Assessing activation stage and employing a "next steps" approach to supporting patient self-management. J Ambul Care Manage 2007;30(1):04479-20070. [Medline: 17170632]
26. Becker ER, Roblin DW. Translating primary care practice climate into patient activation: the role of patient trust in physician. Med Care 2008 Aug;46(8):795-805. [doi: 10.1097/MLR.0b013e31817919c0] [Medline: 18665059]
27. Marshall R, Beach MC, Saha S, Mori T, Loveless MO, Hibbard JH, et al. Patient activation and improved outcomes in HIV-infected patients. J Gen Intern Med 2013 May;28(5):668-674. [doi: 10.1007/s11606-012-2307-y] [Medline: 23288378]
28. Saft HL, Kilaru S, Moore E, Enriquez M, Gross R. The impact of a patient activation measure on asthma outcomes: A pilot study. Chest 2008 Oct 01;134(4_MeetingAbstracts):s2004. [doi: 10.1378/chest.134.4_MeetingAbstracts.s2004]

29. Mayberry R, Willock RJ, Boone L, Lopez P, Qin H, Nicewander D. A high level of patient activation is observed but unrelated to glycemic control among adults with type 2 diabetes. Diabetes Spectrum 2010 Aug 19;23(3):171-176. [doi: 10.2337/diaspect.23.3.171]
30. Munson GW, Wallston KA, Dittus RS, Speroff T, Roumie CL. Activation and perceived expectancies: correlations with health outcomes among veterans with inflammatory bowel disease. J Gen Intern Med 2009 Jul;24(7):809-815. [doi: 10.1007/s11606-009-1002-0]
31. Skolasky RL, Mackenzie EJ, Wegener ST, Riley LH. Patient activation and adherence to physical therapy in persons undergoing spine surgery. Spine (Phila Pa 1976) 2008 Oct 1;33(21):E784-E791. [doi: 10.1097/BRS.0b013e31818027f1] [Medline: 18827683]
32. Stepleman L, Rutter MC, Hibbard J, Johns L, Wright D, Hughes M. Validation of the patient activation measure in a multiple sclerosis clinic sample and implications for care. Disabil Rehabil 2010;32(19):1558-1567. [doi: 10.3109/09638280903567885] [Medline: 20590506]
33. Green CA, Nancy A. Perrin, Michael R. Polen, Michael C. Leo, Judith H. Hibbard, Martin Tusler. Development of the Patient Activation Measure for Mental Health. Adm Policy Ment Health 2010;37:327-333.
34. Hill-Briggs F. Problem solving in diabetes self-management: a model of chronic illness self-management behavior. Ann Behav Med 2003;25(3):182-193. [Medline: 12763713]
35. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
36. Hibbard JH, Cunningham PJ. 2008. Center for Studying Health System Change URL: http://hschange.org/CONTENT/1019/ [accessed 2013-12-30] [WebCite Cache ID 6MFe2zGuN]
37. 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. [doi: 10.1016/j.phr.2004.04.002.S0033354904000573]
38. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 9;288(14):1775-1779. [Medline: 12365965]
39. Remmers C, Hibbard J, Mosen DM, Wagenfield M, Hoye RE, Jones C. Is patient activation associated with future health outcomes and health care utilization among patients with diabetes? J Ambul Care Manage 2009;32(4):320-327. [doi: 10.1097/JAC.0b013e3181ba6e77] [Medline: 19888008]
40. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012 Feb;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
41. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun;33(6):1275-1281 [FREE Full text] [doi: 10.2337/dc09-2153] [Medline: 20299481]
42. Camoni website. URL: http://www.camoni.co.il/ [accessed 2013-12-30] [WebCite Cache ID 6MGL8Ks2U]
43. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
44. Ruggiero KJ, Gros DF, McCauley J, de Arellano MA, Danielson CK. Rural adults' use of health-related information online: data from a 2006 National Online Health Survey. Telemed J E Health 2011 Jun;17(5):329-334 [FREE Full text] [doi: 10.1089/tmj.2010.0195] [Medline: 21524201]
45. Perrin JM, Bloom SR, Gortmaker SL. The increase of childhood chronic conditions in the United States. JAMA 2007 Jun 27;297(24):2755-2759. [doi: 10.1001/jama.297.24.2755] [Medline: 17595277]
46. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The Disease Burden Associated With Overweight and Obesity. JAMA 1999;282(16):1523-1529. [doi: 10.1001/jama.282.16.1523]
47. Taylor K. Paternalism, participation and partnership - the evolution of patient centeredness in the consultation. Patient Educ Couns 2009 Feb;74(2):150-155. [doi: 10.1016/j.pec.2008.08.017] [Medline: 18930624]
48. Falkum E, Førde R. Paternalism, patient autonomy, and moral deliberation in the physician-patient relationship. Attitudes among Norwegian physicians. Soc Sci Med 2001 Jan;52(2):239-248. [Medline: 11144780]
49. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]

Abbreviations

PAM: patient activation measure

Edited by G Eysenbach; submitted 25.03.13; peer-reviewed by J Stockard, Y Zhang; comments to author 28.04.13; revised version received 27.05.13; accepted 21.11.13; published 10.01.14.

Please cite as:
Magnezi R, Bergman YS, Grosberg D
Online Activity and Participation in Treatment Affects the Perceived Efficacy of Social Health Networks Among Patients With Chronic Illness
J Med Internet Res 2014;16(1):e12
URL: http://www.jmir.org/2014/1/e12/
doi:10.2196/jmir.2630
PMID:24413148

Original Paper

Keywords to Recruit Spanish- and English-Speaking Participants: Evidence From an Online Postpartum Depression Randomized Controlled Trial

Alinne Z Barrera[1], PhD; Alex R Kelman[1], BA; Ricardo F Muñoz[1], PhD

Palo Alto University, Palo Alto, CA, United States

Corresponding Author:
Alinne Z Barrera, PhD
Palo Alto University
1791 Arastradero Road
Palo Alto, CA, 94304
United States
Phone: 1 650 433 3854
Fax: 1 650 433 3888
Email: abarrera@paloaltou.edu

Abstract

Background: One of the advantages of Internet-based research is the ability to efficiently recruit large, diverse samples of international participants. Currently, there is a dearth of information on the behind-the-scenes process to setting up successful online recruitment tools.

Objective: The objective of the study was to examine the comparative impact of Spanish- and English-language keywords for a Google AdWords campaign to recruit pregnant women to an Internet intervention and to describe the characteristics of those who enrolled in the trial.

Methods: Spanish- and English-language Google AdWords campaigns were created to advertise and recruit pregnant women to a Web-based randomized controlled trial for the prevention of postpartum depression, the Mothers and Babies/*Mamás y Bebés* Internet Project. Search engine users who clicked on the ads in response to keyword queries (eg, pregnancy, depression and pregnancy) were directed to the fully automated study website. Data on the performance of keywords associated with each Google ad reflect Web user queries from February 2009 to June 2012. Demographic information, self-reported depression symptom scores, major depressive episode status, and Internet use data were collected from enrolled participants before randomization in the intervention study.

Results: The Google ads received high exposure (12,983,196 impressions) and interest (176,295 clicks) from a global sample of Web users; 6745 pregnant women consented to participate and 2575 completed enrollment in the intervention study. Keywords that were descriptive of pregnancy and distress or pregnancy and health resulted in higher consent and enrollment rates (ie, high-performing ads). In both languages, broad keywords (eg, pregnancy) had the highest exposure, more consented participants, and greatest cost per consent (up to US $25.77 per consent). The online ads recruited a predominantly Spanish-speaking sample from Latin America of *Mestizo* racial identity. The English-speaking sample was also diverse with most participants residing in regions of Asia and Africa. Spanish-speaking participants were significantly more likely to be of Latino ethnic background, not married, completed fewer years of formal education, and were more likely to have accessed the Internet for depression information (P<.001).

Conclusions: The Internet is an effective method for reaching an international sample of pregnant women interested in online interventions to manage changes in their mood during the perinatal period. To increase efficiency, Internet advertisements need to be monitored and tailored to reflect the target population's conceptualization of health issues being studied.

Trial Registration: ClinicalTrials.gov NCT00816725; http://clinicaltrials.gov/show/NCT00816725 (Archived by WebCite at http://www.webcitation.org/6LumonjZP).

(J Med Internet Res 2014;16(1):e6) doi:10.2196/jmir.2999

KEYWORDS

Internet intervention; prevention; depression, postpartum; research subject recruitment; women; Spanish speaking

Introduction

The amount of Internet-based research in recent years has increased substantially [1-3] with Internet interventions showing significant promise in both alleviating symptoms and changing behavior [2]. Internet interventions are accessible at any time from any location, can be used anonymously, and may provide a much-needed service to users who feel marginalized or stigmatized [1,2,4]. For researchers, the Internet is an effective method of data collection [5] given that it is cost-effective, efficient at recruiting large sample sizes, and provides an opportunity for sensitive topics to be examined among hard-to-reach populations [6-9].

Most Internet intervention studies have focused on depression and anxiety, but an emphasis on other health issues is gaining interest as further evidence of their effectiveness is demonstrated [2,10]. The characteristics of those who participate in Internet interventions are dependent on the target problems being addressed; however, online study samples repeatedly include a greater proportion of women than men. This trend is not surprising given that the percentage of women using the Internet has steadily increased in recent years [11], and women are reported to be the largest group of Web users searching for health information related to mental illness [12]. The higher incidence of depression among women relative to men is likely a contributing factor to the greater focus on mental health topics, such as depression, among female Internet users.

Women and women-specific health issues are well suited for technology-based interventions with perinatal women garnering much of the attention by researchers (eg, [13-17]). Currently, there are countless websites that provide information on pregnancy and postpartum health. A recent review, however, revealed that the quality of the content exhibited on websites that focused on maternal mental health was partially correct or incomplete at best, and provided few self-help tools and resources to users [18]. Thus, there is a need to develop and empirically test Internet interventions specifically designed for women that address issues that uniquely affect women, such as postpartum depression (PPD). Computer- and Internet-based methods have been used successfully to increase awareness of PPD among adolescent parents [19], to recruit and screen women at risk for PPD [20-22], or to disseminate PPD psychoeducational materials to postpartum women and clinicians, all of whom have responded positively to this platform [22,23]. Recent studies have demonstrated support for behavioral activation [24,25] and cognitive behavioral [26] Internet interventions that aim to reduce depressive symptoms among depressed postpartum women. Two recent reports describe the design, feasibility, and acceptability of Internet interventions to prevent PPD [27,28]. The study presented in this report is based on data collected as part of the Mothers and Babies/*Mamás y Bebés* Internet Project, an Internet-based, 2-condition, pilot randomized controlled trial (RCT) designed to examine the efficacy of a Web-adapted mood management prevention intervention (Mothers and Babies Course/*Curso de Mamás y Bebés*) [29]. Participants were recruited, screened, and randomized to either the mood management Internet intervention or to an information brochure [23]. Preliminary analyses from the prevention trial are currently underway.

Although the Internet offers added benefits and opportunities, there are significant concerns and barriers with recruitment to Internet-based research investigations [30]. Many of these investigations rely on traditional forms of recruitment such as face-to-face recruitment, flyers, community messages, and mass media announcements [31-33]. Earlier studies that used online recruitment strategies struggled with identifying and reaching the target population (eg, [6,34]), low rates of response [35], and possible ethical concerns with recruiting over the Internet [3]. The more common methods of online recruitment are search engine advertisements, social media sites, online forums, Web links, banners, and email lists. The chosen method of recruitment depends on the population being targeted and where their online presence is greatest. However, recent reports suggest that targeted search engine advertisements are the most successful at reaching a wide range of potential participants for Internet research and are cost-effective methods of online recruitment [6,8].

What remains unclear in the literature is how researchers should frame their Internet advertisements so that they are time-efficient and cost-effective (ie, targeted to a specific population who are most likely to enroll in the study). Graham and colleagues [36] provide a good example of delineating the process of tailoring online advertisements to recruit Latino smokers to an online smoking cessation website. As part of their process, the researchers convened a multicultural panel of experts who informed the content of banner ads that were then empirically tested among Latino smokers. The researchers found that tailoring banner ads to portray the cultural value of *familismo* (emphasis on the importance of family influences) was preferable among Spanish-speaking Latino smokers who indicated that the issue of smoking was not just about them, but also about the impact on the relationships they valued [36]. Perhaps more importantly, this study demonstrated that a behind-the-scenes exploration of key factors that influence the success of an online recruitment method is valuable and needs to be empirically tested to maximize the strength of online recruitment. Furthermore, it is recommended that researchers continually monitor the behaviors of users once they are actively engaged in the study website [37].

The current study adds to the growing body of literature describing online recruitment methods for Internet intervention studies. The primary goal was to examine the impact of Google AdWords campaigns to recruit Spanish- and English-speaking pregnant women to an Internet RCT to prevent PPD. Both the Spanish and English ad campaigns used identical keywords and were minimally modified throughout the recruitment period. Given that few Internet intervention studies have focused exclusively on pregnant women, we describe the demographic characteristics of participants who enrolled in the trial.

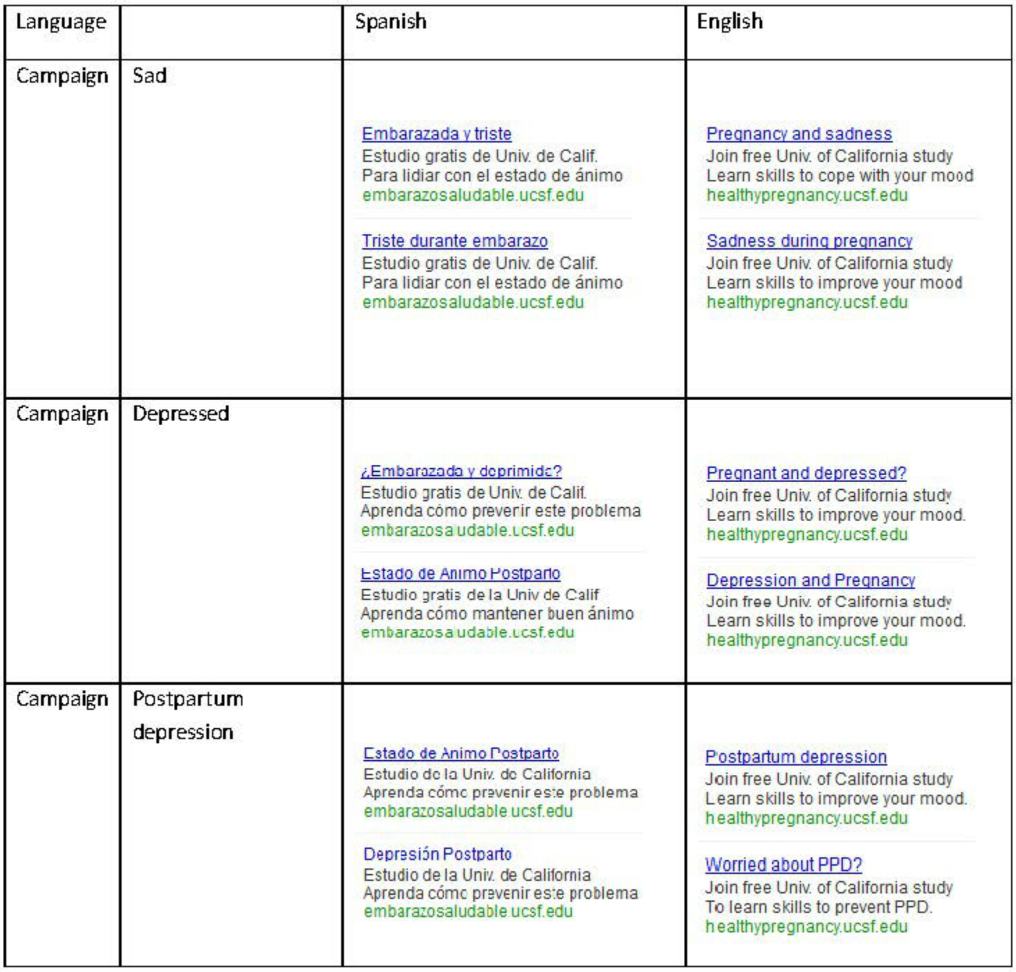
Methods

Internet Search Ads

Google AdWords sponsored link campaigns were created in Spanish and English to advertise the Mothers and Babies/*Mamás y Bebés* Internet Project, a RCT to test the efficacy of an Internet intervention for the prevention of PPD. Google ads consist of a 25-character headline and 2 additional 35-character text lines. In this study, the headlines were descriptive of perinatal depression, whereas the text lines described the study and provided the website link. There were 3 ad campaigns (sad, depressed, and postpartum depression) per language, each containing 2 distinctive ads for a total of 6 ads per language (see Figure 1). Ads were distributed worldwide, and the content and associated keywords of each ad remained unchanged throughout the course of the study recruitment period.

The Google AdWords account management website allows the owner of the ads to manipulate the bidding cost and content of each ad. A Google AdWords grant awarded to our team allowed for a maximum bid of US $1.00 per click. Ad owners can view the number of times the ad is shown (impressions) and the number of Web users who click on the ad (clicks); owners can also use HyperText Markup Language (HTML) code to track the impact of the ads on obtaining the desired action by the target audience (conversion). In this study, the HTML code to detect a conversion was integrated into the study website's consent page (ie, 1 conversion = 1 consent to participate). For the purposes of this paper, we will refer to conversions as "consents" given that it is the desired outcome of online recruitment. Additional information on Google AdWords procedures and metrics can be found elsewhere [38].

Figure 1. Google ads in Spanish and English for a prevention of postpartum depression trial.

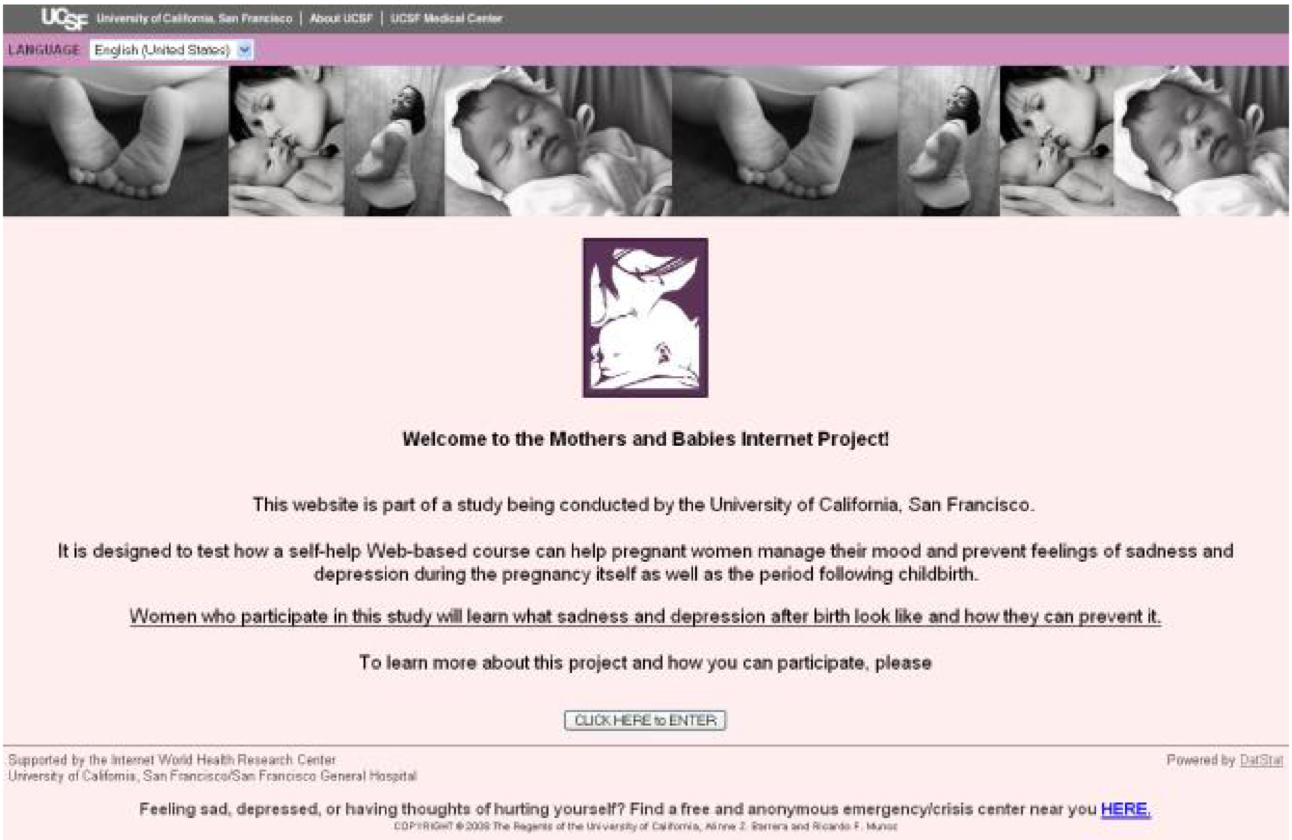


Language		Spanish	English
Campaign	Sad	Embarazada y triste Estudio gratis de Univ. de Calif. Para lidiar con el estado de ánimo embarazosaludable.ucsf.edu Triste durante embarazo Estudio gratis de Univ. de Calif. Para lidiar con el estado de ánimo embarazosaludable.ucsf.edu	Pregnancy and sadness Join free Univ. of California study Learn skills to cope with your mood healthypregnancy.ucsf.edu Sadness during pregnancy Join free Univ. of California study Learn skills to improve your mood healthypregnancy.ucsf.edu
Campaign	Depressed	¿Embarazada y deprimida? Estudio gratis de Univ. de Calif. Aprenda cómo prevenir este problema embarazosaludable.ucsf.edu Estado de Animo Postparto Estudio gratis de la Univ de Calif Aprenda cómo mantener buen ánimo embarazosaludable.ucsf.edu	Pregnant and depressed? Join free Univ. of California study Learn skills to improve your mood. healthypregnancy.ucsf.edu Depression and Pregnancy Join free Univ. of California study Learn skills to improve your mood. healthypregnancy.ucsf.edu
Campaign	Postpartum depression	Estado de Animo Postparto Estudio de la Univ. de California Aprenda cómo prevenir este problema embarazosaludable.ucsf.edu Depresión Postparto Estudio de la Univ. de California Aprenda cómo prevenir este problema embarazosaludable.ucsf.edu	Postpartum depression Join free Univ. of California study Learn skills to improve your mood. healthypregnancy.ucsf.edu Worried about PPD? Join free Univ. of California study To learn skills to prevent PPD. healthypregnancy.ucsf.edu

Participant Recruitment

Web users who searched the Internet between February 3, 2009 and June 15, 2012 with keyword queries associated with the ads were able to click on them and link to the Mothers and Babies/*Mamás y Bebés* Internet Project's website landing page (Figure 2); this initial page briefly described the study and invited those interested to complete the eligibility screener. The study website was fully automated with items that contained logic to determine participant flow through the study. The eligibility screener contained items to assess for eligibility criteria for the RCT: being female, pregnant, age 18 years or older, and interested in the study website for themselves. Eligible women were directed to the baseline assessment, which contained the University of California institutional review board–approved consent form. Participants "signed" the form online by clicking “Yes, I am interested in participating in this study” and by entering a unique password that was generated in real time and provided online to the user. Eligible participants who consented and completed the baseline assessment were considered enrolled in the study regardless of their depression status or stage of pregnancy. Enrolled participants received access to the intervention sites being evaluated and were invited to complete monthly follow-up assessments up to 6 months postpartum.

Figure 2. Screenshot of the study landing page.



Welcome to the Mothers and Babies Internet Project!

This website is part of a study being conducted by the University of California, San Francisco.

It is designed to test how a self-help Web-based course can help pregnant women manage their mood and prevent feelings of sadness and depression during the pregnancy itself as well as the period following childbirth.

Women who participate in this study will learn what sadness and depression after birth look like and how they can prevent it.

To learn more about this project and how you can participate, please

Supported by the Internet World Health Research Center
University of California, San Francisco/San Francisco General Hospital

Feeling sad, depressed, or having thoughts of hurting yourself? Find a free and anonymous emergency/crisis center near you HERE.

Measures

Eligibility Screener

Upon entering the study website, participants indicated their country of residence and preferred language and completed items to determine their eligibility to participate (age, gender, pregnancy status, how they planned to use the website materials). Eligible participants who entered a valid email address were directed to the baseline assessment.

Baseline Assessment

The initial page of the survey was the informed consent that participants were required to complete before proceeding through the study website. Consenting participants were asked to complete questionnaire items on demographic characteristics (eg, country of birth, ethnicity, race, education), Internet use (eg, previous use for health information), pregnancy history (eg, weeks pregnant, pregnancy history), and depression (eg, current symptoms).

Depression status was determined by the Center for Epidemiologic Studies-Depression Scale (CES-D) [39] and the Major Depressive Episode (MDE) Screener-Current/Lifetime version [40]. The CES-D is a 20-item self-report instrument that assesses for the presence of depressive symptoms during the past week. Total scores range from 0-60, with higher scores indicating more severe depressive symptoms. The MDE Screener is an 18-item self-report questionnaire that assesses for the presence of 5 or more MDE symptoms experienced within a 2-week or longer period of time during the past 2 weeks (current MDE) or during any period (other than the past 2 weeks) in their lifetime (past MDE). To screen positive for a MDE, significant impairment, as defined by Criterion C of the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) (*DSM-IV*) must also be present [41]. The MDE Screener is a screening tool that has demonstrated good psychometric properties with diagnostic screeners and clinical interviews [42].

Data Analysis

Data on the performance of search engine ads were extracted from the Google AdWords management website [38]. Participant data for those who visited (eligibility), consented (started baseline), and enrolled (completed baseline) in the RCT were analyzed using SPSS for Windows 20.0 (IBM Corp, Armonk, NY, USA). Descriptive and chi-square analyses were conducted to examine group differences.

Results

Participant Enrollment

Web users interested in the Spanish- and English-language Google ads, as defined by clicking on the ad (ie, clicks), were located in 183 countries and territories; 6745 pregnant women consented to participate with 2517 (37.32%) failing to enter any data in the baseline assessment. Data collected in the eligibility screener indicated that those who consented but did not provide baseline information were older (*P*<.001) and more likely to be English speakers (*P*<.001). Of the remaining 4228 participants, 1653 (24.51% of those who consented) failed to enroll in the intervention study because they did not complete the baseline assessment and, therefore, were excluded from further analyses. The enrolled sample consisted of 2575 participants who met the eligibility criteria, consented to participate, and completed the baseline assessment. Compared to those who consented to participate but did not enroll in the study, participants who enrolled were more likely to be Latino (*P*=.02), older (*P*<.001), employed (*P*=.007), and have attained higher levels of education (*P*<.001). Higher rates of past (11.60% vs 9.51%) and current MDE (19.06% vs 9.69%) were reported for enrolled participants relative to those who did not complete the baseline assessment (*P*<.001).

Impact of Internet Search Ads

The Google ads were active for approximately 40 months during which 12,983,196 impressions and 176,295 clicks were made in response to Web users' search entries (see Table 1). In all, 25.09% of those who clicked on the ad entered the study website after reviewing the brief description provided on the study's landing page. Over 60% of the total impressions (62.27%) were in response to Web users who searched in English, whereas 66.10% of the total clicks were made by those who searched in Spanish. In the Spanish campaign, 116,531 Web users of 4,898,063 who were presented with an ad (ie, an impression) clicked the study ad (2.38%). In the English campaign, there were 59,764 clicks in response to 8,085,133 impressions (0.74%). See Table 2 for detailed data on the performance of each ad campaign by language.

The keywords with the greatest exposure (ie, highest clicks and impressions) and which generated the highest traffic in both languages were pregnancy/*embarazo* and pregnant/*embarazada* (Table 3). For all campaigns, 113,525 clicks of 4,819,662 impressions (2.35%) and 57,197 clicks of 7,836,146 impressions (0.73%) were generated by these keywords in the Spanish and English campaigns, respectively. In addition, those who queried with these keywords were the majority of those who consented to participate—94.53% (4718/4992) and 92.87% (1823/1963), respectively. In contrast, the rate of consent was highest for English-language speakers who searched with keywords that included a reference to both pregnancy and emotions (eg, mental health during pregnancy). A similar pattern was demonstrated in the Spanish campaign with keyword phrases such as "*depresión despues del parto*/depression after birth" or "*embarazada y deprimida*/pregnant and depressed" resulting in a greater proportion of consenting participants. Keywords that were descriptive of pregnancy without a reference to emotions (eg, months pregnant) also resulted in a higher rate of consent. This occurred with more frequency in the Spanish campaign (*nueva mama*/new mother, *meses embarazo*/months pregnant, *semanas embarazo*/weeks pregnant) than in the English campaign (months pregnant).

Table 1. Participant online recruitment for the Mothers and Babies/*Mamás y Bebés* Internet Project.

Ad or user behavior	Language, n		Total, n
	Spanish	English	
Impressions (ads presented)	4,898,063	8,085,133	12,983,196
Clicks (clicks on ads)	116,531	59,764	176,295
Entered site (user proceeded beyond initial page)	28,074	16,157	44,231
Screened (user answered ≥1 eligibility items)	11,620	5349	16,969
Eligible (user met eligibility criteria)	8728	3738	12,466
Consented (user agreed to consent)	4773	1972	6745
Enrolled (user consented and completed baseline assessment)	2012	563	2575

Table 2. Performance of Google AdWords campaigns from February 3, 2009 to June 15, 2012.

Ad or user behavior	Ad performance, n					
	Sad		Depressed		Postpartum depression	
	Spanish	English	Spanish	English	Spanish	English
Consent	4855	1782	136	179	1	2
Clicks	114,902	53,667	1574	5987	55	110
Impressions	4,794,470	7,184,384	92,280	856,634	11,313	44,115

Table 3. Spanish- and English-language ads with the highest-performing keyword(s).

Ad character headline	Keyword(s)	Consents	Cost per consent (US $)	Clicks	Impressions
Spanish					
Embarazada y triste	*Embarazo/embarazada* (pregnancy/pregnant)	4718	$15.98	113,525	4,819,662
Triste durante embarazo	*Meses de embarazo* (months pregnant)	156	$8.55	2130	54,987
¿Embarazada y deprimida?	*Depresión en el embarazo* (depression during pregnancy)	109	$2.89	739	11,716
English					
Sadness during pregnancy	Pregnancy/pregnant	1823	$25.77	57,197	7,836,146
Depression and pregnancy	Pregnancy depression	82	$11.99	1301	97,488
Pregnancy and sadness	Sad pregnant	17	$12.52	290	18,095
Pregnant and depressed?	Depressed pregnant	16	$9.12	193	16,457

Enrolled Participant Characteristics

The final sample of enrolled participants (N=2575) consisted of pregnant women with a mean age of 28.16 years (SD 5.47) (see Table 4). Most completed study materials in Spanish (78.13%), were of Latino/Hispanic ethnic identity (77.89%), and self-identified their racial background as *Mestizo* descent (person of mixed Spanish and Indigenous ancestry; 35.16%) or European descent (30.43%). Most participants were married or living with a partner (66.99%), employed (59.90%), and college educated (70.37%).

There were no differences in pregnancy status or depression history between the Spanish- and English-speaking participants. Participants were mostly in the second trimester of their pregnancy (mean 16.56 weeks, SD 9.59) and most (69.34%) did not meet *DSM-IV* criteria for a MDE. However, more English speakers met the criteria for a current MDE relative to Spanish speakers (20.04% vs 18.81%, respectively), whereas a higher percentage of Spanish speakers met criteria for a past MDE when compared to the English-speaking participants (11.88% vs 10.50%, respectively), although these differences were not statistically significant (*P*=.61). The mean CES-D score was elevated for both samples (mean 27.51, SD 13.77) with Spanish speakers endorsing slightly higher levels of depressive symptoms (*P*=.01). Participants reported accessing the Internet for information related to depression when not pregnant (29.31%) and during the perinatal period (42.64%). Spanish speakers accessed the Internet for depression information at greater proportions than English speakers did, especially when not pregnant or postpartum (32.05% vs 19.35%, *P*<.001).

Group comparisons revealed that the Spanish-speaking women were more likely than English-speaking women to self-identify their ethnic background as Latino or Hispanic (91.45 vs 9.85%, *P*<.001). The racial descent varied significantly (*P*<.001) by language with the Spanish speakers mostly identifying as *Mestizo* (44.10%), European/Caucasian (32.74%), and other (16.42%), whereas the English speakers were of Asian (41.97%), European/Caucasian (21.60%), and African descent (19.75%). Similarly, Spanish speakers resided in Latin America and Spain whereas English speakers were primarily from India (Table 5). A greater proportion of English speakers were married or living with a partner (88.63%) compared to Spanish speakers (60.92%, *P*<.001). English speakers were more likely to have earned advanced educational degrees (28.08% vs 8.75%) whereas Spanish speakers were mostly comprised of college educated (73.38% vs 59.60%, *P*<.001) women.

Table 4. Baseline characteristics of enrolled participants.

Demographic items	Spanish n=2012	English n=563	Total N=2575	*P*
Age, mean (SD)	28.18 (5.68)	28.07 (4.60)	28.16 (5.47)	.69
Latino/Hispanic (n=2343,[a]), n (%)	1818 (91.45)	39 (9.85)	1857 (77.89)	<.001
Race (n=2343), n (%)				<.001
Mestizo [b]	819 (44.10)	5 (1.03)	824 (35.17)	
European/Caucasian descent	608 (32.74)	105 (21.60)	713 (30.43)	
Other	305 (16.42)	73 (15.02)	378 (16.13)	
Asian descent	11 (0.59)	204 (41.97)	215 (9.18)	
African descent	12 (0.65)	96 (19.75)	108 (4.61)	
American Indian/Alaska Native	102 (5.49)	3 (0.62)	105 (4.48)	
Married/live with partner (n=2572), n (%)	1224 (60.92)	499 (88.63)	1723 (67.99)	<.001
Education (n=2528), n (%)				<.001
12 years or less	353 (17.86)	68 (12.32)	421 (16.65)	
University level/degree	1450 (73.38)	329 (59.60)	1779 (70.37)	
Advanced degree	173 (8.75)	155 (28.08)	328 (12.97)	
Employed (n=2561), n (%)	1217 (60.76)	317 (56.81)	1534 (59.90)	.09
Weeks pregnant (n=2564), mean (SD)	16.71 (9.62)	16.01 (9.47)	16.56 (9.59)	.13
MDE history (n=2492), n (%)				.61
None	1371 (69.31)	357 (69.45)	1728 (69.34)	
Current MDE	372 (18.81)	103 (20.04)	475 (19.06)	
Past MDE	235 (11.88)	54 (10.50)	289 (11.60)	
CES-D score (n=2475), mean (SD)	27.86 (13.89)	26.07 (13.22)	27.51 (13.77)	.01
Use Internet for depression information, n (%)				
During perinatal period	873 (43.39)	225 (39.96)	1098 (42.64)	.15
When not pregnant (n=2591)	643 (32.05)	107 (19.35)	750 (29.31)	<.001

[a] Valid percent reflects participants who completed the item.

[b] *Mestizo*: person of mixed Spanish and Indigenous ancestry.

Table 5. Country of residence of enrolled participants.

Country	%[a]
Spanish-speaking participants (2008/2012)	
Chile	16.30
Mexico	15.24
Spain	12.00
Argentina	11.35
Colombia	11.20
Venezuela	11.00
Peru	5.23
Bolivia	3.43
Ecuador	2.79
Dominican Republic	2.54
Uruguay	1.34
Paraguay	1.14
29 Countries/territories with less than 1% each	6.44
English-speaking participants (562/563)	
India	36.30
South Africa	12.81
Pakistan	7.47
United Kingdom	7.12
Iran	2.67
Nigeria	2.31
United States	2.13
Ghana	1.96
Kenya	1.78
United Arab Emirates	1.60
Saudi Arabia, Maldives, Uganda, Ireland (each)	1.07
54 Countries/territories with less than 1% each	19.57

[a] Based on self-reported country of residence data.

Discussion

Principal Findings

This study examined the impact of Spanish- and English-language Google AdWords to recruit an online sample of participants to an Internet RCT for the prevention of PPD. This method of recruitment was effective at exposing a large number of female Web users to the opportunities offered by the study website, such as learning skills to manage changes in mood during the transition to motherhood. During the 3 years that the ads were active, 176,295 Web users clicked on the advertised link and 12,466 pregnant women met eligibility criteria; 54% agreed to participate and 2575 (20.66%) completed sufficient baseline information to enroll in the study. These data suggest that pregnant women from around the world are interested in learning skills to manage their mood during and after pregnancy, are willing to use Web-based resources, and that the Internet is a viable means to reach them. Although a large number of women showed interest in joining the study, a large proportion exited the site immediately after clicking the ad or failed to continue once they were informed of participation details. A possible explanation for the former is the mismatch between what Web users queried and what they eventually found on the study's landing page. That is, the information listed on the site did not correspond with what they were looking for or hoping to find when they searched, clicked, and initially visited the study website. The burden of participation may have been a factor to dissuade eligible participants from continuing their engagement in the study once they learned the details of participation from the consent form. Clearly, we need to make such Internet research sites more interesting and less burdensome so that more of those who are eligible not only consent to participate but also complete all facets of the study.

High Performing Keywords

Our examination of the highest-performing keywords for each of the Google AdWords campaigns revealed 3 primary findings. First, keywords that were more descriptive of the website content (eg, pregnant and depressed) resulted in a higher rate of consent to participate. This pattern of user behavior was evident in both the Spanish and English ads that referenced pregnancy and emotions, and suggests that the biggest gains occur when advertisements closely relate to the product being promoted. This approach is especially important online where there is heavy competition for the user's attention and sustaining interest in one site over another is challenging. The specificity of pairing a reference to pregnancy and emotions in the keywords resulted in a higher consent rate among Spanish versus English users. This keyword type was also the most cost-effective in both languages relative to other high-performing keyword types (eg, pregnancy without referencing emotions). Cost-effectiveness by itself is not the most important measure of usefulness. The most cost-effective keywords yielded the least number of participants. To recruit the most participants in a reasonable period of time, we must be ready to pay for keywords that yield higher cost per consent. Given that most researchers are working within budgetary constraints, being able to maximize the cost of each consented individual is of high priority and importance. Furthermore, for the English ads there were more variations of this keyword type and, therefore, more related activity. Given that both the Spanish and English ads were identical in their content and associated keywords, this finding may suggest that Spanish-speaking women are less likely to search for information related to emotional health during pregnancy than English-speaking women or that the associated keywords do not reflect how Spanish-speaking women conceptualize perinatal distress. In fact, keywords that did not reference emotions or feelings but described pregnancy characteristics (eg, months pregnant) were very high activity ads and yielded a much larger number of impressions, clicks, consents, and enrolled participants. This suggests that women may be initially attracted by information about pregnancy that does not include information on emotional health during pregnancy. Thus, ads that are more generic may result in better recruitment outcomes for more women, of whom a subset may be subsequently intrigued by the focus on such issues as mood.

Keyword Language Differences

Previous reports have suggested that Latinos and Spanish speakers manifest psychological distress through physical complaints that can stem from stigma, cultural barriers, and differences in how psychological issues are conceptualized and experienced [43,44]. This study found that ads associated with broad search terms (eg, pregnancy) had greater reach and enrollment impact. This pattern of user behavior was evident in both languages and is consistent with female patterns of online search behavior which shows that women, more than men, search for health-related issues and that pregnancy-related queries are a major topic of inquiry among childbearing-aged female Web users [45,46]. There is a high likelihood that ads that were linked to broad keywords appeared in a large number of searches, many of which did not result in a visit to the study website. However, Spanish-speaking women who used broad keywords in their queries appeared to be more interested in the ads than English-speaking women as evidenced by the higher number of clicks and consents relative to the number of impressions which were almost twice as high in the English ads. We speculate that there may be fewer such Internet resources in Spanish than in English, therefore drawing a higher number of Spanish speakers to our website. These data indicate that broad, nonspecific keywords were effective at recruiting a large, diverse sample of pregnant women. However, this approach was more costly in the long term than more-specific keywords, which were more cost-effective per consent but resulted in fewer consenting participants. Researchers need to consider what their recruitment goals are and adjust their ad campaigns accordingly. To the best of our knowledge, this is the first fully automated prevention of PPD trial to recruit a large worldwide sample of Spanish- and English-speaking pregnant women. At the start of our recruitment period, we were uncertain of the potential interest of the study among pregnant women. Thus, our initial recruitment goal, which was to reach a large number of potential female Web users across the world, was accomplished.

Participant Characteristic Outcomes

A secondary aim of this study was to examine the characteristics of individuals who consented to participate in a Web-based trial. The ads attracted a mostly Spanish-speaking sample of women who resided in Latin American countries. This was an unexpected outcome of the recruitment efforts given the lower Internet penetration rate in this region of the world relative to Oceania/Australia, Europe, and North America [47]. Although the English-speaking sample was smaller, participants were equally diverse with the largest number of participants indicating that they resided in India and across different regions of Africa. The geographic location of participants mirrors the recent gender shift in Internet access by women in developed and developing nations. Furthermore, it highlights the potential need to develop new technology-based interventions targeting women from all around the world, especially those with current limited access to online resources but whose access is growing at a rapid rate (eg, 1310.8% growth in Latin American and the Caribbean in the past 10 years) [47].

Spanish- and English-speaking participants were relatively similar on demographic characteristics with a few exceptions. The breakdown of the origin of our participants is important within a global mental health perspective, and especially critical when considering maternal mental health needs worldwide. To date, few resources exist in these regions of the world to address the day-to-day needs of pregnant women [48,49], especially as they relate to mental health issues during pregnancy. Furthermore, these women are accessing the Internet in regions of the world where the overall Internet penetration is lower and men significantly outnumber women in their use of the Internet [11,50]. This gender divide is likely to change as the global use of technology continues to grow.

Participants in this study were well educated, with a greater proportion of English-language speakers possessing advanced degrees, whereas Spanish-language speakers had mostly attended university level or fewer years of formal education. It is clear that women with varied levels of education are using the Internet to obtain pregnancy-related information. Spanish- and English-speaking participants differed in their marital status with Spanish speakers more likely to indicate that they were single or without a current partner, thus potentially raising their risk for the onset of postpartum mood disorders [51]. Finally, the high rate of major depression found in this sample is of clinical concern and echoes a recent call-to-action to make maternal mental health a top priority [52]. Nearly one-third (30.66%) of our participants screened positive for a MDE sometime in their lifetime, yet our recruitment efforts did not seek or select participants based on their depression status. This indicates that the Internet is an effective tool to reach pregnant women who are depressed or at high risk for depression. Rates of perinatal depression range from 10%-25% [53,54], with point prevalence rates during pregnancy hovering around 12%-15% [51,55]. The higher rate of MDE found in this study may be attributed to measurement differences, the use of a self-report screening measure and not on clinician diagnostic information, the symptom overlap between pregnancy and depression, or to self-selection. Regardless, high rates of depressive symptoms were reported which is consistent with online samples of postpartum women [22]. These data highlight the need to develop multilingual, culturally considerate Internet resources for pregnant women regardless of depression risk. Many websites provide pregnancy information and many pregnant women are accessing the Internet to help guide their prenatal and postpartum health care [56]. However, the quality and reliability of the sites that include perinatal mental health topics is variable with a majority falling short of relaying complete and accurate information and even fewer providing practical skills to reduce symptoms [18]. Current studies that are in progress, such as the Mothers and Babies/*Mamás y Bebés* Internet Project directed by our team, as well as other prevention of PPD trials [27,28], stand to make a significant contribution to the availability of empirically tested Internet interventions to prevent PPD.

Limitations

Our findings were limited in several ways. First, the Google AdWords grant awarded to our team capped our bids at US $1.00. Second, this study solely reports on the use of minimally modified Google AdWords campaigns. We do not know how the ads would have performed if we had managed the campaigns based on their performance by adjusting the keywords, daily cost, or regional exposure. Third, given how the ads were set up and linked to the study website, we are unable to connect keyword types to participant characteristics. In order to understand how to target the ads to participants in different regions of the world who speak different languages and who conceptualize emotional distress differently, it would be beneficial to explore this further by recording the keyword entered by the user in their online search and examining their characteristics and behaviors on the study website. Finally, our data are only generalizable to Spanish and English speakers who use the Web to obtain information on the perinatal period.

Conclusions

This study contributes to the growing understanding of online recruitment for intervention trials. We provide evidence that pregnant women in many regions of the world are already seeking this type of health information and choosing to engage in Internet interventions by virtue of their participation in this study. Nonconsumable Internet interventions or those that can be reused with minimal added cost have the potential to reduce health disparities globally because they can be used an unlimited number of times from any region of the world without significant increases in cost [1]. The opportunity to share these resources with a wide range of individuals who may lack local resources requires the ability to design and test these interventions with global samples of individuals from diverse ethnic and cultural backgrounds. The health field would do well to use the potential of Internet interventions to increase health resources focused on maternal mental health and to reduce health disparities where there are few resources to protect childbearing women and advance the well-being of mothers and their babies.

Acknowledgments

The original RCT was funded by the National Institute of Mental Health (F32MH077371). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. Additional funding was provided by a Robert Wood Johnson Health Disparities Seed Grant (Nancy Adler, PI) to Dr Barrera, funding from the University of California Committee on Latino Research which funded the UCSF/SFGH Latino Mental Health Research Program (Muñoz, PI), and funds granted to the UCSF Internet World Health Research Center (Muñoz, Director) from the Department of Psychiatry at SFGH while under the direction of Robert Okin, MD. A very special thanks to Google, Inc for awarding the team (Muñoz, PI) an AdWords grant, which provided the ability to recruit worldwide using Google-sponsored links. Online data were collected by using the Illume survey system developed by DatStat, Inc. We thank Drs Lynn Abeita and Eunice Kim for feedback on earlier drafts of this paper. Lastly, we are thankful to all the women who participated in this research.

Conflicts of Interest

None declared.

References

1. Muñoz RF. Using evidence-based Internet interventions to reduce health disparities worldwide. J Med Internet Res 2010;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]
2. Griffiths KM, Christensen H. Review of randomised controlled trials of Internet interventions for mental disorders and related conditions. Clinical Psychologist 2006;10(1):16-29. [doi: 10.1080/13284200500378696]
3. Im EO, Chee W. Recruitment of research participants through the Internet. Comput Inform Nurs 2004;22(5):289-297. [Medline: 15520599]
4. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting Internet intervention research. Cognitive Behaviour Therapy 2011 Jun;40(2):82-97. [doi: 10.1080/16506073.2011.573807]
5. Temple EC, Brown RF. A comparison of Internet-based participant recruitment methods: Experiences engaging the hidden population of cannabis users in research. Journal of Research Practice 2011;7(2):1-20.
6. Jones RB, Goldsmith L, Williams CJ, Kamel Boulos MN. Accuracy of geographically targeted Internet advertisements on Google AdWords for recruitment in a randomized trial. J Med Internet Res 2012;14(3):e84 [FREE Full text] [doi: 10.2196/jmir.1991] [Medline: 22718043]
7. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
8. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the Internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
9. Reips UD, Lengler R. The Web Experiment List: a Web service for the recruitment of participants and archiving of Internet-based experiments. Behav Res Methods 2005 May;37(2):287-292. [Medline: 16171201]
10. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
11. Boland AL, Pauline MM, Vollman A. comScore. 2010 Jun 30. Women on the Web: How women are shaping the Internet URL: http://www.comscore.com/Insights/Presentations_and_Whitepapers/2010/Women_on_the_Web_How_Women_are_Shaping_the_Internet [accessed 2013-12-17] [WebCite Cache ID 6LwNRydrv]
12. Lamberg L. Online empathy for mood disorders: patients turn to Internet support groups. JAMA 2003 Jun 18;289(23):3073-3077. [doi: 10.1001/jama.289.23.3073] [Medline: 12813097]
13. Anderson GD, Nelson-Becker C, Hannigan EV, Berenson AB, Hankins GD. A patient-centered health care delivery system by a university obstetrics and gynecology department. Obstet Gynecol 2005 Jan;105(1):205-210. [doi: 10.1097/01.AOG.0000146288.28195.27] [Medline: 15625165]
14. Britt DW, Norton JD, Hubanks AS, Navidad SA, Perkins RJ, Lowery CL. A two-period assessment of changes in specialist contact in a high-risk pregnancy telemedical program. Telemed J E Health 2006 Feb;12(1):35-41. [doi: 10.1089/tmj.2006.12.35] [Medline: 16478411]
15. Dennis CL, Kingston D. A systematic review of telephone support for women during pregnancy and the early postpartum period. J Obstet Gynecol Neonatal Nurs 2008;37(3):301-314. [doi: 10.1111/j.1552-6909.2008.00235.x] [Medline: 18507601]
16. Hod M, Kerner R. Telemedicine for antenatal surveillance of high-risk pregnancies with ambulatory and home fetal heart rate monitoring--an update. J Perinat Med 2003;31(3):195-200. [doi: 10.1515/JPM.2003.026] [Medline: 12825474]
17. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an Internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
18. Moore D, Ayers S. A review of postnatal mental health websites: help for healthcare professionals and patients. Arch Womens Ment Health 2011 Dec;14(6):443-452. [doi: 10.1007/s00737-011-0245-z] [Medline: 22109827]
19. Logsdon MC, Barone M, Lynch T, Robertson A, Myers J, Morrison D, et al. Testing of a prototype Web based intervention for adolescent mothers on postpartum depression. Appl Nurs Res 2013 Aug;26(3):143-145. [doi: 10.1016/j.apnr.2013.01.005] [Medline: 23473677]
20. Le HN, Perry DF, Sheng X. Using the Internet to screen for postpartum depression. Matern Child Health J 2009 Mar;13(2):213-221. [doi: 10.1007/s10995-008-0322-8] [Medline: 18278545]
21. Kim H, Bracha Y, Tipnis A. Automated depression screening in disadvantaged pregnant women in an urban obstetric clinic. Arch Womens Ment Health 2007;10(4):163-169. [doi: 10.1007/s00737-007-0189-5] [Medline: 17593320]
22. Maloni JA, Przeworski A, Damato EG. Web recruitment and Internet use and preferences reported by women with postpartum depression after pregnancy complications. Arch Psychiatr Nurs 2013 Apr;27(2):90-95. [doi: 10.1016/j.apnu.2012.12.001] [Medline: 23540519]
23. Wisner KL, Logsdon MC, Shanahan BR. Web-based education for postpartum depression: conceptual development and impact. Arch Womens Ment Health 2008 Dec;11(5-6):377-385. [doi: 10.1007/s00737-008-0030-9] [Medline: 18784975]
24. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2013 Oct 23:1-15. [doi: 10.1017/S0033291713002092] [Medline: 24148703]
25. O'Mahen HA, Woodford J, McGinley J, Warren FC, Richards DA, Lynch TR, et al. Internet-based behavioral activation--treatment for postnatal depression (Netmums): a randomized controlled trial. J Affect Disord 2013 Sep 25;150(3):814-822. [doi: 10.1016/j.jad.2013.03.005] [Medline: 23602514]

26. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: Formative research and design of the MomMoodBooster program. JMIR Res Protoc 2012;1(2):e18 [FREE Full text] [doi: 10.2196/resprot.2329] [Medline: 23612274]
27. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a Web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]
28. Jones BA, Griffiths KM, Christensen H, Ellwood D, Bennett K, Bennett A. Online cognitive behaviour training for the prevention of postnatal depression in at-risk mothers: a randomised controlled trial protocol. BMC Psychiatry 2013 Oct 17;13(1):265 [FREE Full text] [doi: 10.1186/1471-244X-13-265] [Medline: 24131528]
29. Muñoz RF, Le H, Ippen CG, Diaz MA, Urizar GG, Soto J, et al. Prevention of postpartum depression in low-income women: Development of the Mamás y Bebés/Mothers and Babies course. Cognitive and Behavioral Practice 2007 Feb;14(1):70-83. [doi: 10.1016/j.cbpra.2006.04.021]
30. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
31. Alexander GL, Divine GW, Couper MP, McClure JB, Stopponi MA, Fortman KK, et al. Effect of incentives and mailing features on online health program enrollment. Am J Prev Med 2008 May;34(5):382-388 [FREE Full text] [doi: 10.1016/j.amepre.2008.01.028] [Medline: 18407004]
32. Bowen A, Williams M, Horvath K. Using the Internet to recruit rural MSM for HIV risk assessment: sampling issues. AIDS Behav 2004 Sep;8(3):311-319 [FREE Full text] [doi: 10.1023/B:AIBE.0000044078.43476.1f] [Medline: 15475678]
33. Hedman E, Ljótsson B, Rück C, Furmark T, Carlbring P, Lindefors N, et al. Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. Computers in Human Behavior 2010 Jul;26(4):736-740. [doi: 10.1016/j.chb.2010.01.010]
34. Jones RB, Goldsmith L, Hewson P, Williams CJ. Recruitment to online therapies for depression: pilot cluster randomized controlled trial. J Med Internet Res 2013;15(3):e45 [FREE Full text] [doi: 10.2196/jmir.2367] [Medline: 23462072]
35. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided Internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011 Sep;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]
36. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an Internet cessation program: impact and costs. J Med Internet Res 2012;14(4):e116 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]
37. Crutzen R, Roosjen JL, Poelman J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. Health Promot Int 2013 Mar;28(1):36-42. [doi: 10.1093/heapro/das008] [Medline: 22377974]
38. Google. Google AdWords: Online advertising by Google URL: http://www.google.com/adwords/?sourceid=awo&subid=ww-et-awhp_nelsontest3_nel_p&clickid [accessed 2013-12-17] [WebCite Cache ID 6JyunxlUT]
39. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
40. Muñoz RF. Preventing major depression by promoting emotion regulation: A conceptual framework and some practical tools. Int J Ment Health Promot 1998 Sep(Inaugural Issue):23-40 [FREE Full text]
41. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition). Washington, DC: American Psychiatric Association; 2000.
42. Vázquez FL, Muñoz RF, Blanco V, López M. Validation of Muñoz's mood screener in a nonclinical Spanish population. European Journal of Psychological Assessment 2008 Jan 21;24(1):57-64. [doi: 10.1027/1015-5759.24.1.57]
43. Nadeem E, Lange JM, Miranda J. Perceived need for care among low-income immigrant and US-born black and Latina women with depression. J Womens Health (Larchmt) 2009 Mar;18(3):369-375 [FREE Full text] [doi: 10.1089/jwh.2008.0898] [Medline: 19281320]
44. Hulme PA. Somatization in Hispanics. J Psychosoc Nurs Ment Health Serv 1996 Mar;34(3):33-37. [Medline: 8676289]
45. Fox S. Health topics: 80% of Internet users look for health information online. Washington, DC: Pew Internet & American Life Project; 2011 Feb 01. URL: http://www.pewInternet.org/~/media//Files/Reports/2011/PIP_Health_Topics.pdf [accessed 2013-12-03] [WebCite Cache ID 6LcmcoBqq]
46. Pogorelc D. Google's top health searches for: cancer, pregnancy symptoms and…hemorrhoid?. 2012. URL: http://medcitynews.com/2012/12/googles-top-health-searches-for-2012-cancer-pregnancy-symptoms-and-hemorrhoid/ [accessed 2013-09-27] [WebCite Cache ID 6JyuyWBRT]
47. Internet World Stats. 2012. World Internet Users Statistics Usage and World Population Stats URL: http://www.Internetworldstats.com/stats.htm [accessed 2006-01-15] [WebCite Cache ID 5CzJrxnr0]
48. Nour NM. An introduction to maternal mortality. Rev Obstet Gynecol 2008;1(2):77-81. [Medline: 18769668]
49. Belizán JM, Cafferata ML, Belizán M, Tomasso G, Chalmers B. Goals in maternal and perinatal care in Latin America and the Caribbean. Birth 2005 Sep;32(3):210-218. [doi: 10.1111/j.0730-7659.2005.00372.x] [Medline: 16128976]

50. Dalberg Global Development Advisors. Women and the Web: Bridging the Internet gap and creating new global opportunities in low and middle-income countries.: Intel Corporation; 2012. URL: http://www.intel.com/content/dam/www/public/us/en/documents/pdf/women-and-the-web.pdf [accessed 2013-12-17] [WebCite Cache ID 6LwO36Rs6]
51. Le Strat Y, Dubertret C, Le Foll B. Prevalence and correlates of major depressive episode in pregnant and postpartum women in the United States. J Affect Disord 2011 Dec;135(1-3):128-138. [doi: 10.1016/j.jad.2011.07.004] [Medline: 21802737]
52. Rahman A, Surkan PJ, Cayetano CE, Rwagatare P, Dickson KE. Grand challenges: integrating maternal mental health into maternal and child health programmes. PLoS Med 2013;10(5):e1001442 [FREE Full text] [doi: 10.1371/journal.pmed.1001442] [Medline: 23667345]
53. O'hara MW, Swain AM. Rates and risk of postpartum depression—a meta-analysis. Int Rev Psychiatry 1996 Jan;8(1):37-54. [doi: 10.3109/09540269609037816]
54. Fisher J, Cabral de Mello M, Patel V, Rahman A, Tran T, Holton S, et al. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bull World Health Organ 2012 Feb 1;90(2):139G-149G [FREE Full text] [doi: 10.2471/BLT.11.091850] [Medline: 22423165]
55. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
56. Romano AM. A changing landscape: Implications of pregnant women's Internet use for childbirth educators. J Perinat Educ 2007;16(4):18-24 [FREE Full text] [doi: 10.1624/105812407X244903] [Medline: 18769519]

Abbreviations

CES-D: Center for Epidemiological Studies-Depression
HTML: HyperText Markup Language
MDE: major depressive episode
PPD: postpartum depression
RCT: randomized controlled trial

Edited by D Mohr; submitted 30.09.13; peer-reviewed by T Donker, B Danaher; comments to author 04.11.13; revised version received 06.12.13; accepted 07.12.13; published 09.01.14.

Please cite as:
Barrera AZ, Kelman AR, Muñoz RF
Keywords to Recruit Spanish- and English-Speaking Participants: Evidence From an Online Postpartum Depression Randomized Controlled Trial
J Med Internet Res 2014;16(1):e6
URL: http://www.jmir.org/2014/1/e6/
doi:10.2196/jmir.2999
PMID:24407163

Original Paper

Using Google AdWords for International Multilingual Recruitment to Health Research Websites

Margaret S Gross[1], BA; Nancy H Liu[1], PhD; Omar Contreras[1], BA; Ricardo F Muñoz[1,2], PhD; Yan Leykin[1], PhD

[1]University of California, San Francisco, CA, United States

[2]Palo Alto University, Palo Alto, CA, United States

Corresponding Author:
Yan Leykin, PhD
University of California
3333 California St
Suite 465
San Francisco, CA, 94143
United States
Phone: 1 415 476 8799
Fax: 1 415 476 7744
Email: yan.leykin@ucsf.edu

Abstract

Background: Google AdWords, the placement of sponsored links in Google search results, is a potent method of recruitment to Internet-based health studies and interventions. However, the performance of Google AdWords varies considerably depending on the language and the location of the target audience.

Objective: Our goal was to describe differences in AdWords performance when recruiting participants to the same study conducted in four languages and to determine whether AdWords campaigns can be optimized in order to increase recruitment while decreasing costs.

Methods: Google AdWords were used to recruit participants to the Mood Screener, a multilingual online depression screening tool available in English, Russian, Spanish, and Chinese. Two distinct recruitment periods are described: (1) "Unmanaged", a 6-month period in which ads were allowed to run using only the AdWords tool itself, with no human intervention, and (2) "Managed", a separate 7-week period during which we systematically sought to optimize our recruitment campaigns.

Results: During 6 months of unmanaged recruitment, our ads were shown over 1.3 million times, resulting in over 60,000 site visits. The average click-through rate (ratio of ads clicked to ads displayed) varied from 1.86% for Chinese ads to 8.48% for Russian ads, as did the average cost-per-click (from US $0.20 for Chinese ads to US $0.50 for English ads). Although Chinese speakers' click-through rate was lowest, their rate of consenting to participate was the highest, at 3.62%, with English speakers exhibiting the lowest consent rate (0.97%). The conversion cost (cost to recruit a consenting participant) varied from US $10.80 for Russian speakers to US $51.88 for English speakers. During the 7 weeks of "managed" recruitment, we attempted to improve AdWords' performance in regards to the consent rate and cost by systematically deleting underperforming ads and adjusting keywords. We were able to increase the number of people who consent after coming to the site by 91.8% while also decreasing per-consent cost by 23.3%.

Conclusions: Our results illustrate the need to linguistically and culturally adapt Google AdWords campaigns and to manage them carefully to ensure the most cost-effective results.

(J Med Internet Res 2014;16(1):e18) doi:10.2196/jmir.2986

KEYWORDS

Internet recruitment; multinational; online; Internet research; recruitment costs

Introduction

The rapid advance of digital media is changing the way individuals interact with and consume information. Digital media is becoming ubiquitous in many parts of the world, quickly outpacing traditional media (eg, newspapers) [1,2]. Given the rapid growth and increasing popularity of digital media, it is important for researchers interested in recruiting representative samples to remain informed about the efficient and effective ways of utilizing the instruments afforded by digital media for recruitment into research studies [3].

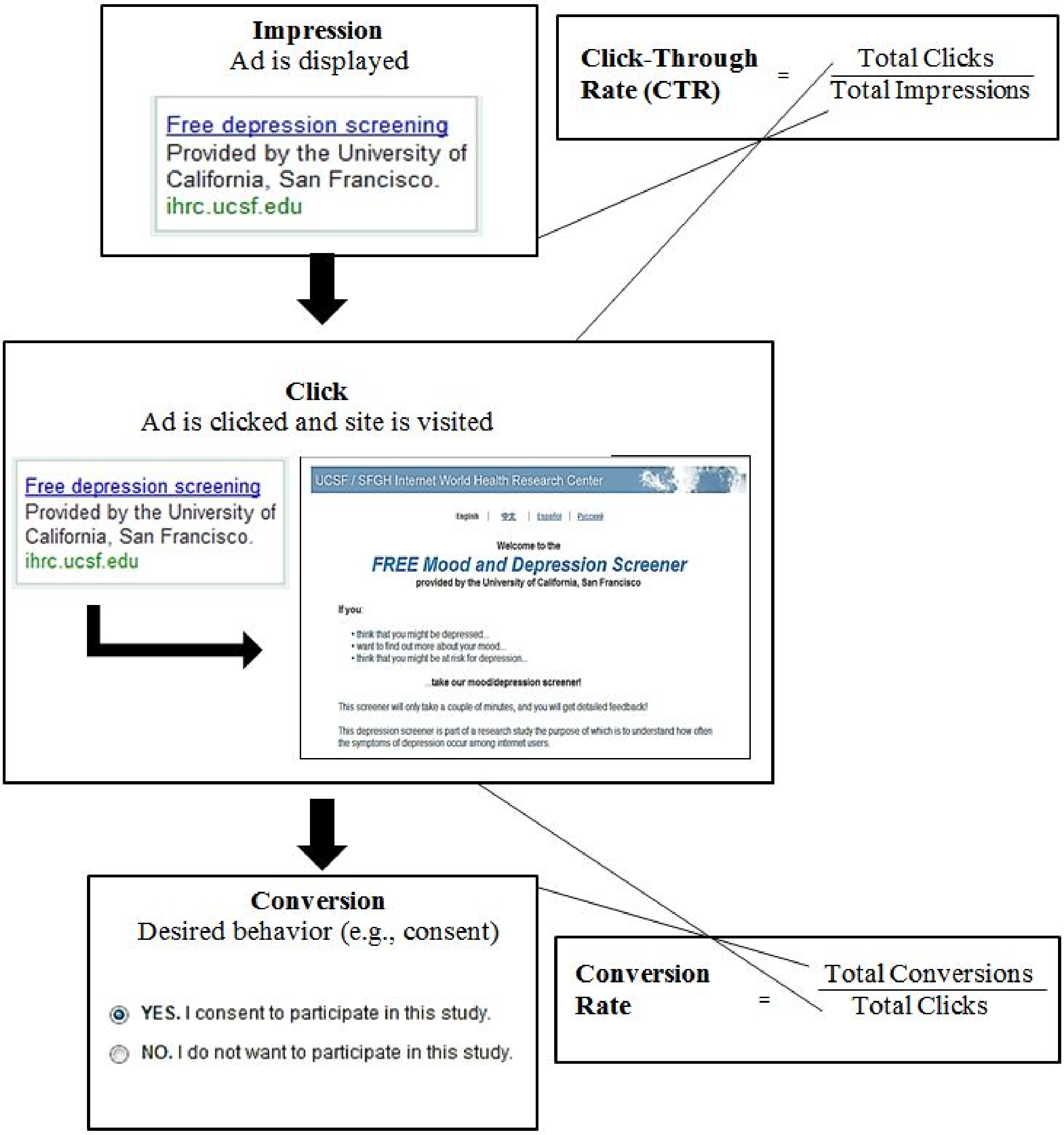
One of the chief advantages of digital media is that much of it is available for free; the costs of media production are offset by revenue that comes largely from advertisement. A company that might be considered a preeminent example of such model is Google, which provides a variety of digital media services for consumers free of charge by having a powerful and successful platform to sell advertisements that are displayed to consumers of these services. Indeed, in the past few years, Google has been consistently ranked as the world's most popular search engine [4]. Google, with over 660 million daily visitors is also one of the world's most popular websites in general [5], used for information gathering, shopping, email, navigation, cloud file storage, social networking, translation, and more. This robust "ecosystem" combined with Google's immense popularity appeals to advertisers, who use Google AdWords—Google's advertising service—to create and carry out their ad campaigns. Google's popularity can also benefit researchers aiming to recruit participants into their studies quickly and inexpensively [6]. With such a large proportion of Internet users relying on Google services, it is possible that using Google AdWords for recruitment might yield samples that are reasonably representative of Internet users [7,8].

Though created for advertising products and services, Google AdWords has a robust set of tools that can facilitate recruitment into research. These include customizations related to keywords that trigger the ads, targeting specific geographical areas, timing of ads, and the ability to conduct multiple campaigns with distinct options concurrently. Thus, researchers may be able to recruit only in neighborhoods of interest, and only among individuals already looking for specific information (eg, depression treatment). To facilitate the evaluation of performance of campaigns (eg, to determine whether a specific recruitment message appeals to the audience of interest), AdWords has a range of tools that evaluate specific ads and specific keywords. These include calculations of *impressions* (instances when an ad is displayed to the user), *clicks* (instances when user clicks on an ad), and *conversions* (instances when a desired behavior occurs, eg, a purchase, or, in case of research, a consent). AdWords automatically calculates the *click-through rate* (CTR: ratio of total clicks per total impressions), and *conversion rate* (CR: ratio of total conversions per total clicks). These terms and concepts are illustrated in Figure 1.

AdWords also has tools to monitor and manage the cost of ad campaigns. Costs are incurred whenever a displayed ad is clicked. Displaying the ad (an impression) does not cost money, however, the position of the ad (and therefore its visibility for users) is determined by the maximum amount of money an advertiser is willing to pay for the ad to be clicked, in an auction bidding system. Thus, if two ads from two advertisers could be displayed for a given keyword, the ad for which an advertiser is willing to pay more will be displayed first, increasing the likelihood of it being clicked (the actual cost of the click is US $0.01 greater than the next highest bid). For popular keywords like "depression", research studies need to compete with advertisers who are willing to expend considerable funds for clicks (eg, pharmaceutical companies). Thus, to make the most of the limited budget of research studies as well as other ad campaigns, AdWords can suggest keywords that are likely to generate interest (ie, clicks and conversions) and will automatically preferentially display the ads to improve click counts or CRs (depending on the goal of the campaigns), based on data from past ad performance and Google's own algorithms. However, AdWords recruitment has several drawbacks—chief among which are the strict limit on the number of characters and the fact that costs are calculated based on clicks (visits) rather than actions (eg, consent). Taken together, these drawbacks can add up to a costly campaign for studies that are difficult to describe concisely, where visitors may feel misled by the advertisement.

This paper describes our experience conducting a multilingual worldwide campaign to recruit participants to a website that screens for symptoms of depression, in four languages. We will present the results from two distinct time periods of recruitment. First, a 6-month "Unmanaged" recruitment period, in which AdWords campaigns were allowed to run only with automated optimizations offered by the AdWords tool, with no human intervention. For this period, we will describe similarities and differences in campaign performance between the four languages of recruitment (English, Spanish, Russian, and Chinese). During the second "Managed" period (a separate 7-week period), we systematically sought to optimize our recruitment campaigns beyond the optimization provided by automated AdWords tools, with the goal of increasing our conversion rate (ie, consent rate) while decreasing our costs.

Figure 1. Google AdWords terms.



Welcome to the

FREE Mood and Depression Screener

provided by the University of California, San Francisco

If you:

- think that you might be depressed...
- want to find out more about your mood...
- think that you might be at risk for depression...

...take our mood/depression screener!

This screener will only take a couple of minutes, and you will get detailed feedback!

This depression screener is part of a research study the purpose of which is to understand how often the symptoms of depression occur among internet users.

● YES. I consent to participate in this study.

○ NO. I do not want to participate in this study.

Methods

Participants

Eligible participants were 18 years of age; no other condition was necessary for eligibility. Participants were speakers of either English, Spanish, Chinese, or Russian. Two distinct waves of recruitment are described: a 6-month “Unmanaged” period and a 7-week “Managed” period. During the “Unmanaged” recruitment period (no management of ads), over 60,000 people visited the Mood Screener website, 30,218 were screened for eligibility, and 26,194 individuals were eligible to participate. Among eligible participants, 3828 were English speakers, 7477 were Spanish speakers, 5395 were Chinese speakers, and 9494 were Russian speakers. Eligible participants were on average 30.9 years of age (SD 11.2), 68% (17,811) were women, and 57% (14,930) identified as white or Caucasian.

During the “Managed” recruitment period (active management of ads), over 19,000 people visited the Mood Screener website, 7786 were screened for eligibility, and 6564 individuals were eligible to participate. Among eligible participants, 1282 were English speakers, 2350 were Spanish speakers, 589 were Chinese speakers, and 2343 were Russian speakers. Eligible participants were on average 31.3 years of age (SD 12.0), 66% (4332) were women, and 47% (3085) identified as white or Caucasian.

Materials

The Demographics - 1 questionnaire asked about participants’ age, gender, race, and country of residence. The Demographics - 2 questionnaire asked about participants’ country of birth (if different from country of residence), whether they lived alone or with others, as well as their education, employment status, marital status, income, subjective social status (perceived position in society, ie, the “social ladder” [7]), and history of seeking treatment for depression.

The MDE Screener [9] is an 18-item self-report measure based on Diagnostic Interview Schedule items [10] designed to screen for the presence of current and past major depressive episodes (MDEs). It rates the presence of nine symptoms of depression according to the DSM-IV [11] over a period of 2 weeks or more and assesses whether Criterion C (significant impairment in functioning) is met within the same time span. Participants are asked about the presence of symptoms over the past 2 weeks (to identify a “Current” MDE), as well as symptoms occurring in any 2-week period in their lifetime, excluding the past 2 weeks (to identify “Lifetime” MDE). The screener has been shown to have good agreement with the PRIME-MD [12,13] and with clinician-administered diagnostic interviews [14].

Overall Study Procedures

Study procedures have been described in detail elsewhere [15]. Briefly, worldwide Google AdWords campaigns were used as the main recruitment source. Recruitment was made possible by a Google grant awarded to our research group, which permitted us to run recruitment campaigns for our investigations, with a maximum bid of US $1 per click. This US $1 limit made our ads less competitive in markets with significant AdWords saturation (eg, Western English-speaking world), but reasonably competitive in other parts of the world.

Our ads were triggered by individuals entering terms related to depression or sad mood into the Google search engine; the ads were keyed to the language of the keyword. Google AdWords limits their ads to 25 characters for the title line and 35 characters for each of two body lines (for Chinese-language ads, the limits are 12 and 17 characters for title and body, respectively, as Chinese characters are considered “double width”). Clicking on the ad forwards the user to the first (“landing”) page of the site in the appropriate language; this page contained information on the nature of the site as well as limits to confidentiality. Those interested in continuing advanced to the Demographics – 1 questionnaire (age, race, gender, and country of residence). Eligible participants (those 18+ years of age) then advanced to the MDE Screener, which was followed by the “honesty question” that asked whether participants’ responses are “accurate” or whether participants were “testing the site”. Participants completing “Current” MDE Screener with no more than 2 missing answers were given personalized feedback (brief explanation of the symptom level as well as additional feedback for those indicating suicidality). Participants were then invited to take part in a monthly rescreening study. Interested individuals read the consent document and signed it with their email address. They then completed the Demographics – 2 questionnaire and the “Lifetime” MDE screener, followed by personalized feedback. Starting in February 2013, consenting participants indicating current or past suicidality also completed a detailed suicidal behaviors assessment. Consenting participants were then contacted via email monthly with a link to return to the site to rescreen their mood.

AdWords Management Procedures

This investigation focuses on two distinct recruitment periods. The first, “Unmanaged”, was the 6-month period from November 15, 2011, to May 15, 2012, when no active ad management was taking place, although automatic Google-based ad management was still in effect (via enabling the setting for showing ads that are expected to maximize conversions). The study was then paused while we transitioned to a different software vendor and was restarted in February 2013. We restarted the ads that were in place during our previous recruitment period and started managing them actively shortly thereafter. Thus, the second period, “Managed”, was the 7 weeks from March 21 to May 8, 2013, when we managed ads actively, focusing on increasing conversion (consent) rates. In order to do this, we systematically examined the performance of each ad and disabled the ones that (1) had produced no or very few conversions, because these ads were unlikely to produce future conversions or (2) had high CTRs, but low CRs, because these ads appeared to be attracting those unlikely to consent.

In both cases, our goal was to minimize the opportunity cost of presenting ads that were unlikely to result in a conversion and retain the ads that were most productive in terms of conversions. For each language, we monitored the campaign weekly performance and suspended (paused) the 2-4 worst performing ads (according to the above criteria) each week.

To further optimize performance, we also managed the keywords during this 7-week period. Similarly to ad management, we focused on the performance of each individual keyword and disabled the ones that produced no or very few conversions or had high click counts, but low CRs. In addition to deleting poorly performing keywords, we continuously added keywords including misspelled variations of successful keywords (eg, “depressd” and “depresed”). This was useful to our campaign due to the high frequency of searches containing misspelled words or phrases combined with much less competition among other Google AdWords campaigns.

This study was approved by the University of California, San Francisco Institutional Review Board (Committee for Human Research).

Results

Overview

For the purpose of this study, ads that had the highest numbers of conversions were considered to be the best performing ads. Conversely, if an ad was being clicked but was not generating conversions, the ad was considered low-performing.

Unmanaged Period

The summary of the ad campaign statistics are presented in Table 1. Chinese ads had by far the highest total impressions (602,634, with other languages ranging from 241,507 to 301,493), which is perhaps indicative of both the number of Chinese speakers on the Internet and the demand for information on depression among the Chinese-speaking Internet community. Despite the high number of impressions, Chinese ads generated relatively few clicks (11,209), resulting in the lowest CTR

(1.86%). However, the consent rate for Chinese speakers was the highest (3.62%, with other languages ranging from 0.97% to 2.46%), resulting in the second highest number of consenting participants (406).

Although the Russian campaign had the second fewest impressions (227,391), it had by far the highest number of clicks (23,514, with other languages ranging from 11,022 to 14,459), resulting in the highest CTR of all languages (8.48%, with other languages ranging from 1.86% to 5.99%). Russian cost-per-click was likewise the lowest (US $0.20, with other languages ranging from US $0.33 to US $0.50), suggesting very low competition for ads for this language. Surprisingly, although this campaign had the highest CTR, it also had the second lowest consent rate (1.87%), which nonetheless (due to the high number of initial clicks) produced 440 consented participants.

The English campaign was mostly notable for its highest costs, both per click (US $0.50) and especially per consent (US $51.88), which was almost 5 times the costs for the other languages, which ranged from US $10.80 to US $13.40. The rate of consent among English speakers was the lowest (0.97%), resulting in a far lower number of consenting participants than other languages (107, whereas other languages ranged from 355 to 440).

The Spanish campaign resulted in 355 consented participants with the second highest consent rate (2.46%). This campaign had 241,507 impressions, 14,459 clicks, and the second highest CTRs (5.99%). The cost-per-consent for the Spanish ads was US $13.40.

We also noted both similarities and differences in the keywords (ie, search terms that trigger ads) that were the most effective in generating the most conversions (see Table 2). The most successful keyword in Russian, Chinese, and Spanish was “depression”, while the English campaign’s use of “depression” produced very few conversions; “depression test” was the most effective keyword. This is likely due to the prevalence of competing English ad campaigns that also use the keyword “depression”. Indeed, “depression test” or its variations (eg, “test de depresión”) was one of the 4 most productive keywords in all languages except Chinese. Chinese campaigns were unique in that the characters for “suicide” and “commit suicide” (Table 2) were 2 of the 4 most productive keywords, whereas no other language had “suicide” as one of the top 4 keywords (Russian keyword “не хочется жить”, “do not want to live”, is perhaps an exception). Finally, Spanish was the only language where “anxiety” (“la ansiedad”) appeared as a top keyword, and Russian was the only language where non–mood-related keyword “психологические тесты” (“psychological tests”) was a top keyword. A list of worst performing keywords is included in Table 3.

Finally, for each language there was a particular ad that significantly outperformed other ads with respect to conversions (see Table 4). The English, Russian, and Spanish campaigns’ most successful ads all contained the word “free” when describing the study. All four languages’ most successful ads contained the name “University of California, San Francisco”, or its abbreviation, and were direct in describing the purpose of the study. Ads that contained less formal descriptions (eg, “Can’t eat, sleep, focus?”) were less successful (Table 5).

Table 1. Statistics from the unmanaged 6-month period.

	English	Russian	Chinese	Spanish	Overall
Impressions, n	301,493	277,391	602,634	241,507	1,373,025
Clicks (visitors), n	11,022	23,514	11,209	14,459	69,204
Click-through rate, %	3.66	8.48	1.86	5.99	4.38
Cost-per-click, US$	0.50	0.20	0.42	0.33	0.33
Consented, n	107	440	406	355	1,308
Consent rate, %	0.97	1.87	3.62	2.46	2.17
Cost per consent, US$	51.88	10.80	11.71	13.40	15.15

Table 2. Best performing keywords.

Language	Best performing keywords	English translation
English	depression test	
	sad	
	symptoms of depression	
	Am I depressed	
Russian	депрессия	depression
	психологические тесты	psychological tests
	депрессия тест	depression test
	не хочеться жить	do not want to live
Chinese		depression
		depression
		suicide
		wanted to commit suicide
Spanish	Depresión	depression
	Test de depresión	depression test
	La ansiedad	anxiety
	Síntomas de depresión	symptoms of depression

Table 3. Worst performing keywords.

Language	Worst performing keywords	English translation
English	major depressive disorder symptoms	
	online depression test	
	depression screening	
	Take a depression test	
Russian	Имею ли я депрессию	Do I have depression
	дипрессия симптомы	depression[1] symptoms
	Depressija	depression[2]
	депрессия тест бесплатно	depression test free
Chinese		depression screening questionnaire
		depressed
		no hope
		excessive pressure
Spanish	Inútil y triste	useless and sad
	Prueba gratuita de depresión	free depression trial
	Deprezion	depression[1]
	Estado de ánimo triste	sad mood

[1] Deliberate misspellings in the original language.

[2] Transliteration.

Table 4. Best performing ads

Language	Best performing ads	English translations
English	Free depression test from the University of California, San Francisco	
Russian	Тест на депрессию. Узнайте есть ли у Вас депрессия. Бесплатный тест настроения из УКСФ	Test for depression. Find out if you have depression. Free mood test from UCSF.
Chinese		Do you think you might have depression? Find out – five minute survey provided by UCSF.
Spanish	Test de depresión gratis proporcionado por la Universidad de California, San Francisco.	Free depression test provided by the University of California, San Francisco.

Table 5. Worst performing ads.

Language	Worst performing ads	English translations
English	Can't eat, sleep, focus? Find out if you are depressed with a free online test from UCSF.	
Russian	Проверка настроения. Узнайте есть ли у Вас депрессия. Бесплатный тест настроения из УКСФ.	Mood checker. Find out if you have depression. Free mood test from UCSF.
Chinese		Feeling sad and hopeless? Check whether your symptoms are related to depression. Free depression questionnaire.
Spanish	Examine su ánimo. Complete gratis una evaluación de la Universidad de California, SF.	Examine your mood. Complete a free evaluation from the University of California, SF.

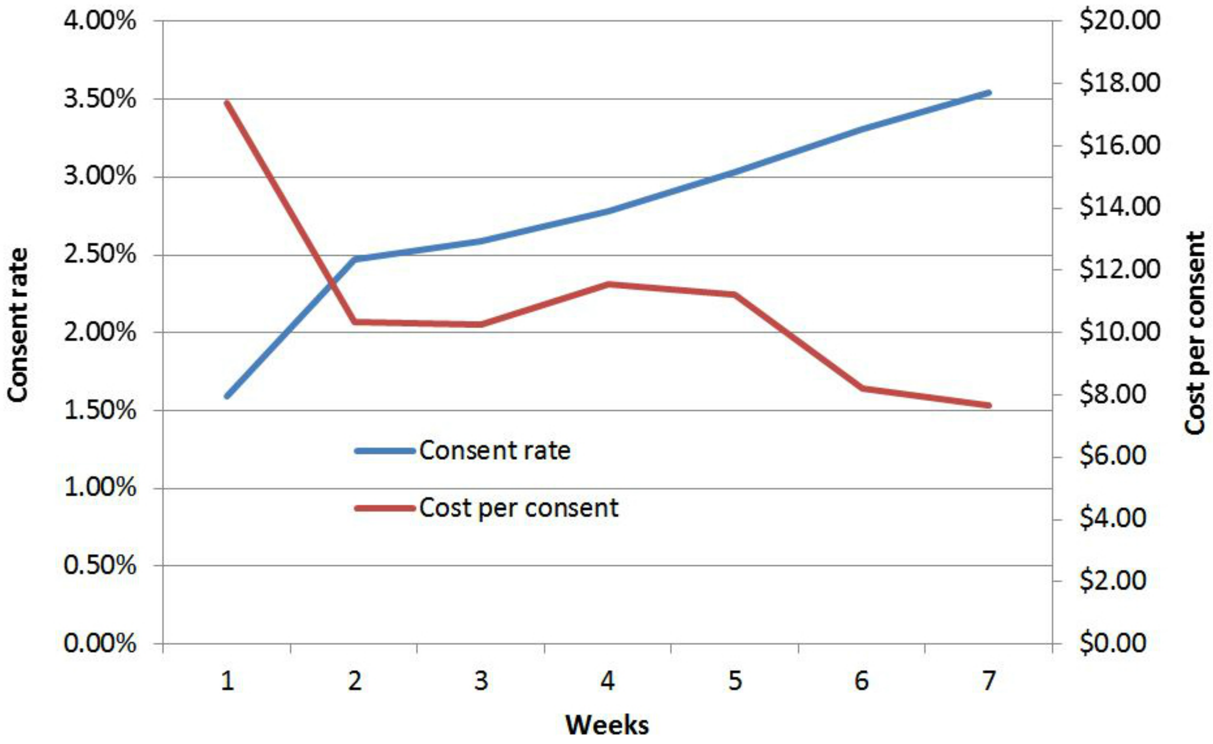
Managed Period

During the 7-week managed period, we monitored the campaigns closely, suspending underperforming ads and keywords and adding new keywords, with the goal of improving consent rate and reducing per-consent cost. Table 6 presents details on the first 2 weeks of the managed period and the last 2 weeks, to illustrate the changes. The number of impressions increased by 38.10%, clicks increased by 34.82%, conversions increased by 91.84% and the CR increased by 42.11%. The cost-per-click and cost-per-consent decreased by 16.13% and 23.31% respectively. The weekly changes in the two metrics most relevant to researchers planning to conduct AdWords campaigns—consent rate and per-consent costs—are illustrated in Figure 2.

Table 6. Changes in ad campaign performance during the managed period.

	First 2 weeks	Last 2 weeks	Change, %
Impressions, n	331,273	457,506	+38.10
Clicks (visitors), n	8,812	11,881	+34.82
Click-through rate, %	2.66	2.60	–2.26
Cost-per-click, $US	0.62	0.52	–16.13
Consented, n	184	353	+91.84
Consent rate, %	2.09	2.97	+42.11
Cost per consent, $US	18.23	13.98	–23.31

Figure 2. Weekly data on consent rate and per-consent cost during the managed period.



Discussion

Principal Findings

The purpose of this study was to describe worldwide Google AdWords campaigns recruiting participants to a health research website consisting of a depression screening study in four languages, to illustrate similarities and differences among the four languages and to demonstrate that active management of AdWords campaigns can substantially increase the recruitment rates while decreasing costs.

Our results offer a compelling illustration of the competitive nature of Google AdWords campaigns. Given the auction system that drives AdWords, popular markets such as the English-speaking market require greater expenditures to be competitive. The low rate of consent along with very high costs likely reflect the high competition for English-language depression-related keywords as well as perhaps the perception among English speakers that multiple options for information and monitoring are available to them. Not only are there more advertisers vying for the top spot in English-language ads, thereby increasing the price, but the market itself is saturated with ads and products that cater to English speakers, which results in users being both more selective and more accustomed to ignoring advertisements. These pressures likely resulted in English-language ads having the highest cost-per-click and by far the highest per-consent cost, which was about five times higher than any of the other three languages. Thus, a researcher wishing to conduct multilingual campaigns must consider not only the relative popularity (and therefore cost) of the keywords they plan to use to trigger the ads, but also the languages of the campaign itself, and to allocate funds appropriately to ensure equal recruitment.

Another difference that appeared to emerge in our data is the timing of individuals' commitment, as evidenced chiefly by differences in the CTRs and CRs between Chinese and Russian ads. Whereas Russian speakers were quite willing to click on ads—their CTR was the highest of all languages—they appeared reluctant to commit to participate, exhibiting a very low consent rate. This suggests that, even though the initial interest in the site among Russian speakers was high, it applied only to the initial screening, and not to any further involvement. Chinese speakers, on the other hand, may be reluctant to visit an unknown site based on a simple ad, but once they visit, they are more likely to continue participating in the study. The results suggest that Chinese-speaking individuals looking for information on depression are quite selective as to which ad they respond to, but once they make a selection they are more likely to maintain their commitment. This suggests that both the sites and the ads may need to be structured differently depending on the language of the intended audience. For instance, a researcher might choose to design an especially attractive and thorough ad for Chinese speakers or might put more effort into the design of the recruitment message for Russian speakers.

There were notable similarities between the languages in regards to ads and keywords that were the most productive. For instance, ads briefly describing the purpose of the study (eg, depression screener) and using the word "free" resulted in highest CTRs and CRs. The ads that listed specific symptoms or ads that promoted self-improvement did not generate sufficient participant flow. This suggests that researchers wishing to simplify their ad campaigns should consider using simple, descriptive ads that are likely to be universally acceptable. Future studies might examine whether individuals with specific characteristics might be more likely to respond to a specific type of an ad. Keywords likewise revealed similarities, with all languages having "depression" or "depression test" as one of the most productive keywords. This observation was expected: individuals are more likely to click on an ad that conforms to the terms of their search. There were some notable departures from this pattern. First, keywords related to suicide generated considerable participant flow among Chinese-speakers (and among Russian-speakers as well). The fact that Chinese speakers who searched for suicide-related terms clicked on an ad for a depression-related site might suggest that these individuals are particularly troubled by suicidal symptoms [16]. Alternatively, this might suggest the paucity of resources available online for Chinese and Russian speakers (for instance, searching for the word "depression" in English produces 62 million results, compared to 15 million results for the Chinese term for depression and less than 3 million for Russian). The same may be true for Spanish speakers (only 1.5 million results for the Spanish word for depression), for whom "la ansiedad" (anxiety) was one of the top keywords that led individuals to click on an ad for a depression screening study.

Importantly, we found that that by actively managing AdWords campaigns, it is possible to both increase the rate of recruitment and decrease costs. Our strategy targeted both ads and keywords. We removed under-performing keywords in an attempt to weed out individuals who may be less likely to participate (based on data from previous visitors using the same keywords to enter the study). We also removed under-performing ads to ensure that individuals would see ads that were more likely to bring them into the study. These efforts resulted in higher traffic and lower per-person costs. In the era of shrinking research budgets, being able to significantly increase the consent rate while decreasing per-consent cost should be a relief for researchers attempting to do large-scale recruitment with Google AdWords. Although Google AdWords has built-in tools to improve ad performance, researchers who wish to maximize the efficiency of their ad campaigns are advised to actively manage and optimize their ad campaigns. No person-hours were spent during the 6-month "Unmanaged" campaign. During the 7-week managed period, a research assistant spent approximately 1.5-2 hours per week examining the campaigns, pausing keywords and ads, and adding keywords. Slightly more time was spent during the beginning weeks of the managed period to make adjustments to the campaigns, and time commitments lessened somewhat during the final weeks. The personnel cost, however, was offset by the greater efficiency and effectiveness of the recruitment campaigns, and we believe that this investment in person-hours is well justified.

Limitations

This study has several limitations. The study describes two distinct recruitment periods that took place 1 year apart, with the managed period being conducted in the spring, while the unmanaged period was conducted over winter and the beginning of spring. It is possible that seasonal differences could have affected the flow and the type of participants searching depression-related terms. The numbers from these periods are not directly comparable given the rapid changes in Internet and its users and in Internet advertisement. To illustrate changes in ad performance, we selected a short management period (7 weeks) to minimize the historical influence of changes in Internet and Internet advertisement. We report changes in ad performance in a "pre-post" design; conducting two campaigns simultaneously, one managed and one unmanaged, would have been a more powerful illustration. However, such a design may introduce considerable undesirable confounds as ads in such two campaigns would be competing against each other. For Chinese and Russian speakers, Google is not the dominant search engine (Baidu and Yandex have larger market shares in China and Russia, respectively), thus our data may not be fully representative of Chinese- and Russian-speaking Internet users. Given the evolving nature of Internet users, the specifics of our results may not reflect future Internet users; however, our conclusions regarding the need to monitor and carefully manage ad campaigns to optimize their performance will most likely remain applicable well into the future.

Conclusions

The Internet made it possible to conduct research in the global community, on populations that would otherwise not be represented in research samples, and to do so quickly, efficiently, and inexpensively. Researchers who wish to take advantage of such tremendous research opportunities can make use of sophisticated tools specifically designed to reach individuals on the Internet and deliver targeted recruitment messages. The enormous diversity of global populations calls for a similar diversity of outreach, both to make research opportunities appealing to as many people as possible, as well as to increase the efficiency of such endeavors. This study demonstrates the potential of a tool such as Google AdWords for recruitment of research participants, illustrates the differences in effectiveness of recruitment campaigns between four different languages, and emphasizes the value of actively managing recruitment. Understanding both differences and similarities of outreach approaches will help advance research practices and, consequently, the research itself.

Acknowledgments

This work was supported by: Robert Wood Johnson Health and Society Scholars Seed Grant (Leykin, principal investigator), an infrastructure grant from the University of California Committee on Latino Research to the University of California/San Francisco General Hospital Latino Mental Health Research Program (Muñoz, principal investigator), NIMH grant 5K08MH091501 (Leykin, principal investigator), and Grant SRG-0-059-11 (Leykin, principal investigator) from the American Foundation for Suicide Prevention (the content is solely the responsibility of the authors and does not necessarily represent the official views of the

American Foundation for Suicide Prevention). We would also like to express our thanks to Google, Inc, for awarding us an AdWords grant (Muñoz, principal investigator), and the Brin-Wojcicki Foundation for their generous support.

Conflicts of Interest

None declared.

References

1. The CMO Survey. 2013. February 2013 Survey Results URL: http://www.cmosurvey.org [accessed 2013-12-05] [WebCite Cache ID 6LeDLEy29]
2. Global Web Index. 2013. Global Media Consumption: The Digital Reality URL: http://blog.globalwebindex.net/online-time-now-exceeds-offline-media-consumption-globally/ [accessed 2013-12-05] [WebCite Cache ID 6LeDyHLKD]
3. McDonnell DD, Lee H, Kazinets G, Moskowitz JM. Online Recruitment of Targeted Populations: Lessons Learned From a Smoking Cessation Study Among Korean Americans. Social Marketing Quarterly 2010 Aug 26;16(3):2-22. [doi: 10.1080/15245004.2010.500441]
4. comScore Releases July 2013 U.S. Search Engine Rankings. 2013. URL: http://www.comscore.com/Insights/Press_Releases/2013/8/comScore_Releases_July_2013_U.S._Search_Engine_Rankings [accessed 2013-09-17] [WebCite Cache ID 6JiAav83f]
5. The top 500 sites on the web. URL: http://www.alexa.com/topsites [accessed 2013-09-17] [WebCite Cache ID 6JiAiUNNS]
6. Jones RB, Goldsmith L, Hewson P, Williams CJ. Recruitment to online therapies for depression: pilot cluster randomized controlled trial. J Med Internet Res 2013;15(3):e45 [FREE Full text] [doi: 10.2196/jmir.2367] [Medline: 23462072]
7. Adler NE, Epel ES, Castellazzo G, Ickovics JR. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. Health Psychol 2000 Nov;19(6):586-592. [Medline: 11129362]
8. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476. [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]
9. Muñoz RF. The Major Depression Episode (MDE) Screener. 2008. URL: http://www.medschool2.ucsf.edu/latino/manuals.aspx [accessed 2014-01-03] [WebCite Cache ID 6MMc7tIge]
10. Robins LN, Cottler L, Bucholz K, Compton W. Diagnostic interview schedule for DSM-IV. St Louis, MO: Washington University School of Medicine; 1995.
11. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edition ed. Washington DC: Author; 1994.
12. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994 Dec 14;272(22):1749-1756. [Medline: 7966923]
13. Muñoz RF, McQuaid JR, González GM, Dimas J, Rosales VA. Depression screening in a women's clinic: using automated Spanish- and English-language voice recognition. J Consult Clin Psychol 1999 Aug;67(4):502-510. [Medline: 10450620]
14. Vázquez FL, Muñoz RF, Blanco V, López M. Validation of Muñoz's Mood Screener in a Nonclinical Spanish Population. European Journal of Psychological Assessment 2008 Jan 21;24(1):57-64. [doi: 10.1027/1015-5759.24.1.57]
15. Leykin Y, Muñoz RF, Contreras O. Are consumers of Internet health information "cyberchondriacs"? Characteristics of 24,965 users of a depression screening site. Depress Anxiety 2012 Jan;29(1):71-77 [FREE Full text] [doi: 10.1002/da.20848] [Medline: 21681872]
16. Liu NH, Contreras O, Muñoz RF, Leykin Y. Suicide attempts and depression among Chinese speakers over the internet. 2013 Presented at: International Society for Research on Internet Interventions; 2013; Chicago, Il.

Abbreviations

CR: conversion rate
CTR: click-through rate
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edition
MDE: major depressive episode
PRIME-MD: Primary Care Evaluation of Mental Disorders

Edited by P Carlbring; submitted 26.09.13; peer-reviewed by M Jones, T Aalbers; comments to author 05.11.13; revised version received 11.12.13; accepted 23.12.13; published 20.01.14.

Please cite as:
Gross MS, Liu NH, Contreras O, Muñoz RF, Leykin Y
Using Google AdWords for International Multilingual Recruitment to Health Research Websites
J Med Internet Res 2014;16(1):e18
URL: http://www.jmir.org/2014/1/e18/
doi:10.2196/jmir.2986
PMID:24446166

Original Paper

Understanding the Usage of Content in a Mental Health Intervention for Depression: An Analysis of Log Data

Julia EWC Van Gemert-Pijnen[1*], PhD; Saskia M Kelders[1*], PhD; Ernst T Bohlmeijer[1*], PhD

University of Twente, Department of Psychology, Health and Technology, Enschede, Netherlands
[*]all authors contributed equally

Corresponding Author:
Julia EWC Van Gemert-Pijnen, PhD
University of Twente
Department of Psychology, Health and Technology
Drienerlolaan 5
Enschede, 7522 NB
Netherlands
Phone: 31 534896050
Fax: 31 534892895
Email: j.vangemert-pijnen@utwente.nl

Abstract

Background: Web-based interventions for the early treatment of depressive symptoms can be considered effective in reducing mental complaints. However, there is a limited understanding of which elements in an intervention contribute to effectiveness. For efficiency and effectiveness of interventions, insight is needed into the use of content and persuasive features.

Objective: The aims of this study were (1) to illustrate how log data can be used to understand the uptake of the content of a Web-based intervention that is based on the acceptance and commitment therapy (ACT) and (2) to discover how log data can be of value for improving the incorporation of content in Web-based interventions.

Methods: Data from 206 participants (out of the 239) who started the first nine lessons of the Web-based intervention, Living to the Full, were used for a secondary analysis of a subset of the log data of the parent study about adherence to the intervention. The log files used in this study were per lesson: login, start mindfulness, download mindfulness, view success story, view feedback message, start multimedia, turn on text-message coach, turn off text-message coach, and view text message. Differences in usage between lessons were explored with repeated measures ANOVAs (analysis of variance). Differences between groups were explored with one-way ANOVAs. To explore the possible predictive value of the login per lesson quartiles on the outcome measures, four linear regressions were used with login quartiles as predictor and with the outcome measures (Center for Epidemiologic Studies—Depression [CES-D] and the Hospital Anxiety and Depression Scale—Anxiety [HADS-A] on post-intervention and follow-up) as dependent variables.

Results: A significant decrease in logins and in the use of content and persuasive features over time was observed. The usage of features varied significantly during the treatment process. The usage of persuasive features increased during the third part of the ACT (commitment to value-based living), which might indicate that at that stage motivational support was relevant. Higher logins over time (9 weeks) corresponded with a higher usage of features (in most cases significant); when predicting depressive symptoms at post-intervention, the linear regression yielded a significant model with login quartile as a significant predictor (explained variance is 2.7%).

Conclusions: A better integration of content and persuasive features in the design of the intervention and a better intra-usability of features within the system are needed to identify which combination of features works best for whom. Pattern recognition can be used to tailor the intervention based on usage patterns from the earlier lessons and to support the uptake of content essential for therapy. An adaptable interface for a modular composition of therapy features supposes a dynamic approach for Web-based treatment; not a predefined path for all, but a flexible way to go through all features that have to be used.

(J Med Internet Res 2014;16(1):e27) doi:10.2196/jmir.2991

KEYWORDS

mental health; depression; Web-based intervention; acceptance commitment therapy

Introduction

Web-based interventions for mental health treatment are promising. Clients prefer these treatments to face-to-face treatments alone, and Web-based interventions are successful in realizing the objectives of treatment [1-4]. Such interventions can be considered effective in access to and reduction of mental complaints comparable to traditional face-to-face therapy [2], although the reach of interventions is quite biased; most users are female and highly educated. To understand the effects of Web-based interventions, it is important to know which elements of an intervention contribute to effectiveness [2]. Many studies focus on whether and which user characteristics explain or predict the effects of an intervention, thereby overlooking the characteristics of the Web-based intervention that can influence the effects. From prior studies we know that interactivity (eg, support from a counselor [5]), dose-response, frequent updates of content, and persuasive technology (eg, reminders) increase adherence to a Web-based intervention [6].

Adherence, the usage of a system as intended, appears to be a mediator for realizing treatment objectives. In this connection, log data are important to get objective and real-time information about the usage of the intervention, how users “walk through the intervention” when they log in, and what actions they carry out during the login period. Log data as such can provide a broad and in-depth insight into adherence or (non)adherence during a treatment process. The number of logins, the number of actions per participant, and the proportion of completed modules (lessons, actions) and time spent in the treatment course have been used as metrics to identify the attrition curve and the differences between user groups, for example adherers, nonadherers, or lower active users [7]. A higher number of completions of activities or modules per login led to a greater benefit of the treatment program than little login activity or medium activity per login [7]. This can imply that support for login can improve outcomes.

Recent studies have shown that adherence to a Web-based intervention can be predicted on the usage patterns of an intervention [8,9]. For example, the number of logins in the first week of a treatment process or the number of active days in the first week can be used as predictors for adherence to the interventions. Understanding how users interact with technology is important to increase the ease of using a Web-based intervention and to avoid dropouts. In particular, log data from the early usage of the intervention can provide valuable prompts for employment of persuasive triggers to remind low-active users to login. For example, log data from a lifestyle (weight loss) Web-based intervention that combines social networking tools (blogs, discussion forum) with an online dietary intervention program [8] showed that high usage of the intervention during the first weeks (early interaction) was highly correlated with retention of the user to the intervention. Usage was a stronger predictor for a return than behavior intentions or demographics. The usage of persuasive features, such as social support, influenced the return and adherence to the intervention.

To improve the ease of use and persuasiveness of the Web-based interventions, insight is needed into not only the number of logins or proportion of completed modules but also the usage of the features of the intervention over time. The study [9] used log data to understand whether high and low adherers differ in the usage of the features of the intervention. The log files contained data about actions taken by each participant: user ID, action type (for example “logged in”), action specification (to know how they “walk” through the intervention, eg, lesson 1, screen 11), and which elements were “visited” (clicks on, eg, feedback), the time, and day. Based on that data, the number of times each participant performed an action in the Web-based intervention was extracted. For logins, this meant that not only the total number of logins per participant during the intervention period was extracted, but furthermore, the number of logins per participant per lesson of the intervention. From this data it was possible to observe, for each participant, which elements in the intervention (lessons) had been reached and which specific actions were undertaken (viewing feedback for example) at what moment. The emergent patterns in usage (low/high visited elements) and differences between users (low/high adherence) can be of value to know the critical moments for employment of persuasive features to support motivation, to explain the risks of dropout, and to stimulate users in a positive way to achieve their goals. The differences in user patterns can be used to motivate low or medium adherers to log in and to remind them to use the content that is crucial for progress in treatment.

To optimize a Web-based intervention, log data can be used to predict what features and what combination of features correlate with adherence and effects. From a previous study [6], we know that Web-based interventions consist of different content features such as education, information, exercises, monitoring, in combination with a variety of persuasive features, such as SMS text messaging (short message service, SMS) and reminders, to support users during the treatment process. In most cases, the content in Web-based interventions is based on therapies to change behaviors or lifestyles, similar to the traditional face-to-face behavior change interventions [2]. Research on the format or design of the content of Web-based interventions is scarce; most research is done to compare Web-based interventions with traditional interventions in terms of perceived helpfulness, satisfaction or reached goals, and effect sizes (meta-analysis) [2]. The content element for a Web-based intervention is often applied straight to a Web-based environment, usually organized in accordance with the traditional face-to-face approach, providing a scheme and schedule for doing homework, exercises, and for providing feedback. The employment of persuasive features (reminders, etc) seems rather ad hoc and intuitive [6].

Log data analysis can provide rich and worthwhile knowledge about how an intervention works in practice and which elements in an intervention (content and system) should be improved in such a way that participants can benefit more from the intervention. In particular, it is important to know how the content of a Web-based therapy should be presented to encourage participants to follow the treatment path. In particular, for content-driven interventions delivered via the Internet, the use of persuasive features, such as reminders, can be important to motivate participants to take up the education, information, and exercise modules from the therapy. This requires that log data are incorporated appropriately in the intervention in order to measure the performance in practice. To do so, a log protocol can be developed to gather information about users (login ID, informed consent), action types (login/logout, start lesson, start exercise, download exercise, view feedback message, etc), action specification (eg, name of text message that was viewed, visited page of system), time, and day [9-11].

The results of log data can be used to optimize the usability of the system and the persuasiveness of the content of the intervention. For example, when certain features are not used as expected, the log files and usage patterns can be considered for investigating problems related to ease of use, for example, features that are not well incorporated in the system and as such could be overlooked. Reminders, for example, might not be well connected to the content, which could result in no actions. As such, log data can provide the prompts for a more dynamic and flexible presentation of content for a Web-based therapy than a “self-help-book on the Internet”. Therefore, the aims of this study were (1) to illustrate how log data can be used to understand a Web-based intervention that is based on acceptance and commitment therapy (ACT) and (2) how log data can be of value for improving the incorporation of content in Web-based interventions.

This study was built on the findings of a prior study [9] using log data to identify differences in usage activities of adherers and nonadherers. The study found that the predictive value of characteristics of participants on adherence is very small and that for realizing effects the uptake of all features of the intervention is important. The aforementioned studies [7,9] found that high active users can benefit more from the intervention, but that more insight is needed in how use of a program of therapy impacts outcomes. Therefore, the focus of this study was on the usage of elements or features of a therapy-based program, that is, ACT, which is a form of cognitive behavioral therapy (CBT) rapidly being implemented in mental health care. Although the content of some exercises and education in ACT may be slightly different from CBT, the generic approach of the intervention is similar. As in CBT in general, we expect that many new Web-based interventions based on ACT will be built.

Methods

Parent Study and Participants

In this study, we carried out a secondary analysis of a subset of the log data of a Web-based treatment aimed at reducing depression from the Web-based “Living to the Full” intervention.

The analyses described in this paper have been performed on a subset of data collected in the parent study on the adherence of the Web-based intervention for the prevention of depression. Participants were adults with mild to moderate depressive symptoms, that is, >9 and <39 on the Center of Epidemiologic Studies—depression scale (CES-D) [12], who completed our online screening procedure. Participants were recruited from the general public (advertisement) and were selected to avoid users who are not a target of therapy. For the current study, data from participants who started the first of nine lessons were used. Therefore, we used the data of 206 out of the 239 participants of the parent study [9].

The Web-Based “Living to the Full” Intervention

Context

The Web-based “Living to the Full” intervention is based on ACT [13] and mindfulness [14,15] and has been published as a self-help book [16]. The intervention has been shown to be effective in reducing depressive and anxiety symptoms as a group and self-help course with email support [17-19]. The Web-based intervention is an individually based self-help program. Participants can access the Web-based intervention at any time, from any place, free of charge. The intended usage is 3 hours per week.

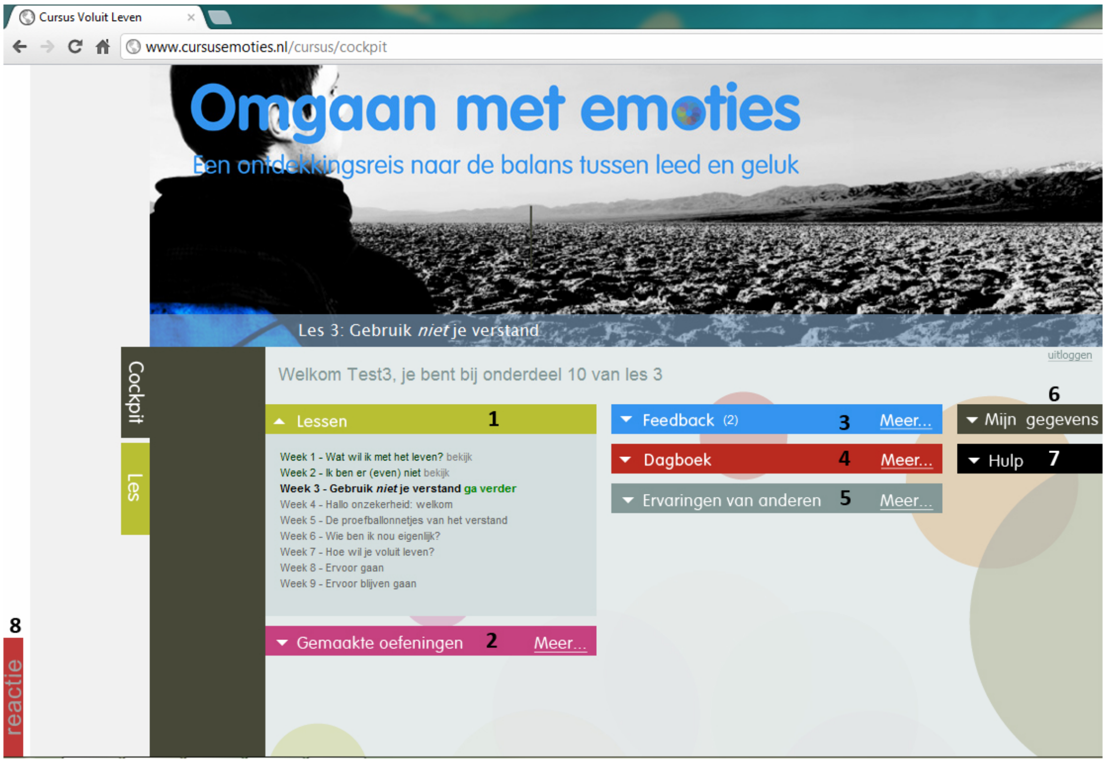
ACT consists of three main processes: open, centered, and engaged responding [20]. The first process, open responding, consists of acceptance of negative emotions and feelings and cognitive defusion. The opposite of acceptance, experiential avoidance (EA), has been defined as the attempt to escape or avoid personal events (emotions, memories, thoughts), even when the attempt to do so causes psychological harm [21]. Avoidance strategies, such as the excessive use of medication, food, or alcohol, or refraining from work or social activities, are common reactions to distress and are often effective in the short term. However, in the long term, EA tends to foster feelings of frustration and escalating psychological distress. Cognitive defusion is the ability to let go of entanglement with negative thoughts by viewing them from a distance in a non-judgmental way. The second process, centered responding, consists of mindfulness. Mindfulness refers to a state of being attentive to and aware of experiences (including physical sensations, emotions, thoughts, imagery) occurring in the present moment in a non-judgmental or accepting way [14,22]. The third process, engaged responding, refers to knowing one’s personal values and to committing one’s self to actions based on these values, even in the presence of undesired feelings and thoughts [23]. From the perspective of ACT, values can be seen as an intrinsic motivating framework for leading a meaningful life. When a person is open, centered, and engaged, they are considered to be psychologically flexible, that is, able to act effectively in accordance with personal values in the presence of negative private experiences [24]. The three main processes of ACT are highly interrelated, and one cannot be psychologically flexible without responding in all three ways [20].

Content

When participants log in to the Web-based intervention, they start in the “cockpit”. This home page of the Web-based intervention is presented in Figure 1. The Web-based intervention included a number of features, which can be grouped as general features, content features, and supportive or persuasive features. General features within the intervention are My Account (Figure 1, number 6), where participants can edit personal information as their username and password; Help

(Figure 1, number 7), where participants can find basic information about the use of the intervention itself; and a React button (Figure 1, number 8), where participants can comment on the intervention. Content features are lessons (Figure 1, number 1) and exercises (Figure 1, number 2) and will be elaborated on in the next section. Persuasive features have been designed to support the uptake of content of the nine lessons and will be discussed in a following section.

Figure 1. The home screen of the Web-based intervention "Living to the Full".



Welkom Test3, je bent bij onderdeel 10 van les 3

Week 1 - Wat wil ik met het leven? bekijk
Week 2 - Ik ben er (even) niet bekijk
Week 3 - Gebruik *niet* je verstand ga verder
Week 4 - Hallo onzekerheid: welkom
Week 5 - De proefballonnetjes van het verstand
Week 6 - Wie ben ik nou eigenlijk?
Week 7 - Hoe wil je voluit leven?
Week 8 - Ervoor gaan
Week 9 - Ervoor blijven gaan

Content Features

The ACT-based intervention consisted of nine lessons (see Figure 1, number 1) that had to be completed in a chronological order in a 12-week period. Participants were free to choose whether they worked through a lesson in one session or in multiple sessions. Participants were expected to follow all nine lessons to receive the full extent of the ACT processes.

Lessons 1, 2, and 3 focused on becoming aware of the negative consequences of EA among participants (so they became aware of the consequences of not responding openly, the first process of ACT). For example, in Lesson 2, participants made a list of all the stresses they were experiencing. They mapped all the coping strategies they used in the past in order to get control of their distress and the investments (eg, time, money) that were involved in applying these control strategies. The participants were then asked to consider the effectiveness of these strategies in the short term and in the long term and to consider that these strategies were not effective in reducing distress in the long term despite the large amount of time and money they invested. This gave rise to a feeling of creative hopelessness [13]. They became aware that the agenda of control has not been working for them, and this awareness created feelings of hopelessness. At the same time, these feelings opened up the possibility of an alternative coping strategy: acceptance of negative emotions, feelings, and thoughts. In Lessons 4, 5, and 6, participants practiced open responding. The participants learned to accept undesired emotions and to defuse negative thoughts and to disidentify with outer world aspects such as status or luxury goods and inner states such as feelings. In Lesson 5 for example, the participants had to write down five thoughts that bothered them a lot. As a second step, they had to write before each thought “I have the thought that…”. This exercise helped them to become aware that what they were thinking were mere thoughts and not reality. Another exercise was a mindfulness exercise in Lesson 4 in which they had to focus on a difficult situation they had experienced. The participants learned to notice and accept unpleasant feelings in their bodies and stay present with any negative emotions that might arise. In Lessons 7, 8, and 9, the central focus was on engaged living (third process of ACT). Participants identified important values in various life domains such as work or personal relationships and identified actions in compliance with these values. For example, in Lesson 7, participants had to clarify their core values in life domains such as relationships, work, and health. They then had to indicate to which extent they were currently living in accordance with these values. They also had to define actions that were congruent with their values. Exercises were done to strengthen the commitment to value-based living. Centered responding is a focus in all lessons.

Important transitions are between Lessons 3 and 4 and between Lessons 6 and 7. Between Lessons 3 and 4, participants had to let go of the agenda of control and avoidance and to adopt the agenda of acceptance of distress. An example of an exercise was the tug-of-war metaphor. Participants were invited to imagine themselves in a tug of war with their unwanted emotions and thoughts. This is an exhausting battle that cannot be won. As an alternative, they could decide to drop the rope and end the tug of war. Between Lessons 6 and 7, participants were invited to make decisions and change their behavior based upon their values even in the presence of unwanted feelings and thoughts. In the last three lessons, participants had to “walk the talk”.

The nine lessons were structured as follows. Each lesson started with an introduction of the topic of that specific lesson related to the aforementioned ACT and lessons. Every lesson incorporated exercises, homework, mindfulness exercises, metaphors (like the tug of war), and information related to that specific lesson. A diary could be used starting in Lesson 2. In each lesson, a short mindfulness meditation exercise was introduced that participants had to practice on a daily basis at home. In Lessons 1, 2, 5, 6, and 8, a new mindfulness exercise was presented in an online audio player, but participants had the opportunity to download the exercise as an MP3 file as well (see Multimedia Appendix 1). In Lesson 3, the participant was asked to repeat the exercises of Lessons 1 and 2. In Lessons 4 and 7, the mindfulness exercise was presented through text instruction. In Lesson 9, no new exercise was introduced. Rather, the participants were advised to integrate mindfulness exercises in their daily life. Users could start an exercise and download the exercise for practicing (see Multimedia Appendix 1).

Persuasive Features

The persuasive features that were available through the nine lessons for all participants were feedback (Figure 1, number 3), a diary (Figure 1, number 4), and success stories (Figure 1, number 5). Participants received feedback via the system after completing a lesson. They had to view the feedback by clicking on the feedback option (see Multimedia Appendix 1). The feedback was provided when a participant had viewed all psychoeducational material and had completed all the exercises. Furthermore, the feedback was sent to the Web-based system at least 5 days after the participant had started the lesson. This was done to ensure that a participant spent enough time on each lesson to be able to fully process the information. All feedback was designed to support the participant in the uptake of the content and was intended to strengthen the ACT principles.

Participants could use the diary feature to monitor their behavior and their learning of the ACT processes. The diary was not an obligatory feature of the intervention and was purely for the participant; the counselor had no access to the diary entries.

The intervention included nine success stories; after the completion of each lesson, a new success story became available for the participants (Multimedia Appendix 1). Each story had the same format in which a fictional earlier participant answered questions on the meaning of the course to them, positive experiences with the course, difficult moments within the course, and what advice they could give the current participant. Each story was different in that it focused on the content of the lesson in which the story was released. The goal of the success stories was to support the uptake of the content of each lesson and for the participants to be able to identify themselves with other users of the intervention.

Apart from persuasive features that were available for all participants, there were two features that were available for only a part of the participants. These features were part of the fractional factorial randomized controlled trial design of the parent study; both features were available for approximately half of the participants. Inclusion of these features was randomized, so a participant could have both features present, a single one, or neither of them. These features were the inclusion of multimedia (ie, video clips) within the lessons and the inclusion of an SMS coach (Multimedia Appendix 1).

For the participants in the scenario that included video, Lessons 1 to 8 included a clip in which an expert in ACT gave a thorough explanation of the ACT process central to that lesson. The videos were intended to support the understanding of the ACT processes, but the same content was also explained in the psychoeducational material available for all participants.

The participants in the scenario that included an SMS coach had the opportunity to turn on the SMS coach. This meant they received three SMS text messages per lesson on their mobile phone. Regardless of whether the SMS coach was turned on or off, for the participants in that scenario, all SMS text messages were presented within the Web-based intervention as well (Multimedia Appendix 1). The messages were intended to remind the participants to continue the lessons and to do the mindfulness exercises, but they were also intended to support the uptake of the ACT processes in daily life. This was done, for example, by posing ACT-related questions tailored to the lesson the participant was working on.

Data Collection and Analysis

In this study, we focused on a subset of the data that was collected in the larger study. Our focus was on the usage of the content and persuasive features per participant per lesson. These data were collected by automatically logging the actions of each participant within the Web-based intervention. From these data, the number of specific actions per participant per lesson were extracted.

Of the content features, the number of logins within a certain lesson was logged. We did not log viewing of the different screens of psychoeducational material and exercises within each lesson because of the obligatory and chronological nature of each lesson (ie, all the screens in each lesson had to be viewed in order for the participant to be able to go to the next lesson). Of the mindfulness exercises, the number of times a participant started and/or downloaded each exercise was logged (Multimedia Appendix 1).

Of the persuasive features that were available for all participants, we logged the number of feedback messages viewed and the number of success stories viewed. Usage of the diary feature was not logged due to the optional nature of this feature and because the diary was purely for the participants themselves.

Of the persuasive features that were available to some participants, we logged the number of video clips viewed, the number of SMS text messages viewed within the Web-based intervention, and the number of times the SMS coach was turned on and off.

To summarize, the log files used in this study were per lesson: login, start mindfulness, download mindfulness, view success story, view feedback message, start video, turn on SMS coach, turn off SMS coach, and view text message. Action specifications were, for example, the name of the mindfulness exercise that was started or which text message was viewed and was used to identify the number of unique actions per lesson (eg, did the participant view the same feedback message multiple times or did the participant view different messages?).

Statistical analysis was done with SPSS 20. For all analysis an alpha of .05 was used as the level of significance. Differences between lessons were explored with repeated measures analysis of variance (ANOVA) where the factor "time" had nine levels (one for each lesson of the intervention). For the trend analyses, we were mainly interested in whether there was a linear trend (eg, whether during the 9 time points, the overall score increased or decreased). However, significant fluctuations over time may also be of interest. Therefore, we also investigated whether there was a quadratic, cubic, or order 4 until order 8 effect. For this paper, significant higher order effects were interpreted as a significant fluctuation in use over the nine lessons. The higher the order, the stronger the fluctuations in usage.

To explore differences between higher and lower active users, a quartile split was used to divide the participants into four groups based on the average number of logins per started lesson. Differences between these groups on usage of the different features were explored with one-way ANOVAs. Differences between the groups on the number of participants who turned the SMS coach on were explored with a chi-square test.

To explore the possible predictive value of the login per lesson quartiles on the outcome measures, four linear regressions were used with login quartiles as predictor and with the outcome measures (CES-D and Hospital Anxiety and Depression Scale—Anxiety [HADS-A] at post intervention and follow-up) as dependent variables. Depressive symptoms were measured with the CES-D (20 items, score 0-60; higher scores mean more depressive symptoms) [12,25] at baseline, post intervention, and follow-up. Anxiety symptoms were measured with the HADS-A (7 items, score 0-21; higher scores mean more anxiety symptoms) [26,27] at baseline, post intervention, and follow-up.

Missing data on clinical measures (CES-D and HADS-A) were imputed with the expectation-maximization method in PASW 18. This method estimates the unmeasured data based on maximum likelihood estimates using observed data in an iterative process [28]. Observed data on CES-D, HADS-A, gender, age, education, lesson reached, and support situation were used for estimation.

Results

Usage of Content and Persuasive Features per Lesson

Of the 238 participants that were randomized and received login information, 206 participants logged in (Lesson 1) and 118 participants completed Lesson 9 (Table 1). The use of the intervention (adherence rate) fluctuated over time; critical moments for dropout appeared in Lessons 3 and 6 [9].

Table 1 presents the mean login and the mean use of features that were available for all participants per lesson. Over time, the mean login activities for the features available for all participants decreased significantly: linear effect, $P<.001$, $F_1=13.656$ (see Figure 1). Table 1 shows that not all features have been used in all the lessons, like mindfulness (unique messages, downloaded, success stories; mean score below 1).

Table 1. Mean logins and mean use of features available for all participants per lesson.

Lesson No. (n)	Login,	Feedback[a], mean (SD)	Unique[b] feedback, mean (SD)	MF[c] started, mean (SD)	Unique MF started,	Unique MF downloaded, mean (SD)	Success stories, mean (SD)	Unique success stories, mean (SD)
1 (206)	4.40 (3.05)	2.04 (2.34)	1.36 (1.00)	1.75 (1.59)	0.88 (0.33)	0.50 (0.50)	1.67 (1.53)	1.28 (0.78)
2 (194)	4.12 (2.82)	2.10 (2.37)	1.53 (1.18)	1.28 (1.31)	0.78 (0.49)	0.51 (0.58)	1.03 (1.40)	0.87 (1.06)
3 (174)	3.92 (2.58)	2.20 (2.94)	1.53 (1.56)	0.83 (1.41)	0.57 (0.77)	0.03 (0.17)	0.84 (1.29)	0.68 (1.00)
4 (159)	4.27 (2.34)	2.44 (3.60)	1.68 (1.78)	0.11 (0.55)	0.06 (0.29)	0.03 (0.22)	0.73 (1.20)	0.63 (0.97)
5 (152)	3.70 (2.00)	1.99 (2.53)	1.50 (1.38)	1.08 (1.01)	0.73 (0.47)	0.40 (0.59)	0.51 (1.05)	0.44 (0.84)
6 (149)	4.26 (3.25)	1.87 (2.54)	1.46 (1.59)	1.03 (1.32)	0.64 (0.56)	0.40 (0.63)	0.58 (1.28)	0.52 (1.08)
7 (135)	3.96 (3.79)	2.29 (3.36)	1.73 (2.09)	0.05 (0.45)	0.02 (0.15)	0.06 (0.40)	0.76 (1.76)	0.58 (1.27)
8 (125)	3.54 (2.90)	2.99 (6.58)	1.98 (3.11)	0.82 (1.13)	0.58 (0.73)	0.50 (0.75)	0.94 (2.31)	0.76 (1.59)
9 (118)	3.95 (4.22)	2.97 (3.97)	2.19 (2.61)	0.23 (1.03)	0.15 (0.63)	0.34 (1.14)	0.95 (2.10)	0.80 (1.63)

[a] Feedback messages viewed.

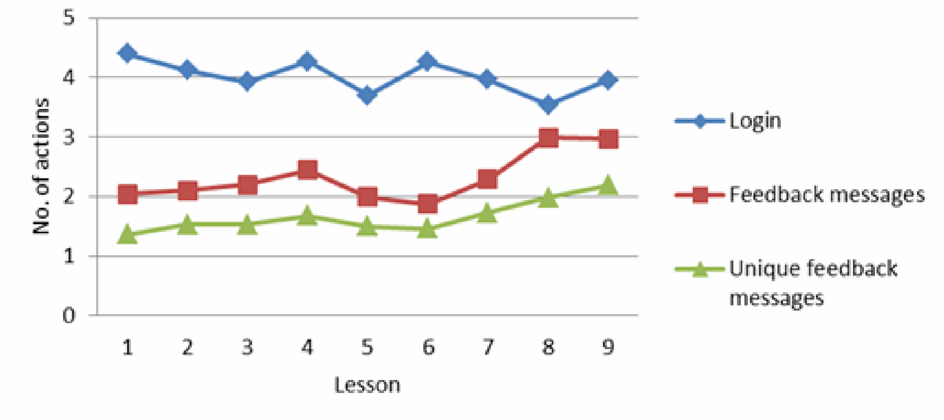
[b] Unique per lesson, that is, the number of different messages or exercises that a participant used in a particular lesson.

[c] MF: mindfulness exercise.

Feedback

Figure 2 shows the feedback per lesson. Participants received feedback after having completed all the exercises of that lesson. The use of feedback messages did not change significantly during the treatment process. By contrast, the number of unique feedback messages viewed increased significantly during the treatment process (linear effect, $P=.025$, $F_1=5.176$). This shows that at the beginning the participants viewed the same messages multiple times (more total messages, but not more unique or different messages), and later on, participants viewed different messages, but each of these messages fewer times (the same number of total messages, but more unique messages).

Figure 2. Login and feedback per lesson.



Mindfulness

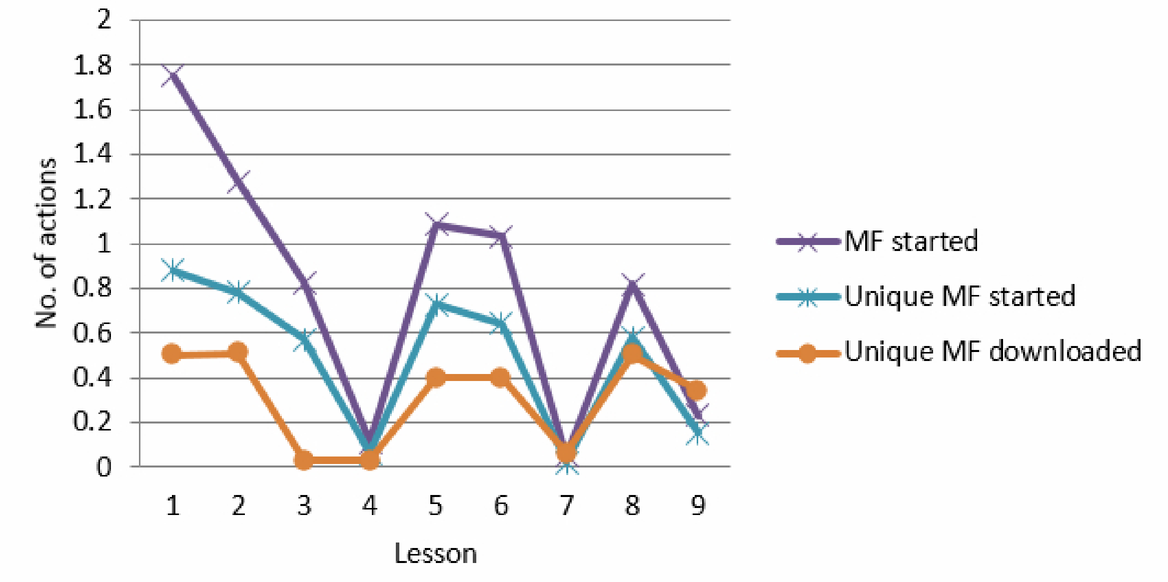
Mindfulness exercises were introduced in each lesson for daily practice. The use of mindfulness exercises decreased significantly over time: linear effect, *P*<.001, F_1=99.029 (see Figure 3). Furthermore, there were differences over time, where Lessons 4 and 7 showed the least use of mindfulness exercises: quadratic (*P*<.001), cubic (*P*<.001), order 5 (*P*=.016), order 6 (*P*<.001), order 7 (*P*<.001), and order 8 (*P*=.025) effect.

There were more participants who started to use the mindfulness exercises than participants who downloaded these exercises (for offline practicing). The pattern for the usage of unique mindfulness exercises is the same. There was a significant decrease over time (linear effect, *P*<.001, F_1=100.042) and a variance in usage over time: quadratic (*P*<.01), cubic (*P*<.001), order 5 (*P*=.015), order 6 (*P*<.001), order 7 (*P*<.001), and order 8 (*P*<.001) effect. Downloading of mindfulness exercises did not increase or decrease over time (linear effect, F_1=0.340, *P*=.561), but there was a significant variance in usage during the treatment process: quadratic (*P*<.001), cubic (*P*<.01), order 5 (*P*<.001), order 6 (*P*<.001), and order 7 (*P*<.001) effect (see Figure 3).

Lesson 3 mindfulness exercises were repeated from Lessons 1 and 2, and Lessons 4 and 7 had only text as content for the mindfulness exercises. The use of started (linear effect, *P*<.001, F_1=41.511) and unique started (*P*<.001, F_1=24.975) online mindfulness exercises in Lessons 1, 2, 5, 6, and 8 decreased significantly.

A comparison of Lessons 1, 2, 3, 5, 6, and 8 showed a similar pattern for started (linear effect, *P*<.001, F_1=36.787) and unique started (linear effect, *P*<.001, F_1=22.485) exercises. There was a decrease of unique downloaded mindfulness exercises for Lesson 3 (order 4 and 5; *P*<.001, F_1=34.152).

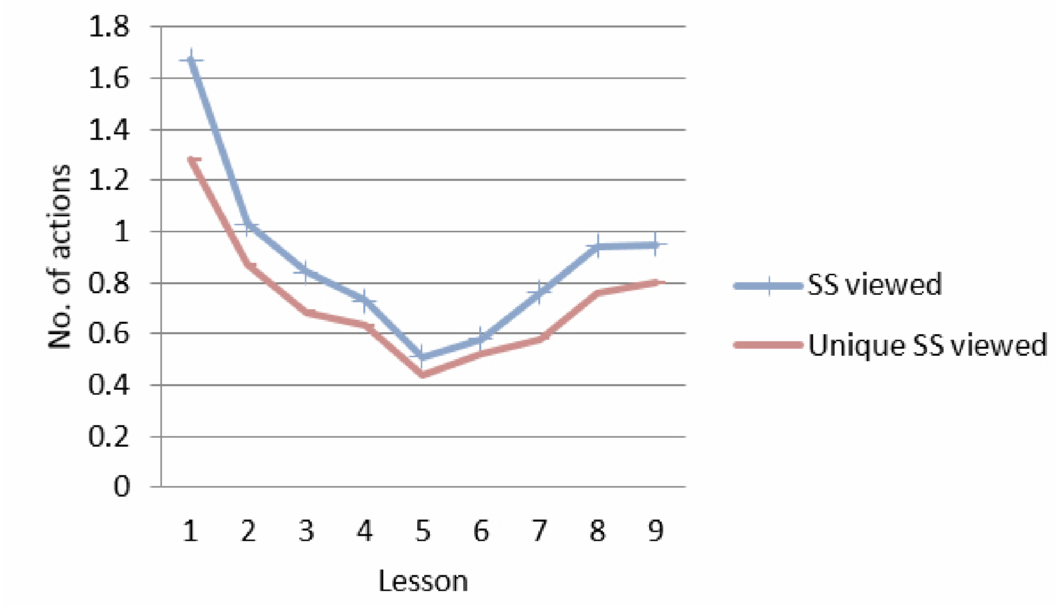
Figure 3. Mindfulness per lesson.



Success Stories

After completion of each lesson, a new success story was introduced. The use of success stories decreased significantly over time (viewed, linear effect, $P<.001$, $F_1=17.020$; and unique viewed, $P<.001$, $F_1=13.745$). A significant fluctuation over time was observed: quadratic effect, $P<.001$ for both viewed and unique viewed (see Figure 4).

Figure 4. Success stories per lesson.



Usage of Text Messages and Multimedia per Lesson

Table 2 presents the number of participants that received the SMS and video features and the mean use of these features, per lesson. The SMS coach was available in the Web-based system, and participants were also able to turn on the SMS coach on their mobile phone. All SMS text messages were presented within the Web-based system, regardless of whether the mobile phone option was activated or not. Table 2 shows that only a few used the mobile phone option for SMS messages.

Table 2. Mean use of features, per lesson for the features SMS and video.

Lesson No.	SMS feature[a], n	SMS on[b], n	SMS off[b], n	SMS viewed[c], mean (SD)	Unique SMS viewed[d], mean (SD)	Video feature[a], n	Videos viewed, mean (SD)	Unique videos viewed, mean (SD)
1	105	13	3	1.44 (2.35)	0.95 (1.27)	116	0.59 (0.94)	0.40 (0.49)
2	98	2	0	1.32 (2.43)	0.99 (1.62)	109	0.44 (0.87)	0.28 (0.45)
3	85	0	0	1.09 (2.29)	0.86 (1.68)	96	0.61 (1.32)	0.39 (0.49)
4	79	0	0	1.54 (3.39)	1.19 (2.39)	91	0.74 (1.05)	0.45 (0.50)
5	74	0	0	0.96 (2.62)	0.84 (2.19)	90	0.66 (0.96)	0.46 (0.54)
6	73	1	0	1.21 (3.86)	0.99 (2.55)	87	0.49 (0.54)	0.40 (0.56)
7	71	1	0	0.86 (2.26)	0.68 (1.80)	79	0.66 (1.32)	0.42 (0.52)
8	66	0	0	2.44 (2.14)	2.14 (5.95)	71	0.66 (0.99)	0.48 (0.61)
9	63	0	2	1.92 (6.05)	1.48 (4.56)	64	0.05 (0.38)	0.03 (0.25)

[a] Number of participants who received the intervention with the feature included.

[b] Number of participants who turned the SMS coach on/off in a particular lesson.

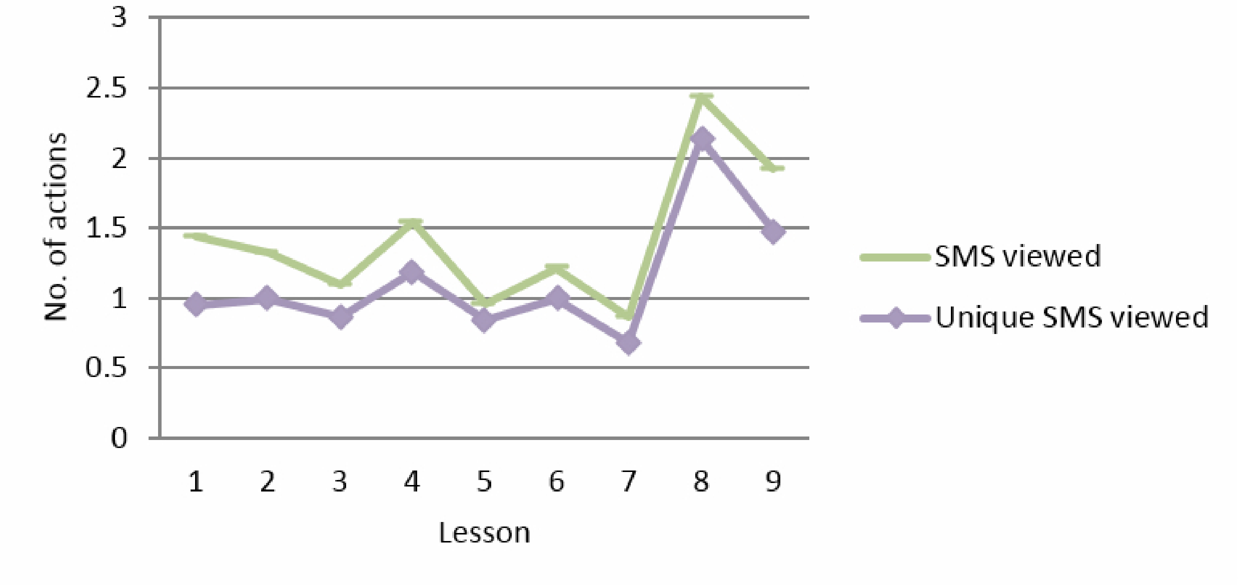
[c] Number of SMS messages viewed within the Web-based intervention.

[d] Unique per lesson, that is, the number of different messages that a participant used in a particular lesson.

SMS Coach

The SMS coach provided messages per lesson to remind participants to do the exercises and to support them via tailored messages during the lessons. Figure 5 shows the viewed and unique viewed SMS messages, per lesson. The number of viewed SMS messages over nine lessons did not increase or decrease over time. However, there was a significant fluctuation in usage: SMS viewed (linear effect, $P=.579$, $F_1=0.311$; quadratic effect, $P=.032$; and order 8 effect, $P=.018$) and unique SMS viewed (linear effect, $P=.274$, $F_1=1.219$; and order 8 effect, $P=.004$).

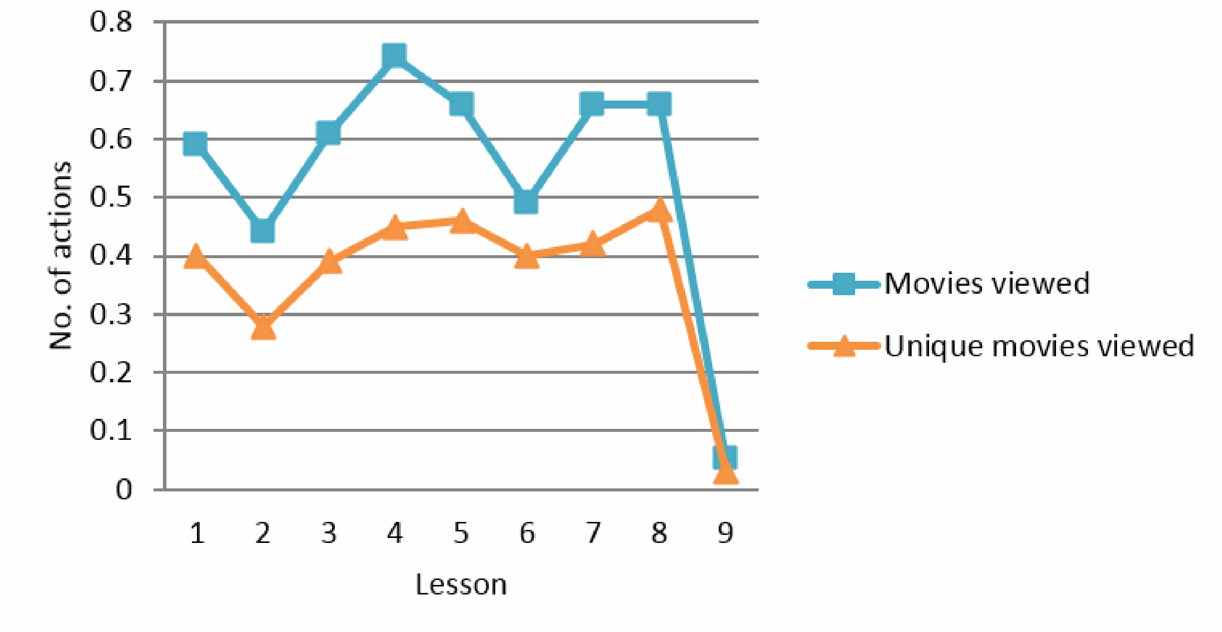
Figure 5. SMS per lesson.



Multimedia

Videos to explain the treatment and to understand the content of a particular lesson were provided from Lessons 1 to 8. The videos viewed (linear effect, *P*=.01, F_1=7.133) and unique videos viewed (linear effect, *P*<.001, F_1=15.935) over nine lessons showed a significant decrease and a fluctuation in usage (quadratic) over time (videos viewed: quadratic effect, *P*<.001; order 5 effect, *P*<.01, and unique videos viewed: quadratic effect, *P*<.001; cubic effect, *P*<.001; order 4 effect, *P*=.041; order 5 effect, *P*<.01) (see Figure 6). Without Lesson 9 (because there was no video for Lesson 9), there was no significant decrease in videos viewed and unique videos viewed (videos viewed: linear effect, *P*=.977, F_1=0.001; and unique videos viewed: linear effect *P*=.244, F_1=1.382). Only a difference in usage over time was observed for unique videos viewed (order 4 effect, *P*<.1).

Figure 6. Movies per lesson.



Use of Content and Persuasive Features for Different Users

Table 3 shows the average actions per started lesson (mean use of features), by level of logins per started lesson (quartile of logins). The average number of logins per started lesson was used as a measure to divide users into higher (3rd, 4th quartile) and lower active users (1st, 2nd). Higher logins over time (9 weeks) corresponded with a higher usage of features (in most cases significant; see Table 3) and vice versa.

Table 3. Mean logins and mean use of features, for different users (1st-4th quartile, n=206).

	1st quartile logins (n=51)	2nd quartile logins (n=57)	3rd quartile logins (n=49)	4th quartile logins (n=49)	*F* (df)	*P*
Login, mean (SD)	1.95 (0.46)	3.18 (0.31)	4.24 (0.38)	6.55 (2.30)	139.376 (3, 202)	<.001
Feedback, mean (SD)	0.97 (0.77)	1.63 (1.07)	2.19 (1.62)	3.34 (2.36)	20.867 (3, 202)	<.001
Unique feedback, mean (SD)	0.75 (0.43)	1.08 (049)	1.17 (0.56)	1.54 (0.71)	17.090 (3, 202)	<.001
MF started, mean (SD)	0.74 (0.66)	0.75 (0.53)	0.95 (0.74)	1.15 (0.71)	4.464 (3, 202)	.005
Percentage MF started, mean (SD)	70.1 (34.7)	66.6 (30.3)	76.3 (27.7)	80.2 (30.0)	2.030 (3, 202)	.111
Percentage MF downloaded, mean (SD)	24.0 (33.0)	48.9 (40.9)	47.3 (40.1)	62.8 (40.5)	8.652 (3, 202)	<.001
Success stories, mean (SD)	0.61 (0.59)	0.92 (0.68)	0.83 (0.65)	1.29 (0.96)	7.474 (3, 202)	<.001
Percentage success stories, mean (SD)	49.1 (47.3)	61.8 (34.6)	54.9 (31.9)	70.4 (28.5)	3.189 (3, 202)	.025
SMS feature, n	25	30	25	25		
SMS turned on, %	16.0	36.7	12.0	32.0	Chi-square: 6.201 (df=1)	.102
SMS viewed, mean (SD)	0.42 (0.77)	0.73 (1.04)	1.36 (1.72)	2.37 (2.86)	6.111 (3, 101)	.001
Percentage SMS viewed, mean (SD)	11.8 (23.4)	18.5 (22.9)	28.1 (30.1)	41.1 (37.0)	4.992 (3, 101)	.003
Multimedia feature, n	22	34	30	30		
Videos viewed, mean (SD)	0.34 (0.54)	0.38 (0.56)	0.45 (0.53)	0.80 (0.83)	3.236 (3, 112)	.024
Percentage videos viewed, mean (SD)	24.3 (36.0)	26.5 (35.8)	35.7 (40.1)	43.4 (41.7)	1.508 (3, 112)	.216

Usage and Impact

To explore the possible predictive value of the login per lesson quartiles on the outcome measures, we performed exploratory linear regressions with the login quartiles as predictor and with the outcome measures (depressive and anxiety symptoms on post intervention and on follow-up) as dependent variables. Tables 4 and 5 show that only when predicting depressive symptoms at post intervention, the linear regression yielded a significant model with login quartile as a significant predictor. However, the percentage of explained variance was only 2.7%.

Table 4. Linear regression with login quartiles as predictor.

Model	Included	B (SE)	*P*	Odds ratio (95% CI)
Predicting CES-D at post intervention				
	Constant	20.8 (0.90)	<.001	19.0-22.6
	Login quartile	–1.17 (0.49)	.018	–2.14 to –0.20
Predicting CES-D at follow-up				
	Constant	18.1 (0.95)	<.001	16.2-20.0
	Login quartile	–0.63 (0.52)	.229	–1.65 to 0.40
Predicting HADS-A at post intervention				
	Constant	8.16 (0.35)	<.001	7.46-8.86
	Login quartile	–0.37 (0.19)	.057	–0.72 to 0.01
Predicting HADS-A at follow-up				
	Constant	7.44 (0.33)	<.001	0.78-8.10
	Login quartile	–0.23 (0.18)	.210	–0.59 to 0.13

Table 5. Characteristics of the tested linear regression models.

Model	*F* (df=1, 204)	*P*	R^2
Predicting CES-D at post intervention	5.697	.018	.027
Predicting CES-D at follow-up	1.457	.229	.007
Predicting HADS-A at post intervention	3.660	.057	.018
Predicting HADS-A at follow-up	1.582	.210	.008

Discussion

Principal Results

The aims of this study were (1) to illustrate how log data can be used to understand the uptake of the content of a Web-based intervention based on ACT and (2) to show how log data can be of value for improving the incorporation of content in Web-based interventions.

The log files used in this study were per lesson: login, and use of features by means of starting a feature (click) and/or downloading a feature (click) for practicing. We were interested in which features of the intervention were used, in what way (viewing and/or downloading for practicing), and in what intensity (did the participant view the same features multiple times or did the participant view different features?). In addition, we wanted to know the impact of logins (high/low active uses) on outcomes and the impact of the features on outcomes. The results were of relevance for tailoring the incorporation of the content of the Web-based intervention to the stages of the program of therapy to reduce depression. The log data of this study showed that the uptake of the content is not quite in accordance with the underlying principles of therapy, and as such, it provides insights for improvement to the program.

Overall, the pattern that emerged from the log data showed a decrease in logins and a decrease in the use of content and persuasive features over time (nine lessons). The decrease in logins might be a learning effect; from the parent study [9] it appeared that adherers needed fewer sessions (logins) to complete a lesson at the end of the treatment than during the first lessons of the treatment program. The pattern of the usage of content and persuasive features varied significantly (quadratic, cubic, and in high order) over time, except for the usage of Feedback messages. The use of Feedback did not change over time, in contrast with the use of unique Feedback messages. This might be related to a need for tailored feedback during the end of the treatment process, when commitment to "value-based living" is the focus of therapy (Lessons 7, 8, and 9). The use of success stories fluctuated over time and increased after Lesson 5. The content of each story was tailored to the essence of a particular lesson and stage of therapy. Lessons 6 to 8 were part of a transition process to change behaviors based upon personal values for engaged living [21,24]. This may have influenced the need for stories with experiences of others and their advice. The use of SMS text messages fluctuated over time and increased at the end of the treatment process, after Lesson 7. An explanation might be that at the end of the treatment process, more SMS messages were available than in the earlier lessons.

The quadratic, cubic, and high order effects in usage were important to understanding the connection between the uptake of features per lesson and the three main processes of ACT that guide the lessons of the intervention (content). According to Hayes [24], these main processes are highly interrelated, and one has to go through all three stages. The underlying principle of the "Living to the Full" intervention is that participants benefit from therapy when they use all features related to the three main processes of ACT. In addition, the lessons must be done in chronological order. Mindfulness is the focus of all lessons and refers in particular to the centered responding process. The decline in mindfulness in Lessons 3, 4, and 7 was related to the mindfulness content in those particular lessons (repetition of Lessons 1 and 2, and text only in Lesson 7). In addition, the suboptimal use of online and offline features (downloaded mindfulness for practicing) indicated that participants did not benefit from all the mindfulness features that were available, which is relevant given that mindfulness is an integral part of the intervention and a central feature of ACT-based interventions. At the end of the course (Lessons 8 and 9), users were asked to integrate mindfulness exercises in their daily lives. This may have resulted in a more intensive use, which aligns with the increase in unique downloaded mindfulness exercises. Moreover, at the end, users received an overview of all the exercises done. After week 9, they were not allowed to return to the intervention with all features. So it may be that users wanted to complete all possible exercises at the end of the lessons because they knew they could not return to the features after week 9. Overall, the increase of persuasive features at the end of treatment, the engaged responding stage, might indicate that participants needed motivational support to comply with the "value-based-living" stage.

The log data from this study can be useful for understanding how the incorporation of content in a Web-based therapy can be improved, in such a way that participants can benefit more from the intervention. The study revealed that higher logins over time correlate with a higher usage of most features. Higher active users consume all features more than low active users, which may indicate a need to stimulate users to log in to stick to the program of therapy. The login per lesson did not predict the outcomes; only when predicting depressive symptoms at post intervention was the login a significant predictor, although the explained variance was low. This sheds light on the dose-response relationship: a higher active usage per lesson might lead to better results in terms of outcomes. When entering the use of the content and persuasive features per lesson as predictors (a separate model per lesson), none of the models significantly predicted reduction of depression. These findings questioned the importance of the uptake of *all* features or completing *all* modules to realize the objectives of the therapy. From our prior study [9], we know that adding a single feature to the intervention, except feedback, does not impact adherence or outcomes. Also, a variance in intensity of features (more extensive use of a single feature compared to a lower extensive use of that feature) did not influence the adherence or outcomes [29]. It might be that not a single feature but a *combination* of content and action-tailored persuasive features can play a supportive role *during* the intervention period, resulting in higher involvement and more learning effects. Are there certain combinations of features that predict adherence, return rates, or outcomes? Although the ACT-based therapy requires a fixed programming of lessons, the high (quadratic, cubic, high order) fluctuation in usage of content and persuasive features requires more flexibility in treatment, if we take into account differences in learning styles (eg, fewer sessions needed to complete the lessons) and allow for differences in paths and tempo when “walking through the intervention” (shorter paths and higher activities, for example).

In addition, the use of action-related persuasive features can strengthen the effects on adherence. A different study showed that exposure to an action-related persuasive feature (social support to reduce weight) influenced the persistence with the Web-based intervention [8]; not the action per se but the repercussions of the action influenced persistence [8]. These results are preliminary, but promising for Web-based interventions to tailor content and persuasive features to therapy and involved actions, for example, using online and offline exercises like mindfulness. Via the log data, we observed that participants, for example, started with viewing an exercise but never downloaded the exercise. This is relevant information for the developers of the intervention (identification of action-related triggers for extra support to download exercises). Log data are an objective registration of online usage. In combination with surveys, self-reports, and interviews during the intervention period, it is possible to gain insight into the reasons for activities carried out or not and in the evaluation of the system and content. The react button, one of the features of the intervention of this study, shows that most of the remarks were on the quality of the system (bugs, confusion about how the system works) and service (availability of intervention after completion) [29]. For improving the incorporation of content that plays a major role in therapy, insight into the capacities of persuasive features to activate or to engage users is important.

To benefit from log data, content experts (eg, therapists) have to communicate with designers in an early stage of intervention development about which features should be incorporated into the Web-based system and in which logical order (paths, combinations). A well-thought out activity log protocol that describes which features of an intervention should be logged and how log data can be incorporated in the intervention should guide this debate, in order to capture data for the purpose of identifying which combinations and paths (routing) facilitate the uptake of content and which combination of features (content and persuasive) works best for whom. From our research [29-31], we know that such a debate is fruitful but often complicated. First, the collaboration between developers and content experts is often problematic because developers and content experts use different “mental spaces” and different “vocabularies”. Second, the incorporation of log data can be laborious. Third, advanced data analytics are needed to identify and predict usage patterns. The debate involves an inquiry on several levels. At the level of content, we need to decide which features are integral parts of the therapy and lessons, what kind of support (guided, unguided) is intended, and what kind of persuasive features can motivate the uptake of content and offline exercises. At a system level, it is a matter of considering the logical order of content and persuasive features to be available during the lessons, as well as discussing flexibility in the “walking-path-route” through the system (eg, have participants go through the content in a chronological order, through all lessons once). At the level of users, what is the supposed eHealth literacy? At the level of evidence, what data should be collected to be able to identify factors that influence adherence, outcomes, and retention? These questions are relevant before the design of the intervention system starts. To guide such an early stage development debate, we developed guidelines with an accompanying toolkit [30,31] and a protocol for incorporating log data into Web-based interventions [9,11,29].

Limitations

There are several limitations regarding this study. First, log data studies are in their infancy, and more insight is needed into real-time usage data to learn what works in practice and for whom. Some studies are available that report on the uptake of content in percentages of modules that are used (eg, [32]), and some have preliminary results about which features support persistence (eg, [8]). Advanced analysis techniques are needed to develop interventions that are dynamic and flexible. For example, machine learning techniques are needed to search for paths for success, to tailor the routing in the intervention to learning profiles, and to know how much effort is needed for success in outcomes (eg, the minimal, maximal amount of sessions to complete the therapy). Markov models might be of value to identify successful combinations of features. Furthermore, presentation techniques are needed to visualize the data in such a way that therapists and designers can discuss it. For the development of theory-driven Web-based interventions, an advanced login system should be built in the intervention to be able to gather data that meet the requirements for advanced analytics (machine learning, Markov models).

The other limitation we have is a reduced sample of participants that used the SMS coach and multimedia, due to a fractional factorial design in the prior study [9]. In this study, we used only log data to evaluate how the content and persuasive features have been used. The opinion of participants about the added value of these features has not been reported in this study; we will report on this in another study.

Future Research

The findings of this study can be used for the development of a log protocol to enable advanced analytics to measure the uptake of content of a Web-based intervention and for improving the incorporation of content and persuasive features.

To develop Web-based interventions, a protocol should be developed that explains what kind of log files to build into the system to get data that are manageable and that make sense for usage analysis. Such a protocol can be published as part of the description of the Web-based intervention, for reasons of transparency and replications. In current studies, the use of log data is not reported comprehensively. A log protocol, for example, can include information about login (reach of an intervention), the time, moments of usage (exposure), the activities performed (which features are used), the intensity of usage (how many activities in 1 session), the description of a session (interaction time period after login, eg, 30 minutes, to avoid calculating when participants are inactive), the reactions (eg, a reaction on feedback), and so on. An example of this can be found in [11,29].

For an in-depth analysis of how participants can benefit more from the intervention and which combinations of features matter most to whom, advanced methods are needed to perform log data analyses and to recognize and predict usage patterns automatically. Pattern recognition can be relevant to guide participants through the intervention based on their usage and return patterns from earlier lessons and to support them in the uptake of content essential for therapy, like mindfulness. In a study on Web-based interventions for depression [33], further investigation was suggested into the relationship between viewing interactive content features aimed at improving engagement and the usage of other features of the program. In our study, no simple linear relationship was found between features and the use of content. To be able to reveal other nonlinear relationships, we will use more complex models or more advanced techniques (machine learning) in future research.

By means of log data, a better incorporation of content and persuasive features in the intervention is possible and a better intra-usability of content and persuasive features within the system (home page and accompanied subpages). The findings of this study revealed that the incorporation of content features and action-related persuasive features can be improved for a better connection of the therapy program with lessons and exercises, and to foster the practicing of exercises (not just viewing, also downloading for practicing). Furthermore, based on the findings, a discussion is needed between designers and content experts about the pacing of the therapy program. For example, do the participants need to go through the content in a chronological order, through all the lessons once? How should flexibility be built into the program for coping with the new behaviors and for self-reflection? Although the therapy program is obligatory, this does not imply a one-size-fits-all approach. Participants may differ in needs for support at different moments and might have different learning and e-learning styles and needs for self-reflection during a treatment process. The fluctuation in high orders (above 4) can be used to review the principles of treatment taking into account a variable and momentary use of features instead of completing lessons and exercises equally.

An interactive and adaptable interface for a modular composition of therapy features supposes a new concept of the Web-based treatment approach: not a predefined path for all but a flexible way to go through major features that had to be used to benefit from therapy. In terms of technology, such a flexible presentation of content is not a problem, but it first requires a debate among content experts and designers to review the benefits and drawbacks of such an approach and its implications for therapy (eg, online vs blended formats). The log data revealed that the integration of online and offline activities should be considered for improvements. There is a lack of balance between the started exercises and the downloaded exercises for practice. A better integration with mobile support (eg, to download exercises) and the incorporation of action-related persuasive triggers could be an option. The integration with mobile services has been addressed in a study about a Web-based program for depression because of the underuse of SMS text messages to create updates via mobile phone [33].

To identify the potential features for supporting better interaction with the system, we think the Persuasive System Design (PSD) model [34] can be valuable. The PSD model provides a taxonomy for incorporation of persuasive features into the system, to support tasks (eg, exercises), dialogue, social support, and to support the credibility of the system. These persuasive features can be tested in an experimental design (like the Multi-Phase Optimization Strategy [MOST] [35]) in order to assess the added values for the uptake of content of the Web-based intervention. This helps avoid a more intuitively based choice of persuasive features.

In the end, technology can support therapy. But to achieve success, the development of theory-driven content should be interwoven with the development of technology. The process of development of Web-based interventions and built-in log data should be described in a way that enables replications.

Conclusions

It can be concluded that the log data are valuable for gaining insight into the usage of content of the Web-based “Living to the Full” intervention. Based on the log data, it is possible to tailor the content and persuasive features of the intervention to the main processes of acceptance and commitment therapy. To gain more detailed insight into how participants can benefit more from the intervention, advanced analytic methods are needed to identify which navigation route facilitates the uptake of content and which combination of features (content and persuasive) works best for whom.

The pattern that emerged from the log data showed a decrease in logins and a decrease in the use of content and persuasive features over time (nine lessons). The usage of content and persuasive features varied significantly (quadratic) during the treatment process. The uptake of mindfulness, which is an integral part of ACT, varied substantially during the treatment process in contrast to feedback. The usage of persuasive features like SMS, multimedia, and success stories and feedback increased during the third part of the ACT (commitment to value-based living), which might indicate that at this stage, motivational support is relevant.

To improve the uptake of content of the intervention, a better integration of lessons and persuasive features in the intervention and a better intra-usability of features within the system are needed. It is preferable to have an overarching view of the composition of features that are an integral part of the ACT intervention and features that are supportive. A well-planned activity log protocol that describes which features of an intervention should be logged and how log data can be incorporated into the intervention should be developed in the early stage of design. Such a protocol enables us to identify which navigation route in the system facilitates the uptake of content and which combination of features (content and persuasive) works best for whom.

Conflicts of Interest

Two of the authors were involved in the development of the Web-based “Living to the Full” intervention.

Multimedia Appendix 1

Web-based Living to the Full intervention.

[PDF File (Adobe PDF File), 890KB - jmir_v16i1e27_app1.pdf]

References

1. 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]
2. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of Internet-based psychotherapeutic interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
3. Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):iii, xi-ixiv, 1 [FREE Full text] [Medline: 16959169]
4. Robertson L, Smith M, Castle D, Tannenbaum D. Using the Internet to enhance the treatment of depression. Australas Psychiatry 2006 Dec;14(4):413-417. [doi: 10.1111/j.1440-1665.2006.02315.x] [Medline: 17116083]
5. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
6. 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]
7. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
8. Freyne J, Saunders I, Brindal E, Berkovsky S, Smith G. Factors associated with persistent participation in an online diet intervention. 2012 May 05 Presented at: CHI'12 Extended Abstracts on Human Factors in Computing Systems; May 5, 2012; Austin, Texas. [doi: 10.1145/2212776.2223805]
9. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a Web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
10. Han JY. Transaction logfile analysis in health communication research: challenges and opportunities. Patient Educ Couns 2011 Mar;82(3):307-312. [doi: 10.1016/j.pec.2010.12.018] [Medline: 21277146]
11. Kelders S, Van Gemert-Pijnen J. Using log-data as a starting point to make eHealth more persuasive. In: Berkovsky S, Freyne J, editors. PERSUASIVE, LNCS 7822. Berlin Heidelberg: Springer-Verlag; 2013:99-109.
12. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
13. Hayes SC, Strosahl K, Wilson KG. Acceptance and commitment therapy: An experiential approach to behavior change. New York: The Guilford Press; 2003.
14. Kabat-Zinn J. Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness. New York, N.Y: Delacorte Press; 1990.
15. Kabat-Zinn J. Wherever you go, there you are: Mindfulness meditation in everyday life. New York: Hyperion; 1994.
16. Bohlmeijer ET, Hulsbergen M. Voluit Leven. Mindfulness of de kunst van het ervaren, nu als praktisch hulpboek. [Living to the full. Mindfulness or the art of acceptance, now as a practical help book]. Amsterdam: Boom; 2008.
17. Bohlmeijer ET, Fledderus M, Rokx TA, Pieterse ME. Efficacy of an early intervention based on acceptance and commitment therapy for adults with depressive symptomatology: Evaluation in a randomized controlled trial. Behav Res Ther 2011 Jan;49(1):62-67. [doi: 10.1016/j.brat.2010.10.003] [Medline: 21074752]

18. Fledderus M, Bohlmeijer ET, Pieterse ME, Schreurs KM. Acceptance and commitment therapy as guided self-help for psychological distress and positive mental health: a randomized controlled trial. Psychol Med 2012 Mar;42(3):485-495. [doi: 10.1017/S0033291711001206] [Medline: 21740624]
19. Fledderus M, Bohlmeijer ET, Smit F, Westerhof GJ. Mental health promotion as a new goal in public mental health care: a randomized controlled trial of an intervention enhancing psychological flexibility. Am J Public Health 2010 Dec;100(12):2372. [doi: 10.2105/AJPH.2010.196196] [Medline: 20966360]
20. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy, Second Edition: The Process and Practice of Mindful Change. New York: The Guilford Press; 2011.
21. Hayes SC, Wilson KG, Gifford EV, Follette VM, Strosahl K. Experimental avoidance and behavioral disorders: a functional dimensional approach to diagnosis and treatment. J Consult Clin Psychol 1996 Dec;64(6):1152-1168. [Medline: 8991302]
22. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
23. Trompetter HR, Ten Klooster PM, Schreurs KM, Fledderus M, Westerhof GJ, Bohlmeijer ET. Measuring values and committed action with the Engaged Living Scale (ELS): Psychometric evaluation in a nonclinical sample and a chronic pain sample. Psychol Assess 2013 Dec;25(4):1235-1246. [doi: 10.1037/a0033813] [Medline: 23914955]
24. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
25. Haringsma R, Engels GI, Beekman AT, Spinhoven P. The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. Int J Geriatr Psychiatry 2004 Jun;19(6):558-563. [doi: 10.1002/gps.1130] [Medline: 15211536]
26. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
27. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med 1997 Mar;27(2):363-370. [Medline: 9089829]
28. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B Stat Methodol 1977:1-38.
29. Kelders SM. Understanding adherence to web-based interventions [Dissertation]. Enschede, the Netherlands: University of Twente; 2012.
30. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
31. Van Gemert-Pijnen JEWC, Peters O, Ossebaard HC. Improving eHealth. The Hague: Boom; 2013.
32. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
33. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. J Med Internet Res 2013 Jun;15(6):e121 [FREE Full text] [doi: 10.2196/jmir.2248] [Medline: 23807565]
34. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. Commun Assoc Inf Syst 2009;24(1):28.
35. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]

Abbreviations

ACT: acceptance commitment therapy
CBT: cognitive behavioral therapy
CES-D: Center of Epidemiologic Studies—depression scale
EA: experiential avoidance
HADS-A: Hospital Anxiety and Depression Scale—Anxiety
MOST: Multi-Phase Optimization Strategy
PSD: persuasive system design
SMS: short message service

Edited by D Mohr; submitted 29.09.13; peer-reviewed by J Freyne, P VandeVen, F Schneider; comments to author 07.11.13; revised version received 20.12.13; accepted 25.12.13; published 31.01.14.

Please cite as:
Van Gemert-Pijnen JEWC, Kelders SM, Bohlmeijer ET
Understanding the Usage of Content in a Mental Health Intervention for Depression: An Analysis of Log Data
J Med Internet Res 2014;16(1):e27
URL: http://www.jmir.org/2014/1/e27/
doi:10.2196/jmir.2991
PMID:24486914

Corrigenda and Addenda

Title Correction: Sources of Information and Behavioral Patterns in Online Health Forums: Observational Study

Fabian Sudau[1], MSc; Tim Friede[2], PhD; Jens Grabowski[1], PhD; Janka Koschack[3], PhD; Philip Makedonski[1], MSc; Wolfgang Himmel[3], PhD

[1]Institute of Computer Science, Georg-August-University Göttingen, Göttingen, Germany
[2]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[3]Department of General Practice, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Fabian Sudau, MSc
Institute of Computer Science
Georg-August-University Göttingen
Goldschmidtstraße 7
Göttingen, 37077
Germany
Phone: 49 551 39 17202
Fax: 49 551 39 14415
Email: fabian.sudau@cs.uni-goettingen.de

Related Article:

Correction of: http://www.jmir.org/2014/1/e10

(J Med Internet Res 2014;16(1):e39) doi:10.2196/jmir.3280

The authors of the paper originally entitled "Sources of Information and Behavioral Patterns in Online Health Forums: Qualitative Study" (J Med Internet Res 2014 Jan 14;16(1):e10) wish to correct the title of their publication. The subtitle "Qualitative Study" is misleading because the analysis was primarily quantitative. The title has now been changed to read "Sources of Information and Behavioral Patterns in Online Health Forums: Observational Study". This error has been corrected in the online version of the paper on the JMIR website on January 31, 2014, together with publishing this correction notice. There are no changes to the contents of the paper. A correction notice has been sent to PubMed. This was done before submission to Pubmed Central and other full-text repositories.

Edited by G Eysenbach; submitted 28.01.14; this is a non–peer-reviewed article;accepted 28.01.14; published 31.01.14.

Please cite as:
Sudau F, Friede T, Grabowski J, Koschack J, Makedonski P, Himmel W
Title Correction: Sources of Information and Behavioral Patterns in Online Health Forums: Observational Study
J Med Internet Res 2014;16(1):e39
URL: http://www.jmir.org/2014/1/e39/
doi:10.2196/jmir.3280
PMID:30583454

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org